Probabilistic inference of ecohydrological parameters using observations from point to satellite scales

Maoya Bassiouni[1], Chad W Higgins[1], Christopher J Still[2], and Stephen P Good[1]

[1]Department of Biological and Ecological Engineering, Oregon State University, Corvallis, OR 97333, USA
[2]College of Forestry, Oregon State University, Corvallis, OR 97333, USA

Correspondence to: Maoya Bassiouni (bassioum@oregonstate.edu)

Abstract. Vegetation controls on soil moisture dynamics are generally not measured directly and not easy to translate into scale and site-specific ecohydrological parameters for simple soil water balance models. We hypothesize that empirical probability density functions (pdfs) of relative soil moisture or soil saturation encodes sufficient information to determine these ecohydrological parameters, and that these parameters can be estimated through inverse modelling of the commonly used stochastic soil water balance. We developed a generalizable Bayesian inference approach to estimate soil saturation thresholds at which plants control soil water losses, based only on soil texture, rainfall and soil moisture data at point, footprint, and satellite scales. The Nash-Sutcliffe efficiencies between empirical pdfs derived from a year of observations and the optimal analytical soil saturation pdfs assuming a steady state or wet and dry seasonal dynamics ranged from 0.89 to 0.99. The coefficient of variation of posteriori parameter distributions ranged from <1 to 15 %. The parameter identifiability was not significantly improved in the more complex seasonal model, however small differences in parameter values indicates that the steady state model may have absorbed dry season dynamics. Parameter estimates were most constrained for scales and locations at which soil water dynamics are more sensitive to the fitted ecohydrological parameters of interest. In these cases, convergence of the model inversion was attained less rapidly but ultimately provided better goodness of fit and lower uncertainty. Results were robust using as little as 100 daily observations randomly sampled from the full records, demonstrating the advantage of analyzing soil saturation pdfs instead of time series to estimate ecohydrological parameters from sparse records. This work combined modelling and empirical approaches in ecohydrology and provided a simple framework to obtain analytical descriptions of soil moisture dynamics at a range of spatial scales that are consistent with soil moisture observations.

1 Introduction

The movement of water from soils, through plants, and back to the atmosphere via transpiration, is a critical component of local and global hydrologic cycles and is the largest surface-to-atmosphere water pathway (Good et al., 2015). A realistic analytical description of soil moisture dynamics is key to understanding ecohydrological processes that regulate the productivity of natural and managed ecosystems. Rodriguez-Iturbe et al. (1999) introduced a conceptually simple framework using a bucket model of soil-column hydrology forced with stochastic precipitation inputs, where soil water losses are only a function of soil saturation. Given this ecohydrological framework, the probability density function (pdf) of soil moisture and the mean components of the soil water balance (rainfall, runoff, evapotranspiration, and leakage losses) are analytically derived and depend on simple abiotic characteristics such as average climate and soil texture, and biotic characteristics including soil saturation thresholds at which vegetation can influence soil water losses. However, the shapes of analytical soil moisture pdfs are generally not consistent with observations when literature values for model parameters are used (Miller et al., 2007). Also, because of simplifications made to describe soil water loss processes in the model, some parameters such as field capacity and the wilting point do not correspond to conventional definitions and need to be calibrated (Dralle and Thomspon, 2016). Analytical soil moisture pdfs have never been

directly compared to empirical pdfs derived from measurements beyond the point scale. Observation networks provide freely available point scale, spatially integrated soil moisture observations, while remotely sensed soil moisture observations are available through satellite products. These data sources create an opportunity to: 1) evaluate whether analytical soil saturation pdfs are consistent with observations across a range of scales; and 2) determine average ecohydrological parameters relevant to each scale.

Estimates of ecohydrological parameters are relevant to a large range of applications for which the stochastic soil water balance framework has been used and adapted, including: the effects of climate, soil and vegetation on soil moisture dynamics (Laio et al, 2001a; Rodriguez-Iturbe et al., 2001; Porporato et al., 2004), ecohydrological factors driving spatial and structural characteristics of vegetation (Caylor et al., 2005; Manfreda et al., 2017), soil salinization dynamics (Suweis et al., 2010), biological soil crusts (Whitney et al., 2017), vegetation stress, optimum plant water use strategies and plant hydraulic failure (Laio et al., 2001b; Manzoni et al. 2014; Feng et al., 2017), vertical root distributions (Laio et al., 2006), plant pathogen risk (Thomspon et al., 2013), streamflow persistence in seasonally dry landscapes (Dralle et al., 2016), and soil water balance partitioning (Good et al., 2014 ; Good et al., 2017. A survey of close to 400 ecohydrology publications found that 40% relied heavily on simulation, rarely integrated empirical measurements, and were almost never coupled with experimental studies, suggesting a critical need to combine modelling and empirical approaches in ecohydrology (King and Caylor, 2011). A few studies have directly confronted the governing equations of the stochastic soil water balance model with observed soil moisture data and fewer have attempted to optimize model parameters to best fit soil moisture observations. Miller et al., (2007) calibrated soil moisture pdfs to project vegetation stress in a changing climate. Dralle and Thompson (2016) developed an analytical expression for annually integrated soil moisture pdfs under seasonal climates and calibrated soil moisture thresholds between which evapotranspiration is maximum and zero to compare the model to soil moisture observations at a savanna site. Chen et al., (2008) related evapotranspiration observations at the stand scale to soil moisture values using a Bayesian inversion approach, and Volo et al., (2014) calibrated the soil moisture loss curve to investigate effects of irrigation scheduling and precipitation on soil moisture dynamics and plant stress. The functional form of the soil moisture losses was approximated using conditionally averaged precipitation (Salvucci, 2001; Saleem and Salvucci, 2002) and remotely sensed data (Tuttle and Salvucci, 2014). The time scale of soil moisture dry downs, derived from the soil moisture loss equations, were parameterized using evapotranspiration measured at micro-meteorological stations (Teuling et al., 2006) and space-borne near-surface soil moisture observations (McColl et al., 2017). These studies indicate that the ecohydrological soil water balance framework is consistent with ground and larger scale remotely sensed measurements.

This study expands upon previous work and presents a general method for the inference of ecohydrological parameters and associated uncertainty, from observed soil moisture pdfs at a range of scales. The inference approach was applied to co-located and concurrent soil moisture observations from a range of biomes at the point, footprint, and satellite scales. Parameters that are representative of larger scale observations are necessary to characterize ecohydrological processes at ecosystem scales and are more relevant to ecohydrological modelling. These larger scale parameters integrate a range of ecohydrological interactions that are poorly understood and difficult to measure directly. The presented inference framework provides a means to quantify and compare the sensitivity of soil moisture dynamics at varying scales through estimates of these simple ecohydrological parameters. Non-biological controlling factors on the soil water balance including rainfall and soil texture can generally be assessed from readily available data, including site measurements, regionalized maps, and satellite observations. Vegetation controls on soil water dynamics are largely unknown and difficult to measure at hydrologically meaningful scales (Li et. al., 2017). Vegetation water-use traits are generally observed at the species level and are not easily translated to the simple parameters necessary in soil-water balance models. The rate of soil water losses from the near-surface soil layer, in which soil moisture measurements are generally

made, do not precisely correspond to evapotranspiration observed or calculated from meteorological stations. We thus focused on estimating parameters that are not directly observable, in particular the soil saturation thresholds at which vegetation controls soil water losses and the maximum rate of evapotranspiration from a near-surface soil layer. We use an inverse modelling approach and data that are commonly collected at environmental monitoring sites or measured from satellites.

We assume that if a sufficient range of soil moisture values are observed at a site, then the shape of the empirical soil saturation pdf is constrained by the ecohydrological factors driving soil moisture dynamics. We hypothesize that key information required to determine these ecohydrological factors is encoded in empirical soil saturation pdfs, and that this information can be extracted by calculating the inverse of the commonly used stochastic soil water balance. The analysis of soil saturation pdfs is a more robust and integrated approach to investigate ecohydrological factors of soil water dynamics than time series analysis. Soil saturation pdfs are less sensitive to the many sources of uncertainty, sensor noise, and common gaps in soil moisture observations and do not require high quality co-located and concurrent hydrologic measurements that are often lacking. Two key assumptions which are imbedded in the proposed method are tested: (1) The analytical pdf models properly describe empirical soil moisture pdfs observed from annual data at each scale and location. Annual soil moisture records can be affected by transitional dynamics between wet and dry seasons and the appropriate level of model complexity must be used. We will compare parameter identifiability using a steady-state and a seasonal formulation of the analytical model for soil saturation pdfs (2) The whole range of realizable soil moistures values is captured by the selected time series and the soil moisture pdf determined from these observations is not truncated. We will determine whether the inference method based on soil saturation pdfs is robust against reduced data availability by repeating the model inversions on subsets of the soil moisture time series and show that the method can be applied to sparse datasets.

A number of studies have combined inverse modelling approaches with ground and remotely sensed soil moisture data to successfully extract meaningful hydrologic information (Xu et al., 2006; Miller et al, 2007; Chen et al., 2008; Volo et al., 2014; Wang et al., 2016; Baldwin et al., 2017). In particular, Bayesian inference methods are effective in relating prior pdfs of observations to posterior estimates of model parameters (Xu et al., 2006; Chen et al., 2008; Baldwin et al., 2017). The soil water balance model provides a direct analytical equation for soil moisture pdfs that is convenient to use with the Bayesian paradigm because it is a low parameter model with few data inputs. In this study, we developed a Bayesian inversion approach to directly estimate soil water balance model parameters that best fit soil moisture pdfs derived from observations at point, footprint, and satellite scales. We selected a Bayesian inversion approach instead of a least-squares or maximum likelihood approach because it quantifies the inference uncertainty directly and improves upon the work of Miller et al. (2007), which used a least-squares approach to calibrate soil saturation pdfs. In addition, measures of inference uncertainty and parameter convergence diagnostics provided by the Bayesian approach can be used to evaluate the validity of model inversion and develop criteria to generalize the presented framework.

The goal of this study was to match empirical soil moisture pdfs derived from point-, footprint-, and satellite-scale observations to a commonly used analytical model. We demonstrate the use of a Bayesian inversion framework to calibrate the ecohydrological parameters of a simple stochastic soil water balance model that best fit empirical soil moisture pdfs. We first present data sources, define the analytical model for soil moisture pdfs including parameter assumptions, and detail the algorithm used in the Bayesian inversion. Then, we present a summary of the goodness of fit of optimal analytical soil moisture pdfs and estimated parameter uncertainty. Results were evaluated to test key method assumptions including model complexity and data availability. Finally, we discuss the potential of the approach to provide a simple means to investigate variability in ecohydrological controlling factors at

varying spatial scales. This work combines modelling and empirical approaches in ecohydrology to provide more realistic analytical descriptions of soil moisture dynamics. Estimates of ecohydrological parameters that are consistent with observed soil moisture pdfs, from point to ecosystem scales, are needed to better characterize site-specific ecohydrological processes.

2. Data and Methods

2.1 Data analysed

Daily soil moisture observations from three data products at three different spatial scales were used in this study. Point-scale soil moisture at 10 cm depth was taken from the FLUXNET2015 data product (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/). Footprint-scale soil moisture was taken from the Cosmic-ray Soil Moisture Observing System (COSMOS) (http://cosmos.hwr.arizona.edu/Probes/probelist.html). The COSMOS soil moisture footprint measures soil moisture at an average depth of 20 cm with a radius ranging from 130 to 240 m, depending on site characteristics (Köhli et al., 2015). Near-surface soil moisture observations at a spatial resolution of 0.25˚ were taken from the European Space Agency's (ESA) Climate change Initiative (CCI) project. The combined soil moisture product (ECV-SM, version 0.2.2) that merges soil moisture retrievals from four passive (SMMR, SMM/I, TMI, and ASMR-E) and two active (AMI and ASCAT) coarse resolution microwave sensors was used (Liu et al., 2011; Liu et al., 2012; Wagner, 2012). Although the ECV-SM sensing depth is less than 5 centimetres, it has been shown to have a close relation to ground-based observations of soil moisture in the upper 10 centimetres (Dorigo et al., 2015). Daily rainfall time series were compiled from the FLUXNET2015 dataset for the point-and footprint-scale analysis, and the National Aeronautics and Space Administration's (NASA) Tropical Rainfall Measuring Mission (TRMM) dataset (Huffman et al., 2007) for the satellite-scale analysis.

A total of 4 sites with data available during the 2012 calendar year for each soil moisture and rainfall product were selected for this analysis (Table 1). Selected sites span a range of land cover types including crop and grasslands, oak savanna, deciduous forest and pine forest. For each site, the dominant soil texture of the upper soil layer was determined from the Harmonized World Soil Database (HWSD) (version 1.2) (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012). Soil porosity values, derived from the HWSD available as ancillary data through the ESA-CCI data product were used for the satellite-scale analysis. For point- and footprint-scale data products, the maximum soil moisture observation during the year 2012 was used as a site-specific soil porosity estimate. Soil porosity for each site was applied to compute the relative soil moisture content or soil saturation ($0 \leq s \leq 1$) from each observed soil moisture value. Soil saturation and rainfall data at each scale and for each site during the selected analysis period are presented in Fig. 1 and summary statistics are reported in Table 1. The difference in data quality between data sources and sites is not expected to significantly affect empirical soil saturation pdfs and resulting parameter estimates in this study. All sites had full records of daily point-, and footprint-scale observations except for US-Me2, which had 55 missing footprint-scale observations during the winter months when the ground is was saturated and frozen. The number of daily satellite-scale observations in the 2012 records ranged from 202 and 283.

2.2 Analytical model for soil saturation probability density functions (pdfs)

2.2.1 Model definition

The framework used in this study is based on a standard bucket model of soil column hydrology at a point forced with stochastic precipitation inputs and in which soil water losses are a function of soil saturation. We follow the simple formulation of soil water

losses in Laio et al. (2001a). We apply two associated analytical formulation for the pdf detailed below. The first is a steady state solution and the second takes into account wet and dry season dynamics. However, the methodology described in Sect. 2.3 can be customized to characterize site-specific parameters and test consistency between observed and analytical soil saturation pdfs for any application or adaptation of the stochastic ecohydrological framework.

The soil water balance model is defined at a point scale and a daily time scale, for a soil with porosity n, and assumes soil saturation is uniform in the considered soil column depth Z. Rainfall, the only input to the soil water balance, is treated as a Poisson process characterized by an average event frequency, λ, and average event intensity, α. For simplification, we assume that the rainfall applied is equal to the amount reaching the ground surface and do not account for rainfall intercepted by vegetation. Interception may be a significant component of the soil water balance at forested sites and may need to be accounted for in other studies. The daily soil water balance is written as the difference between φ, the rate of infiltration from rainfall and χ, the rate of soil moisture losses:

$$nZ\frac{ds(t)}{dt} = \varphi[s(t);t] - \chi[s(t)] \quad (1)$$

$\varphi[s(t);t]$ is a stochastic process controlled by rainfall and is also a state-dependent process, because excess rainfall relative to available soil storage is converted to surface runoff. $\chi[s(t)]$, the soil moisture loss curve, is summarized in Fig. 2a and includes leakage losses due to gravity and evapotranspiration and is described in stages determined by five soil saturation thresholds (Laio et al., 2001a). These stages are: (1) the saturation point ($s = 1$), at which all pores are filled with water; (2) the field capacity (s_{fc}), at which soil-gravity drainage becomes negligible compared to evaporation; (3) the point of incipient stomata closure (s^*), at which plants begin to reduce transpiration from water stress; (4) the wilting point (s_w), at which plants cease to transpire; and (5) the hydroscopic point (s_h), at which water is bound to the soil matrix. Soil water losses are controlled by physical soil properties for saturation states above s_{fc}. The rate of leakage due to gravity is assumed maximum when the soil is saturated (K_s) and decays exponentially to a value of 0 at s_{fc} (Brooks and Corey, 1964). Soil water losses are controlled by micro-meteorological conditions for saturation states between s_{fc} and s^*. The rate of evapotranspiration is assumed to occur at a maximum rate (E_{max}), which is independent of the saturation state. Soil water losses are controlled primarily by vegetation for saturation states between s^* and s_w. Plants close their stomata in response to soil water deficits that drive leaf water potential gradients, as well as to atmospheric vapor pressure deficits, and evapotranspiration decreases linearly from E_{max} to E_w at s_w. Soil water losses are controlled by soil diffusivity for soil saturation states below s_w, and soil evaporation decreases linearly from E_w to 0 at s_h. Soil water losses are negligible for soil saturation states below s_h. The piece-wise linear relation between soil saturation and evapotranspiration is a simplifying assumption commonly used in soil water balance models.

For this simplified theoretical description of the soil water loss curve and stochastic rainfall forcing, the analytical solution of the steady-state probability distributions of soil saturation, $p(s)$ given by Laio et al. (2001a) is:

$$p(s)=\begin{cases}0, & 0<s\le s_h,\\ \frac{C}{\eta_w}\left(\frac{s-s_h}{s_w-s_h}\right)^{\frac{\lambda(s_w-s_h)}{\eta_w}-1}e^{-\gamma s}, & s_h<s\le s_w,\\ \frac{C}{\eta_w}\left[1+\left(\frac{\eta}{\eta_w}-1\right)\left(\frac{s-s_w}{s^*-s_w}\right)\right]^{\frac{\lambda(s^*-s_w)}{\eta-\eta_w}-1}e^{-\gamma s}, & s_w<s\le s^*,\\ \frac{C}{\eta}e^{-\gamma s+\frac{\lambda}{\eta}(s-s^*)}\frac{\eta}{\eta_w}^{\frac{\lambda(s^*-s_w)}{\eta-\eta_w}}, & s^*<s\le s_{fc},\\ \frac{C}{\eta}e^{-(\beta+\gamma)s+\beta s_{fc}}\left(\frac{\eta e^{\beta s}}{(\eta-m)e^{\beta s_{fc}}+me^{\beta s}}\right)^{\frac{\lambda}{\beta(\eta-m)}+1}\frac{\eta}{\eta_w}^{\frac{\lambda(s^*-s_w)}{\eta-\eta_w}}e^{\frac{\lambda}{\eta}\left(s_{fc}-s^*\right)}, & s_{fc}<s\le 1,\end{cases} \tag{2}$$

where

$\frac{1}{\gamma}=\frac{\alpha}{nZ}$,

$\eta_w=\frac{E_w}{nZ}$,

$\eta=\frac{E_{max}}{nZ}$,

$m=\frac{K_s}{nZ(e^{\beta\left(1-s_{fc}\right)}-1)}$,

$\beta=2b-4$.

where b, is an experimentally determined parameter used in the Clapp and Hornberger, (1978) soil water retention curve and the constant C can be obtained numerically to ensure the integral of $p(s)$ is equal to 1. This framework was derived under the assumption of steady state, wherein parameters are constant for a given period of time. We used a simplifying relation $E_w=0.05E_{max}$ to reduce the number of parameters.

To account for transient dynamics between wet and dry seasons we adopt the framework in Dralle and Thompson (2016). The dry season is defined as a period of duration t_d, in which precipitation is negligible and does not contribute to soil moisture. During the dry season, soil saturation decays from an initial value s_0 to $s(t_d,s_0)$. For simplification in this study, t_d is identified using rainfall records at a monthly step (see Sec 2.2.2) and s_0 is the soil saturation value on the last day of the wet season and does not imply that s_0 is the soil saturation following the last significant storm of the wet season. The annual soil saturation pdf, (p_{wd}(s)) is then calculated as the weighted sum of the wet and dry season pdfs.

$$p_{wd}(\mathrm{s})=\left(1-\frac{t_d}{365}\right)p_w(s)+\frac{t_d}{365}p_d(s) \tag{3}$$

The wet season pdf, $p_w(s)$ is the steady-state solution in Eq. 2. The dry season pdf, $p_d(s)$ is numerically determined by

$$p_d(s)=\int_{s_0}p_{S_d|S_0}(s,s_0)p_0(s_0)ds_0 \tag{4}$$

where $p_0(s_0)$ is the pdf the initial dry season soil saturation, equal to $p_w(s)$ and $p_{S_d|S_0}(s,s_0)$ is the pdf of dry season soil saturation given an initial condition s_0.

$$p_{S_d|S_0}(s,s_0)=\frac{C_d}{t_d}\begin{cases}\frac{e^{\beta(s_0-s)}}{(\eta^d-m)e^{\beta(s_0-s)}-\eta^d+m+me^{\beta(s_0-s_{fc})}}, & s_{fc}<\mathrm{s}\le 1,\\ \frac{1}{\eta^d}, & s^*<\mathrm{s}\le s_{fc},\\ \frac{1}{\eta^d-{\eta_w}^d}\left(\frac{s^*-s_w}{(\eta^d-{\eta_w}^d)(s-s_w)+{\eta_w}^d(s^*-s_w)}\right), & s_w<\mathrm{s}\le s^*\\ \frac{1}{{\eta_w}^d}\left(\frac{s_w-s_h}{s-s_h}\right), & s_h<\mathrm{s}\le s_w\\ 0, & s_0\le \mathrm{s},\\ 0, & \mathrm{s}\le s_h,\\ 0, & \mathrm{s}\le s(t_d,s_0)\end{cases} \tag{5}$$

where η^d and ${\eta_w}^d$ are equivalent to η and η_w relative to ${E_{max}}^d$ the maximum evapotranspiration rate in the dry season climate and C_d is a normalization constant. The expression for $p_{S_d|S_0}(s, s_0)$ used in this study was derived following the framework in Dralle and Thompson (2016) but using the analytical expression for soil saturation decay, $s(t, s_0)$ in absence of rainfall given by Laio et al., 2001. In this study we will evaluate whether annual soil saturation data from point, footprint, and satellite scales are consistent with the assumptions in the steady state solution in Eq (2) and the second seasonal solution in Eq (3). This comparison will determine how the level of model complexity affects the identifiability of the ecohydrological parameters of interest.

2.2.2 Climate, soil and vegetation parameter characterization

The rainfall characteristics (λ and α), the length of the dry period t_d, and physical soil parameters (s_{fc}, s_h, K_s, and b) used in Eq. (2) and (3) are based on readily available data. We chose values based on our best estimates of the driving climate and physical soil controls on the soil water balance. We thus focused on estimating the ecohydrological parameters s^*, s_w, and E_{max} which describe vegetation controls on soil water losses and are not easily observable. We acknowledge that the pre-defined rainfall characteristics and physical soil parameters based on observations or literature values may not be perfectly representative of the processes at each location or scale and could create biases and uncertainties in our fitted parameters of interest.

Rainfall characteristics λ and α associated with the annual record and the wet season months for at each sites were calculated for each site from the FLUXNET2015 and TRMM rainfall records following Rodriguez-Iturbe et al. (1984) and listed in Table 1. The FLUXNET2015 rainfall characteristics were used for the point- and footprint-scale analysis, while the TRMM rainfall characteristics were used for the satellite-scale analysis. For each location, the FLUXNET2015 rainfall depth for each month of the year was evaluated and consecutive months contributing to less than 5 percent of the site's annual rainfall were categorized as dry season months. The TRMM rainfall records were generally consistent with the ground-based measurements. The length of the dry period, t_d was then calculated as the number of days in these months and the same dry season months and value for t_d were used for the point, footprint, and satellites scales. The dry season period for each site is shaded in grey in Fig 1. Physical soil characteristics for soil textures associated with each site, s_h, K_s, and b were taken from Rawls et al. (1982) and are listed for each site in Table 1. To be most consistent with the assumption that drainage losses are generally insignificant compared to evapotranspiration losses the day following a rain event, s_{fc} was estimated from each soil saturation record and listed in Table 1. All days in the 2012 record following an observed decrease in soil saturation were identified and s_{fc} was estimated as the 95th percentile of the soil saturation value of these selected days. Daily soil moisture states below s_w and above s_{fc} are rare (Laio et al., 2001) therefore we do not expect the average soil texture values for s_h and K_s to significantly affect results. The soil depth considered corresponded to the measurement sensing depths of 10, 20, and 5 cm for the point, footprint, and satellite scales, respectively. Because the soil depth Z is shallower than the rooting depth, E_{max} is only a fraction of the atmospheric moisture demand (or potential evapotranspiration) contributed by that soil depth and therefore unknown. The framework we present thus considers 4 (or 3 if seasonality is ignored) unknown soil water balance parameters, s^*, s_w, E_{max}, and ${E_{max}}^d$. Our goal is estimate these parameters, as defined over the following intervals:

$$\begin{cases} s_h \le s^* \le s_{fc}, \\ s_h \le s_w \le s_{fc}, \\ 0 \le E_{max} \le 10, \\ 0 \le {E_{max}}^d \le 10 \end{cases} \quad (6)$$

where 10 mm day^{-1} is the pre-defined upper possible bounds for potential evapotranspiration. Estimates of s^* and s_w can be converted to soil matrix potential if soil water retention parameters are well known. The Clapp and Hornberger, (1978) soil water retention curve is highly non-linear and estimates of soil water potential at which stomata fully are open or closed were not evaluated in this study.

A key assumption in this analysis is that the whole range of realizable soil moistures values is captured by the selected time series and the soil moisture pdf determined from these observations is not truncated. In these conditions, the shape of the soil saturation pdf is controlled by the actual physical constraints that parameterizes the analytical solution and these parameters can be determined with certainty. We expect that estimated soil saturation thresholds will have greater certainty if the empirical soil saturation pdf is most defined around those values and greater uncertainty if there are relatively fewer soil saturations values observed around the thresholds. If the range of observed values is not representative of the soil moisture pdf because it is truncated or affected by noise in the data, parameter estimates may be biased. The minimum and maximum observed soil saturation values during 2012 are reported in Table 1 to indicate the range of observed soil saturation values used to estimate ecohydrological parameters in this study. We will determine whether the inference method based on soil saturation pdfs is robust against reduced data availability by repeating the model inversions on subsets of the soil moisture time series and will show that the method can be applied to sparse datasets.

2.3 Bayesian inversion approach

2.3.1 Application of the Bayes theorem

Bayes' theorem, Eq. (7) is used to relate $p(S)$, the empirical soil saturation pdf of $j = [1, \dots, m]$ soil saturation observations (s_j) and the analytical soil saturation pdfs in Eq. (3), derived from the simple soil water balance model in Eq. (1), with 4 unknown soil water balance parameters $\theta = [s^*, s_w, E_{max}, {E_{max}}^d]$.

$$p(\theta|S) = \frac{p(S|\theta)\, p(\theta)}{p(S)} \quad (7)$$

The posterior distribution, $p(\theta|S)$, is the solution of the inverse problem and describes the probability of model parameters θ given the set $S = [s_1, s_2, \dots s_m]$ of soil saturation observations. Assuming uninformed prior knowledge, the prior distribution of model parameters θ, $p(\theta)$, are defined by uniform distributions over the intervals in Eq. (6). The conditional probability of observations S given model parameters θ, $p(\mathrm{S}|\theta)$, is the likelihood function of model parameters θ.

2.3.2 Parameter estimation

The Metropolis-Hasting Markov chain Monte Carlo (MH-MCMC) technique is used to estimate the posterior distribution of $p(\theta|S)$ by drawing random model samples θ_i from $p(\theta)$ and evaluating $p(\mathrm{S}|\theta_i)$ (Metropolis et al., 1953; Hastings, 1970; Xu et al., 2006). The likelihood function of a model i, $p(S|\theta_i)$ defined by

$$p(S|\theta_i) = \prod_{j=1}^{m} p\left(s_j|\theta_i\right) \quad (8)$$

where $p\left(s_j|\theta_i\right)$ is the probability of observation s_j given the model in Eq. (2) or Eq (3) using parameters θ_i.

The MH-MCMC technique converges to a stationary distribution according to the ergodicity theorem in Markov chain theory. The sampling algorithm consists of repeating two steps: (1) a proposing step, in which, the algorithm generates a new model $\theta_i{}'$ using

a random function that is symmetric about the previously accepted model θ_i, and (2) a moving step, in which, $\theta_i{}'$ is tested against the Metropolis criterion (a) to estimate if it should be accepted or rejected.

$$a = \frac{p(S|\theta_i{}')}{p(S|\theta_i)} \tag{9}$$

If $a > 1$, then θ_i is accepted and $\theta_{i+1} = \theta_i{}'$ is used for the next sample. If $a < 1$, a random number $p_* \in [0,1]$ is drawn from a uniform distribution and compared to a. If $p_* < a$, then $\theta_i{}'$ is accepted and $\theta_{i+1} = \theta_i{}'$ is used for the next sample. If $p_* > a$, $\theta_i{}'$ is rejected and $\theta_{i+1} = \theta_i$ is used for the next sample. If $\theta_i{}'$ is an inconsistent model in which the soil saturation thresholds (s_w, s_*) are ranked incorrectly or any of the soil water balance parameters (s^*, s_w, E_{max}, ${E_{max}}^d$) are outside of their defined physical bounds, the model likelihood is 0 and $\theta_i{}'$ is never accepted. In this study, the log-likelihood was more convenient to compute than the likelihood. The symmetric function used in the proposing step was a Gaussian distribution with a mean value equal to the accepted model θ_i and a standard deviation of 1 percent of interval range for which each parameter is defined in Eq. (6).

The value of the standard deviation of each model parameter was set after a number of test runs to generally ensure an acceptance rate between 20 and 50% (Robert and Rosenthal, 1998). Statistics of the estimated parameters in θ are obtained from the union of 3 run samples of 20 thousand simulations each. The burn-in period is the number of simulations after which the running mean and standard deviation are stabilized. We considered a burn-in period of 10 thousand simulations, which were discarded for each run sample. If the acceptance rate of a run sample is below 1% or greater than 90% after the burn-in period, the run was discarded and we concluded that the algorithm got stuck in a local minimum that may be physically impossible. Convergence was evaluated by the Gelman-Rubin (GR) diagnostic (Gelman and Rubin, 1992) on the run samples. The GR diagnostic determines that the algorithm reaches convergence when the within-run variability (σ_w) is roughly equal to the between-run variability (σ_b), i.e. σ_w/σ_b approaches 1. We verified that the GR diagnostic for each estimated parameter was lower than 1.1. If the GR diagnostic did not indicate that the 3 run samples converged, the run with the lowest likelihood was discarded and a new run sample was re-initiated until convergence was attained. The number of attempts was counted and quantifies how rapidly converging results are obtained. The mean and standard deviation of each parameter were computed from the total of 30 thousand simulations of θ resulting from the 3 converging run samples. A mean analytical model of soil saturation pdf was determined by applying Eq. (6) with the mean values of the 30 thousand posteriori parameter estimates

2.4 Evaluation of model inversion

Parameters estimated through the Bayesian inversion methods do not have direct measurement against which they can be validated. We therefore analyse the goodness of fit between the empirical and analytical soil saturation pdfs and uncertainty metrics of the model inversion to evaluate the identifiability of the ecohydrological parameters. The model inversion was evaluated by the following criteria.

(1) Convergence of the Bayesian inversion: a GR diagnostic below 1.1 for all unknown parameters is obtained from the union of 3 run samples and within a maximum of 10 run samples.

(2) Goodness of fit: a quantile-level Nash-Sutcliffe efficiency (NSE) (Müller et al., 2016) between the optimum analytical pdf derived from the mean parameter estimates and the empirical pdfs derived from observations greater than 0.85 and a Kolmogorov-Smirnov statistic below 0.2.

(3) Low uncertainty in parameter estimates: the posterior distributions of parameter estimates are physically plausible and have coefficients of variations below 20%.

This study investigates questions of model complexity, uncertainty in parameter estimation, data availability, and scales of applicability through the following analysis.

(1) We applied the inversion framework to variations of the analytical model for soil saturation pdfs of increasing complexity. The first is the steady-state model in Eq. 3 and the second is the seasonality model in Eq. 6. The annual soil moisture records are affected by transitional dynamics between wet and dry seasons. We determined whether the added complexity of the dry season pdf increases the identifiability of ecohydrological parameters or if the simpler steady state solution is sufficiently consistent with annual empirical soil saturation pdfs.

(2) We performed the model inversion using subsets of each soil saturation record by randomly resampling fractions of the data down to 10 % of the annual timeseries and goodness of fit statistics were computed between the resulting analytical models and the empirical models based on the full annual record. We determined the number of data points necessary to infer converging model parameters that best match observations and whether the proposed inference method based on soil saturation pdf can be reliably used to identify ecohydrological parameters from sparse datasets.

(3) We compared co-located parameter estimates and their uncertainty at a range of scales for each site. We determine whether the soil saturation pdf model inversion framework is applicable to point, footprint, and satellite-scale observations and whether inferred parameters can be appropriate for ecohydrological modelling at all scales and locations.

3. Results and discussion

3.1 Level of model complexity

For each of the 4 selected locations, optimal analytical soil saturation pdfs consistent with empirical pdfs derived from soil saturation observations were obtained through the Bayesian inversion framework and using a MH-MCMC algorithm. The model inversions for each site, scale, and for both the steady state and seasonal models met the evaluation criteria listed in Section 2.4. Posteriori probability distributions of soil water balance parameters (s_w, s^*, E_{max}) were overall well constrained. The parameter estimates and their coefficient of variation as well as the model goodness of fit statistics are summarized in Table 2. Figures 2 through 5 present a comparison between empirical and analytical pdfs with associated quantile-quantile plots for point, footprint, and satellite scales at the 4 study sites and for both the steady state and seasonal models. The goodness of fit between the empirical pdfs and the analytical models was only slightly better for the seasonal model compared to the steady state model. However, the coefficient of variation of the posteriori parameter distributions was smaller for the steady state model and convergence was attained more rapidly. The Bayesian inversion of the steady state model is therefore more computationally efficient. The parameter identifiability was not greatly improved by the more complex seasonal model. The estimated soil moisture thresholds s_w was consistently smaller for the steady state model than for the seasonal model and the s^* was often higher. This may indicate that the s_w and s^* parameters in the steady state model could be biased and have absorbed dry season dynamics. Previous studies have calibrated soil saturation pdf models and found ecohydrological parameters values that can be compared to those in Table 2. For example, using point-scale observations at US-Ton, best fit values of s_w and s_{fc} were 0.26 and 0.82 (Dralle and Thompson, 2016) and best-fit values of s^* and E_{max} were 0.3 and 1.9 mm d^{-1} (Miller et. al. 2007). We conclude that although the seasonal model is conceptually more appropriate and consistent with our physical understanding of annual soil water dynamics, the steady state model provides satisfactory results and is generally matches annual empirical pdf at each site considered in this study.

3.2 Data availability

For each spatial scale and site, the steady state model was inversed, using random subsamples of 100 to 10 percent of the 2012 time series and results were summarized in Fig 6. For all sites and scales the number of observations did not significantly impact model inference. The NSE, Kolmogorov-Smirnov statistic and parameter estimates were relatively stable down to about 100 observations. Model parameter values and the variability of parameter estimates between the 10 repeats in each subsample fraction were not sensitive to the number of observations used. Results indicate the identifiability of ecohydrological parameters through the inversion of the analytical model of soil moisture pdfs was robust because the mean and standard deviation of the randomly selected subsets of annual data were generally representative of the full record. There was no correlation between the small differences in the mean and standard deviations of the subsamples and the model goodness of fit. We conclude that the proposed inference method based on soil saturation pdf can be reliably used to identify ecohydrological parameters from sparse datasets. This is particularly relevant to large scale soil moisture measurement such as satellite products that are not continuous.

3.3 Site and scale considerations

Parameter estimates were most constrained for scales and locations at which soil water dynamics are more sensitive to the fitted ecohydrological parameters of interest. In these cases, convergence of the model inversion was generally attained less rapidly but ultimately provided better goodness of fit. Soil saturation states at drier sites may be more controlled by soil water loss parameters, while soil saturation states at wetter sites may be more controlled by rainfall characteristics. The estimated soil saturation thresholds had greater certainty if the empirical soil saturation pdf were most defined around those values and greater uncertainty if there are relatively fewer soil saturations values observed around the thresholds. For example, the uncertainty of s_w was greater for the humid subtropical deciduous forest site (US-MMS) than for the Mediterranean savanna sites (US-Ton) and the uncertainty of s^* was greater for US-Ton than US-MMS. Similarly, soil saturation states representing larger spatial scales are less sensitive to specific site characteristics. In this study parameter uncertainty for satellite and footprint scales was greater than for the point scale. Estimates of larger scale soil water balance parameters are more relevant to regional ecohydrological dynamics. Differences in parameter estimates between scales within a site may be associated with differences in soil texture properties, such as porosity and field capacity, that were determined separately for each record. Figure 2 through 5 also show that co-located and concurrent soil saturation pdfs are different at each scale and suggest variability in observed soil water dynamics at each scale. Differences in controlling processes between scales were specifically determined from the model inversion for each scale and provided robust scale-specific parameters for ecohydrological modelling.

4. Conclusions

Empirical pdfs derived from soil saturation observations provided key information to determine unknown ecohydrological parameters s^*, s_w, and E_{max}. This study documented a generalizable Bayesian inversion framework to infer parameters of the stochastic soil water balance model and their associated uncertainty using freely available rainfall and soil moisture observations at point, footprint and satellite scales. Model assumptions were appropriately met and optimal analytical soil saturation pdfs were consistent with empirical pdfs. Uncertainty in parameter estimates were small. Stable results were obtained using sparse subsets of the datasets, demonstrating the advantage of analyzing soil saturation pdfs instead of time series and the robustness of the proposed framework when only sparse datasets are available. The model inversion results were also used to evaluate the sensitivity of the

soil water balance model to ecohydrological parameters at varying scales and locations. We demonstrated that the form of the simple ecohydrological model for soil saturation pdfs was in agreement with observations from point, footprint, and satellite scales; however optimal parameters were different at each scale because co-located and concurrent soil saturation pdfs are different at each scale and may result from spatial heterogeneity in soil water dynamics. Methods developed in this study can be applied in future studies to better understand differences in soil water dynamics at different scales and improve the scaling of ecohydrological processes. Results demonstrated the value of large scale near-surface soil moisture observations to improve the characterization of soil water dynamics at ecosystem scales. The relation between the soil moisture threshold values inferred from the near surface soil moisture data with dynamics in the full active rooting zone are unknown. This study provided a method to estimate ecohydrological characteristics that are not directly observable, and for which established estimation methods are not available. The datasets used in this study are freely available from sensor networks and global satellite products and methods can therefore be applied to a large range of sites or to full global datasets to improve understanding of spatial patterns in ecohydrological parameters relevant for local and global water cycle analyses.

Data and code availability

All datasets used in this study were downloaded from publicly available sources: point-scale soil moisture and rainfall data are available through FLUXNET2015 (http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/); footprint-scale soil moisture data are available through COSMOS (http://cosmos.hwr.arizona.edu/Probes/probelist.html); remotely-sensed soil moisture data are available through ESA CCI (http://www.esa-soilmoisture-cci.org/node/145); remotely sensed rainfall data are available through NASA TRMM (https://pmm.nasa.gov/data-access/downloads/trmm); global soil texture data are available through FAO HWSD (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). Custom scripts in the Python computing language associated with this analysis are available upon request through a private gitHub repository and will be made publicly available after revisions of this manuscript. (Citation and doi TBD).

Competing interests

Authors declare that they have no conflict of interest.

Acknowledgments

We thank Minghui Zhang, Marc Müller, David Dralle, and Xue Feng and the Editor Sally Thomspon for their thoughtful reviews and useful feedback on the earlier draft of this manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1314109-DGE. S.P.G. acknowledges the financial support of the Unites States National Aeronautics and Space Administration (NNX16AN13G). This work used the Extreme Science and Engineering Discovery Environment (XSEDE) via allocation DEB160018, which is supported by National Science Foundation grant number ACI-1548562. This work used data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices.

References

Baldocchi, D.: AmeriFlux US-Ton Tonzi Ranch, doi:10.17190/AMF/1245971, 2001-.

Baldwin, D., Manfreda, S., Keller, K. and Smithwick, E. A. H.: Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States, Journal of Hydrology, 546, 393–404, doi:10.1016/j.jhydrol.2017.01.020, 2017.

Biraud, S.: Ameriflux US-ARM ARM Southern Great Plains site-Lamont, doi:10.17190/AMF/1246027, 2002-.

Brooks, R. and Corey, T.: Hydraulic properties of porous media, hydrology papers, Colorado State University, 1964.

Caylor, K. K., D'Odorico, P. and Rodriguez-Iturbe, I.: On the ecohydrology of structurally heterogeneous semiarid landscapes, Water Resources Research, 42(7), doi:10.1029/2005WR004683, 2006.

Chen, X., Rubin, Y., Ma, S. and Baldocchi, D.: Observations and stochastic modeling of soil moisture control on evapotranspiration in a Californian oak savanna: soil moisture control on ET, Water Resources Research, 44(8), doi:10.1029/2007WR006646, 2008.

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

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M. and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sensing of Environment, 162, 380–395, doi:10.1016/j.rse.2014.07.023, 2015.

Dralle, D. N., Karst, N. J. and Thompson, S. E.: Dry season streamflow persistence in seasonal climates, Water Resources Research, 52(1), 90–107, doi:10.1002/2015WR017752, 2016.

Dralle and Thompson (2016)

Feng, X., Dawson, T. E., Ackerly, D. D., Santiago, L. S. and Thompson, S. E.: Reconciling seasonal hydraulic risk and plant water use through probabilistic soil-plant dynamics, Global Change Biology, 23(9), 3758–3769, doi:10.1111/gcb.13640, 2017.

Gelman, A. and Rubin, D. B.: Inference from iterative simulation using multiple sequences, Statistical science, 457–472, 1992.

Good, S. P., Soderberg, K., Guan, K., King, E. G., Scanlon, T. M. and Caylor, K. K.: δ 2 H isotopic flux partitioning of evapotranspiration over a grass field following a water pulse and subsequent dry down, Water Resources Research, 50(2), 1410–1432, doi:10.1002/2013WR014333, 2014.

Good, S. P., Noone, D. and Bowen, G.: Hydrologic connectivity constrains partitioning of global terrestrial water fluxes, Science, 349(6244), 175–177, 2015.

Good, S. P., Moore, G. W., & Miralles, D. G.: A mesic maximum in biological water use demarcates biome sensitivity to aridity shifts. Nature Ecology & Evolution 1, 1883–1888, doi:10.1038/s41559-017-0371-8, 2017.

Hastings, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57(1), 97, doi:10.2307/2334940, 1970.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, Journal of Hydrometeorology, 8(1), 38–55, doi:10.1175/JHM560.1, 2007.

Li, Y., Guan, K., Gentine, P., Konings, A. G., Meinzer, F. C., Kimball, J. S., Xu, X., Anderegg, W. R. L., McDowell, N. G., Martínez-Vilalta, J., Long, D. G. and Good, S. P.: Estimating global ecosystem iso/anisohydry using active and passive microwave satellite data: Estimate global ecosystem iso/anisohydry, Journal of Geophysical Research: Biogeosciences, doi:10.1002/2017JG003958, 2017.

King, E. G. and Caylor, K. K.: Ecohydrology in practice: strengths, conveniences, and opportunities, Ecohydrology, 4(4), 608–612, doi:10.1002/eco.248, 2011.

Köhli, M., Schrön, M., Zreda, M., Schmidt, U., Dietrich, P. and Zacharias, S.: Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons, Water Resources Research, 51(7), 5772–5790, doi:10.1002/2015WR017169, 2015.

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

Laio, F., Porporato, A., Fernandez-Illescas, C. P. and Rodriguez-Iturbe, I.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: IV. Discussion of real cases, Advances in Water Resources, 24(7), 745–762, 2001b.

Laio, F., D'Odorico, P. and Ridolfi, L.: An analytical model to relate the vertical root distribution to climate and soil properties: vertical root distribution, Geophysical Research Letters, 33(18), n/a-n/a, doi:10.1029/2006GL027331, 2006.

Law, B.: AmeriFlux US-Me2 Metolius mature ponderosa pine, doi:10.17190/AMF/1246076, 2002-.

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., McCabe, M. F. and Evans, J. P.: Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals, Hydrology and Earth System Sciences, 15(2), 425–436, doi:10.5194/hess-15-425-2011, 2011.

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., Evans, J. P. and van Dijk, A. I. J. M.: Trend-preserving blending of passive and active microwave soil moisture retrievals, Remote Sensing of Environment, 123, 280–297, doi:10.1016/j.rse.2012.03.014, 2012.

Manfreda, S., Caylor, K. K. and Good, S. P.: An ecohydrological framework to explain shifts in vegetation organization across climatological gradients: Vegetation pattern in dry environments, Ecohydrology, 10(3), e1809, doi:10.1002/eco.1809, 2017.

Manzoni, S., Vico, G., Katul, G., Palmroth, S. and Porporato, A.: Optimal plant water-use strategies under stochastic rainfall, Water Resources Research, 50(7), 5379–5394, doi:10.1002/2014WR015375, 2014.

McColl, K. A., Wang, W., Peng, B., Akbar, R., Short Gianotti, D. J., Lu, H., Pan, M. and Entekhabi, D.: Global characterization of surface soil moisture drydowns:, Geophysical Research Letters, 44(8), 3682–3690, doi:10.1002/2017GL072819, 2017.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: Equation of State Calculations by Fast Computing Machines, The Journal of Chemical Physics, 21(6), 1087–1092, doi:10.1063/1.1699114, 1953.

Miller, G. R., Baldocchi, D. D., Law, B. E. and Meyers, T.: An analysis of soil moisture dynamics using multi-year data from a network of micrometeorological observation sites, Advances in Water Resources, 30(5), 1065–1081, doi:10.1016/j.advwatres.2006.10.002, 2007.

Novick, K., Phillips, R.: AmeriFlux US-MMS Morgan Monroe State Forest, doi:10.17190/AMF/1246080, 1999-.

Porporato, A., Daly, E. and Rodriguez-Iturbe, I.: Soil water balance and ecosystem response to climate change, The American Naturalist, 164(5), 625–632, 2004.

Rawls, W. J., Brakensiek, D. L. and Saxtonn, K. E.: Estimation of soil water properties, Transactions of the ASAE, 25(5), 1316–1320, 1982.

Roberts, G. O. and Rosenthal, J. S.: Optimal Scaling for Various Metropolis–Hastings Algorithms, 1998.

Rodriguez-Iturbe, I., Gupta, V. K. and Waymire, E.: Scale considerations in the modeling of temporal rainfall, Water Resources Research, 20(11), 1611–1619, 1984.

Rodriguez-Iturbe, I., Porporato, A., Ridolfi, L., Isham, V. and Coxi, D. R.: Probabilistic modelling of water balance at a point: the role of climate, soil and vegetation, in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 455, pp. 3789–3805, The Royal Society., 1999.

Rodriguez-Iturbe, I., Porporato, A., Laio, F. and Ridolfi, L.: Intensive or extensive use of soil moisture: plant strategies to cope with stochastic water availability, Geophysical Research Letters, 28(23), 4495–4497, 2001.

Saleem, J. A. and Salvucci, G. D.: Comparison of soil wetness indices for inducing functional similarity of hydrologic response across sites in Illinois, Journal of Hydrometeorology, 3(1), 80–91, 2002.

Salvucci, G. D.: Estimating the moisture dependence of root zone water loss using conditionally averaged precipitation, Water resources research, 37(5), 1357–1365, 2001.

Suweis, S., Rinaldo, A., Van der Zee, S. E. A. T. M., Daly, E., Maritan, A. and Porporato, A.: Stochastic modeling of soil salinity, Geophysical Research Letters, 37(7), doi:10.1029/2010GL042495, 2010.

Teuling, A. J., Seneviratne, S. I., Williams, C. and Troch, P. A.: Observed timescales of evapotranspiration response to soil moisture, Geophysical Research Letters, 33(23), doi:10.1029/2006GL028178, 2006.

Thompson, S., Levin, S. and Rodriguez-Iturbe, I.: Linking Plant Disease Risk and Precipitation Drivers: A Dynamical Systems Framework, The American Naturalist, 181(1), E1–E16, doi:10.1086/668572, 2013.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D. and Peterson, G. D.: XSEDE: accelerating scientific discovery, Computing in Science & Engineering, 16(5), 62–74, 2014.

Tuttle, S. E. and Salvucci, G. D.: A new approach for validating satellite estimates of soil moisture using large-scale precipitation: Comparing AMSR-E products, Remote Sensing of Environment, 142, 207–222, doi:10.1016/j.rse.2013.12.002, 2014.

Volo, T. J., Vivoni, E. R., Martin, C. A., Earl, S. and Ruddell, B. L.: Modelling soil moisture, water partitioning, and plant water stress under irrigated conditions in desert urban areas, Ecohydrology, doi:10.1002/eco.1457, 2014.

Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E. and Ertl, M.: Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Annals), 7, 315–321, 2012.

Wang, T., Franz, T. E., Yue, W., Szilagyi, J., Zlotnik, V. A., You, J., Chen, X., Shulski, M. D. and Young, A.: Feasibility analysis of using inverse modeling for estimating natural groundwater recharge from a large-scale soil moisture monitoring network, Journal of Hydrology, 533, 250–265, doi:10.1016/j.jhydrol.2015.12.019, 2016.

Whitney, K. M., Vivoni, E. R., Duniway, M. C., Bradford, J. B., Reed, S. C. and Belnap, J.: Ecohydrological role of biological soil crusts across a gradient in levels of development, Ecohydrology, 10(7), e1875, doi:10.1002/eco.1875, 2017.

Xu, T., White, L., Hui, D. and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, Global Biogeochemical Cycles, 20(2), doi:10.1029/2005GB002468, 2006.

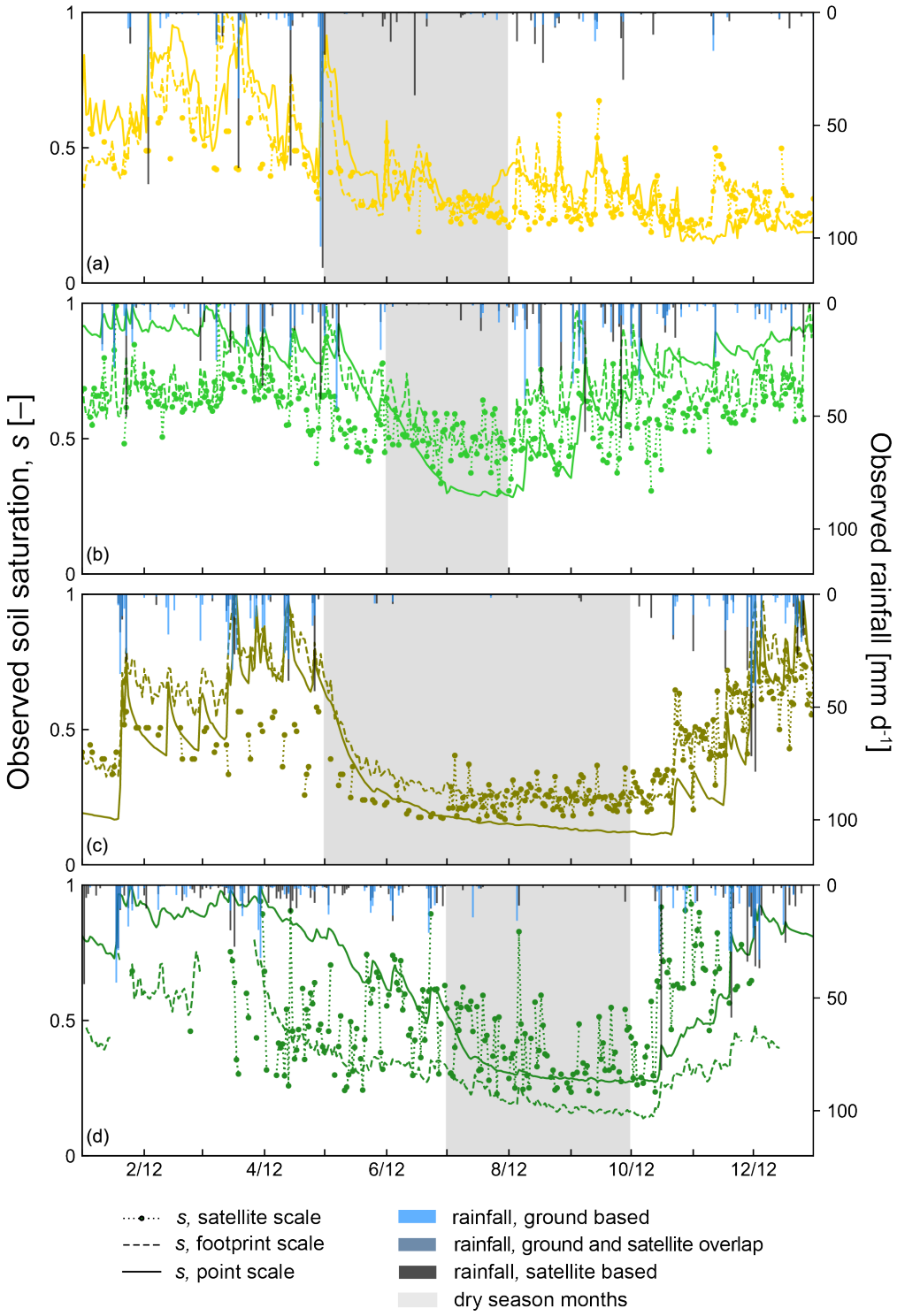

Figure 1: Soil saturation and rainfall time series from (a) US-ARM, (b) US-MMS, (c) US-Ton, and (d) US-Me2.

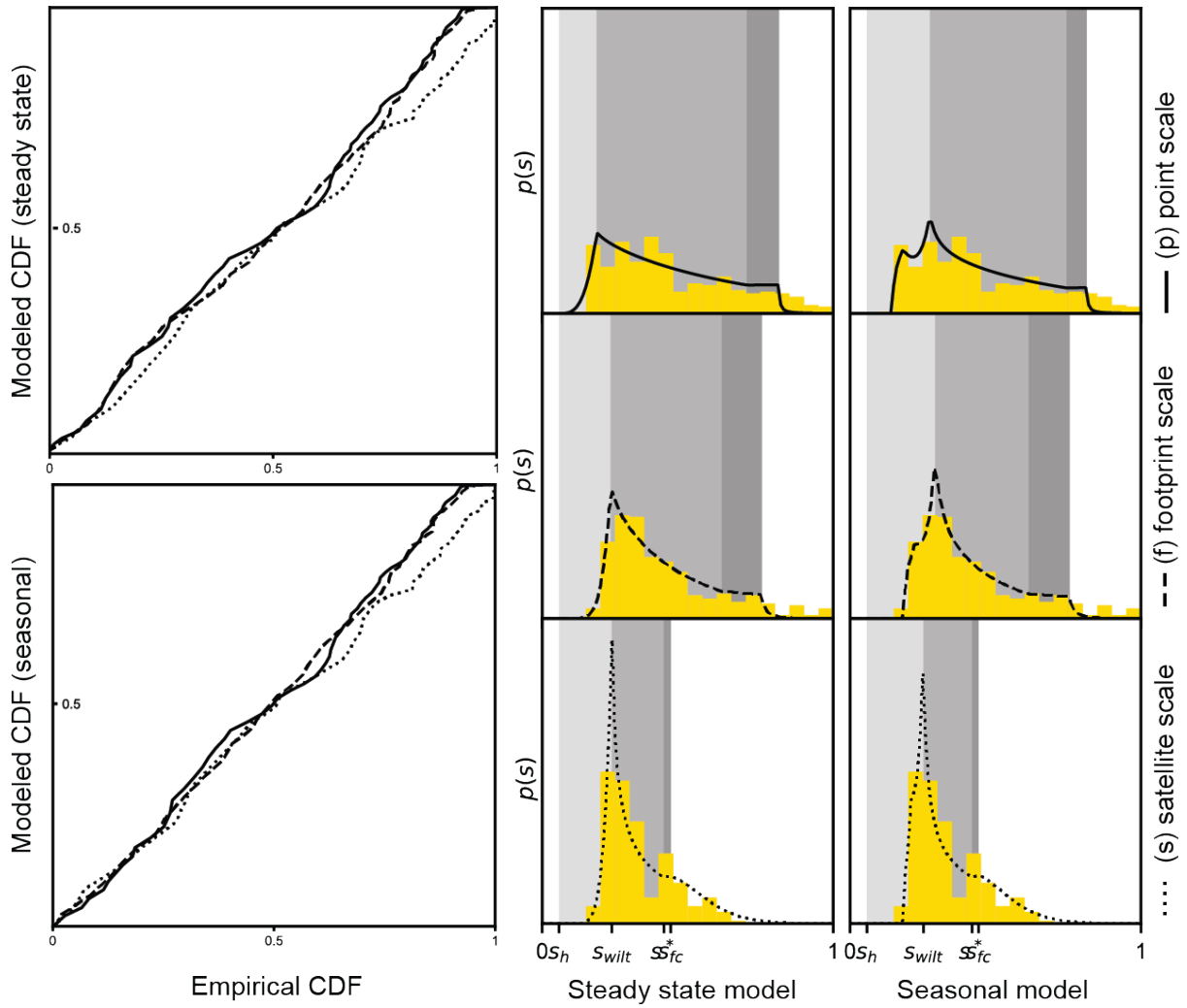

Figure 2: Empirical versus modelled cumulative density functions (CDF) and soil saturation probability distribution (p(s)) for US-ARM. The mean values of the posteriori parameter distributions were used with the analytical model in Eq (3) in the steady state model and Eq (6) in the seasonal model. The grey shaded areas correspond to the soil saturation thresholds (s_h, s_w, s^*, s_{fc}) in the water balance model.

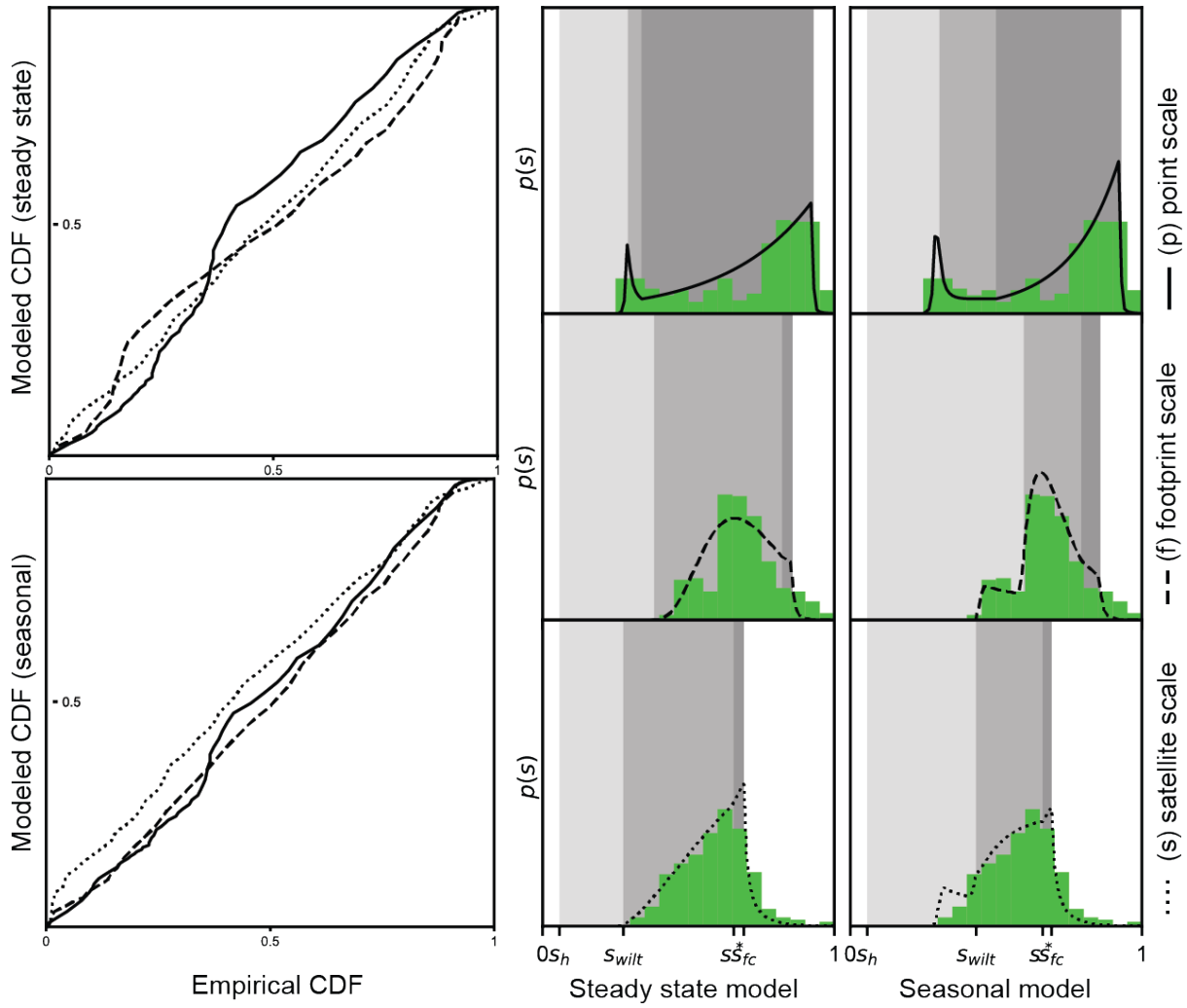

Figure 3: Empirical versus modelled cumulative density functions (CDF) and soil saturation probability distribution (p(s)) for US-MMS. The mean values of the posteriori parameter distributions were used with the analytical model in Eq (3) in the steady state model and Eq (6) in the seasonal model. The grey shaded areas correspond to the soil saturation thresholds (s_h, s_w, s^*, s_{fc}) in the water balance model.

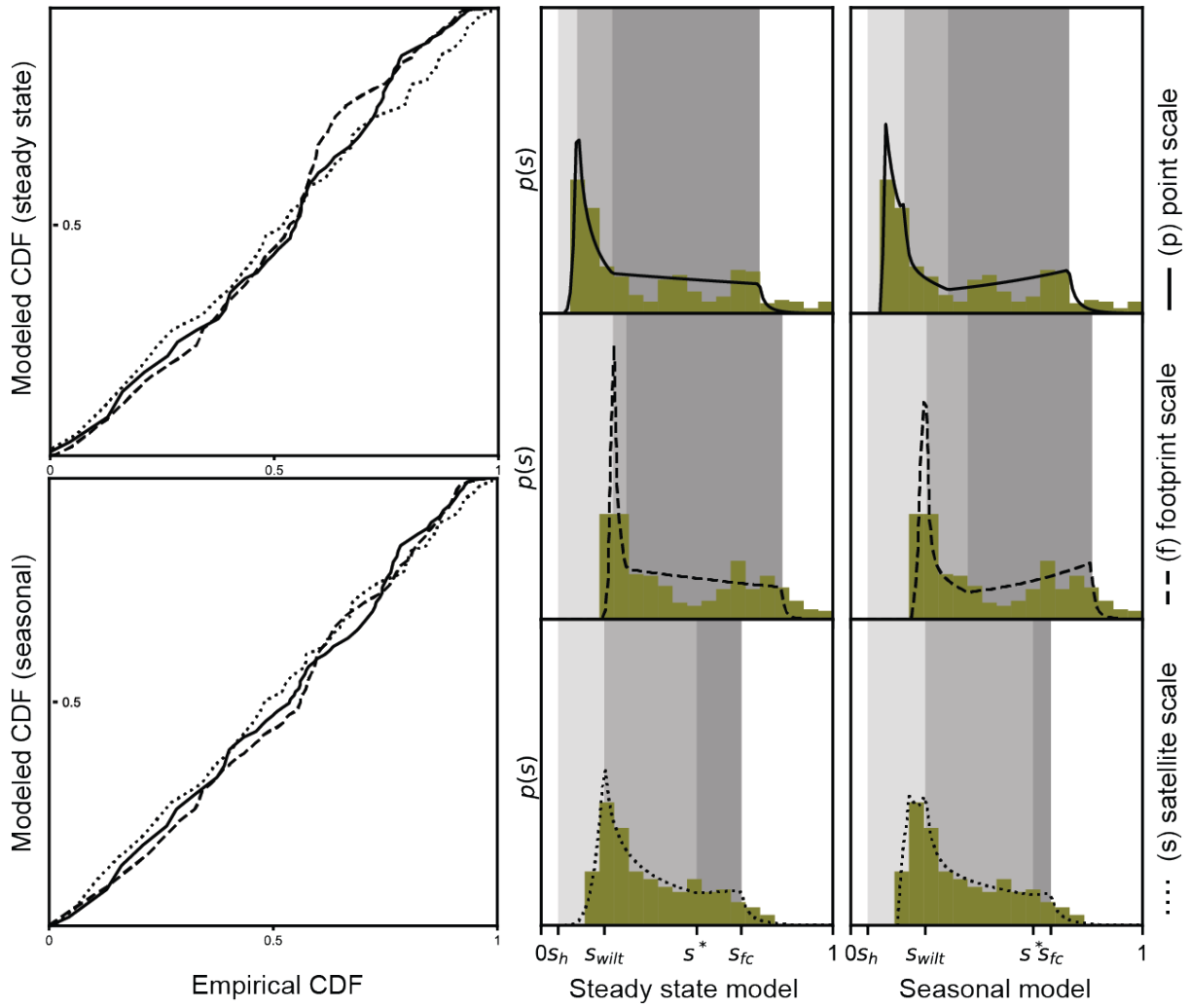

Figure 4: Empirical versus modelled cumulative density functions (CDF) and soil saturation probability distribution (p(s)) for US-Ton. The mean values of the posteriori parameter distributions were used with the analytical model in Eq (3) in the steady state model and Eq (6) in the seasonal model. The grey shaded areas correspond to the soil saturation thresholds (s_h, s_w, s^*, s_{fc}) in the water balance model.

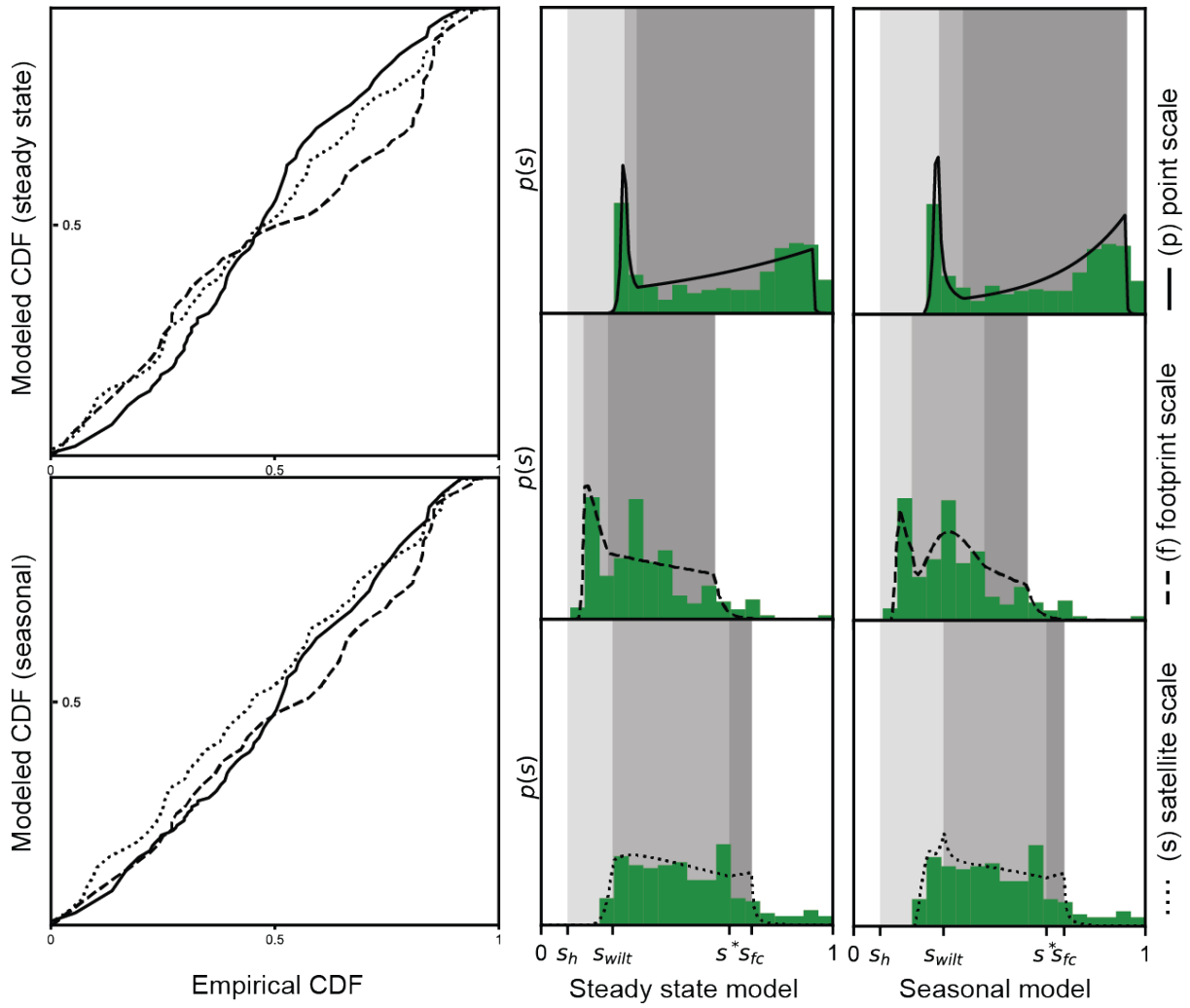

Figure 5: Empirical versus modelled cumulative density functions (CDF) and soil saturation probability distribution (p(s)) for US-Me2. The mean values of the posteriori parameter distributions were used with the analytical model in Eq (3) in the steady state model and Eq (6) in the seasonal model. The grey shaded areas correspond to the soil saturation thresholds (s_h, s_w, s^*, s_{fc}) in the water balance model.

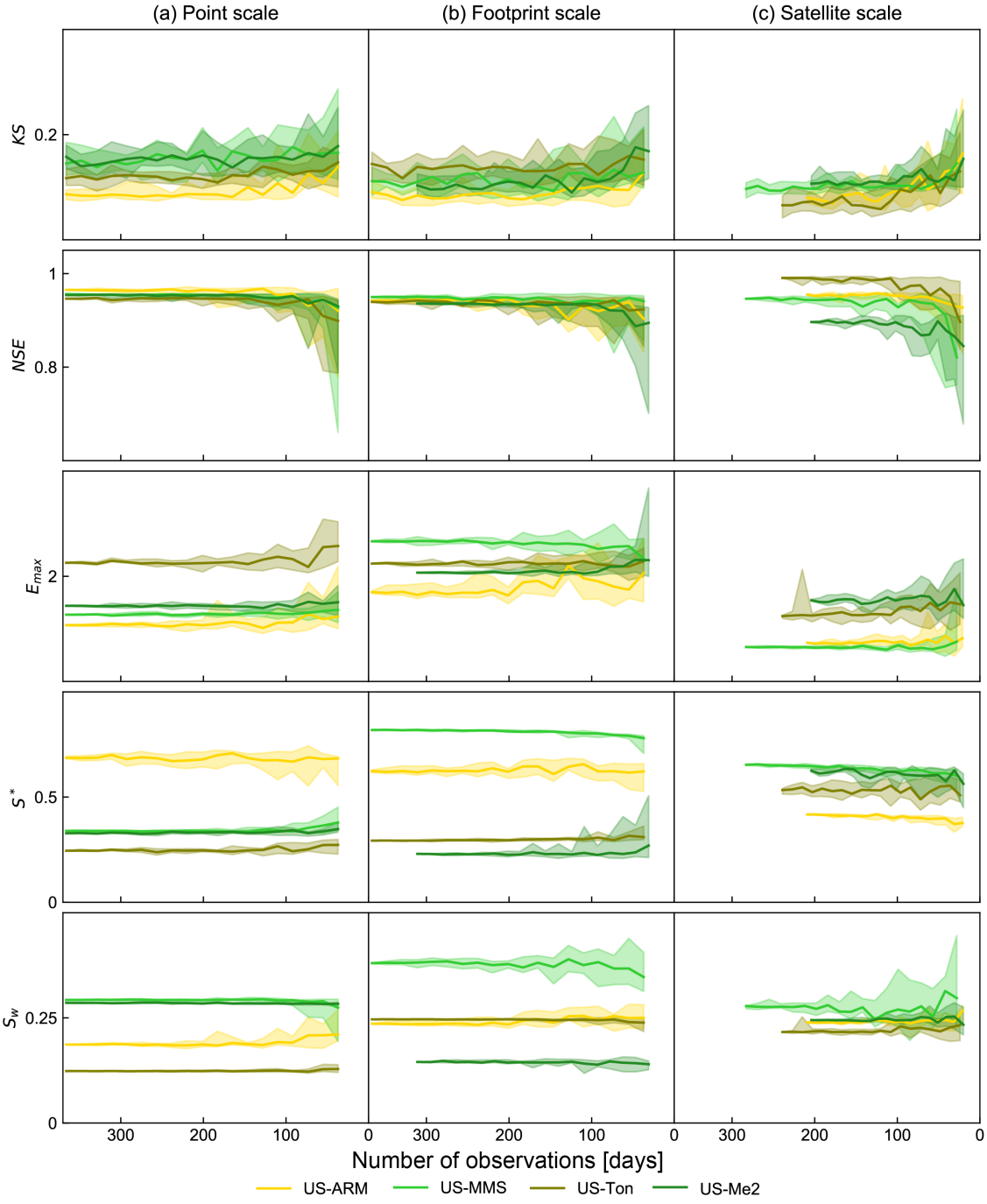

Figure 6 - Goodness of fit and ecohydrological parameters inferred with decreasing number of soil saturation observations (steady state model). For each subsample category, the median results of 10 repeats are plotted and results between the 90th and 10th percentiles are shaded. Colors correspond to the four sites in the legend. KS, Kolmogorov Smirnov statistic; NSE, quantile-level Nash Sutcliffe efficiency; E_{max}, maximum evapotranspiration in mm d^{-1}; s^*, point of incipient stomatal closure; s_w, wilting point.

Table 1 – Selected study sites

Site Name	ARM Southern Great Plains	Morgan Monroe State Forest	Tonzi Ranch	Metolius Mature Ponderosa Pine
FLUXNET2015 ID	US-ARM	US-MMS	US-Ton	US-ME2
COSMOS ID	15	27	32	38
Latitude	36.6058 (36.625)	39.3232 (39.375)	38.4316 (38.375)	44.4523 (44.375)
Longitude	-97.4888 (-97.375)	-86.4131 (-86.375)	-120.966 (-120.87)	-97.4888 (-97.375)
Elevation [m]	314	275	177	1253
Vegetation	Crops and grassland	Deciduous forest	Oak savanna	Ponderosa pine forest
Soil Texture	Loam	Loam	Loam	Sandy Loam
MAT [°C]	14.8	10.9	15.8	6.3
MAP [mm]	843	1032	559	523
α [mm day^{-1}]	21.0$^{(p, f)}$, 24.4$^{(s)}$	9.04$^{(p, f)}$, 11.8$^{(s)}$	9.3$^{(p, f)}$, 16.9$^{(s)}$	8.1$^{(p, f)}$, 11.6$^{s)}$
α_w [mm day^{-1}]	21.4$^{(p, f)}$, 26.8$^{(s)}$	9.1$^{(p, f)}$, 11.9$^{(s)}$	8.7$^{(p, f)}$, 16.7$^{(s)}$	7.9$^{(p, f)}$, 11.6$^{(s)}$
λ [day^{-1}]	0.05$^{(p, f)}$, 0.08$^{(s)}$	0.24$^{(p, f)}$, 0.20$^{(s)}$	0.22$^{(p, f)}$, 0.10$^{(s)}$	0.24 $^{(p, f)}$, 0.21$^{(s)}$
λ_w [day^{-1}]	0.07$^{(p, f)}$, 0.08$^{(s)}$	0.27$^{(p, f)}$, 0.23$^{(s)}$	0.39$^{(p, f)}$, 0.17$^{(s)}$	0.31$^{(p, f)}$, 0.27$^{(s)}$
t_d [days]	92	61	153	92
***n* [-]**	0.35$^{(p)}$, 0.34$^{(f)}$, 0.46$^{(s)}$	0.46$^{(p)}$, 0.66$^{(f)}$, 0.43$^{(s)}$	0.53$^{(p)}$, 0.39$^{(f)}$, 0.43$^{(s)}$	0.36$^{(p)}$, 0.59$^{(f)}$, 0.41$^{(s)}$
K_s [mm day^{-1}]	317	317	317	622
***b* [-]**	4.55	4.55	4.55	3.11
s_h [-]	0.06	0.06	0.06	0.09
s_{fc} [-]	0.81$^{(p)}$, 0.75$^{(f)}$, 0.44$^{(s)}$	0.93$^{(p)}$, 0.86$^{(f)}$, 0.69$^{(s)}$	0.75$^{(p)}$, 0.83$^{(f)}$, 0.69$^{(s)}$	0.94$^{(p)}$, 0.60$^{(f)}$, 0.72$^{(s)}$
s_{min} [-]	0.15$^{(p)}$, 0.19$^{(f)}$, 0.19 $^{(s)}$	0.28$^{(p)}$, 0.44$^{(f)}$, 0.30 $^{(s)}$	0.11$^{(p)}$, 0.22$^{(f)}$, 0.17 $^{(s)}$	0.27$^{(p)}$, 0.14$^{(f)}$, 0.23 $^{(s)}$
s_{max} [-]	1.0$^{(p)}$, 1.0$^{(f)}$, 0.67 $^{(s)}$	1.0 $^{(p)}$, 1.0 $^{(f)}$, 1.0 $^{(s)}$	1.0$^{(p)}$, 1.0$^{(f)}$, 0.80 $^{(s)}$	1.0$^{(p)}$, 1.0$^{(f)}$, 1.0$^{(s)}$
Mean *s* [-]	0.44$^{(p)}$, 0.42$^{(f)}$, 0.33 $^{(s)}$	0.71$^{(p)}$, 0.68$^{(f)}$, 0.59 $^{(s)}$	0.38$^{(p)}$, 0.49$^{(f)}$, 0.38 $^{(s)}$	0.64$^{(p)}$, 0.35$^{(f)}$, 0.50 $^{(s)}$
Standard deviation *s* [-]	0.21$^{(p)}$, 0.19$^{(f)}$, 0.11 $^{(s)}$	0.21$^{(p)}$, 0.11$^{(f)}$, 0.12 $^{(s)}$	0.25$^{(p)}$, 0.23$^{(f)}$, 0.17 $^{(s)}$	0.25 $^{(p)}$, 0.16$^{(f)}$, 0.18 $^{(s)}$

Latitude and longitude in parenthesis correspond the centroid of the satellite area associated with the site location; MAT, mean annual temperature from long-term FLUXNET2015 data; MAP, mean annual precipitation from long-term FLUXNET2015 data; soil texture taken from the HWSD; *n,* porosity; K_s, saturated soil hydraulic conductivity; *b,* pore size distribution index; s_h, hydroscopic point; s_{fc}, field capacity; α, observed average daily rainfall depth in 2012, the subscript *w* indicates that α was computed for only the wet season months; λ, observed average daily rainfall frequency in 2012, the subscript *w* indicates that λ was computed for only the wet season months;; t_d, number of days in the dry season; superscripts $^{(p)}$, $^{(f)}$, and $^{(s)}$ correspond to values used for the point-, footprint-, and satellite scale analysis. Citations for each FLUXNET2015 site: Sebastien Biraud (2002–) AmeriFlux US-ARM ARM Southern Great Plains site- Lamont, 10.17190/AMF/1246027; Kim Novick, Rich Phillips (1999–) AmeriFlux US-MMS Morgan Monroe State Forest, 10.17190/AMF/1246080; Bev Law (2002–) AmeriFlux US-Me2 Metolius mature ponderosa pine, 10.17190/AMF/1246076; Dennis Baldocchi (2001–) AmeriFlux US-Ton Tonzi Ranch, 10.17190/AMF/1245971

Table 2 Estimated ecohydrological parameters and goodness of fit of analytical soil saturation pdfs

Site name	Scale	N		NSE		KS		E_{max}		s^*		s_w	
		p	p_{wd}	p	p_{wd}	p	p_{wd}	p	p_{wd}	p	p_{wd}	p	p_{wd}
US-ARM	point	4	4	0.96	0.96	0.07	0.07	1.1 (11)	1.3 (14)	0.7 (8)	0.74 (5)	0.19 (4)	0.27 (7)
	footprint	3	3	0.94	0.94	0.08	0.06	1.7 (11)	2 (12)	0.62 (7)	0.61 (9)	0.24 (3)	0.29 (2)
	satellite	3	3	0.96	0.97	0.08	0.09	0.7 (13)	0.5 (13)	0.42 (4)	0.42 (4)	0.24 (3)	0.25 (2)
US-Ton	point	3	4	0.95	0.97	0.09	0.08	2.3 (4)	1.9 (10)	0.24 (6)	0.33 (7)	0.12 (1)	0.18 (6)
	footprint	3	3	0.94	0.98	0.13	0.08	2.2 (3)	1.8 (8)	0.29 (2)	0.4 (10)	0.25 (0)	0.26 (1)
	satellite	3	9	0.99	0.99	0.06	0.07	1.2 (15)	1 (13)	0.53 (12)	0.62 (6)	0.22 (3)	0.26 (3)
US-MMS	point	3	4	0.96	0.98	0.12	0.08	1.3 (3)	1.1 (6)	0.34 (3)	0.5 (8)	0.29 (0)	0.31 (2)
	footprint	3	3	0.95	0.95	0.13	0.08	2.7 (6)	4.5 (10)	0.82 (2)	0.79 (3)	0.38 (5)	0.59 (1)
	satellite	3	6	0.95	0.88	0.1	0.14	0.7 (8)	0.9 (10)	0.65 (4)	0.66 (3)	0.28 (9)	0.43 (2)
US-Me2	point	3	8	0.95	0.97	0.16	0.1	1.4 (3)	1.1 (7)	0.33 (3)	0.37 (8)	0.29 (0)	0.29 (1)
	footprint	3	6	0.94	0.94	0.09	0.1	2.1 (2)	2.9 (10)	0.23 (4)	0.45 (5)	0.15 (2)	0.2 (6)
	satellite	3	4	0.89	0.89	0.12	0.1	1.6 (12)	1.4 (15)	0.64 (8)	0.66 (8)	0.25 (3)	0.31 (4)

Values in parenthesis correspond to the coefficient of variation of the posteriori parameter estimates in percentage.
p, analytical model for the soil saturation pdf without seasons, p_{wd}, analytical model for the soil saturation pdf including wet and dry seasons; N, number of needed to obtain 3 converging results (see Sect. 2.3.2); NSE, quantile-level Nash Sutcliffe efficiency; KS, Kolmogorov Smirnov statistic; E_{max}, maximum ev d^{-1} (the weighted average wet and dry season E_{max} is reported for the p_{wd} model) ; s^*, point of incipient stomatal closure; s_w, wilting point.