

A changepoint approach to modelling non-stationary soil moisture dynamics

Mengyi Gong^{*1}, Rebecca Killick¹, Christopher Nemeth¹, and John Quinton²

¹School of Mathematical Sciences, Lancaster University, Lancaster, U.K.

²Lancaster Environment Centre, Lancaster University, Lancaster, U.K.

Abstract

Soil moisture dynamics provide an indicator of soil health that scientists model via drydown curves. The typical modelling process requires the soil moisture time series to be manually separated into drydown segments and then exponential decay models are fitted to them independently. Sensor development in recent years means that experiments that were previously conducted over a few field campaigns can now be scaled to months or years at a higher sampling rate. To better meet the challenge of increasing data size, this paper proposes a novel changepoint-based approach to automatically identify structural changes in the soil drying process and simultaneously estimate the drydown parameters that are of interest to soil scientists. A simulation study is carried out to demonstrate the performance of the method in detecting changes and retrieving model parameters. Practical aspects of the method such as adding covariates and penalty learning are discussed. The method is applied to hourly soil moisture time series from the NEON data portal to investigate the temporal dynamics of soil moisture drydown. We recover known relationships previously identified manually, alongside delivering new insights into the temporal variability across soil types and locations.

Keywords: soil moisture drydown, changepoint detection, PELT, temporal dynamics

1 Introduction

Healthy soil plays a critical, yet underappreciated, role in storing and filtering water, sustaining biodiversity, maintaining food production, and mitigating climate change through soil organic carbon sequestration (Lehmann *et al.*, 2020; Frelih-Larsen *et al.*, 2022). It is estimated that nearly 80% of carbon in the terrestrial ecosystems of the planet is found in soil (Ontl & Schulte, 2012). Soil water is an important component of soil health, crucial to the supply of water to plants, and is therefore fundamental to agricultural production. It is also a key component in the hydrological cycle, regulating the recharge of groundwater and the flow of water to surface water bodies - both are critical for ecosystem function and human health (McCull *et al.*, 2017). In addition, soil moisture is intricately connected to large-scale climate models (McCull *et al.*, 2017; Salvia *et al.*, 2018). For example, the soil moisture observations and the identified drydowns were used to evaluate different processes as well as calibrate the associated parameters in the ORCHIDEE (<https://orchidee.ipsl.fr/>) land surface model (Raoult *et al.*, 2021).

^{*}Corresponding address: School of Mathematical Sciences, Lancaster University, Lancaster, U.K., LA1 4YF, E-mail: m.gong1@lancaster.ac.uk

1.1 Soil moisture drydown analysis

The last few decades have seen a growth in research on soil moisture dynamics (Vereecken *et al.*, 2014). Soil drydown modelling is one area that has drawn the attention of scientists as more data from underground sensors and satellites have become available. According to McColl *et al.* (2017), the dynamics of soil moisture is governed by the vertically integrated water budget. Typically, water in the soil comes from precipitation and is lost via drainage, runoff, and evapotranspiration (i.e., the water loss through plant transpiration and soil and plant evaporation). These three pathways consist of the water loss function. The water loss function is encoded in the shape of the soil moisture drydown curve: the soil moisture time series directly following a precipitation event, during which there is no further infiltration input, e.g., further rainfall, at the soil surface. The soil moisture drydown curve, which is identified from a soil moisture time series, is usually modeled as an exponential decay process (McColl *et al.*, 2017; Shellito *et al.*, 2018; Salvia *et al.*, 2018),

$$\theta(t) = \Delta\theta \exp\left(-\frac{t}{\omega}\right) + \theta_f \quad (1)$$

where $\theta(t)$ represents the soil moisture content at time t of the current drydown period, $\Delta\theta$ is the increase of soil moisture content from the end of the last drydown period, θ_f is the estimated lower bound of the soil moisture observations (McColl *et al.*, 2017; Shellito *et al.*, 2016), and ω reflects the exponential drying rate of the soil and is sometimes referred to as the ‘temporal e-folding decay’ of soil moisture. The temporal (and spatial) variation of the e-folding decay parameter ω is one of the aspects that soil scientists are interested in. For example, Salvia *et al.* (2018) and Ruscica *et al.* (2020) investigated the seasonal dynamics and the spatial patterns in the e-folding decay in southeastern South America, leading to a discussion on the importance of effective sampling frequency. Dong *et al.* (2022) used the drying rate (which is related to the e-folding decay) estimated from the drydown models to investigate the transition of evapotranspiration regimes. Both Tso *et al.* (2023) and Sinha *et al.* (2024) aimed to develop a way of validating various soil moisture products, e.g., satellite data, in-situ data, and data generated from statistical and hydrological models, using the estimated decay parameters and their seasonal patterns.

Typically, soil drydown modelling requires the soil moisture time series to be manually separated into segments representing the drydown process. These segments are referred to as drydown curves. An exponential decay model is fitted to each of them (McColl *et al.*, 2017; Salvia *et al.*, 2018). The results are often snapshot views of the drydown property, characterized by the temporal e-folding decay parameter, or in an inter-seasonal study, a set of snapshots reflecting the temporal variations of different seasons (Salvia *et al.*, 2018; Sehgal *et al.*, 2021). Advancements in sensor technology allow scientists to obtain higher frequency time series of soil moisture over longer periods at lower costs, meaning it is now possible to monitor the changes in soil properties through time. Ecological observatories, such as NEON (National Ecological Observatory Network, <https://data.neonscience.org/>), are collecting large volumes of such data offering the potential to ask questions about how soil properties, such as those associated with drydown curves, change dynamically. The large volumes of data and continuous monitoring present challenges to conventional modelling approaches that rely on the manual extraction of soil dynamics.

Motivated by the current practice of data segmentation, this paper proposes a novel changepoint-based approach to automatically identify the drydown patterns in the soil drying process. In

a nutshell, changes caused by sudden increases in soil moisture over a long time series are captured automatically, and the parameters characterizing the drying processes following the sudden increases are estimated simultaneously. Specifically, each segment following a changepoint is modeled using an exponential decay model similar to model (1) with segment-specific parameters. This allows the model to capture the temporal variations in the drying process and complements conventional soil drydown modelling. It requires little data pre-processing and can be applied to a soil moisture time series directly, which is attractive when working with large data sets. To identify changepoints, we extend the penalised exact linear time (PELT) method (Killick *et al.*, 2012) to estimate the soil moisture model parameters simultaneously.

Similar patterns to soil moisture time series (Figure 1) are seen in calcium imaging data in neuroscience. A key problem in computational neuroscience is the inference of the exact times the neuron spiked based on the noisy calcium fluorescence trace. Jewell & Witten (2018) treat this problem as a changepoint problem and improvements to their initial approach are made in Jewell *et al.* (2020). There are major differences between the neuroscience and soil moisture problems. Whereas neuroscientists are interested in the timing of the spikes, soil scientists are interested in the characteristics of the exponential drying process. Thus whilst the decay parameter is a nuisance in calcium imaging and is integrated out in Jewell *et al.* (2020), the decay is a key parameter characterizing the drying of the soil. Furthermore, the decay parameter in soil may display temporal dynamics as suggested in Salvia *et al.* (2018); Sehgal *et al.* (2021) so there is interest in monitoring it across segments. Alongside this, within the soil moisture drydown models, we may wish to include covariate information e.g., precipitation, temperature, or vegetation, that may allow us to improve the segmentation and better model the individual segments.

1.2 NEON data

The development of the changepoint-based modelling approach is motivated by the soil moisture time series from NEON soil water and salinity data product (NEON, 2021). NEON, or National Ecological Observatory Network, is a continental-scale observation facility designed to collect long-term open-access ecological data to better understand how U.S. ecosystems are changing. The soil moisture data were collected using the Sentek TriSCAN sensors and available as 1-minute and 30-minute interval time series products (Ayres & Roberti, 2018). Figure 1 shows some examples of the hourly (sub-sampled) soil moisture time series recorded at different field sites, including the time series from June 2018 to June 2020 at Santa Rita Experimental Range (SRER) in Arizona, the time series from July 2017 to November 2018 at Smithsonian Environmental Research Center (SERC) in Maryland and the time series from February 2018 to January 2019 at Talladega National Forest (TALL) in Alabama. All these time series display sudden increases in soil moisture followed by drying processes at potentially different decay speeds over time. Modelling the drydown characteristics at the three sites has the potential to highlight differences between the soils and their response to different climates and vegetation. Manually identifying and extracting the drydown curves would be very laborious and difficult to deploy rapidly to further sites. The changepoint-based approach we propose provides a solution. The estimated changepoints and parameters provide a dynamic summary of the data, as opposed to the static view from conventional drydown analyses. We hypothesize that, with a sufficiently long time series, the method has the potential to identify long-term changes in the drydown parameters, which could be essential to the study of changing soil health.

The remainder of the paper is divided into five sections. Section 2 introduces the proposed

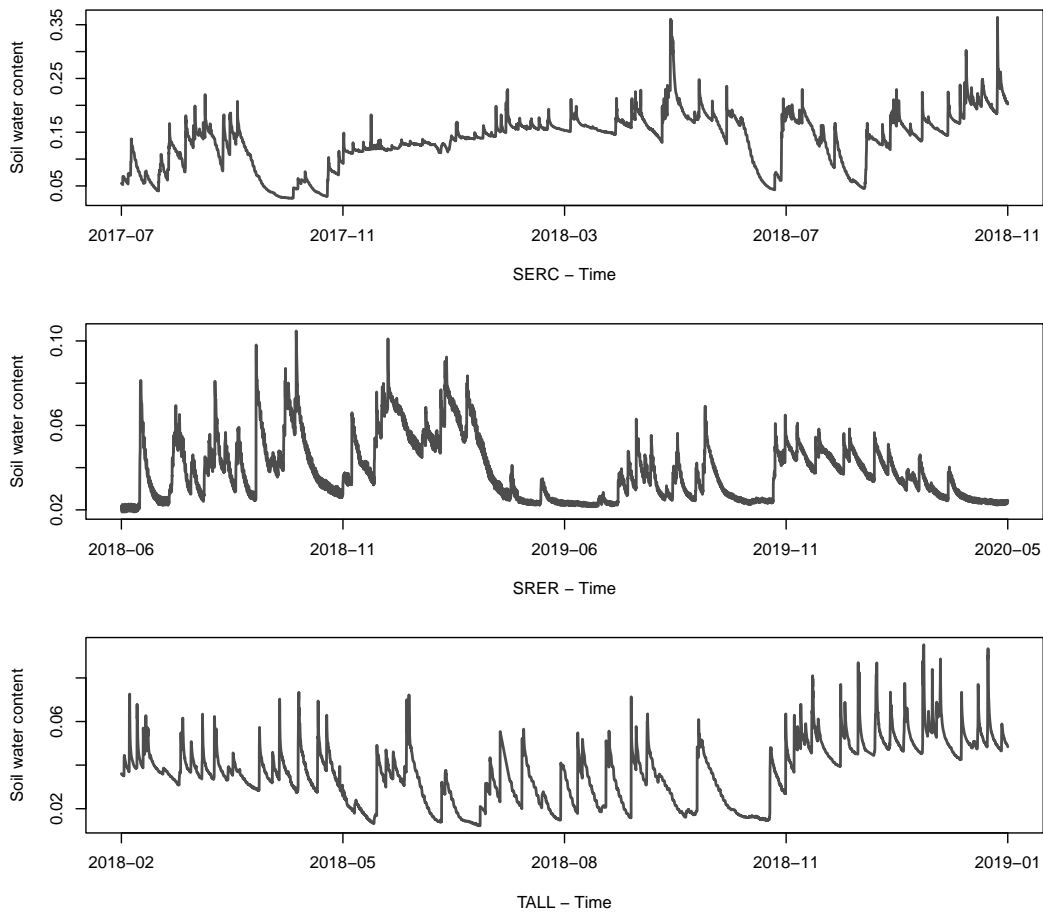


Figure 1: Hourly soil volumetric water content time series recorded at field sites SERC, SRER and TALL from the NEON data portal.

change point-based method to model soil moisture time series and the algorithm to implement the method. Section 3 presents the simulation study to assess the performance of the method. Section 4 discusses the practical aspects of the method, including the use of covariates and penalty learning. Section 5 applies the method to the soil moisture time series from the NEON data portal. Section 6 concludes the paper and discusses some directions for future work.

2 The change point method for soil moisture time series

2.1 The proposed model for soil moisture dynamics

Denote the observed soil moisture content at time t by Y_t , and denote the set of time points right before the sudden increases in soil moisture content by $\tau_i, i = 1, \dots, k$. The dynamics in soil moisture can be described using

$$\begin{aligned} Y_t &= \theta_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \\ \theta_t &= \phi_t \theta_{t-1} + \Delta_t, \end{aligned} \tag{2}$$

where θ_t is the underlying soil moisture at time t , ϕ_t ($0 < \phi_t < 1$) is the soil moisture decay parameter, and Δ_t is the increase in soil moisture, which takes a positive value at $t = \tau_i + 1$, $i = 1, \dots, k$ and 0 otherwise. The second equation in model (2) translates to an exponential decay model for the segment between two sudden increases. That is, for $t \in (\tau_i + 1, \tau_{i+1})$, where $\tau_i + 1$ is the start of the current segment and τ_{i+1} is the end of the current segment,

$$Y_t = \alpha_{0i} + \alpha_{1i} \phi_i^{(t-\tau_i)} + \epsilon_t. \quad (3)$$

Here α_{0i} ($\alpha_{0i} > 0$) is the asymptotic soil moisture content and $\alpha_{0i} + \alpha_{1i}$ ($\alpha_{1i} > 0$) is the peak soil moisture content right after the end of the last segment at τ_i . Unlike the decay of calcium concentration, soil moisture decreases at different speeds over different periods. This is a result of the temporal variation in the elements that affect the speed at which the soil loses water, e.g., temperature and vegetation. It carries interesting information on soil moisture dynamics. To reflect this feature, a segment specific decay parameter is used in the model, i.e., $\phi_t = \phi_i$ for $t \in (\tau_i + 1, \tau_{i+1})$, $i = 1, \dots, k$. Depending on the properties of the soil, the asymptotic soil moisture can be fixed throughout a time series or be segment-specific.

To make the fitting of the exponential decay model (3) easier, a re-parameterisation $\phi_i = \exp\{-\exp(\gamma_i)\}$ is used, giving

$$Y_t = \alpha_{0i} + \alpha_{1i} \exp\{-\exp(\gamma_i)(t - \tau_i)\} + \epsilon_t. \quad (4)$$

This removes the constraint on ϕ_i so that $\exp\{-\exp(\gamma_i)\} \in (0, 1)$ for all $\gamma_i \in \mathbb{R}$. Note that γ_i is essentially a reparameterisation of the e-folding decay parameter ω in the soil drydown model (1). In other words, $1/\exp(\gamma)$ is equivalent to ω if t in model (1) and (4) have the same unit.

The parameters of the exponential decay model (4) can be estimated by minimising the non-linear least square (NLS) fit. Iterative algorithms, such as Gauss-Newton, Newton-Raphson, and Levenberg-Marquardt can be used to solve the optimisation problem (Bates & Watts, 1988). These algorithms can be implemented using various R (R Core Team, 2023) functions, e.g., the `nls` function which implements the Gauss-Newton algorithm and the `port` algorithm (Dennis *et. al.*, 1988), or `nlfm` function from package `nlmrt` (Nash, 2016) which uses the Nash variant of the Levenberg-Marquardt algorithm (Nash, 2014).

The negative log-likelihood of the estimated model (4) is used as the cost function of the changepoint detection problem. Note that the cost function is a function of multiple parameters, α_{0i} , α_{1i} and γ_i . As a result, functional pruning, which was developed for uni-parameter cost functions, is not an appropriate choice (Rigaill, 2015; Maidstone *et. al.*, 2017). The additional effort required to identify the multi-dimensional region where the multivariate cost function attains its minimum at each step undermines the computational efficiency of functional pruning. Therefore, this paper chooses to develop a changepoint detection procedure based on the penalised exact linear time (PELT) method (Killick *et. al.*, 2012). The PELT method is flexible and computationally efficient, making it suitable for the soil moisture time series, which typically consists of 10,000 to 20,000 time points. Details of the PELT method and its applicability to the problem in this paper are described in section 2.2.

In addition, lower and upper limits may be applied to the asymptotic soil moisture parameter α_{0i} and the increase parameter α_{1i} , to ensure valid soil moisture values and a positive increase. The positive constraint on the increase of soil moisture requires $\alpha_{0i} + \alpha_{1i} > \alpha_{0i+1}$. These constraints on parameters are treated as lower and upper bounds in NLS optimisation.

2.2 Model estimation using PELT

The optimisation goal is to identify the set of changepoints, $0 = \tau_0 < \tau_1 < \dots < \tau_k < \tau_{k+1} = n$, that minimises the overall penalised cost function

$$\sum_{i=0}^k \mathcal{C}(Y_{(\tau_i+1):\tau_{i+1}}) + \lambda f(k), \quad (5)$$

where the cost function $\mathcal{C}(Y_{(\tau_i+1):\tau_{i+1}})$ is

$$(\tau_{i+1} - \tau_i) \left\{ \log(2\pi) + 1 + \sum_{t=\tau_i+1}^{\tau_{i+1}} \{Y_t - \hat{\alpha}_{0i} - \hat{\alpha}_{1i} \exp(-\exp(\hat{\gamma}_i)(t - \tau_i))\}^2 \right\},$$

which is twice the negative log-likelihood of the exponential decay model (4) fitted to the segment $Y_{(\tau_i+1)}, \dots, Y_{\tau_{i+1}}$, and the penalty function is $f(k) = k$, the number of changepoints. Other types of penalties are available, e.g., the modified Bayesian information criterion (BIC) (Zhang & Siegmund, 2007).

The PELT method by Killick *et al.* (2012) starts with the recursive computation of the overall cost function of the data up to time point s , $s = 1, \dots, n$. Denoted this cost function as $F(s)$, the recursion is

$$F(s) = \min_{\boldsymbol{\tau} \in \mathcal{T}_s} \left\{ \sum_{i=0}^m \mathcal{C}(Y_{(\tau_i+1):\tau_{i+1}}) + \lambda k \right\} = \min_{0 \leq \tau < s} \{F(\tau) + \mathcal{C}(Y_{(\tau+1):s}) + \lambda\},$$

where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$, \mathcal{T}_s is the set of $\{\boldsymbol{\tau} : 0 = \tau_0 < \tau_1 < \dots < \tau_m < \tau_{m+1} = s\}$, and τ is the last changepoint before s . Instead of searching through all candidate time points $0 \leq \tau < s$ for the optimal solution to τ , the algorithm prunes the candidate time points that can never be the last optimal changepoint for data $Y_{1:s}$, and searches only within a reduced set of candidate time points. Specifically, the pruning criterion (Killick *et al.*, 2012) is, for all $t < t' < s$ satisfying

$$\mathcal{C}(Y_{(t+1):t'}) + \mathcal{C}(Y_{(t'+1):s}) + K \leq \mathcal{C}(Y_{(t+1):s}) \quad (6)$$

for some constant K , the time point t can never be the last optimal changepoint prior to time point s if

$$F(t) + \mathcal{C}(Y_{(t+1):t'}) + K \geq F(t'). \quad (7)$$

Consequently, all time points t that satisfy condition (7) can be removed from the search and the computational cost is reduced.

Under the i.i.d. normal distribution assumption of ϵ_t in the exponential decay model (4), two times the negative log-likelihood of the model satisfies the inequality (6) with $K = 0$ (Killick *et al.*, 2012). Intuitively, consider adding a changepoint at t' between t and s while keeping the parameters in the exponential decay model unchanged. This gives the equivalence condition of (6) with $K = 0$. Any updated parameters that reduce either $\mathcal{C}(Y_{(t+1):t'})$ or $\mathcal{C}(Y_{(t'+1):s})$ will reduce the overall cost. Hence, a changepoint detection procedure based on PELT can be established, which is depicted in Algorithm 1.

Note that the set S_t in Algorithm 1 contains the changepoints that cannot be pruned due to the minimum segment length. Sometimes, the non-linear least square estimation of the model (4) does not produce a converged result. In these situations, the cost function of the segment is set to a very large value to represent an infinite cost, and the PELT iteration is modified slightly.

Algorithm 1: PELT for detecting changepoints in soil moisture time series

Input : data Y_1, \dots, Y_n ;
cost function $\mathcal{C}(Y_{t:s})$;
minimum segment length l ($l \geq 3$);
penalty parameter λ ;
constant K that satisfies the inequality (6);

Initialise ;
set $F(0) = -\lambda$;
for $t = 1, \dots, 2l - 1$ **do**
 1. fit the exponential decay model to $Y_{1:t}$ and compute the cost function
 $F(t) = \mathcal{C}(Y_{1:t}) + \lambda$;
 2. initialise the set of changepoint $\text{cpt}(t) = \{0\}$;
 3. initialise the set of candidate changepoints $R_t = \{0\}$;
set $R_{2l} = \{0, l\}$;

Iterate ;
for $t \leftarrow 2l$ **to** n **do**
 1. fit the exponential decay model to $Y_{(\tau+1):t}$, $\tau \in R_t$, and compute
 $F(t) = \min_{\tau \in R_t} \{F(\tau) + \mathcal{C}(Y_{(\tau+1):t}) + \lambda\}$;
 2. find $\tau^* = \arg \min_{\tau \in R_t} \{F(\tau) + \mathcal{C}(Y_{(\tau+1):t}) + \lambda\}$;
 3. update $\text{cpt}(t) = \text{cpt}(\tau^*) \cup \{\tau^*\}$;
 4. update $R_{t+1} = \{t - l + 1\} \cup \{\tau \in R_t : F(\tau) + \mathcal{C}(Y_{(\tau+1):t}) + K \leq F(t)\} \cup S_t$, where
 $S_t = \{\tau \in R_s, s = t, \dots, t - l + 2 : F(\tau) + \mathcal{C}(Y_{(\tau+1):s}) + K > F(s)\}$;

Output: a set of changepoints $\text{cpt}(n) = \{0, \tau_1, \dots, \tau_k, n\}$;
estimated exponential decay model parameters $\hat{\alpha}_{0i}, \hat{\alpha}_{1i}, \hat{\gamma}_i, i = 1, \dots, k + 1$.

During the iteration, both $F(\tau)$ and $\mathcal{C}(Y_{(\tau+1):t})$ could be infinite for a candidate changepoint τ . When the “historical” cost $F(\tau)$ is infinite, then τ can never be the last optimal changepoint prior to t and hence it should be pruned. On the contrary, when $F(\tau)$ is finite, but $\mathcal{C}(Y_{(\tau+1):t})$ is infinite, there is a possibility that the model fitted to the segment starting from $\tau + 1$ will converge when more observations are added to the segment. That is, $\mathcal{C}(Y_{(\tau+1):(t+\delta)})$ may be finite for some $\delta \geq 1$. Therefore, no pruning is applied to those τ values and they are all kept for the next iteration. The modified iteration is given in the supplemental document.

3 Simulation study

To investigate the performance of the method developed above, a simulation study is carried out. Two problems of particular interest in this case are, (1) whether the algorithm can identify the locations of the sudden increases in the time series in different scenarios, and (2) whether the method can produce a reasonable estimation of the model parameters, in particular γ .

3.1 Simulation design

The observed soil moisture time series sometimes display temporal patterns in the frequency of the sudden increases, e.g., more frequent increases during the rainy summer season than the dry winter period. The sudden increases in soil moisture may also appear at very different scales.

For example, there can be a series of smaller increases during a long large-scale drying process as shown in the figure in the supplemental document. Based on these features, three scenarios are considered in terms of the frequency of the sudden increases, (S1) randomly distributed over time, (S2) following a temporal pattern where one part of the time series has more frequent increases than the rest, (S3) large scale sudden increases randomly distributed over time, along with small scale increases over a long drying period. Each of the scenarios will be paired with two noise levels, giving six scenarios in total. For convenience, we use “S1a” to refer to scenario S1 with small noise level, “S2b” to refer to scenario S2 with large noise level, so on and so forth.

There may be temporal variation in the drying rate in a long time series, which can be attributed to e.g., seasons, vegetation, and human activities. This is reflected in the simulated time series by alternating slow-drying and fast-drying periods. Despite the interest in the decay parameter, this type of variation does not affect the changepoint detection procedure. Therefore, the temporal variation of the decay parameter is fixed across all scenarios.

Time series of length 5000 are generated using the following steps. First, the changepoints are simulated from different Poisson distributions. In particular, changepoints in scenarios S1a and S1b are generated from a single Poisson distribution. Changepoints in scenarios S2a and S2b are simulated from two Poisson distributions with different intensity parameters to reflect the temporal patterns. Changepoints in scenarios S3a and S3b are simulated by two nested Poisson processes, one over the entire time span $n = 5000$ with lower intensity, and the other over a long decaying period with higher intensity. Then the drying rates, the spikes and the asymptotic parameters are simulated from various uniform distributions. Finally, Gaussian random noises are added to the time series. A summary of the specifications of the six scenarios is given in Table 1. Details of the simulation procedure and an example of the simulated time series from each of the six scenarios are given in the supplemental document.

Table 1: The specifications of the simulation scenarios.

	Spikes	Drying rate	Noises	Replicates
S1a	1 Poisson	2 uniform	$\sigma = 0.0005$	200
S2a	2 Poisson	2 uniform	$\sigma = 0.0005$	200
S3a	2 Poisson at 2 scales	2 uniform at 2 scales	$\sigma = 0.0005$	100×2 penalties
S1b	1 Poisson process	2 uniform distributions	$\sigma = 0.001$	200
S2b	2 Poisson processes	2 uniform distributions	$\sigma = 0.001$	200
S3b	2 Poisson at 2 scales	2 uniform at 2 scales	$\sigma = 0.001$	100×2 penalties

The simulation was implemented in R using the code developed by the authors. The penalty values were selected in advance based on trial runs and were fixed for all replicates. In particular, two different penalties were used in scenarios S3a and S3b to retrieve the large scale and small scale changepoints respectively, and each of the penalties were applied to half of the replicates. The following statistics were computed to investigate the performance of the method, (a) the true positive (TP) rate and false positive (FP) rate of the changepoint detection, (b) the distance between the set of estimated changepoints and the set of true changepoints, (c) the difference between the fitted time series and the true simulated time series quantified as the root mean squared errors (RMSE), (d) the difference between the estimated drying rates and the true simulated drying rates quantified as the RMSE.

3.2 Summary of simulation results

Table 2 shows the averaged true positive rates and false positive rates over all simulation replicates for six scenarios. Here “small λ ” and “large λ ” in brackets refer to the use of a small penalty or a large penalty, respectively. The true positive rates reached over 90% for scenarios S1a, S1b, S3a (large λ) and S3b (large λ). The true positive rates are relatively lower (82.39% and 85.96% respectively) in scenarios S3a (small λ) and S3b (small λ). This is due to the challenges in estimating smaller-scale changes as the signals are much weaker. The results improved when a relaxed version of true positive is considered, i.e., an estimated changepoint $\hat{\tau}_j$ has a match with a true changepoint τ_i if $\hat{\tau}_j \in (\tau_i - 10, \tau_i + 10)$. Increasing noise standard errors did not appear to affect the changepoint detection, which may be explained by the fact that the gaps between small and large noise levels are not distinctive enough to cause major differences. The false positive rates are low across all scenarios with the majority of replicates smaller than 0.1%, regardless of counting the exact match or a match within the $(\tau_i - 10, \tau_i + 10)$ intervals. Over 1/3 of the replicates in scenarios S1a, S2a, S1b, and S2b have false positive rates of 0. This is slightly lower in scenarios S3a (small λ) and S3b (small λ) as the signals are weaker and may be confused with noises.

Table 2: Average true positive rates and false positive rates (in %) calculated under the exact matching condition and the relaxed condition (i.e., within ± 10 time points of the true changepoint)

	TP	FP	TP (± 10)	FP (± 10)
S1a	91.96%	0.02%	94.40%	0.01%
S2a	89.71%	0.02%	92.36%	0.01%
S3a (small λ)	82.39%	0.05%	86.12%	0.03%
S3a (large λ)	95.76%	0.04%	96.45%	0.04%
S1b	92.05%	0.02%	94.77%	0.01%
S2b	89.71%	0.02%	92.51%	0.01%
S3b (small λ)	85.96%	0.04%	87.89%	0.02%
S3b (large λ)	95.91%	0.02%	96.35%	0.02%

For a more comprehensive comparison of the estimated and true changepoints, a distance developed by Shi *et. al.* (2022) was computed to investigate the dissimilarity between the two configurations, e.g. the true changepoints $\boldsymbol{\tau} = \{\tau_1, \dots, \tau_m\}$ and the estimated changepoints $\boldsymbol{\eta} = \{\eta_1, \dots, \eta_k\}$. It is defined as

$$d(\boldsymbol{\tau}, \boldsymbol{\eta}) = |m - k| + \min\{\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\eta})\},$$

where m and k are the number of changepoints in each set, and

$$\mathcal{A}(\boldsymbol{\tau}, \boldsymbol{\eta}) = \sum_{i=1}^m \sum_{j=1}^k \frac{|\tau_i - \eta_j|}{n} I_{ij},$$

is the overall cost of assigning η_j to τ_i , $j = 1, \dots, k$, $i = 1, \dots, m$. To be specific, $I_{ij} = 1$ if η_j is paired with τ_i and $I_{ij} = 0$ otherwise, following a linear assignment problem. Note that when $m \neq k$, not all τ_i and η_j are paired; when there is a perfect match between the two sets of changepoints, $d(\boldsymbol{\tau}, \boldsymbol{\eta}) = 0$. Such a distance accounts for the dissimilarity in both the number and the locations of the true and estimated changepoints, thus providing a more comprehensive quantification of the differences. The distances are presented in Table 3. It appears that most

of the scenarios have relatively small distances, apart from scenario S3b (small λ) where some smaller-scale change-points are missed when the noise level is higher.

The RMSE of the fitted time series is computed as a measure of the overall fit of the estimated model. Alongside this, the RMSE of the estimated decay parameter γ is also computed to investigate how the method retrieved the key parameter in the exponential decay model. The results are shown in Table 3. The overall fit of the model was reasonable for all scenarios. The RMSE of scenarios S3a (large λ) and S3b (large λ) are among the highest, which is expected as the large penalty was designed to capture only the large-scale increases. It also seems difficult to retrieve the decay parameters γ from the small-scale increases in scenarios S3a (small λ) and S3b (small λ), which again is as expected. The RMSEs of γ are an order of magnitude smaller in the rest of the scenarios than the two most challenging ones.

Table 3: Summary statistics of the distance metric between two sets of change-points, the root mean squared errors (RMSE) of the fitted time series and the RMSE of the estimated decay parameter γ .

	Distance			RMSE			RMSE γ		
	10%	median	90%	10%	median	90%	10%	median	90%
S1a	0	0.0015	1.2325	0.0005	0.0023	0.0068	0.0001	0.0016	0.0936
S2a	0	0.0227	2.1863	0.0005	0.0022	0.0053	0.0002	0.0124	0.0926
S3a (small λ)	0.0053	1.1093	5.0979	0.0006	0.0009	0.0047	0.0147	0.0844	0.1341
S3a (large λ)	0	1.0019	4.0000	0.0019	0.0033	0.0055	0.0006	0.0063	0.0401
S1b	0	0.0053	1.2023	0.0009	0.0024	0.0073	0.0002	0.0023	0.0980
S2b	0	0.0343	2.1302	0.0009	0.0024	0.0054	0.0003	0.0147	0.0934
S3b (small λ)	0.0128	3.0703	23.0000	0.0010	0.0013	0.0049	0.0110	0.0713	0.1379
S3b (large λ)	0	1.0000	2.0912	0.0024	0.0036	0.0062	0.0006	0.0032	0.0174

Finally, a few examples of the replicates from different scenarios are presented to give intuition to the averages presented in the tables. Examples of replicates with high true positive rates and small mean squared errors from scenarios S1a, S2a and S3a are given in Figure 2. They represent the performance of the proposed method in the majority of the replicates. There are a few situations where the method failed to achieve a satisfactory fit, either in terms of the true positive rate or the estimated decay parameter. An example of the replicate with low true positive rates is shown in the left panel of Figure 3. Here the lack of fit was the result of the minimum segment length used (24) being larger than the distance between two adjacent increases. Hence, the method missed the change-points and created a knock-on effect on some later time points. This can be improved by simply reducing the minimum segment length, see for example the right panel of Figure 3, where the minimum segment length is 12. The lack of fit due to the difficulty in capturing the smaller scale patterns in scenarios S3a (small λ) and S3b (small λ) may be improved by changing the penalty λ in the optimisation problem. In the simulation study, both the penalty and the minimum segment length were fixed; whereas in real application, these values will be tuned according to the problem.

4 Practical considerations

The simulation study in section 3 demonstrates the ability of the proposed method to detect structural changes and exponential decay parameters in different types of time series of repeated sudden increases and decays. In real applications, however, there could be complications that

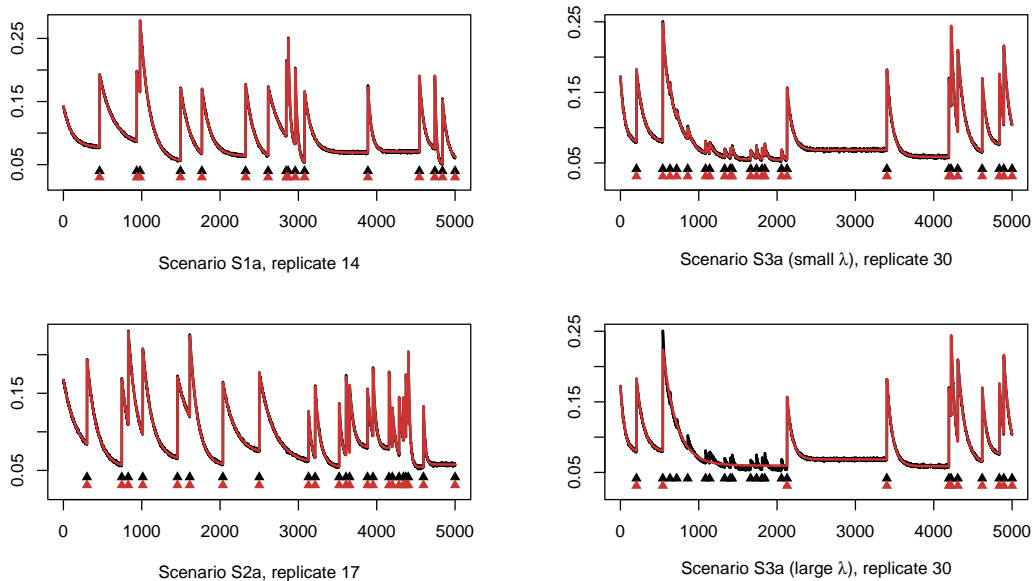


Figure 2: Examples of simulation replicates with high true positive rate from scenarios S1a, S2a and S3a. The black curve and the black triangles represent the simulated time series and the true changepoints. The red curve and the red triangles represent the fitted time series and the estimated changepoints.

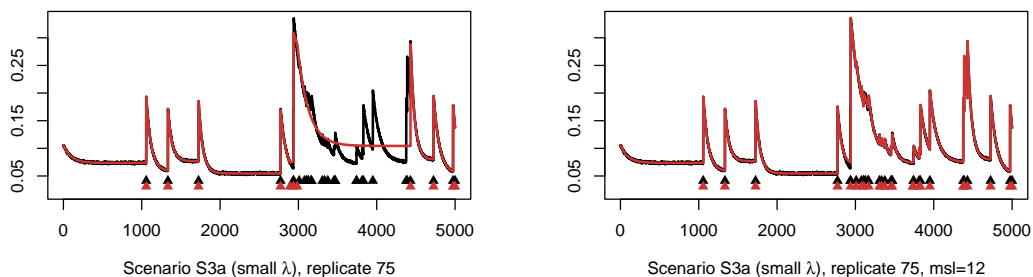


Figure 3: (Left) An example of a simulation replicate with low true positive rate from scenario S3a. (Right) The improved result using a smaller minimum segment length. The black curve and the black triangles represent the simulated time series and the true changepoints. The red curve and the red triangles represent the fitted time series and the estimated changepoints.

require careful consideration. Changepoint detection may be improved by taking into account additional information. In this section, several aspects of the proposed method are discussed, including the possibility of adding covariates to the model and the selection of penalty parameters.

Precipitation is considered to be one of the most important drivers of soil moisture increase. Although it is not equivalent to the amount of water infiltrating into the soil due to other pathways of water loss, such as surface run-off (Hillel, 2003; McColl *et al.*, 2017), it is expected to correlate with the locations and frequencies of the peaks. When relevant precipitation data are available, one could consider including precipitation as a covariate of the exponential decay model. For example, the precipitation time series X_t may be used directly to help model the

peaks of soil moisture as

$$Y_t = \alpha_{0i} + (\alpha_{1i} + \beta_i X_t) \exp\{-\exp(\gamma_i)(t - \tau_i)\} + \epsilon_t .$$

Alternatively, it may be converted into an indicator variable for the presence or absence of rainfall and included in the model as

$$Y_t = \alpha_{0i} + (\alpha_{1i} + \mathbf{Z}_{ti}^\top \boldsymbol{\beta}_i) \exp\{-\exp(\gamma_i)(t - \tau_i)\} + \epsilon_t ,$$

where $\mathbf{Z}_{ti}^\top = (Z_{ti}^1, \dots, Z_{ti}^m)$, with $Z_{ti}^j = \mathbb{1}(t \geq z_j)$ for $j = 1, \dots, m$ and z_1, \dots, z_m being the timings of rainfall instances within segment i . This changes the content of the changepoints from the times when soil moisture peaks to the times when the decay rate or the asymptotic level changes, or the time when the moisture level peaks without rainfall. However, they are still suitable for the investigation of the temporal dynamics of soil moisture drydown. For this approach to work, the precipitation time series needs to reflect the rainfall-infiltration pattern relatively accurately, which is challenging for field data. In addition, sometimes the precipitation becomes surface run-off, which does not contribute to the change of soil water content.

The proposed method relies on tuning parameters, such as the penalty function, the parameter and the minimum segment length. The penalty function used in sections 2 and 3 is the number of changepoints. Other choices are available and equally theoretically valid. In the simulation study, the penalty parameter was fixed within each scenario. In practice, the model does not reflect the full reality of the process and thus the penalty parameter is not optimal and may need to be selected based on the feature of the time series. For a systematic approach, Haynes *et. al.* (2017) introduces the CROPS algorithm, which efficiently computes the changepoint problem for a range of penalties. Different types of adaptive penalties have also been developed in the literature, such as Hocking *et. al.* (2013) and Truong *et. al.* (2017) where the penalty parameter is estimated to match the annotations of changepoints by experts. Such annotations are usually not available for the soil moisture time series. However, one may use climate data to create pseudo annotations and carry out penalty learning. One potential choice is precipitation.

To implement this we could assume that the locations and frequencies of the peaks correlate with the timings and frequencies of *some* rainfall events. Then the precipitation time series, after applying a certain threshold or transformation, may be used as the experts' annotations. Comparing the detected numbers and/or locations of changepoints under different penalty parameters to the annotated changepoints can provide some indication of the choice of an appropriate penalty parameter. For example, the penalty learning method in Hocking *et. al.* (2013) compares the possible number of changepoints annotated by experts in different regions of the time series to the number of changepoints detected by the algorithm given penalty λ in corresponding regions. The loss function to minimise is

$$\mathcal{E}(\lambda) = \sum_{(r,a) \in \{R,A\}} \mathbb{1} \left(\left| \boldsymbol{\tau}^{(\lambda)} \cap r \right| \notin a \right) ,$$

where $\boldsymbol{\tau}^{(\lambda)}$ is the optimal set of changepoints given λ , R is a set of regions on the time axis and A is a set of annotations on the possible number of changepoints within each region in set R . The regions do not necessarily cover the entire span of the time series, which means the missing gaps in the precipitation time series are not problematic. Alternatively, the method in Truong

et. al. (2017) selects the appropriate penalty λ through minimising the excess penalised risk,

$$\mathcal{E}(\lambda) = \mathcal{R}(\lambda; Y_{1:n}, \tau^{xpt}) - \mathcal{R}(\lambda; Y_{1:n}, \tau^{(\lambda)}) ,$$

where $\mathcal{R}(\lambda; Y_{1:n}, \tau)$ is the risk function (i.e., overall cost) of segmenting the data according to τ , and τ^{xpt} is the collection of the changepoints labelled by the experts. The challenge of implementing such a penalty learning method is the creation of the annotations, i.e., converting the precipitation time series into experts' annotations. The conversion may be case-specific and may require additional information about the field site. However, it does not rely on distinctive rainfall-infiltration patterns. It only requires the rainfall instances to correlate with the occurrences of peaks to some extent. Therefore, precipitation data that are not suitable as covariate data may still be fit for annotations. An example of creating annotations and implementing the two penalty learning methods using NEON soil moisture and precipitation data is given in the supplemental document.

The minimum segment length is used in the proposed method to put a realistic lower bound on the soil drying time and to allow enough data in the NLS estimation of the exponential decay model. In the soil drydown literature, the minimum length of the drydown period is often specified by considering a combination of sampling frequency, soil hydrological properties and how the drydown curves are identified (McColl *et al.*, 2017; Shellito *et al.*, 2016; Raoult *et al.*, 2021; Tso *et. al.*, 2023). The motivation is to identify the smooth drydown curves that will give a good estimation of the decay parameter, not necessarily all the drydown curves from the time series. For example, in Raoult *et al.* (2021), a minimum of 5 days was required for any drydown period, due to the sampling frequency (i.e., daily time series) and the way in which the drydown curves were identified via precipitation. McColl *et al.* (2017) described the reason on their choice of a 9-day threshold with a discussion on sampling frequency and soil types.

The NEON soil moisture time series have a higher temporal resolution than most of the satellite or in-situ data products. Hence there are more data available to make it easier to estimate the decay parameter, and using a similar threshold of e.g., 5-day, as in the literature, is unnecessary. In addition, there may be several wetting and drying events within such a period, and setting the minimum segment length too large would affect the estimation of the changepoints, as demonstrated in the simulation. Whether it is possible to further reduce the minimum segment length would depend on how the exponential decay model is estimated and how noisy the data are. Multiple sudden increases within a very short window may be associated with sensor noise so it is not always appropriate to try to recover these. In summary, the choice of minimum segment length is a balance between having enough data to fit the exponential decay model and not being too large so that it hinders changepoint detection.

5 Analysing NEON soil moisture time series

In this section, the changepoint model was applied to the soil moisture time series from the NEON data portal. Soil water and salinity data have been collected in 46 field sites across the U.S.. Here three terrestrial field sites with contrasting features: the Smithsonian Environmental Research Center (SERC) in Maryland, the Santa Rita Experimental Range (SRER) in Arizona, and the Talladega National Forest (TALL) in Alabama, are investigated.

A full description of the three field sites can be found at <https://www.neonscience.org/field-sites>, but their characteristics are summarised here for the readers' convenience. The Smithsonian Environmental Research Center is located in, Maryland on the Rhode River, a sub-

estuary of the Chesapeake Bay. The climate is temperate and humid, with an average annual temperature of 13.6°C (56.5°F) and a mean annual precipitation of 1075mm. Soils are formed into fluvial marine deposits with some areas of overlying alluvium and loess and the vegetation is dominated by coastal hardwood forests and cropland. The Santa Rita Experimental Range, located in the Sonoran Desert, Arizona, is characterized by a semi-arid, hot climate. The mean annual temperature is 19.3°C (67°F). The Sonoran Desert is wetter than most deserts with a mean annual precipitation of 346.2mm each year which is distributed in two wet periods. Diurnal temperature swings of up to 32°C (89.6°F) are common. The soils found at SRER are those typical of desert regions - they are mostly composed of alluvial deposits from the Santa Rita Mountains. Vegetation at the site is dominated by drought-resistant, thorny species. The Talladega National Forest is located in west-central Alabama. It has a subtropical climate with hot summers, mild winters, and year-round precipitation. This warm, moist air contributes to the formation of convection storms and thunderstorms in the region, causing major precipitation pulses and flooding. The area is subject to tornadoes and hurricanes. The average annual temperature is 17.2°C (62.9°F) and the average annual precipitation is about 1380mm. The soils in TALL are primarily sand, clay, and mudstone formed from undifferentiated marine segments. The vegetation at TALL is dominated by conifers, with some areas of intermixed conifers, hardwoods, bottomland hardwoods, and wetlands.

Soil moisture measurements are made in vertical profiles consisting of up to eight depths in five instrumented soil plots at each site. The data are presented as 1-minute and 30-minute averages. Here the 30-minute data product is used and the data are further sub-sampled to 1-hour time series for the changepoint analysis¹.

The location with the fewest missing observations was selected from each field site, and the period with no large missing gap was selected. These are, 1 July 2017 to 30 November 2018 at location 1 for field site SERC, 1 June 2018 to 31 May 2020 at location 4 for field site SRER, and 1 February 2018 to 31 January 2019 at location 5 for field site TALL (see Figure 1). Linear interpolation was applied to fill the small amounts of missing data within each time series.

To begin with, the constraints on model parameters are set and the values of the tuning parameters are selected. (a) An upper cap of 0.4 was applied to the soil moisture time series from the NEON data portal. As a result, the same upper bound was introduced to the parameters α_{0i} and α_{1i} in model (4) during the optimisation of the non-linear least squares. (b) The minimum segment length was chosen to be 24 hours. (c) It was required that the size of the sudden increases be greater than 0.001. This value was used to filter out the sensor noise, which is on the scale of 10^{-4} (Ayres & Roberti, 2018). (d) Considering jointly the overall cost of the fitted models and the experimental result from penalty learning using the precipitation time series, it appeared that a penalty around 250 would be appropriate for the soil moisture time series from site SERC, a penalty around 200 for site SRER, and a penalty around 200 for site TALL. Although three different penalties are used for the three field sites, the procedures of selecting the penalty values remained roughly the same. The algorithm was first run on a series of penalty values. Then the penalised overall cost of the final configuration, the mean squared errors, and the loss function of the penalty learning problem were evaluated. Finally, an appropriate penalty was selected based on the minimiser of one or more of these quantities. Therefore, it is possible to create a set of rules and automate the penalty selection procedure when the method is applied to a larger number of time series from different field sites.

¹In soil drydown modelling based on satellite data or in-situ data, the temporal resolutions of the data are usually lower. For example, satellite data typically have one observation per day or one observation every 2, 3 days.

The identified changepoints (the black triangles) and the fitted time series (the red curves) are presented in the top panels of Figure 4 to 6 corresponding to field site SERC, SRER and TALL respectively. The estimated asymptotic soil moisture parameter α_1 and the exponential decay parameter γ (the black lines) for each segment along with the uncertainty bands (the light grey bands) are presented in the 2nd and the 3rd panels. To be precise, α_1 and γ are plotted as piecewise constant functions over the span of the corresponding segments. This helps to visualise the temporal dynamics in the estimated parameters. Due to the lower and upper limits used in the NLS optimisation, the standard errors of parameters at the boundaries tend to be very large. Therefore, the range of the y-axis was decreased, and the very large standard errors are shown as light grey bands stretching from the bottom to the top of the canvas. The e-folding decay parameter ω (in days) in the soil drydown model (1) is also computed from the estimated exponential decay parameter γ and the results are given in the figures in the supplemental document.

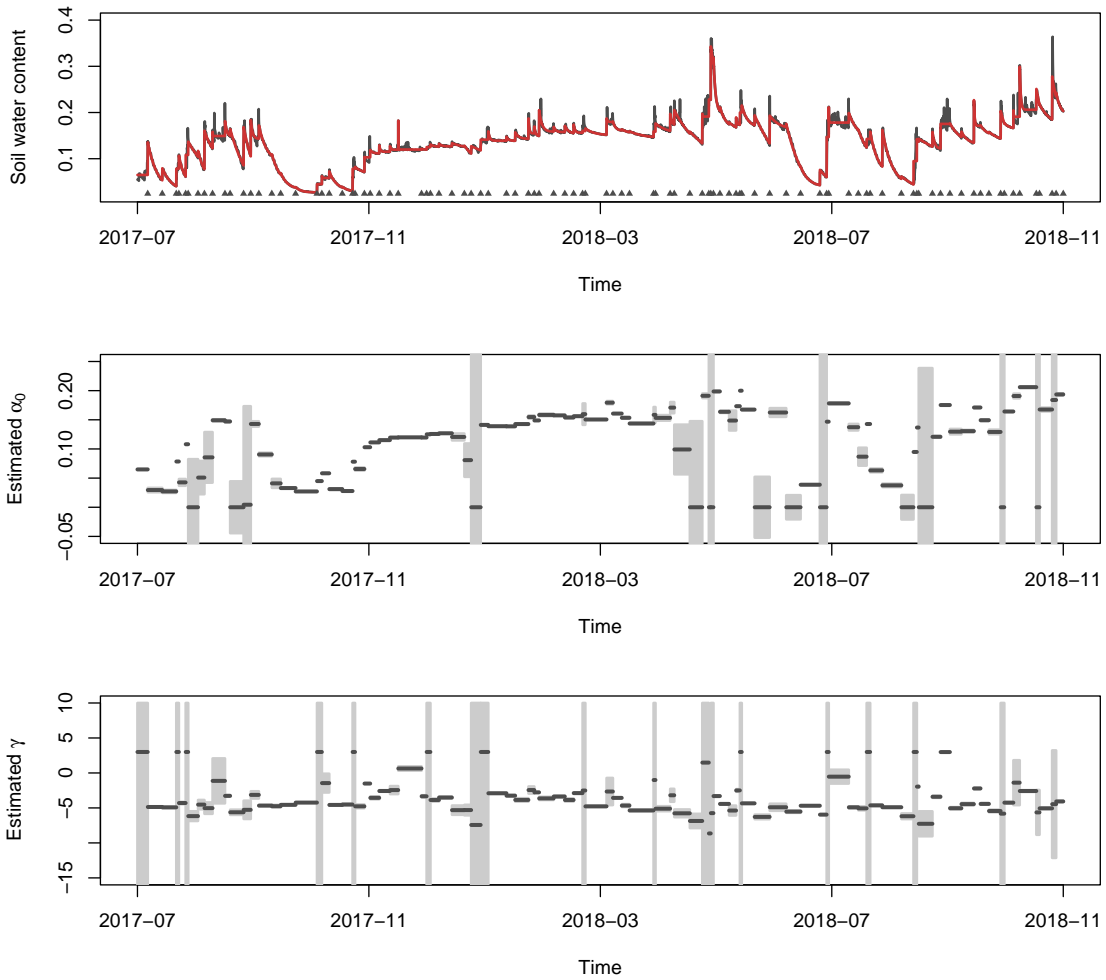


Figure 4: (Top) The soil moisture time series and the estimated changepoints (black triangles) at location 1, field site SERC. (Middle) The estimated asymptotic parameter α_0 over time with a light grey uncertainty band. (Bottom) The estimated decay parameter γ over time with coloured uncertainty band.

All three models achieved a reasonable visual fit, where the red curves in the top panel of

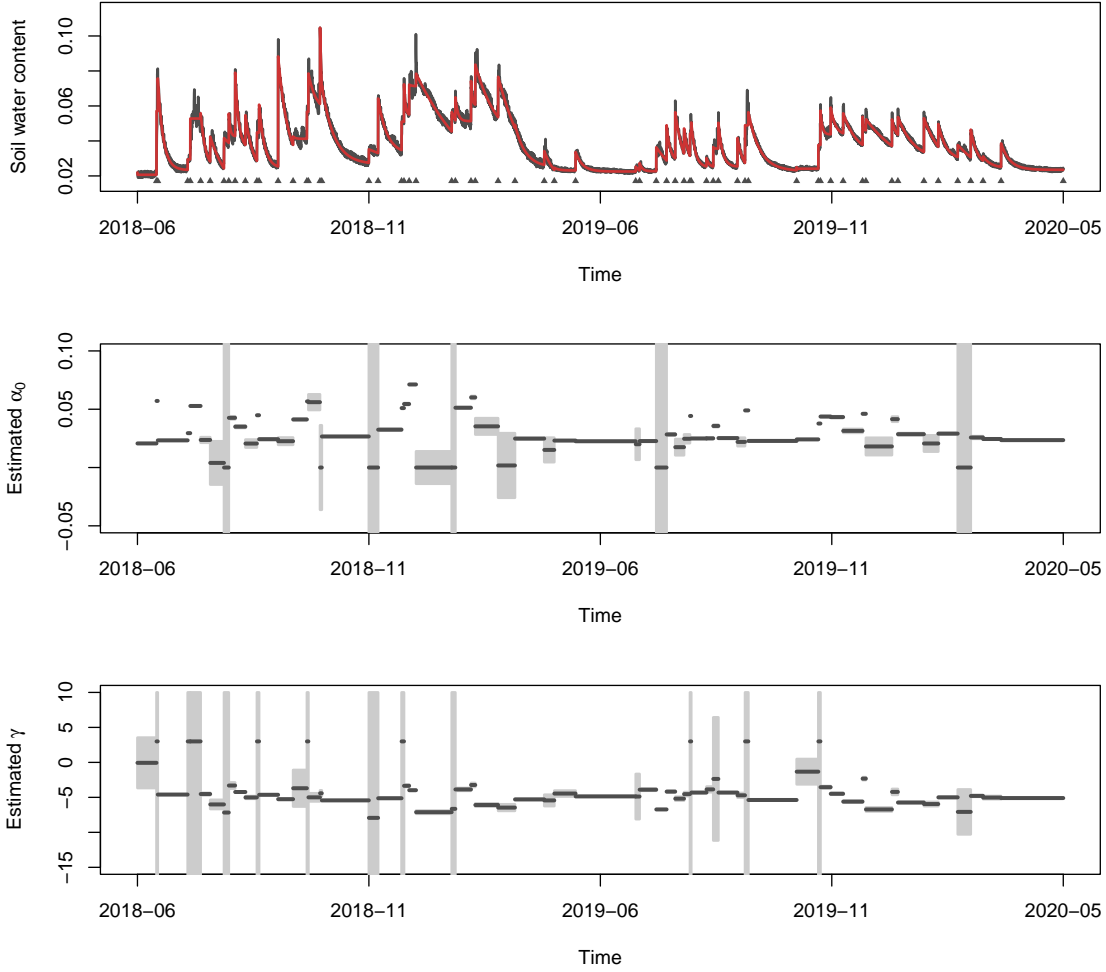


Figure 5: (Top) The soil moisture time series and the estimated changepoints (black triangles) at location 4 in field site SRER. (Middle) The estimated asymptotic parameter α_0 over time with a light grey uncertainty band. (Bottom) The estimated decay parameter γ over time with coloured uncertainty band.

Figure 4 to 6 captured the main temporal patterns in the data. Some of the sudden increases are missed due to their short distance to adjacent increases. The lack of fit in some parts of the time series, e.g., around July 2018 in site SERC, was associated with the relatively slow increase in the soil moisture which contradicts the assumption of a sudden increase. However, these are not common features. There is no distinctive temporal pattern in the occurrence of the changepoints in this case.

There is a clear difference in the estimated asymptotic soil moisture α_0 between the field sites. The asymptotic soil moisture in site SERC is generally higher than that in SRER and TALL (see also the histograms in the supplemental document). This is understandable as site SRER experiences a desert-like climate and site TALL, though humid, has high temperatures year-round. There appear to be some temporal variations in α_0 as well. For example, for site SERC, the α_0 during the winter 2018 period behaved slightly differently from the summer period. This suggests that the approach taken in conventional soil drydown modelling where the asymptotic soil moisture is fixed throughout time may not be appropriate. Allowing α_0

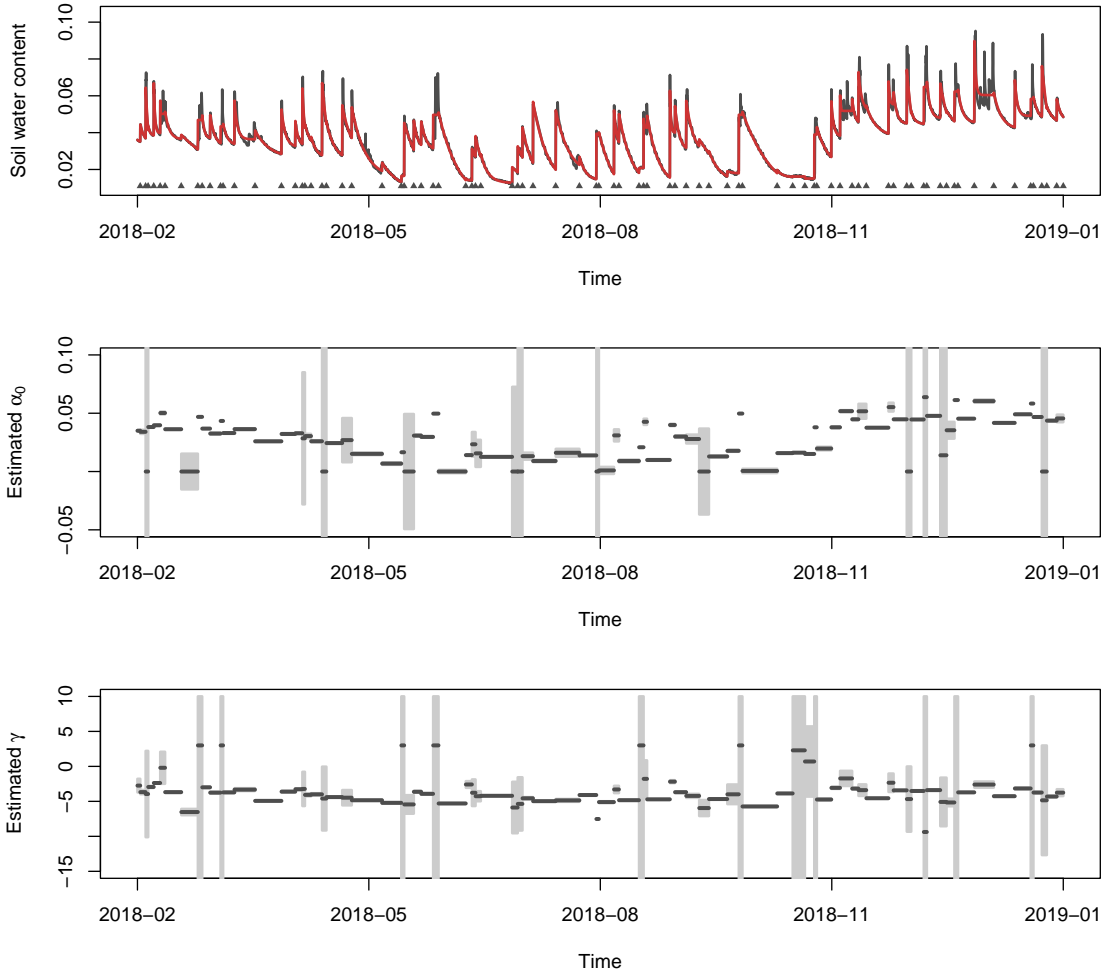


Figure 6: (Top) The soil moisture time series and the estimated change points (black triangles) at location 5, field site TALL. (Middle) The estimated asymptotic parameter α_0 over time with a light grey uncertainty band). (Bottom) The estimated decay parameter γ over time with coloured uncertainty band.

to change over time has the potential to improve the estimation of other parameters in the drydown model.

The differences in the scale of the estimated exponential decay parameter γ are less distinctive. The estimated γ for site SRER suggests a slightly slower drying rate than the other two sites. This can also be seen from the figures displaying the e-folding decay parameter ω in the supplemental document. This could be explained by the sparse desert vegetation at site SRER extracting water more slowly or the low unsaturated hydraulic conductivity of the dry soil at SRER, which results in little drainage to lower soil layers. Although there are temporal variations in the estimated decay parameter, there is no clear trend or seasonal pattern in the result here. A longer time series would potentially reveal more interesting features in the dynamics of soil moisture.

Fundamentally, conventional soil drydown modelling relies on additional hydrological or physical information; in contrast, the proposed method is data-driven. In particular, all parameters of the drydown model are allowed to vary over time. As a result, there will be differences

between the parameters estimated using the changepoint method and those from soil drydown literature. Investigating these differences may lead to a better understanding of soil moisture drydown. To summarise, the proposed method provides a different insight into the soil moisture dynamics which is not available using the conventional modelling approach.

6 Discussion

This paper proposed a changepoint-based method to investigate the temporal dynamics in the soil moisture time series. The method aims to identify the structural changes in the form of sudden increases in soil moisture and estimate the parameters characterising the drying process that follows the sudden increase. The method is related to the soil drydown modelling but takes a different approach. It does not rely on the manual identification of soil drydown curves from a soil moisture time series. Instead, it applies a changepoint detection algorithm directly on the soil moisture time series, which automatically identifies the segments representing the exponential decay of soil moisture. The estimation of the soil moisture decay parameters is carried out simultaneously. In addition, the method can be applied to soil moisture time series with little data pre-processing. Thus, when compared to conventional soil drydown modelling, the proposed method has the advantage of easy implementation to a large data set with minimal data preparation. The method also has the flexibility to make use of relevant information, e.g., the precipitation time series data, to improve the segmentation. The simulations and data examples demonstrated the ability of the proposed approach to recover important features of soil moisture drydown.

Unlike the simulated time series, the real soil moisture time series can display patterns beyond the simple exponential decay. For example, when the soil is saturated during wet seasons or when it is frozen during the winter, the soil moisture time series can show very different patterns from a drydown curve. In practice, it may be sensible to focus on the periods when drydown processes are dominating. Introducing additional information, e.g., temperature or seasonal regimes, may also help to capture different drying patterns. Alternatively, it may be helpful to develop methods that do not rely on the exponential decay assumption and use more flexible models to describe different drying patterns.

Evaluating the uncertainty of the estimated changepoints, though important in some applications, is difficult in a multiple changepoints detection problem. [Chen *et al.* \(2021\)](#) proposed a method to compute the confidence intervals for the changepoints in the calcium time series by finding the maximum disturbance that generates the same changepoint. However, the efficient computation of the confidence intervals is tailored to the functional pruning algorithm and therefore is not suitable for use with PELT. Other approaches to quantifying uncertainty in the literature, such as bootstrapped confidence intervals ([Huskova & Kirch, 2008](#); [Hollaway *et al.*, 2021](#)) and posterior distributions of the changepoint numbers/locations ([Fearnhead, 2006](#); [Nam *et al.*, 2015](#)), do not generalise to the changepoint problem in this paper easily. Therefore, the uncertainty of the identified changepoints is not considered here.

Extending the current method to cover these situations will be a piece of important future work. For example, a useful extension would be to relax the model assumptions and use more flexible models to describe soil moisture dynamics in various scenarios, such as dry and saturated conditions. Potential solutions include re-formatting the problem as a state space model where different states represent different scenarios. Such a model, when estimated within a Bayesian framework, may also provide uncertainty measures to the estimated changepoints.

Finally, the changepoint method described in section 2 shares some similarities with the so-called shot noise model, which involves a compound Poisson process describing the intensity of the shots and an impulse-response function describing the decay pattern (Xiao & Lund, 2006; Eliazar & Klafter, 2005). In the classical shot noise model, the decay pattern is often modelled as an exponential decay, and it has been used in Tsakiris *et. al.* (1988) to estimate the rainfall infiltration, which contributes to the soil moisture dynamics. The shot noise model does not rely on the arrival times of the shots, and hence there is no need to identify the changepoints. However, parameters in the compound Poisson process and the impulse-response function are required. It can be difficult to estimate decay patterns that are changing over time, which is a key feature that the changepoint-based method seeks to expose.

Acknowledgements

The authors gratefully acknowledge the support from the UKRI-funded project Signals in the Soil (Grant No. NE/T012307/1). CN was supported by the Engineering and Physical Sciences Research Council (EPSRC), grant numbers EP/V022636/1, EP/S00159X/1 and EP/Y028783/1.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Code and data availability

The changepoint detection algorithm and the analysis of the soil moisture time series are implemented in R (version 4.3.1). The code can be accessed from the GitHub repository <https://github.com/GMY2018/Changepoint4soil>.

The data used are publicly available from the United States National Ecological Observation Network (NEON, <https://data.neonscience.org/>). The soil moisture time series data can be accessed from <https://data.neonscience.org/data-products/DP1.00094.001>. The precipitation data can be accessed from <https://data.neonscience.org/data-products/DP1.00006.001>.

References

- Ayres, E., Roberti, J., 2018. NEON algorithm theoretical basis document (ATBD): TIS soil moisture and water salinity. Manuscript accessed from <https://data.neonscience.org/data-products/DP1.00094.001> in May, 2021.
- Bates, D. M., Watts, D. G., 1988. *Nonlinear Regression Analysis and Its Applications*, John Wiley & Sons.
- Chen, Y. T., Jewell, S., Witten, D., 2021. Quantifying uncertainty in spikes estimated from calcium imaging data. *Biostatistics*, 24(2), 481–501
- Dennis, J. E., Gay, D. M., Welsch, R. E., 1988. An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7(3), 348–368.
- Dong, J., Akbar, R., Short Gianotti, D. J., Feldman, A. F., Crow, W. T., Entekhabi, D., 2022. Can surface soil moisture information identify evapotranspiration regime transitions? *Geophysical Research Letters*, 49(7), <https://doi.org/10.1029/2021GL097697>
- Eliazar, I., Klafter, J., 2005. On the nonlinear modelling of shot noise. *The Proceedings of the National Academy of Sciences*, 102(39), 13779–13782.
- Fearnhead, P., 2006. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16(2), 203–213.
- Frelih-Larsen, A., Riedel, A., Hobeika, M., Scheid, A., Gatteringer, A., Niether, W., Siemons, A., 2022. Interim report: Role of soils in climate change mitigation. *German Environment Agency*, Last accessed from <https://www.ecologic.eu/sites/default/files/publication/2023/50061-role-of-soils-in-climate-change-mitigation.pdf> in October 2024.
- Haynes, K., Eckley, I. A., Fearnhead, P., 2017. Computationally efficient changepoint detection for a range of penalties. *Journal of Computational and Graphical Statistics*, 126(1), 134–143.
- Hillel, D., 2003. Chapter 14: Water enter into soil, *Introduction to Environmental Soil Physics*, Elsevier Academic Press.
- Hollaway, M. J., Henrys, P. A., Killick, R., Leeson, A., Watkins, J., 2021. Evaluating the ability of numerical models to capture important shifts in environmental time series: A fuzzy change point approach *Environmental Modelling and Software*, 139 (2021) 104993.
- Hocking, T. D., Rigaiil, G., Bach, F., Vert, J., 2013. Learning sparse penalties for change-point detection using max margin interval regression. *Proceedings of 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013. JMLR: W&CP volumn 28.
- Huskova, M., Kirch, C., 2008. Bootstrapping confidence intervals for the change-point of time series. *Journal of Time Series Analysis*, 29 (6), 947–972.
- Jewell, S., Witten, D., 2018. Exact spike train inference via \mathcal{L}_0 optimization. *Annals of Applied Statistics*, 12(4), 2457–2482.
- Jewell, S., Hocking, T. D., Fearnhead, P., Witten, D., 2020. Fast nonconvex deconvolution of calcium imaging data. *Biostatistics*, 21(4), 709–726.

- Killick, R., Fearnhead, P., Eckley, I. A., 2012. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500), 1590–1598.
- Lehmann, J., Bossio, D. A., Kögel-Knabner, I., Rillig, M. C., 2020. The concept and future prospects of soil health. *Nature Reviews Earth & Environment*, 1, 544–553.
- Maidstone, R., Hocking, T., Rigaiil, G., Fearnhead, P., 2017. On optimal multiple changepoint algorithms for large data. *Statistics and Computing*, 27, 519–533.
- McColl, K. A., Wang, W., Peng, Bin., Akbar, R., Gianott, D. J. S., Lu, H., Pan, M., Entekhabi, D., 2017. Global characterization of surface soil moisture drydowns. *Geographical Research Letters*, 44, 3682–3690.
- Nam, C. F. H., Aston, J. A. D., Eckley, I. A., Killick, R., 2015 The uncertainty of storm season changes: quantifying the uncertainty of autocovariance changepoints. *Technometrics*, 57(2), 194–206
- Nash, J. C., 2014. *Nonlinear Parameter Optimization Using R Tools*. John Wiley & Sons, Incorporated, New York.
- Nash, J. C., 2016. nlmrt: Functions for Nonlinear Least Squares Solutions. R package version 2016.3.2, <https://CRAN.R-project.org/package=nlmrt>.
- NEON (National Ecological Observatory Network). Soil water content and water salinity (DP1.00094.001). <https://doi.org/10.48443/bakr-aj85>. Dataset accessed from <https://data.neonscience.org> in October, 2021
- Ontl, T. A., Schulte, L. A., 2012. Soil Carbon Storage. *Nature Education Knowledge* 3(10):35. Last accessed from <https://www.nature.com/scitable/knowledge/library/soil-carbon-storage-84223790/> in October 2024.
- R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- Raoult, N., Ottlè, C., Peylin, P., Bastrikov, V., Maugis, P., 2021. Evaluating and optimizing surface soil moisture drydowns in the ORCHIDEE land surface model at in situ locations. *Journal of Hydrometeorology*, 22, 1025–1043.
- Rigaiil, G., 2015. A pruned dynamic programming algorithm to recover the best segmentations with 1 to K_{max} changepoints. *Journal de la Société Française de Statistique*, 156, 180–205.
- Ruscica, R., Polcher, J., Salvia, M., Sorensson, A., Piles, M., Jobbágy, E., Karszenbaum, H., 2020. Spatio-temporal soil drying in southeastern South America: the importance of effective sampling frequency and observational errors on drydown time scale estimates. *International Journal of Remote Sensing*, 41:20, 7958-7992.
- Salvia, M., Ruscica, R., Sorensson, A., Polcher, J., Piles, M., Karszenbaum, H., 2018. Seasonal analysis of surface soil moisture dry-downs in a land-atmosphere hotspot as seen by LSM and satellite products. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 5521-5524.
- Sehgal, V., Gaur, N., Mohanty, B. P., 2021. Global surface soil moisture drydown patterns. *Water Resources Research*, 57.

- Shellito, P. J., Small, E. E., Colliander, A., Bindlish, R., Cosh, M. H., Berg, A. A., Bosch, D. D., Caldwell, T. G., Goodrich, D. C., McNairn, H., Prueger, J. H., Starks, P. J., van der Velde, R., Walker, J. P., 2016. SMAP soil moisture drying more rapid than observed in situ following rainfall events *Geographical Research Letters*, 43, 8068–8075
- Shellito, P. J., Small, E. E., Livneh, B., 2018. Controls on surface soil drying rates observed by SMAP and simulated by the Noah land surface model. *Hydrology and Earth System Science*, 22, 1649–1663.
- Shi, X., Gallagher, C., Lund, R., Killick, R., 2021. A comparison of single and multiple change-point techniques for time series data. *Computational Statistics and Data Analysis*, 170 (2022) 107433.
- Sinha, J., Sharma, A., Marshall, L., Kim, S., 2024. Characterizing satellite soil moisture dry-down: a bivariate filtering approach. *Water Resources Research*, 60(7), <https://doi.org/10.1029/2022WR034019>.
- Truong, C., Oudre, L., Vayatis, N., 2017. Penalty learning for Changepoint detection. *2017 25th European Signal Processing Conference (EUSIPCO)*, 1614–1618
- Tsakiris, G., Agrafiotis, G., Kiountonzi, E., 1988. A shot noise model for the generation of soil moisture data. *Stochastic Hydrology and Hydraulics*, 2, 51–59.
- Tso, C-H. M., Blyth, E., Tanguy, M., Levy, P. E., Robinson, E., Bell, V., Zha, Y., Fry, M., 2023. Multiproduct characterization of surface soil moisture drydowns in the United Kingdom. *Journal of Hydrometeorology*, 24(12), 2299–2319.
- Vereecken, H., Huisman, J. A., Pachepsky, Y., Montzka, C., van der Kruk, J., Bogaen, H., Weihermüller, L., Herbst, M., Martinez, G., Vanderborght, J., 2014. On the spatio-temporal dynamics of soil moisture at the field scale. *Journal of Hydrology*, 516 (2014), 76–96.
- Xiao, Y., Lund, R., 2006. Inference for shot noise. *Statistical Inference for Stochastic Processes*, 9, 77–96.
- Zhang, N. R., Siegmund, D. O., 2007. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63, 22–32.

List of figure legends

Figure 1: Hourly soil volumetric water content time series recorded at field sites SERC, SRER and TALL from the NEON data portal.

Figure 2: Examples of simulation replicates with high true positive rates from scenarios S1a, S2a and S3a. The black curve and the black triangles represent the simulated time series and the true changepoints. The red curve and the red triangles represent the fitted time series and the estimated changepoints.

Figure 3: (Left) An example of a simulation replicate with low true positive rate from scenario S3a. (Right) The improved result using a smaller minimum segment length. The black curve and the black triangles represent the simulated time series and the true changepoints. The red curve and the red triangles represent the fitted time series and the estimated changepoints.

Figure 4: (Top) The soil moisture time series and the estimated changepoints (black triangles) at location 1, field site SERC. (Middle) The estimated asymptotic parameter α_0 over time with a light grey uncertainty band. (Bottom) The estimated decay parameter γ over time with coloured uncertainty band.

Figure 5: (Top) The soil moisture time series and the estimated changepoints (black triangles) at location 4 in field site SRER. (Middle) The estimated asymptotic parameter α_0 over time with a light grey uncertainty band. (Bottom) The estimated decay parameter γ over time with coloured uncertainty band.

Figure 6: (Top) The soil moisture time series and the estimated changepoints (black triangles) at location 5, field site TALL. (Middle) The estimated asymptotic parameter α_0 over time with a light grey uncertainty band. (Bottom) The estimated decay parameter γ over time with coloured uncertainty band.