2018 IEEE 5th International Conference on Data Science and Advanced Analytics

Wetting and Drying of Soil: From Data to Understandable Models for Prediction

Aniruddha Basak
Carnegie Mellon University
abasak@cmu.edu

Ole J. Mengshoel
Carnegie Mellon University
ole.mengshoel@sv.cmu.edu

Kevin Schmidt
U. S. Geological Survey
kschmidt@usgs.gov

Chinmay Kulkarni
Carnegie Mellon University
chinmaymkulkarni@gmail.com

***Abstract*—Soil moisture is critical to agriculture, ecology, and certain natural disasters. Existing soil moisture models often fail to predict soil moisture accurately for time periods greater than a few hours. To tackle this problem, we introduce in this paper two novel models, the Naive Accumulative Representation (NAR) and the Additive Exponential Accumulative Representation (AEAR). The parameters in these models reflect hydrological redistribution processes of gravity and suction. We validate our models using soil moisture and rainfall time series data collected from a steep gradient post-wildfire site in Southern California. Data analysis is challenging, since rapid landscape change in steep, burned hillslopes is typically observed in response to even small to moderate rain events. We found that the AEAR model fits the data well for three distinct soil textures at different depths below the ground surface (at 5cm, 15cm, and 30cm). Similar strong results are demonstrated in controlled soil moisture experiments. Our recommended AEAR model has been validated as effective and useful by earth scientists, giving better forecasts than existing models for time horizons of 10 to 24 hours.**

***Index Terms*—soil moisture; earth science; time series; forecasting; exponential models; stochastic optimization; evolutionary algorithms.**

I. Introduction

Soil moisture[1] plays an essential role in agriculture, ecology, and natural disasters such as drought and flooding [15], [28]. Soil moisture may change rapidly over time and can show substantial variation with depth within a soil column as well as laterally through space.

Agricultural soils need to be sufficiently drained to minimize saturation and control high salinity. While managed irrigation is typically used to control soil moisture, root water uptake below the wilting point of a plant results in decreased crop yields. Improved prediction of soil moisture response to rainfall, as studied in this paper, would enable water managers of agricultural applications to further optimize irrigation schedules, plant-water uptake, and cost.

Soil moisture also has a key role in certain natural disasters. The National Oceanic and Atmospheric Administration (NOAA) and the United States Geological Survey (USGS) established a demonstration flash-flood and debris-flow early-warning system[2] for recently burned areas in Southern California. This system covers eight counties within southern California, and utilizes the National Weather Service's (NWS) Flash Flood Monitoring and Prediction (FFMP) system. FFMP identifies when both flash floods and debris flows are likely to occur based on comparisons between radar precipitation estimates and rainfall intensity-duration threshold values.

At a terrestrial level, soil moisture is measured by using probes inserted into the soil column. These probes measure dielectric properties of the soil to estimate volumetric water content (VWC) and are controlled by local data loggers or via telemetering to a server. In the resulting soil moisture time series data sets, the wetting and drying cycles can be clearly observed, but remain challenging to predict. Such data sets form the basis for our analysis in this paper.

Fig. 1: *Left*: Photograph of burned hillslope with rainfall, overland flow, and soil moisture monitoring instrumentation after the Canyon fire on the Pepperdine University campus. *Right*: Photograph from December 19, 2007 of a small post-fire debris flow and flood on the Pepperdine University campus following minor rainfall.

While theoretical models exist to quantify rainfall and runoff (*e.g.*, Richards' equation [27]), site-specific field conditions and hysteresis preclude their simple use. Taking a more data-driven approach, attempts have been made to predict soil moisture using time-series forecasting models (*e.g.*, autoregressive integrated moving average or ARIMA [2]). Unfortunately, these statistical models do not result in models that reflect the hydrological processes in soil. The Antecedent Water Index (AWI) model [9], [38], though, strikes the delicate balance of fitting soil moisture time-series data while providing meaningful information to geophysicists by expressing hydrologic parameters estimated from data. However, we establish in this paper that the AWI model is limited when predicting soil moisture for time horizons exceeding a few hours.

Inspired by the AWI model, we study in this paper two

[1]Three different terms are often used to describe the same characteristic: "soil moisture," "volumetric water content (VWC)," and "volumetric soil moisture." We prefer the first term in this paper.

[2]NOAA/USGS Demonstration Flash-Flood and Debris-Flow Early-Warning System http://landslides.usgs.gov/hazards/warningsys.php

978-1-5386-5090-5/18/$31.00 ©2018 IEEE
DOI 10.1109/DSAA.2018.00041

novel soil moisture models, the Naive Accumulative Representation (NAR) and the Additive Exponential Accumulative Representation (AEAR). The NAR model is a stepping stone for the AEAR model. These models accumulate rainfall over a time interval and can fit a diverse range of wetting and drying curves. Model parameters are estimated from data, but at the same time the models are designed to resemble hydrological models of soil processes. The NAR and AEAR models have been validated by earth scientists at USGS. Consequently, they promise to be more meaningful to the earth science community than traditional time series models (ARIMA and ARMAX). Moreover, these models can be used recursively to forecast soil moisture only using initial soil moisture measurement.

We validate our NAR and AEAR models on a challenging post-fire soil moisture time series data set from Southern California [30], where rapid sediment erosion and deposition was observed following a wildfire (see Figure 1). We find that our models can fit moisture data from different soil depths with distinct soil types. The models can be used for prediction and can also explain various soil properties including drying and wetting rates. Further, we evaluate our models' predictions using another soil moisture data set obtained from controlled experiments. Overall, the best predictions were obtained with the novel AEAR model, which is also data-driven and understandable to earth scientists.[3]

Even though we focus on soil moisture prediction in this paper, our methods may apply to other complex time-series modeling and forecasting problems. As an example, consider workload bursts for network and computing resources. Workload prediction is important in workload rebalancing and autoscaling for the cloud [35]. In extreme cases, such workload bursts can be considered flash events (notice the similar language to "flash flooding"), and include major breaking news and sporting events of global interest, such as the soccer World Cup [3], [5]. Web site traffic can also change dramatically over time due to the varying habits of social network users [23].

The rest of this paper is structured as follows. First, we describe the modeling goals and requirements in Section II. Next, we discuss related work in time series modeling and prediction in Section III. We discuss the soil moisture data sets in Section IV. In Section V, we introduce our two novel soil moisture models. The data analysis process and results for our data sets are presented in Section VI. A more qualitative evaluation, or model validation, by earth scientists is performed in Section VII.

II. Goal and Requirements

Our goal is to develop mathematical soil moisture models. After first introducing notation and terminology, we discuss in this section three model requirements.

We consider a time series or sequence of records, $(r_0, r_1, \ldots, r_i, \ldots)$, where each record consists of a time stamp t and a measurement value v: $r = (t, v)$ or for simplicity v_t. For a particular data set, we consider the following sequences:

- A sequence (or time series) of soil moisture measurements: $\mathscr{M} = (M_0, \ldots, M_i, \ldots)$.
- A sequence of soil moisture predictions: $\hat{\mathscr{M}} = \left(\hat{M}_0, \ldots, \hat{M}_i, \ldots\right)$.
- A sequence of rainfall measurements: $\mathscr{I} = (I_0, \ldots, I_i, \ldots)$.
- A sequence of rainfall predictions: $\hat{\mathscr{I}} = \left(\hat{I}_{t^*}, \ldots, \hat{I}_{t^*+i}, \ldots\right)$.

When predicting, we are at a particular time point t^*, and make soil moisture predictions for (future) times up to a time horizon of τ, thus: $\hat{\mathscr{M}}_{t^*+1:t^*+\tau} = \left(\hat{M}_{t^*+1}, \ldots, \hat{M}_{t^*+\tau}\right)$, where $t^* > 0$. We assume future rainfall values are available by rainfall forecasting models. Our focus is not on rainfall forecasting, as several well-established methods exist [29].

A. Three Model Requirements

A model needs to meet all of the following requirements:

- *Understandability:* The model needs to make sense to earth scientists, by having parameters readily tied to measurable physical processes. A model with parameters that can be easily interpreted from the earth science perspective, and clearly relate to or directly map to soil or hydrological properties, is much to be preferred.
- *Data-driven:* It must be easy to compute model parameters from soil moisture data that is being collected by current sensor technology. A model with parameters that cannot be easily estimated from data is less useful.
- *Accurate predictions:* The model must provide predictions that are accurate in the medium term, which we here define to be predictions with a prediction horizon 5 $\leq \tau \leq$ 24 hours, aligned with the timeframe of accurate weather forecasts. Humans can act on this time scale. For example, if an area is threatened by mud slides, due to high soil moisture, people can be warned, leave the area, and avoid the danger.

All of the above three requirements need to be met; it is not sufficient to meet just two of them.[4]

B. Prediction Approaches

Suppose we are at time t^*. A model should predict soil moisture $\hat{M}_{t^*+\tau}$, given a prediction horizon τ. In our case, τ is on the order of hours. Predictions can be performed using at least two approaches, and it is a requirement that both are supported. There is at time t^* either (i) an actual soil moisture measurement M_{t^*} or (ii) a predicted soil moisture value $\hat{M}_{t^*}$. In addition, there are rainfall measurements in $\mathscr{I}$ or forecasts in $\hat{\mathscr{I}}$ up to time $t^* + \tau$. The following two definitions reflect (i) and (ii) respectively.

Definition If *regular* (or frequent) soil moisture measurements are available, the soil moisture observation M_{t^*} can be used to predict $\hat{M}_{t^*+\tau}$ via a model f:

$$\hat{M}_{t^*+\tau} = f(M_{t^*}, I_{t^*}, \hat{I}_{t^*+1}, \ldots, \hat{I}_{t^*+\tau}), \quad \forall\, t^* > 0. \tag{1}$$

[3]This paper has a Web presence here: https://github.com/olemengshoel/wetting-and-drying-of-soil, with the source code of our novel models.

[4]As an example, machine learning models are obviously data-driven and may provide accurate medium-term predictions. However, some of them are not easily understandable to scientists. For example, it may be possible to train a deep neural network from soil moisture data and use it to compute accurate medium-term predictions. Due to the large number of parameters, deep neural networks are not easily understood by (earth) scientists.

If measurements are regularly or "continuously" stored in a data-logger or telemetered, (1) can be used.

In the case of irregular measurements, the measurement M_{t^*} of (1) is missing (else we have the regular case). Thus, we estimate M_{t^*} using $\hat{M}_{t^*}$ in the prediction model:

Definition For *irregular* (infrequent or single point observation M_0) soil moisture measurements, predicted soil moisture value $\hat{M}_{t^*}$ is used for predicting $\hat{M}_{t^*+\tau}$:

$$\begin{aligned} \hat{M}_\tau &= g(M_0, I_0, \ldots, I_\tau), \\ \hat{M}_{t^*+\tau} &= g(\hat{M}_{t^*}, I_{t^*}, \hat{I}_{t^*+1}, \ldots, \hat{I}_{t^*+\tau}), \quad \forall\, t^* > 0. \end{aligned} \tag{2}$$

The irregular situation in (2) arises, for example, when measurements can only be performed during occasional field visits or by over-passing satellites or aircraft.

III. Related Work

Considering the requirements identified in Section II as well as the soil moisture data presented in Section IV, we now examine related work. The focus is on models that potentially meet the three model requirements from Section II.

A. Simple Exponential Model (SEM)

The Antecedent Water Index (AWI) is considered to be proportional to soil moisture, and AWI is widely used to model soil moisture response [9], [38]. In 1995, Wilson and Wieczorek [38] developed the AWI model based on an analogy of water flow through a leaky bucket:

$$M_t^S = M_{t-1}^S e^{-k_d \Delta t} + \frac{I_t}{k_d}(1 - e^{-k_d \Delta t}), \tag{3}$$

where the superscript S indicates a simple exponential model (SEM). It is a simple sum of two exponentials where the first term in (3) represents recession of AWI after rain ceases and the second term represents increase in AWI due to rainfall. An instantaneous rainfall measurement at time t is denoted by I_t. There is a drainage coefficient, k_d, as a single exponential parameter present in both terms of the equation. This model is inspired by a water balance equation (4), where ΔM_t^S is the change in soil water content at time t; I_t is the mean precipitation; E_t is the mean evapotranspiration; R_t is the net streamflow divergence; and G_t is the net groundwater loss.

$$\Delta M_t^S = I_t - E_t - R_t - G_t. \tag{4}$$

Essentially, (4) relates precipitation, evapotranspiration, and water loss to the rate of change of soil moisture. It has been shown that an approximation of this equation leads to an exponential decay model (under no precipitation) [12], which is very similar to the AWI model (3).

NOAA prediction of soil moisture uses surface hydrology as defined by (4) [37]. Hence, the AWI model is closely related to a state-of-the-art hydrology-inspired soil moisture model. The AWI model was also used to predict shallow landslides in Seattle, Washington, based on rainfall data [9].

However, the SEM model (3) has just one parameter, k_d. Thus it has the same drainage coefficient k_d for both increase and decrease in soil moisture. This is a quite strong assumption, which is not necessarily consistent with the science.

Further, when fitting models to soil moisture data, we found that the SEM model (3) predicts near-surface soil moisture with reasonable accuracy, but fails to explain the more complex behavior of soil moisture in deeper layers. Typically, the soil moisture variations in deeper soil (≥ 15 cm) do not resemble an exponential curve as used in (3).[5]

B. Soil Moisture Indicators

The Variable Infiltration Capacity model with 3 layers (VIC-3L) is widely used to simulate soil moisture data across the world [36]. The VIC-3L model was developed as a generalization of the single layer VIC hydrological model [20]. It is a complicated model involving many parameters accounting for sub-grid variability in soil moisture, land surface vegetation, precipitation, and topography. The Palmer Drought Severity Index (PDSI) and Standardized Precipitation Index (SPI) have also been used to estimate soil moisture [33].

These indicators are not appropriate for either the space- or time-scales of interest to us. A primary limitation of the PDSI is that it cannot be correlated with site-specific water resources such as soil moisture or runoff. Rather, it is a unitless generalized index applicable at the spatial scales of states or counties. Both the PDSI and the SPI rely on data at monthly intervals for long-term (monthly to yearly) assessments of available moisture, whereas we are interested in timescales of hours to asses local runoff and heightened soil moisture and pore-water pressure in the context of landslide susceptibility.

C. Time Series Prediction Models

Aljoumani *et al.* [2] investigated the impact of irrigation on soil water content in silty loam soil using an autoregressive integrated moving average (ARIMA) model. Under the normality assumption, however, the ARIMA model could not properly explain the effect of variable interval irrigation. To remedy this, outlier detection and intervention analysis were used. Unfortunately, their model does not explain physical soil processes in great detail.

Khaertidova and Longobardi [16] analyzed soil moisture dynamics in inter-storm periods. Given their focus on inter-storm periods, there is no analysis of soil moisture increases and a simple exponential decay model is sufficient.

1) Baseline Time Series Model with ARMAX: We use an autoregressive model as a baseline time series prediction method [2]. In particular, we use autoregressive moving average model with exogenous inputs (ARMAX) model, accounting for precipitation as exogenous input terms. In hydrology, ARMAX models have been used for data-driven rainfall-runoff modeling [26].

[5] Experimental results are in Section VI.

Location	Depth (cm)	Dataset Id
Field: Canyon Fire	5	Canyon-5
	15	Canyon-15
	30	Canyon-30
Controlled Experiment: Bucket	10	Bucket-10
	20	Bucket-20
	28	Bucket-28

TABLE I: Summary of six soil moisture data sets used in our data analysis; see Section IV for additional information.

An ARMAX(p,q,b) model expresses the dependence of past soil moisture values $\{M_{t-i}\}$ and rainfall $\{I_{t-l}\}$ on present soil moisture M_t according to the following equation:

$$M_t = \sum_{i=1}^{p} \phi_i M_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \sum_{l=1}^{b} \eta_l I_{t-l} \qquad (5)$$

where (p,q,b) represents the time-delay in autoregressive, moving average, and exogenous input terms respectively. And $\{\phi_i\},\{\theta_j\}$, and $\{\eta_l\}$ are their corresponding weights.

IV. Data Collection and Data Sets

Recent technology improvements have increased the availability of affordable sensors and data loggers that aid in measurement and recording of soil moisture. We now present the soil moisture data sets summarized in Table I.

A. Data Set from the Field

The field experiment data set is composed of soil moisture ($\mathscr{M}$) and rainfall ($\mathscr{I}$) measurements from the Santa Monica mountains near the town of Malibu and Pepperdine University in Southern California [11], [31]. This area was the site of the 2007 Canyon fire. Prior to the fire, the area was covered by chaparral vegetation. However, the fire disturbance removed almost all of the vegetation and changed the soil infiltration properties. Hillslopes within the site are steep with gradients up to 0.9. The colluvial soils are generally less than 0.5m thick, overlying sedimentary rock. These soils have much higher infiltration rates than the underlying bedrock which is composed of Miocene sediments. The sediments produced a range of soil types with a median grain size of 40% to 60% sand. Due to the steep landscape, rapid response from fast hydrologic redistribution is observed following rain events. These factors promote rapid wetting and drying of soil.

Moisture measurements were collected using probes (from Decagon Devices Inc.[6]) measuring volumetric water content (VWC) by estimating the dielectric constant of the media using capacitance/frequency domain technology [17]. These probes were placed at three different soil depths (5cm, 15cm, and 30cm), to represent different soil horizons. Soil moisture measurements were logged every 2 minutes. Tipping-bucket rain gauges provided precipitation data on an irregular schedule in response to rainfall. Soil moisture and rainfall data from December 2007 to April 2008 are shown in Figure 2.

It should be noted, that we only focus on post-fire soil in this paper and do not attempt to model the soil moisture variations in other types of soil. Although, we only discuss one location due to limited space, we tested our methods on soil moisture data collected from a different pit is a nearby location.

[6]http://www.decagon.com/products/soils/volumetric-water-content-sensors/

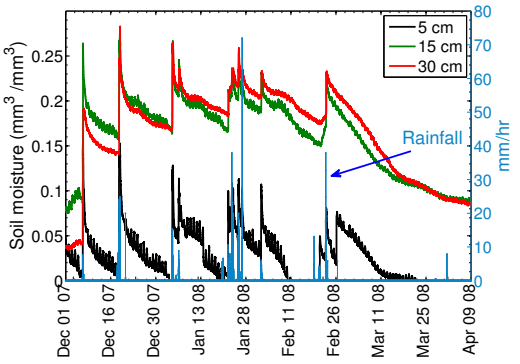


Fig. 2: Rainfall and soil moisture measurements from the Canyon fire data set.

B. Data Set from Controlled Experiment

To study soil moisture dynamics within a controlled experiment, we filled a bucket with sand and gravel. To imitate the post-fire scenario, we also started with very dry material in this experiment.

We performed the experiments in an open space exposed to normal weather conditions. We filled a 5 gallon plastic bucket with sand (grain size < 1mm). To record soil moisture $\mathscr{M}$, we placed three VWC sensors (identical to the Decagon probes used in the field experiment, see Section IV-A), at 10cm, 20cm, and 28cm depths. Two types of data records make up the "rainfall" sequence $\mathscr{I}$. First, we manually added measured amounts of water to the bucket in intervals to bring the soil to a near-saturated condition. Second, some natural rainfall events during the time of the experiment also added water to the bucket. We found the natural rainfall intensities from Weather Underground[7] historical data. For this experiment, there was no vegetation in the soil and the rainfall events were moderate. Overall, the changes in soil moisture are less abrupt compared to the Canyon fire dataset.

V. Soil Moisture Models

The NAR and AEAR models were developed by a multi-disciplinary team consisting of earth scientist and data scientists. The goals of these models are to be physics-based, data-driven, and produce accurate predictions (see the three requirements in Section II). The parameters of these models are estimated from data as reported in the Section VI.

A. Existing Model: Simple Exponential Model (SEM)

We fit the SEM model in (3) to the training data by solving a mean squared error minimization problem:

$$\hat{k}_d = \min_{k_d} \sum_{t=1}^{T} (M_t - M_t^S(k_d; M_{t-1}, I_t))^2. \qquad (6)$$

[7]https://www.wunderground.com/history/

Here, the M- and I-values are in the data, and we optimize to find $\hat{k}_d$ corresponding to k_d in (3). We solve this optimization problem via stochastic methods, see Section VI-B.

The SEM model has certain limitations as discussed in Section V-A. To address these limitations, we now introduce our novel models in Section V-B and Section V-C.

B. Novel Model: Naive Accumulative Representation (NAR)

Tipping bucket rain gauges are commonly used for measuring rainfall. A time series created based on consecutive measurements by this tool is typically discrete and can have multiple values (including 0s) during a continuous rainfall event. Thus, using instantaneous rainfall values as in (3) can create artificial "bumps" in the soil moisture sequence. To remedy this issue, we introduce a novel model which accumulates rainfall values over the prediction horizon τ.

Definition The Naive Accumulative-Representation (**NAR**) model is:

$$M_t^N = M_{t-\tau}^N e^{-k_d \tau} + \sum_{j=0}^{\tau} \Big[\frac{I_{t-j}}{\eta} (1 - e^{-k_w j}) e^{-k_d j} \Big], \tag{7}$$

where k_d is the drying rate and k_w the wetting rate. Thus, we have two constants k_d and k_w in contrast to just one constant in the SEM model. Since rainfall and soil moisture are at different scales, we introduce η as an unknown proportionality constant. Although all rainfall measurements in the prediction horizon τ contribute to an increase in soil moisture, simultaneous evapotranspiration and hydraulic redistribution mechanisms in soil reduce the effect of the rain water. Thus, we add a drying factor $e^{-k_d j}$ in the second term of the equation as well.

To fit the novel **NAR** model to data according to (2), we used the following optimization formulation.

$$\begin{aligned}(\hat{k}_d, \hat{k}_w, \hat{\eta}) = \min_{k_d, k_w, \eta} \sum_{t=1}^{T} (M_t - \\ M_t^N(k_d, k_w, \eta; M_{t-\tau}^N, I_t, \cdots, I_{t-\tau}))^2.\end{aligned} \tag{8}$$

Again, the M- and I-values are in the data, and we optimize to find $\hat{k}_d$, $\hat{k}_w$, and $\hat{\eta}$. In the absence of convexity guarantees, we employ stochastic optimization methods (see Section VI-B).

However, this model has some limitations of (discussed in Section VI). Although the NAR model can fit moisture variations of a shallow soil layer (5cm depth), it performs poorly for the deeper layers (15cm and 30cm depths), which are shielded from evaporation and have finer grained soil types.

C. Novel Model: Additive Exponential Accumulative Representation (AEAR)

We now introduce a sum of exponential functions model, AEAR, which can model a wider range of soil moisture measurements. To obtain the AEAR model in (9), we substitute the single exponential in the first term of the NAR model by a weighted sum of two exponentials:[8]

[8]Using more than two exponentials in the weighted sum was also investigated, but that did not give better results in our case.

Definition The Additive Exponential Accumulative Rainfall (**AEAR**) model is:

$$\begin{aligned}M_t^A = & M_{t-\tau}^A \Big[\alpha e^{-k_s \tau} + (1 - \alpha) e^{-k_g \tau} \Big] \\ & + \sum_{j=0}^{\tau} \Big[\frac{I_{t-j}}{\eta} (1 - e^{-k_w j}) e^{-k_s j} \Big].\end{aligned} \tag{9}$$

The first exponential in (9), with drainage coefficient k_s, represents the steep redistribution decay from the combination of strong suction gradients between wet and dry soil and gravity. The second exponential in (9), with k_g, accounts for the gradual (lower) redistribution decay from low suction gradients dominated by gravity. The relative weighting of these two terms is controlled by a time-varying weight α defined by:

$$\alpha = \sum_{j=0}^{\tau} \Big[\frac{I_{t-j}}{\eta} (1 - e^{-k_w j}) e^{-k_g j} \Big]. \tag{10}$$

After rainfall, the soil suction gradients are strong. But they tend to weaken during a long absence of rain. Thus α needs to be proportional to the rainfall amount. To make the variation of α smooth over time, we use the accumulative rainfall term in (10) instead of instantaneous rainfall values.

The following results illustrates an interesting relationship between the AEAR and ARMAX models.

Theorem 5.1: The AEAR model (Equation 9) reduces to an ARMAX model if α is time-invariant.

Proof Equation 9 can be written as

$$M_t^A = \sum_{i=1}^{\tau} \phi_i M_{t-i}^A + \sum_{j=0}^{\tau} \gamma_j I_{t-j}, \tag{11}$$

where $\phi_1 = \phi_2 = \cdots = \phi_{t-\tau+1} = 0$, and $\phi_{t-\tau} = [\alpha e^{-k_s \tau} + (1 - \alpha) e^{-k_g \tau}]$ are the parameters of the ARMAX autoregressive terms. Here rainfall corresponds to the exogenous input, and the parameters are $\gamma_j = (1 - e^{-k_w j}) e^{-k_s j} / \eta$. If α is time-invariant, $\phi_{t-\tau}$ reduces to an unknown constant. In this case, Equation 11 is identical to the ARMAX model.

However, the reformulation of the AEAR model as an ARMAX model results in absorption of the AEAR parameters k_s and k_g into one time-invariant ARMAX term. Once these parameters are aggregated, the model will lose resemblance to the hydrological processes, where k_s and k_g have distinct meanings as discussed above.

The novel **AEAR** model was also fitted to the data using a mean squared error minimization:

$$\begin{aligned}(\hat{k}_s, \hat{k}_g, \hat{k}_w, \hat{\eta}) = \min_{k_d, k_w, \eta} \sum_{t=1}^{T} (M_t - \\ M_t^A(k_s, k_g, k_w, \eta; M_{t-\tau}^N, I_t, \cdots, I_{t-\tau}))^2, \\ \text{subject to } k_s > k_g.\end{aligned} \tag{12}$$

The M- and I-values are in the data as before, and we estimate $\hat{k}_s$, $\hat{k}_g$, $\hat{k}_w$, and $\hat{\eta}$. We impose the constraint $k_s > k_g$ to reflect the hydrological phenomenon that a suction gradient k_s is stronger than the gradual redistribution k_g. With no convexity guarantees, we use stochastic optimization methods (see Section VI-B) to solve this problem.

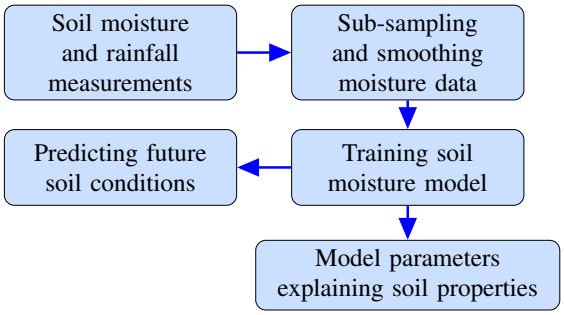


Fig. 3: Steps and outcomes of our soil moisture data analysis; see Section VI-A for further explanations.

VI. Analysis and Prediction Results for Models

In our data analysis, parameters of the mathematical models discussed in Section V were estimated, using the process from Section VI-A and the data discussed in Section IV. Unless otherwise noted, our results are for the Canyon data set.

A. Data Analysis Process

In a wide range of geomorphic, hydrologic, and ecosystem projects, there is a need to derive analytical methods and tools to characterize soil moisture response. Figure 3 shows the steps and data flows of our data analysis process, from the soil moisture measurements, via soil moisture models, to the models' outputs. We now briefly discuss these steps.

The pre-processing step, namely *sub-sampling* followed by *smoothing*, serves multiple purposes. First, it allows us to lower the analysis computational cost via sub-sampling to 20 minutes, since the data collection sampling rate of 2 minutes is higher than needed for our analysis. (This rate was sufficient for fast completion of all analysis and did not lose significant information content in the soil moisture signal.) Second, smoothing reduces the diurnal and other variations while retaining important signals, *i.e.*, the peaks (local maxima) and valleys (local minima).[9] For predicting landslides and runoff-driven erosion, the peaks carry critical information. Third, smooth data obviates the need to prevent fitting of high-frequency diurnal variations as well as other sensor noise and variation. We applied the HyperSTL extrema-preserving smoothing technique [18], [4] to smooth the soil moisture time series without distorting or diminishing the extreme values (peaks and valleys) in the time series.[10]

Next we perform *training of soil moisture models*, using the sub-sampled smooth data. A detailed description of the different models, including our NAR and AEAR models, is presented in Section V. For model fitting, we split a sub-sampled and smoothed soil moisture sequence $\mathscr{M}$ into a training sequence $\mathscr{M}_{\mathscr{T}}$ and a prediction (or test) sequence $\mathscr{M}_{\mathscr{P}}$ as follows: $\mathscr{M}_{\mathscr{T}} = (M_0, \ldots, M_i)$ and $\mathscr{M}_{\mathscr{P}} = (M_{i+1}, \ldots, M_k)$. The wetting and drying of soil is driven by rainfall events. We therefore split the data into $\mathscr{M}_{\mathscr{T}}$ and $\mathscr{M}_{\mathscr{P}}$ as per major rainfall events. We keep the first two wetting and drying cycles in $\mathscr{M}_{\mathscr{T}}$, and the rest in $\mathscr{M}_{\mathscr{P}}$. We assume that accurate rainfall information $\mathscr{I}$ is available throughout.

We train the model of interest by first optimizing and then doing predictions. For example, to estimate the parameters of the AEAR model (9) for a particular data set, AEAR optimization (12) based on the training sequences $\mathscr{M}_{\mathscr{T}}$ and $\mathscr{I}_{\mathscr{T}}$ is first done. Then, we use the trained AEAR model to predict a sequence $\hat{\mathscr{M}}_{\mathscr{P}}$ and evaluate the predictions using the sequence $\mathscr{M}_{\mathscr{P}}$. Our empirical results below show strong prediction results under varying settings.

As the AEAR model is inspired by hydrologic processes in soil, the trained model parameters are *explaining soil properties* like reduction of soil water content by suction gradients and gravity. Section VII discusses such connections between AEAR parameters and soil properties.

B. Optimization Study

In general, there are no convexity guarantees in nonlinear regression. Hence we use stochastic optimization algorithms to fit the models in (3), (7), and (9) to the training data. We study three general optimization algorithms: Real Coded Genetic Algorithm (RGA) [10], Simulated Annealing (SA) [6], and Differential Evolution (DE) [34]. We also study the statistical parameter tuning method IRACE [22].

Evolutionary algorithms including DE and RGA have been successfully employed in hard optimization problems without introducing strong assumptions such as convexity. For example, DE is used for parameter estimation of non-linear models [39]. SA is a probabilistic search technique, and in contrast to DE and RGA it is not population-based. To prevent getting stuck at local optima, SA accepts random neighbors with a small probability. It has been successfully used to solve many difficult data mining problems, such as influence maximization in social networks [13]. IRACE performs iterated racing to automatically tune parameter configurations of an algorithm. IRACE has been successfully used in multiple time-series analysis applications [1].

We use R implementations of the algorithms, present in the packages GA [32] (RGA), GenSA [40] (Generalized SA), DEOptim [24] (DE), and irace [22] (IRACE) respectively. With random initialization and parameters shown in Table II, we ran each of the four optimization algorithms 30 times independently to find the AEAR parameters for the Canyon 30cm soil moisture data set. These parameter settings are a result of tuning experiments designed for efficient optimization. We set the prediction horizon to $\tau = 24$ hours for (6), (8), and (12). We recorded the prediction errors (standard error and maximum absolute error) of the trained models on the test data along

[9] Unfortunately, using dielectric sensors for determining near-surface VWC results in temperature-induced fluctuations for probes inserted into shallow soils. These diurnal variations, along with other disturbances, add a layer of complexity in estimating moisture response functions.

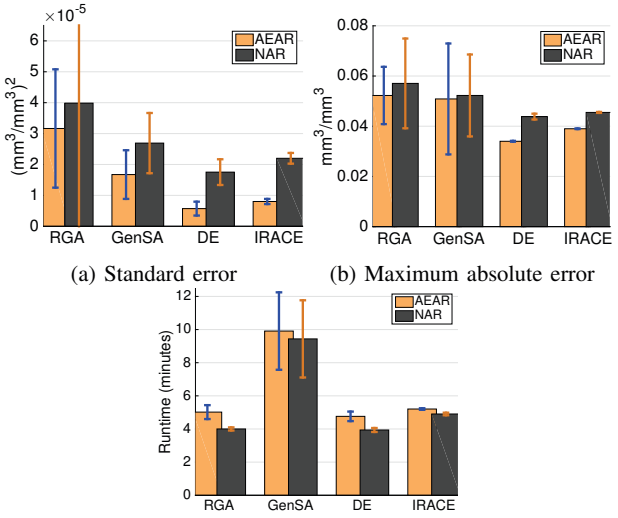
[10] The HyperSTL smoothing technique is based on STL, a seasonal-trend decomposition method using local regression [7]. STL decomposes a time series into trend, seasonal, and remainder components. STL has been used in several areas of science, for example to analyze seasonal patterns in suicides [25] as well as for earth science time series data analysis [4].

Parameter	RGA	GenSA	DE	IRACE
Population size (n_P)	50	1	50	1
Num. generations (n_G)	100	5000	100	5000
Crossover prob. (p_{CR})	0.8	N/A	0.5	N/A
Mutation prob. (p_M)	0.1	N/A	0.8	N/A

TABLE II: Parameter settings for the optimization algorithms.

with the training time. We used a 64 bit Linux computer with an Intel Xeon CPU 3.2GHz processor.

Figure 4 shows the comparative results, in the form of means and standard deviations, for three performance metrics: standard error, maximum absolute error, and training time. We observe that the trained models obtained by DE have minimum prediction errors. Also, the training time is consistently low for DE runs. We conclude that DE performs comparatively better than the other optimization algorithms for this problem, and thus we use DE in the later experiments.



(a) Standard error (b) Maximum absolute error

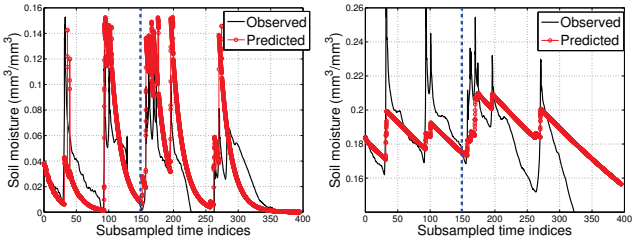
(c) Training time

Fig. 4: Results using 30cm soil moisture test data for AEAR models trained by four different optimization algorithms. The comparison are of prediction error (in (4a) and (4b)) and runtime of training (in (4c)), The mean and variance of all performance metrics are computed over 30 independent runs.

C. Irregular Measurements: NAR and AEAR

Two NAR models were fit separately to 5cm and 15cm soil moisture data. The model predictions are shown in Figure 5. Although the NAR model fits the shallow moisture data (5cm) quite well, it fails for the deeper levels (15cm and 30cm).[11] Apparently, soil moisture variation at deeper levels does not resemble a simple exponential curve. This suggests that the approach is inconsistent with soil moisture response at deeper levels, which respond to additional water transported through the soil column above and possibly from the bedrock below.

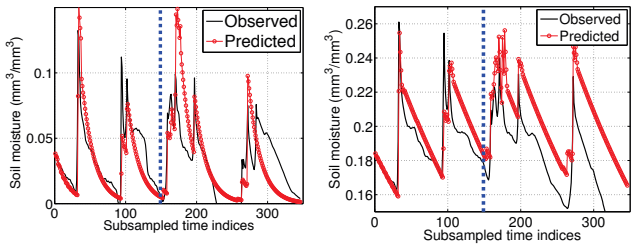
[11] We present only the 15cm results here, and leave out the 30cm results for space reasons.



(a) Predictions at 5cm depth. (b) Predictions at 15cm depth.

Fig. 5: Soil moisture predictions using the NAR model. The points to the left of the blue dashed vertical line are in the training set, and the points to the right are in the test set.

When estimating the AEAR parameters according to (12), the constraint $k_s < k_g$ was implemented by introducing a ratio k_g/k_s and setting $k_g/k_s > 1, k_s > 0$ during optimization. Predicted soil moisture values for the 5cm and 15cm levels using trained AEAR models are shown in Figure 6. The AEAR model fits the 5cm data equally well as the NAR model, but gives much better predictions in the 15cm case.



(a) predictions at 5cm. (b) predictions at 15cm depth.

Fig. 6: Soil moisture predictions using the AEAR model, based on a one-point measurement at $t = 0$.

D. Regular and Irregular Measurements: Prediction Results

The goal here is to compare the SEM, NAR, and AEAR models for both regular (1) and irregular (2) measurements. We also study three ARMAX models of increasing complexity.

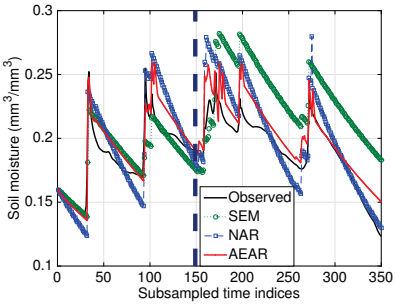
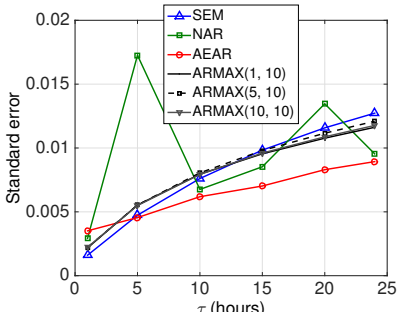
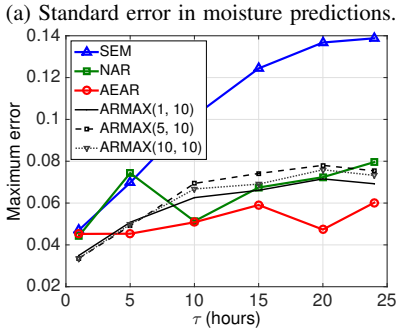


Fig. 7: Comparison of different model predictions (SEM or AWI, NAR, AEAR) at 30cm depth. The standard error in these predictions are 0.037, 0.022, and 0.020 respectively.



(a) Standard error in moisture predictions.



(b) Maximum absolute error in moisture predictions.

Fig. 8: Prediction errors as a function of varying prediction horizon (τ) in regular measurements.

1) Regular Measurements: The prediction horizon τ is a critical parameter in time-series predictions with regular measurements. We trained the previous three models and three new ARMAX models with increasing prediction horizons τ = 1, 5, 10, 15, 20, and 24 hours.

Assuming equal availability of historical soil moisture and rainfall data, we keep the orders of the autoregressive and exogenous input polynomials (as in Equation 5) the same, *i.e.*, $p = b$. Hence we suppress the third order parameter of ARMAX(p,q,p) model and denote it as ARMAX(p,q). We use standard system identification methods [21] to learn the parameters from our training data.[12]

We present the error on the test data of trained models for different prediction horizons (τ) in Figure 8, which contains a key message of this paper. Clearly, the standard error of existing SEM model is smallest for $\tau = 1$, however for this short time horizon, all the models perform quite well. The more complicated task is predicting for longer time horizons τ, and SEM's standard error increases with τ at a much higher rate compared to AEAR's error. The ARMAX models show similar error-growth as the SEM model. For grater time horizons, $\tau \geq 10$, our AEAR model shows minimum standard error among all models.

A clear divergence between the maximum absolute-errors (with increasing τ) of the existing SEM model and the novel accumulative rainfall models (NAR and AEAR) is observed in Figure 8b. The maximum error of the SEM model exceeds 0.12 at τ = 15. As the variations of soil moisture in our dataset lies within 0.3 $mm^3\,mm^{-3}$, more than 0.1 $mm^3\,mm^{-3}$ off predictions imply higher than 33% error. Moreover, a 0.3 $mm^3\,mm^{-3}$ difference in soil suggests very different soil conditions [8]. In contrast, AEAR's error curve is much more flat and it remains within 0.06 $mm^3\,mm^{-3}$ while predicting moisture values a day ahead ($\tau = 24$). This experiment suggests a strong capability of the AEAR model to forecast soil moisture at time horizons $\tau = 5$ to $\tau = 24$ hours, another major modeling improvement.

Although the ARMAX models perform better than the SEM and NAR models in terms of maximum absolute error, the error growth-rates are higher than for AEAR. Hence the AEAR model, with parameters directly related to hydrological processes, shows superiority over more complicated ARMAX models with a higher number of parameters.

2) Irregular Measurements: A comparison of all three trained models with 30cm soil moisture data is shown in Figure 7.[13] Although the existing SEM model fits the training data $\mathscr{M}_{\mathscr{T}}$ fairly well, it does not adequately represent the moisture variation in the test data $\mathscr{M}_{\mathscr{P}}$. The NAR model attempts to find an average exponential model around the observed variations. In contrast, the AEAR model neither overfits the training set, nor deviates from the observed moisture substantially in the test data. The prediction error of the AEAR at the moisture peaks is bounded by 0.04 $mm^3\,mm^{-3}$, which falls within the range of typical accuracy of various volumetric water content sensors using a factory calibration [14], [17].

E. Bucket Data: Prediction Results

We now consider data resulting from the controlled experiment, and compare models using regular measurements. Among the ARMAX models we choose ARMAX(10, 10), which performed best for the Canyon data set.

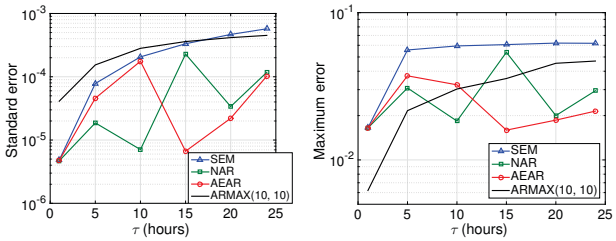
Similar to Section VI-D1, we use the trained SEM, NAR, AEAR, and ARMAX models to predict the soil moisture at 10cm depth. Figure 9 shows the standard error and maximum absolute error of all four models on the test dataset. We observe that the growth rates of NAR and AEAR models' errors are lower than SEM and ARMAX models. Although ARMAX shows low (maximum) error for $\tau = 1,5$ hours, in case of larger prediction horizons $\tau \geq 10$ hours NAR and AEAR models achieve lower error. Although the performance of these two accumulative representations are similar, AEAR predicts slightly more accurately for $\tau \geq 15$ hours.

VII. Validation of Models by Earth Scientists

Both the AEAR and NAR models are intended to represent hydrologic processes in soil and fit well to the soil moisture data. The experimental results in Section VI suggest that the AEAR better models the data sets across multiple layers of soil and also gives better predictions. Consequently, we

[12]Our experiments use the functions implemented in the *tsa* R package.

[13]The irregular setting is not appropriate for ARMAX models, due to high computational complexity of model fitting. Therefore, we do not compare with ARMAX here, unlike what was done for the regular measurements.



(a) Standard error in moisture predictions. (b) Maximum absolute error in moisture predictions.

Fig. 9: Prediction errors as a function of varying τ (in regular measurements) for experimental data.

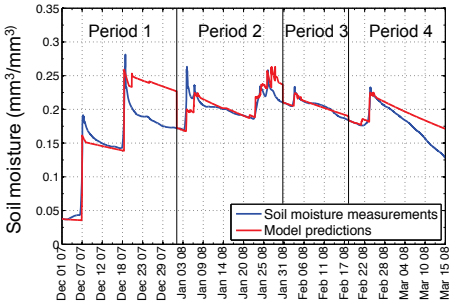
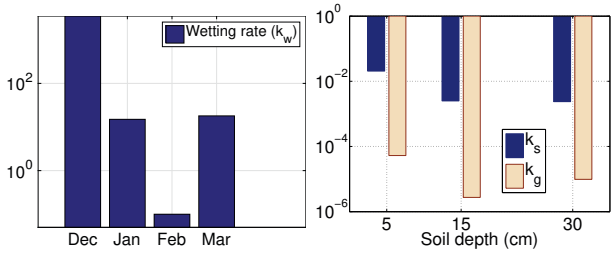


Fig. 10: Split of Canyon Soil moisture data from 30cm deep soil into periods and predictions of trained AEAR models within each period.



(a) Monthly variation of wetting rates. (b) Variation of drying rates with soil depth.

Fig. 11: Wetting and drying properties of soil.

recommend the AEAR model, and study in this section the variation of the estimated AEAR parameters with season and soil depth for Canyon data. These variations reflect the varying nature of wetting and drying cycles of soils, hysteresis of additive cycles, and changing ground surface conditions.

A. Variation with Time

Right after the Canyon fire, the soil was extremely dry. After two rain events in December, the soil properties changed significantly. Thus the moisture response for the first two rainfall events are hypothesized to be too different to be captured by one average model. We therefore split the data into four time periods, see Figure 10, bracketing rainfall from storms and related increases and subsequent soil moisture decreases. Each period covers about one month. However, Period 3 ("February") ends early, in order not to interfere with a storm event near February 23.

We trained separate models for each period, see Figure 10 and Figure 11. A large variation in the wetting rate k_w is observed in the four periods following the wildfire (Figure 11a). At the beginning of Period 1, extremely dry soil absorbs rain almost instantaneously, leading to a very high wetting rate k_w. As soil pores get filled with water and fine-grained sediment, the wetting rate decreases and eventually goes below 1 in Period 3. Then k_w increases in Period 4 when the soil starts drying again, and post-fire vegetation regrowth decreases soil moisture by evapotranspiration.

B. Variation with Soil Depth

Drying of soil at different depths varies due to various environmental factors such as solar radiation penetration, air temperature, and water absorption by vegetation. The AEAR parameters k_s and k_g of trained models from three different soil depths can potentially capture these effects. Figure 11b shows the empirical variation of k_s and k_g with soil depth. We observe that the drying rates at the shallowest layer (5cm) are much higher compared to the deeper layers. This reflects coarser textured soils, high solar radiation, and atmospheric effects near the ground surface. Moreover, k_s and k_g differ by at least two orders of magnitude consistently over all soil depths. This suggests that our choice of two distinct drying terms in the AEAR model (Equation 9) is well-justified.

We also attempted to use black box machine learning models, such as Gaussian Processes, for predicting soil moisture. Based on initial experiments, we concluded that these models tend to overfit quickly due to high model complexity and thus perform poorly when forecasting $\tau = 12$ hours or longer ahead.

VIII. Conclusion and Future Work

We modeled soil moisture response with respect to rainfall from natural storms and determined that the existing AWI based approach cannot model moisture response for time periods greater than 5–10 hours, depending on the data set. Moreover, it does not perform well for deeper soil layers in a post-fire setting, because soil moisture variations in the deeper layers do not follow a simple exponential curve. We developed a novel AEAR model which can fit moisture data from three different soil layers with distinct soil textures. Our AEAR model can be trained with both regular and irregular time series measurements of soil moisture and predicts well in both cases. Moreover, this model is consistent with hydrological processes in soil, as validated by earth scientists, and provides a means to estimate the rates of soil wetting during rain storms and how fast soils may dry in-between storms.

We plan to extend our methods to better incorporate soil moisture variations in different layers of soil, *i.e.,* variation with depth. We will attempt to build upon the model to predict moisture in a broad area of soil, creating a moisture map. Combined with field and remote sensing of landscapes, as

well as storm intensity and duration forecasts, this will enable improvements in forecasts, including landslide forecasts. We can also use grammar-based models to generate interpretations of the predicted time-series and the model parameters [19].

ACKNOWLEDGMENT

This research was, in part, supported by USGS Innovation Center[14] and Cooperative Agreement G15AC00002 between Carnegie Mellon University and U.S. Geological Survey.

REFERENCES

[1] H.-G. Acosta-Mesa, F. Rechy-Ramírez, E. Mezura-Montes, N. Cruz-Ramírez, and R. H. Jiménez. Application of time series discretization using evolutionary programming for classification of precancerous cervical lesions. *Journal of biomedical informatics*, 49:73–83, 2014.

[2] B. Aljoumani, J. A. Sànchez-Espigares, N. Cañameras, R. Josa, and J. Monserrat. Time series outlier and intervention analysis: Irrigation management influences on soil water content in silty loam soil. *Agricultural water management*, 111:105–114, 2012.

[3] M. Arlitt and T. Jin. A workload characterization study of the 1998 world cup web site. *IEEE Network*, 14(3):30–37, May 2000.

[4] A. Basak, O. J. Mengshoel, C. Kulkarni, K. Schmidt, P. Shastry, and R. Rapeta. Optimizing the decomposition of time series using evolutionary algorithms: soil moisture analytics. In *GECCO*, pages 1073–1080, 2017.

[5] S. Bhatia, G. Mohay, D. Schmidt, and A. Tickle. Modelling web-server flash events. In *11th IEEE International Symposium on Network Computing and Applications (NCA-12)*, pages 79–86, 2012.

[6] I. O. Bohachevsky, M. E. Johnson, and M. L. Stein. Generalized simulated annealing for function optimization. *Technometrics*, 28(3):209–217, 1986.

[7] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

[8] S. R. Evett, R. C. Schwartz, J. A. Tolk, and T. A. Howell. Soil profile water content determination: spatiotemporal variability of electromagnetic and neutron probe sensors in access tubes. *Vadose Zone Journal*, 8(4):926–941, 2009.

[9] J. W. Godt, R. L. Baum, and A. F. Chleborad. Rainfall characteristics for shallow landsliding in Seattle, Washington, USA. *Earth Surface Processes and Landforms*, 31(1):97–110, 2006.

[10] D. E. Golberg. *Genetic algorithms in search, optimization, and machine learning*. Addion Wesley, 1989.

[11] M. Hanshaw, K. Schmidt, D. Jorgensen, and J. Stock. By air and land: estimating post-fire debris-flow susceptibility through high-resolution radar reflectivity and tipping-bucket gage rainfall. In *AGU Fall Meeting Abstracts*, pages H51D–00850, 2008.

[12] J. Huang, H. M. van den Dool, and K. P. Georgarakos. Analysis of model-calculated soil moisture over the united states (1931-1993) and applications to long-range temperature forecasts. *Journal of Climate*, 9(6):1350–1362, 1996.

[13] Q. Jiang, G. Song, G. Cong, Y. Wang, W. Si, and K. Xie. Simulated annealing based influence maximization in social networks. In *AAAI*, volume 11, pages 127–132, 2011.

[14] A. L. Kaleita, J. L. Heitman, and S. D. Logsdon. Field calibration of the theta probe for des moines lobe soils. *Applied engineering in agriculture*, 21(5):865, 2005.

[15] K. Kersting, Z. Xu, M. Wahabzada, C. Bauckhage, C. Thurau, C. Roemer, A. Ballvora, U. Rascher, J. Leon, and L. Pluemer. Pre-Symptomatic Prediction of Plant Drought Stress Using Dirichlet-Aggregation Regression on Hyperspectral Images. In *AAAI - Computational Sustainability and AI track*, 2012.

[16] E. Khaertidova and A. Longobardi. Analysis of inter-storm period soil moisture dynamics. *Procedia Environmental Sciences*, 19:208–216, 2013.

[17] F. Kizito, C. Campbell, G. Campbell, D. Cobos, B. Teare, B. Carter, and J. Hopmans. Frequency, electrical conductivity and temperature analysis of a low-cost capacitance soil moisture sensor. *Journal of Hydrology*, 352(3):367–378, 2008.

[18] C. Kulkarni, O. Mengshoel, A. Basak, and K. Schmidt. Optimizing the decomposition of soil moisture time-series data using genetic algorithms. In *AGU Fall Meeting Abstracts*, pages IN23C–1741, 2015.

[19] R. Lee, M. J. Kochenderfer, O. J. Mengshoel, and J. Silbermann. Interpretable categorization of heterogeneous time series data. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pages 216–224. SIAM, 2018.

[20] X. Liang, D. P. Lettenmaier, E. F. Wood, and S. J. Burges. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal Of Geophysical Research*, 99:14–415, 1994.

[21] L. Ljung. System identification: theory for the user. *Englewood Cliffs*, 1987.

[22] M. López-Ibánez, J. Dubois-Lacoste, T. Stützle, and M. Birattari. The irace package, iterated race for automatic algorithm configuration. Technical report, Citeseer, 2011.

[23] O. J. Mengshoel, R. Desai, A. Chen, and B. Tran. Will we connect again? machine learning for link prediction in mobile social networks. In *Proc. of Eleventh Workshop on Mining and Learning with Graphs*, Chicago, IL, August 2013.

[24] K. Mullen, D. Ardia, D. L. Gil, D. Windover, and J. Cline. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011.

[25] T. Partonen, J. Haukka, H. Nevanlinna, and L˙ Analysis of the seasonal pattern in suicide. *Journal of Affective Disorders*, 81(2):133–139, 2004.

[26] R. Remesan and J. Mathew. *Hydrological data driven modelling: a case study approach*, volume 1. Springer, 2014.

[27] L. A. Richards. Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, 1(5):318–333, 1931.

[28] D. Robinson, C. Campbell, J. Hopmans, B. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth. Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7(1):358–389, 2008.

[29] C. Roncoli, K. Ingram, and P. Kirshen. Reading the rains: local knowledge and rainfall forecasting in burkina faso. *Society &Natural Resources*, 15(5):409–427, 2002.

[30] K. Schmidt, M. Hanshaw, J. Howle, J. Kean, D. Staley, J. Stock, and G. Bawden. Hydrologic conditions and terrestrial laser scanning of post-fire debris flows in the San Gabriel Mountains, CA, USA. In *Proceedings of the fifth international conference on debris flow hazards mitigation/mechanics, prediction, and assessment, Padua, Italy*, pages 583–593, 2011.

[31] K. Schmidt, J. Stock, M. Hanshaw, and G. Bawden. Constraining diffusivity and critical slope from post-fire sediment flux of the day, Canyon, and Corral Fires, California. In *AGU Fall Meeting Abstracts*, page 1079, 2008.

[32] L. Scrucca. GA: a package for genetic algorithms in R. *Journal of Statistical Software*, 53(4):1–37, 2012.

[33] A. P. Sims, S. Raman, et al. Adopting drought indices for estimating soil moisture: A North Carolina case study. *Geophysical Research Letters*, 29(8):24–1, 2002.

[34] R. Storn and K. Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[35] P. K. Sundararajan, E. Feller, J. Forgeat, and O. J. Mengshoel. A constrained genetic algorithm for rebalancing of services in cloud data centers. In *2015 IEEE 8th International Conference on Cloud Computing*, pages 653–660, June–July 2015.

[36] C. Tang and T. C. Piechota. Spatial and temporal soil moisture and drought variability in the upper Colorado river basin. *Journal of Hydrology*, 379(1):122–135, 2009.

[37] H. Van den Dool, J. Huang, and Y. Fan. Performance and analysis of the constructed analogue method applied to US soil moisture over 1981–2001. *Journal of Geophysical Research: Atmospheres*, 108(D16), 2003.

[38] R. C. Wilson and G. F. Wieczorek. Rainfall thresholds for the initiation of debris flows at La Honda, California. *Environmental & Engineering Geoscience*, 1(1):11–27, 1995.

[39] H. Wu, H. Zhu, H. Miao, and A. S. Perelson. Parameter identifiability and estimation of HIV/AIDS dynamic models. *Bulletin of mathematical biology*, 70(3):785–799, 2008.

[40] Y. Xiang, S. Gubian, B. Suomela, and J. Hoeng. Generalized Simulated Annealing for Global Optimization: the GenSA Package. 2013.

[14] https://geography.wr.usgs.gov/InnovationCenter/