Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-121-RC2, 2019

Hydrology and Earth System Sciences
Discussions
Open Access

Interactive comment on "Characterization of Hillslope Hydrologic Events Using Machine Learning Algorithms" *by* Eunhyung Lee and Sanghyun Kim

Anonymous Referee #2

Received and published: 14 June 2019

The authors propose the use of machine learning methods to identify clusters in hillslope scale recharge patterns during rainfall events at forested hillslope in Korea. The idea is to explain those by different hydrological process patterns and to identify monitoring sites which are most representative for the recharge clusters. The latter are defined as those with the highest potential for unsupervised machine learning. The underlying data base consists of 10 years hourly through fall and half hourly soil moisture observations in 10, 30 and 60 cm depths in five replicated upslope and downslope profiles.

While I am very positive about the core idea underlying this study, I regret to say that

Printer-friendly version

Discussion paper

present implementation leaves quite a few open doors. These need to be closed by improving the presentation quality of the manuscript but also the scientific quality of the underlying analysis.

1) The title should better reflect that manuscript content. The focus is on soil moisture recharge events, not on rainfall runoff events, which was my initial expectation.

2) Sorry to be pedantic, but I think authors need to define a recharge event in their soil moisture data set. Particularly, the endpoint is not so easy to define. And I would love to see the soil water content time series to get a feeling for the events, and how they look like.

3) Is a dataset of roughly 400 events observed at 30 different locations big enough for machine learning? In this respect I wonder how the inferred clusters will change when reducing the length of their data set?

4) I am not sure whether I properly understood Eq.1. Does the spatial average relate to a constant depth? In this case it describes spatial fluctuations around a spatial average at a fixed time based on a sample size of 10. The estimation variance of the average still pretty large. Do the authors regard this sample as large enough to characterize the average soil water content of the hillslope?

5) Or does Eq. 1 relate to all sensors in all depth? This makes for me not too much sense as the soil water content in different depths belongs to different ensembles! The soil is a low pass filter. I also wonder whether up and downslope soil water content belong to the same ensemble. As water flows downslope, downslope sensors might experience wetting by subsurface flow and infiltration, the upslope ones not. This speaks for different ensembles. This also comes into my mind when looking at Figure 3. DO 5-30 is surely another ensemble than UP 3-10.

6) The robustness of this index requires that the measurement errors of individual observations and the error of the average do not overlap? This is also important for

judging whether clusters are properly separated or not.

7) I am not sure, whether the MRDi does what it should. It can be zero for a sensor if it alternately shows positive and negative deviation from the spatial average, which are of the same magnitude. This surrogates stability, which is not there.

8) For me the most stable sensor has the smallest dTheta/dt. The presented indices are tailored to pick the sensor closed to the spatial mean. This is not the same, particularly if the average is seasonally changing. As said, it would be very helpful to provide soil moisture time series for the observation period. This would also allow the reader to judge the temporal stability of soil water content visually.

9) Reusser et al. (2009) used SOMS to cluster error measures into groups. What was very helpful to interpret the SOMS areas was their use of well-defined errors, which were classified into different parts of the SOM. Based on that the authors came up with a prosaic description of the errors/clusters (peak to small, recession to long, negative volume bias etc). Maybe the authors can show typical events in the different cluster, to make those much more intuitive to the readers.

Technical points: - Line 64: “can be differently appeared”- please reformulate

- Line 64: “The functional relationship between 64 rainfall and soil water storage had been studied (Brocca et al., 2005; Castillo et al., 2003; Xie and Yang, 2013), but how the rainfall features such as rainfall amount, intensity, duration and antecedent soil moisture condition influence hydrological processes and their distributions at the hillslope scale had not been explored yet”. Please be precise hydrological processes is too broad.

- Line 69: This is Wienhöfer and Zehe (2014),

- Machine learning techniques, particularly SOMS have been extensively used in the field of model diagnostics (e.g. Herbst and Casper HESS 2008; Reusser et al. HESS 2009).

- Please make sure that figure captions are informative, currently the information content is often too low.

- Line 272: typo higher

- Figure 3: DO5 30 has soil moisture values larger than 50%, what is the porosity of the soil?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-121, 2019.