Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-121-AC2, 2019

Hydrology and Earth System Sciences
Open Access
Discussions

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

Eunhyung Lee and Sanghyun Kim

kimsangh@pusan.ac.kr

Received and published: 6 July 2019

Thanks for your comments. Please find our responses for two reviewer's comments as follows.

Responses to Reviewer 1's comments 1. Stability indices: The study performs an analysis of temporal soil moisture variability. Thus the "Index of Temporal Stability" (ITS) is introduced. Unfortunately, it is defined in a really counterintuitive and misleading way: It measures exactly the opposite of what the term suggests (l. 155-158). Please rename to "Instability index" or the like.

The term "Index of Temporal Stability (ITS)" had been widely used by many researchers who had worked on spatial distribution of soil moisture measurements. Following papers (16) used ITS for soil moisture variability.

Zhao, Y., S. Peth, X. Y. Wang, H. Lin, and R. Horn (2010), Controls of surface soil moisture spatial patterns and their temporal stability in a semi-arid steppe, Hydrol. Process, 24, 2507-2519. Zhao, W., Z. Cui, J. Zhang, and J. Jin, 2017, Temporal stability and variability of soil-water content in a gravel-mulched field in northwestern China, J. Hydrol., 552, 249-257. Zhang, P., and M. Shao (2013), Temporal stability of surface soil moisture in a desert area of northwestern China, J.Hydrol., 505, 91-101. Penna, D., L. Brocca, M. Borga, and G. D. Fontana (2013), Soil moisture temporal stability at different depths on two alpine hillslopes during wet and dry periods, J. Hydrol. 477, 55-71. Biswas, A., and B. C. Si (2011), Scales and locations of time stability of soil water storage in a hummocky landscape, J. Hydrol., 408, (1-2), 100-112. Gao, L., and M. Shao (2012a) Temporal stability of soil water storage in diverse soil layers, Catena, 95, 24-31. Gao, L., and M. Shao (2012b) Temporal stability of shallow soil water content for three adjacent transects on a hillslope, Agri. Water Manag., 110, 41-54. Gao, L., M. Shao, X. Peng, and D. She (2015) Spatio-temporal variability and temporal stability of water contents distributed within soil profiles at a hillslope scale, Catena, 132, 29-36. Gomez-Plaza, A., J. Alvarez-Rogel, J. Albaladejo, and V. Castillo (2000), Spatial patterns and temporal stability of soil moisture across a range of scales in a semiarid environment, Hydrol. Process., 14, 1261-1277. Li, D. F., and M. Shao (2014), Temporal stability analysis for estimating spatial mean soil water storage and deep percolation in irrigated maize crops, Agric. Water Manage., 144, 140-149. Junquira Junior, J. A., C. R. Mello, P. R. Owens, J. M. Mello, N. Curi, and G. J. Alves (2017), Time-stability of soil water content in an atlantic forest-latosol site, Geoderma, 288, 64-78. Lee, E., and S. Kim (2017), Pattern similarity based soil moisture analysis for three seasons on a steep hillslope, J. Hydrol., 551, 484-494. Vachaud, G., P.P. de Silans, P. Balabanis, M. Vauclin, 1985, Temporal stability of spatially measured soil water probability density function, Soil Sci. Soc. Am. J., 49, 822-828. Brocca, L., F. Melone, T. Moramarco, and

R. Morbidelli, 2009, Soil moisture temporal stability over experimental areas in Central Italy, Geoderma, 148(3-4), 364-374. Xiaoxu, J., M. Shao, X. Wei, and Y. Wang, 2013, Hillslope scale temporal stability of soil water storage in diverse soil layers, J. Hydrol., 498, 254-264. Hu, W., L.K. Tallon, and B.C. Si, 2012, Evaluation of time stability indices for soil water storage upscaling, J. Hydrol, 475, 229-241. Z. Liu, Y. Wang, P. Yu, A. Tian, Y. Wang, W. Xiong, and L. Xu, 2018, Spatial pattern and temporal stability of root-zone soil moisture during growing season on a larch plantation hillslope in Northwest China, Forests, 9, 68.

Therefore, changing ITS as "instability index" seems not appropriate simply because majority of hydrology community used ITS for many years.

Secondly, the mathematical definition is unnecessarily complex and hardly comprehensible (equation 2). Why do you add squared mean and variance of normalized differences (δ)? What information does that index provide that would not already given, e.g., by the mean of normalized differences δ? In addition, both ITS and the normalized difference δ (equation 1) can easily be mixed up with the "soil moisture difference index" $\Delta\theta$ (equation 4). The latter seems to have been used only for the SOM classification. Why do you need that many different indices for soil moisture variation? This is very confusing and not comprehensible for the reader.

We referred widely used indices, such as ITS and SDRD, for soil moisture representation. Of course, we used equation (4) for SOM. We presented Its or other terms because we will compare the performance of SOM in delineating the representative point and those of other existing approaches. We think presenting new development should be done based on the context of existing studies. Of course, we used it for delineating the representative point for antecedent soil moisture content, which is an important input for SOM application. Through this way, we may highlight what is new and how our approach can contribute in analyzing soil moisture data.

2. Classification of hydrological events: Assignment of hydrological events is a crucial

point in this study, as clusters were assigned to processes. Unfortunately, the clustering is a very weak point of the study. First of all, I do not understand why an SOM has been performed prior the cluster analysis. I acknowledge that SOM is a nice tool to visualize multivariate data sets (cf., Fig. 5). However, assigning data to the nodes of an SOM is not without loss of information, as even slightly different instances of the data set are assigned to the same node. Thus I would highly recommend performing the cluster analysis directly on the data rather than on their aggregated representation in an SOM.

Considering complex soil moisture dynamics for numerous rainfall events in hillslope area, two primary objectives in applying self-organizing map to rainfall and soil moisture dataset are as follows; one is the characterization of non-linear relationships among hydrologic variables and clustering of rainfall events based on trained self-organizing map.

Firstly, the visualization capability of self-organizing map, which you mentioned above, is much better than other conventional approaches such as PCA (Principle component analysis) and MDS (Multi-dimensional scaling) and other alternatives, such as linear tools which previously frequently used (Reusch et al., 2005; Liu et al., 2006). Especially, previous researches reported that 2-dimensional space of self-organizing map provides more reliable data representation compared to the other linear dimension reduction approach (Brosse et al., 2001; Chon et al., 1996; Hilton and Salakhutdinov, 2006; Pinto et al., 2008; Reusch et al., 2007).

Secondly, the clustering of rainfall events based on trained self-organizing map is the other main objective. The representative direct clustering method consists of two parts like Hierarchical clustering (Single linkage, Average linkage, Centroid method, Ward's method, Kth Neighbor) and Non-hierarchical clustering(K-means algorithm, K-median algorithm).

Comparison of performance between self-organizing map and direct clustering methods have frequently been reported in previous researches. The self-organizing network is superior to hierarchical clustering in both sensitiveness for noisy dataset and accuracy aspects (Mangiameli et al., 1996; Vesanto and Alhoniemi, 2000; Samsonova et al., 2005; Abbas, 2008; Sivasankari et al., 2014). Using hierarchical clustering approach with dendrogram, it was very difficult to characterize the cluster characteristics as well as to select optimized cluster number for researcher's objective (Samsonova et al., 2005).

The figure (below) is the agglomerative hierarchical clustering results (ward method) dendrogram which was suggested by reviewer. Extracting or finding any useful explanation from this figure is almost impossible. The cluster number obtained from the coefficient delta (see Table) was two, which is almost useless in terms of hydrological interpretation. Therefore, the self-organizing map is comparably feasible to recognize the relationship among rainfall events. This was pointed out by many researchers (Mangiameli et al., 1996; Vesanto and Alhoniemi, 2000; Samsonova et al., 2005; Abbas, 2008; Sivasankari et al., 2014).

Figure for cluster dendrogram

Table for cluster number and delta

Furthermore, many previous researches reported that even non-hierarchical approaches showed worse performances in clustering accuracy compared to self-organizing map (Jassar and Dhindsa, 2015; Dhingra et al., 2013; Toor and Singh, 2013; Abbas, 2008; Kiang and Kumar, 2001). The performance of non-hierarchical method is very sensitive to the heuristic selected initial setting (Jassar and Dhindsa, 2015), but self-organizing map is less prone to the change in initial network setting (Kiang and Kumar, 2001). According to characteristics of dataset, the comparison of performance among clustering method can be changed, but many relevant scientists have proved both applicability and higher performance of self-organizing in many disciplines. This means that the clustering with the SOM instead of directly clustering the data is not only computationally effective approach but also have better or comparable performance with significantly reduced uncertainty and much better robustness compared to direct clustering method (Vesanto and Alhoniemi, 2000).

Brosse, S., J.L. Giraudel, and S. Lek, 2001, Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages, Ecological Modelling, 146, 159-166. Reusch, D.B., R.B. Alley, and B.C. Hewitson, 2005, Relative performance of self-organizing maps and principal component analysis in pattern extraction from synthetic climatological data, Polar Geogr, 29(3), 227-251. Hinton, G.E., and R.R. Salakhutdinov, 2006, Reducing the dimensionality of data with neural networks, Science, 313(5786), 504-507. Liu, Y., R.H. Weisberg, and C.N.K. Mooers, 2006, Performance evaluation of the self-organizing map for feature extraction, J. Geophys. Res., 111, C05018. Pinto, N., D.D. Cox, and J.J. DiCarlo, 2008, Why is real world visual object recognition hard?, PLoS Computational Biology, 4(1), 0151-0156. Chon, T.S., Y.S. Park, K.H. Moon, E. Cha, and Y. Pa, 1996, Patternizing communities by using an artificial neural network, Ecological Modelling, 90, 1403-1409. Reusch, D.B., R.B. Alley, and B.C. Hewitson, 2007, North Atlantic climate variability from a self-organizing map perspective, Journal of geophysical research, 112, D02104, doi:10.1029/2006JD007460. Abbas, O.A., 2008, Comparison between data clustering algorithms, the international arab journal of information technology, 5(3), 320-325. Mangiameli, P., S.K. Chen, and D. West, 1996, A comparison of SOM neural network and hierarchical clustering methods, European Journal of Operational Research, 93, 402-417. Toor, A.K. and A. Singh, 2013, Analysis of clustering algorithms based on number of clusters, error rate, computation time and map topology on large dataset, International Journal of Emerging Trends & Technology in Computer Science, 2(6), 94-98. Samsonova, E., T. Back, J.N. Kok, and A.P. IJzerman, 2005, Reliable hierarchical clustering with the self-organizing map, in Proc. 6th International Symposium on Intelligent Data Analysis, 397-413. Dhingra, S., R. Gilhotra, and Ranishanker, 2013, Comparative analysis of kohonen-SOM and k-means data mining algorithms based on academic activities, International Journal of Computers & Technology, 6(1), 237-241. Vesanto, J., and E. Alhoniemi, 2000, Clustering of the self-organizing map, IEEE TRANSACTIONS ON

NEURAL NETWORKS, 11(3), 586-600. Kiang, M.Y., and A. Kumar, 2001, An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications, Information systems research, 12(2), 177-194. Sivasankari, A., S. Sudarvizhi, and S.R.A. Bai, 2014, comparative study of different clustering and decision tree for data mining algorithm, International Journal of Computer Science and Information Technology Research, 2(3), 221-232. Jassar, K.K., and K.S. Dhindsa, 2015, Comparative study and performance analysis of clustering algorithms, International Journal of Computer Applications, 0975-8887, 1-6.

Secondly, the number of clusters was selected in "a heuristic approach aiming to achieve a hydrologically meaningful classification of events and parsimonious clustering" (l. 289-290) without any clearly defined and understandable criterion. Likewise, no criterion is given for deciding on the superiority of one cluster approach compared to another (l. 211-213). Thus arbitrary decisions seem to have a major effect on the assignment to clusters, and subsequently to hydrological processes. Thanks for your comment. One important criterion was the parsimony in clustering with hydrologically meaningful event classification. We need to make 7 clusters to identify difference between cluster 5 and 6. As explained in the context, the identification of cluster 5 provide completely distinct interpretation (preferential flow) in hillslope hydrological process. We can divide SOM maps into 3 parts or 7 parts depending on the distinction of vertical flow and lateral flow in upslope and downslope (see Tables in paper). But classifying SOM maps as 7 numbers was found to extracting a meaningful hydrologic information about hydrological process characteristics in this hillslope. Further classification looks redundancy in terms of physically meaningful interpretation. The criterion cannot be given as an explicit formula (because natural process can be formulated) but existences of distinct hydrological processes (vertical flow and lateral flow) either upslope or downslope as criterion and interpretation for as illustrated in Table

2. The distribution of soil moisture indices and soil water storage (Table 1 and Table 2) and its statistical (volumetric soil moisture) results in Figure 8 were made from soil

moisture measurements. These are not postulation but evidence. These results are not arbitrary decision but the summary of comprehensive analysis of soil moisture and rainfall for 396 events.

Of course, this is heuristic approach, simply because all clustering algorithms are heuristic approach. As we mentioned, both hierarchical and non-hierarchical method also need heuristic approach for selecting cluster number in dendrogram and initial setting.

Last but not least, according to Fig. 5, and except for antecedent soil moisture and rainfall (?) duration. there is a rather smooth and continuous increase of the soil moisture difference indices. The only exception concerns cluster 5 and 6. In cluster 5, the soil moisture difference index at the upslope sites below 10 cm depth is clearly less compared to the downslope sites, whereas the soil moisture difference index is high both at upslope and downslope sites in cluster 6. Except for these two clusters the cluster analysis obviously subdivides the data set more or less arbitrarily along a single continuous gradient rather than identifying clearly distinct groups (cf., Fig. 8).

Thanks for comment. We partially agreed to your comment for continuous increase of soil moisture difference index. What you pointed out is somewhat natural in SOM classification simply because the distance between the input vector and weighting vector is mainly used for SOM classification (Actually this trend is common for all SOM application). The response of soil moisture on rainfall events consists of rainfall characteristics, antecedent soil water condition, and soil moisture difference index. Although components in same category have inter-connection, every component individually represents each characteristic which we would like to show multiple relationships among components. The dendrogram with dissimilarity matrix among components vectors in SOM map shows distinctively the available number of clusters. Among every cases of clusters, our choice considering hydrological processes was also heuristic, but the optimized number of cluster was 7 for identifying the difference between cluster 5 and cluster 6.

Identification of representative soil monitoring points: A major topic of the study is identification of “representative soil moisture monitoring points” (e.g., l. 30-31). It remains unclear in what regard these sites should be representative. According to l. 50-51 “high stability is an important criterion for determining the best location for the monitoring spatially averaged soil moisture of a given area”. The study focuses on the temporal dynamics. But here the term “representativeness” cannot refer to the temporal dynamics because “high stability points” would systematically underestimate that dynamics. On the other hand, “high stability” points could not be representative for the spatial mean either. Highest temporal stability likely occurs at sites where the soil is close to saturation all the time, that is, at sites with the highest soil moisture values. Or do you mean “representative” in regard to ascription to hydrological event clusters (L. 114-116, l. 485 et seq., l. 511 et seq.)? But then temporal stability of soil moisture would not be a relevant criterion.

Following quotations are from one of best papers in this topic (Penna et al. 2013) “A positive value of MRD for a certain sampling location indicates that the location is wetter than the hillslope average whereas a negative value of MRD indicates that the location is drier than the hillslope average. The SDRD specifies how variable such as estimate is. The sampling location with the highest temporal stability shows the lowest value of ITS and is selected as representative point. Whereas points with high values of ITS are identified as the wettest and driest field location (Penna et al., 2013).” Penna, D., L. Brocca, M. Borga, and G. D. Fontana (2013), Soil moisture temporal stability at different depths on two alpine hillslopes during wet and dry periods, J. Hydrol. 477, 55-71.

Highest temporal stability is not about least variation of soil moisture but about the relationships between points. Vachaud et al. (1985) defined the persistence in time series relationships between measuring locations based on the relative difference of soil moisture and its mean overall sampling points collected at the same time. This is an indicator of surface soil moisture variation in the field (Silva Junior et al. 2016). Both

MRD and SDRD, which are components of ITS, addressed the differences in sptio-temporal variability among the locations sampled and provided information about the representative location(Gao et al., 2015). The representative point for the highest temporal stability had been used widely (See Penna et al., 2013) and we expand existing concept to more general cases to further addressing hillslope hydrological processes. Several studies also reported that the most temporally stable points were significantly related to the soil texture (Jacobs et al., 2004; Zhang and Shao, 2013; Xiaoxu et al., 2013) or topographic indexes (Brocca et al., 2009; Lee and Kim, 2017), and the other studies also noted that the time-stable points were poorly correlated to topography and soil properties (Tallon and Si, 2004). These various findings indicate that the controlling factors of the time-stable points (the representative points) are very complicated and dependent on the characteristics of the study area (e.g., climate, soil, and vegetation). An increase in temporal stability with depth was expected due to the reduced dependence on the climatic, biological, and hydrological factors that determined the SWC dynamics (Martínez-Fernández and Ceballos, 2003; Hu et al., 2009), which was also observed by Cassel et al. (2000) for cropland and by Lin (2006) for forest watersheds. Zhang, P., and M. Shao (2013), Temporal stability of surface soil moisture in a desert area of northwestern China, J.Hydrol., 505, 91-101. Xiaoxu, J., M. Shao, X. Wei, and Y. Wang, 2013, Hillslope scale temporal stability of soil water storage in diverse soil layers, J. Hydrol., 498, 254-264. Brocca, L., F. Melone, T. Moramarco, and R. Morbidelli, 2009, Soil moisture temporal stability over experimental areas in Central Italy, Geoderma, 148(3-4), 364-374. Tallon, L.K., and B.C. Si, Representative soil water benchmarking for environmental monitoring, J. Environ. Inf, 4, 31-39. Jacobs, J.M., B.P. Mohanty, E.C. Hsu, and M. Choi, 2010, Time stability and variability of electronically scanned thinned array radiometer soil moisture during Southern Great Plains hydrology experiments, Hydrol. Process, 24, 2807-2819.

In this study, we expand the representative point from existing definition “the point showing least variation between points” to the point which address efficient characterization of hydrologic event as well as traditional temporal stability. This is why we presented ITS, SDRD as so on.

Assigning clusters to processes: Results from the upslope and downslope sites, respectively, are aggregated (per depth), and differences between these groups are interpreted in terms of systematic differences between these groups. However, it has not even been tested whether these groups differ significantly at all. Data of all single sites are presented in Fig. 2 only, revealing substantial heterogeneity even within the upslope or downslope sites, respectively, and suggesting more small scale variability (consistent with my own experience) rather than systematic differences.

We afraid that reviewer 1 may not understand the reason of machine learning application. We have 30 points soil moisture data for 396 events. One solid reason for the analysis of this big data is to understand or efficient perform the task about how the dataset can be simplified through dimensional reduction in a systematic approach. Exploring local heterogeneity seems completely opposite way will absolutely result into extremely complex patterns which is impossible to understand (Think about locality of soil moisture in 30 points multiplied 396 events). Further completely different observations such as tracer data or stem flow are out of scope of this paper.

The main objective of this study is the analysis of data through SOM under the context of hillslope hydrological processes. Therefore, the difference of hydrological processes between upslope and downslope was the main topic and the small scale variability is associated with the local heterogeneity, which is not the scope of this study. Besides, it is almost impossible to characterize different non-linear patterns through SOM for every single points and interpret complexity and heterogeneity depending on measuring locations and measuring depths. Characterization of soil moisture for two parts (upslope and downslope) is consistent to several previous researches of the study area and many other hillslope studies (Haga et al., 2005; Kim, 2009, 2016; Lee and Kim. 2017, 2019; Tromp and McDonnell, 2004; Uchida et al., 2004). Thanks for your understanding.

We leave out Figure 2 in revised paper.

Haga, H., Y. Matsumoto, J. Matutani, M. Fujita, K. Nishida, and Y. Sakamoto, 2005: Flow paths, rainfall properties, and antecedent soil moisture controlling lags to peak discharge in a granite unchanneled catchment. Water Resour. Res., 41, W12410, doi:10.1029/2005WR004236. Kim, S.: Characterization of soil moisture responses on a hillslope to sequential rainfall events during late autumn and spring, Water Resour. Res. 45, W09425, https://doi.org/ 10.1029/2008WR007239, 2009. Kim, S.: Time series modeling of soil moisture dynamics on a steep mountainous hillside, J. Hydrol., 536, 37-49, https://doi.org/ 10.1016/j.jhydrol.2016.02.027, 2016. Lee, E., and Kim, S.: Pattern Similarity Based Soil Moisture Analysis for Three Seasons on a Steep Hillslope, J. Hydrol., 551, 484-494, https://doi.org/ 10.1016/jhydrol.2017.06.028, 2017 Lee, E. and Kim, S.: Wavelet analysis of soil moisture measurement for hillslope hydrological processes, J. Hydrol, https://doi.org/10.1016/j.jhydrol.2019.05.023 Tromp van Meerveld, I., and McDonnell, J.J.: Comment to “Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes, J.Hydrol., 286, 113-134”, J.Hydrol., 303, 307-312, https://doi.org/10.1016/j.jhydrol.2004.09.002, 2005. Uchida, T., Y. Asano, T. Mizuyama, and J. J. McDonnell, 2004: Role of upslope soil pore pressure on lateral subsurface storm flow dynamics. Water Resour. Res., 40, W12401.

At least there seems to be substantial overlap between upslope and downslope sites. The study aims at assigning single hydrological events, characterized by meteorological and soil moisture data based indices, to clusters, which in turn are interpreted in terms of hydrological processes (l. 90, l. 111-113). The latter step is of fundamental importance for the study. Unfortunately, that step remains completely obscure to me even after having studied the manuscript again and again. There is no clear and comprehensible reasoning at all. How do the indices relate to the respective processes? Relationships are postulated but without sound justification (e.g., l. 317-319, l. 322-324, l. 334-l.338, l. 402-462). Assigning differences in soil moisture at a scale of a few hours to lateral subsurface flow over a distance of roughly 50 m (Fig. 1) would require fairly high lateral flow velocities. Is there any additional evidence (e.g., tracer experiments) for that? Did you account for the spatial heterogeneity of throughfall and stemflow? What about surface runoff that might have re-infiltrated on its way downslope?

This is about interpretation of hydrological processes using results of analysis. Table 1, Table 2 and Figure 8 are not postulation but the evidences obtained from field measurement. The soil moisture difference index is made to present soil moisture variation impact in terms of normalized impact for antecedent soil moisture. This is an effective way to express the impact of rainfall to soil moisture for big data (30 points in 396 events).

l 317-319: The soil moisture difference index was made from soil moisture measurements. Table 1 showed the vertical distribution of soil moisture difference indices are different between upslope and downslope. In case of upslope, UP10 > UP30 > UP60 for clusters 1 to 6 indicating dominance in vertical infiltration, but DO10 > DO60 > DO30 for clusters 3, 4 and 7 and DO60 > DO10 > DO30 for clusters 5 and 6 in downslope. This indicated that the generation of lateral subsurface flow in downslope. Table 1 is not a postulation but the evidence of different hydrological processes between upslope and downslope.

l 322-324: The soil moisture difference index (192.6%) was obtained from saturation data of soil moisture (48% - 50 %) in surface of downslope during rainfall event. This is the evidence of surface saturation for extreme events (cluster 7) which most likely introduce overland flow in downslope. This analysis was done based on field measurement of 64 events.

l 334-339: The analysis appeared in Table 2 was based on field measurements. The water storage analysis for clusters 5, 6 and 7 showed that the storage changes (47,40, 116mm) in deeper depth(60cm) were greater than those (19.6, 11.6, 49.4 mm) for in-

termediate depth (30cm) at downslope area. Considering scales of throughflow (20, 35, 85.5), the storage changes in deeper depth (60cm) can be only explained by subsurface lateral flow over bedrock from upslope. Again this is not postulation but results from field measurements.

We did not measured the stem flow for table 2 but there is no systematic difference in vegetation between upslope and downslope and throughfall can be assumed to almost uniform. Furthermore, the impact of canopy interception for significant events (clusters 5,6 and 7) can be negligible.

l 402-462: The generation of lateral flow in deep soil layers was supported by results from Table 1 and Table 2, which were obtained from soil moisture measurements. Supplementary explanations of component planes (Figure 5) also support spatial distributions of hydrological processes between upslope and downslope. Discussion with Figure 8(made from measurements) and other references (Kim, 2009; Lai et al., 2016; Uchida et al., 2001; Weinjofer and Zehe 2013; Haga et al., 2001; Feng and Liu, 2015) also support hydrological process interpretation for significant events (cluster 5, 6, and 7). Reviewer asked further field measurements (tracer experiment and stem flow). The evidence of soil moisture analysis (Table 1, Table 2 and Figure 8) indicated distinct hydrological processes between upslope and downslope. Unfortunately, we did not collect tracer data and stem flow for 396 events and presentation or analysis of these additional data seems not the scope of this paper. Besides data like stem flow and tracer are not the scope this study. The scope of paper is the application of SOM into big dataset of rainfall and soil moistures and interpretations of hillslope hydrological processes about classification. Thanks for your understanding.

The manuscript requires substantial language editing. Technical corrections:

we corrected English from native speaker. Thanks,

L. 23: Please be more precise. According to Fig. 2 soil moisture was measured at 10 sites but at three different depths each.

=> We corrected it into “30 points (three depths in 10 locations)”

L. 142: Were the trenches re-filled after installation of the soil moisture probes? If not, how to deal with resulting artefacts?

=> Of course, we refilled all locations. Otherwise, how can 10 years monitoring can be sustainable?

L. 186-188: Please be more precise. Logarithm transformation is one out of a set of Box-Cox transformations. Did you apply other Box-Cox transformations as well? If so, for which observables, and how? Besides, neither the logarithm nor other Box-Cox transformations centralize the means of the variables to zero. Instead, these transformations are usually applied when Gaussian distribution is required. However, that is not the case for any of the applied approaches.

=> Of course, we explored Box-Cox transformation and we adapted logarithmic transformation simply because it provides the best result in terms of centralization (such as smallest skewness). This is widely used transformation for centralization (see Salas et al., 1988). Salas, J. D., Delleur, J. W., Yevjeuch, V., Lane, W. L., 1988. Applied Modeling of Hydrologic Time Series. Water Resource Publication. Chelsea, Michigan.

L. 279-281: I do not understand why you select soil moisture measured at one single point “as the representative soil moisture before the event for the SOM analysis”. That introduces an unnecessary bias. Why not taking the mean of the values measured at the different sites?

=> The selection of representative points based on temporal stability method had been widely used and applied in the relationship among soil moisture points in many paper. The measurement data for 10 years cannot be 100% perfect, there can be some missing data and the perfect mean of every soil moisture points is not available for all rainfall events. Therefore, the soil moisture of representative point was replaced for calculating antecedent soil water condition. Again the temporal stability analysis is widely

accepted approach for evaluation of mean soil moisture variation. However, we used to this concept to delineate the representative point for antecedent soil moisture.

L. 574: “https://www.re3data.org/” is not a repository but gives an overview over numerous repositories. Please be more precise: Where will the data be published?

Of course, if the paper is accepted, the data will be uploaded in open repository.

References: Some references are out of alphabetical order (Minet et al. 2013, Montero and Vilar 2014, Zhu et al. 2014).

=> We corrected it. Thanks.

Fig. 5: Figure caption: Missing explanation of the lower panels (cf., caption of Tab. 1).

=> We largely revised captions in many figures. Figures 5. (a)–(j) Component planes of variable weightings for the rainfall amount (a), antecedent soil moisture (b), soil moisture difference indices for the upslope at depths of 10(c), 30(d), and 60 cm(e), rainfall duration(f), rainfall intensity(g), and soil moisture difference indices for the downslope at depths of 10(h), 30(i), and 60 cm(j).

Response to comments from Anonymous Referee #2

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 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.

Thanks for your comments on title. “Characterization of hillslope soil moisture recharge events using machine learning algorithm”

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.

Thanks for your suggestion. The time series of representative rainfall event of 7 clusters are shown in following figures. We assumed the endpoint of recharge event when soil moisture was varied less than 2 % (error bound of measurement) after the rainfall events. This is not noted in the paper because this was not used in SOM application.

These figures will be posted supplementary part.

Figure for representative recharge events for 7 clusters

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?

In section for Results 3.5, we showed the robustness of cluster delineation in SOM algorithm with partial dataset except for each year. More than 90% agreements were found between the projected cluster number in the SOM of the complete dataset and that of the partial dataset (Figure 6). Cross-validation is a widely used method for the validation of SOM classification(Chang et al., 2013; Huang et al., 2015; Kim et al, 2015).

Chang, F. W. Tsai, H. Chen, R. Yam, and E.E. Herricks, 2013, A self-organizing radial basis network for estimating riverine fish diversity, Journal of Hydrology, 476, 280-289, doi:10.1016/j.jhydrol.2012.10.038.

Huang, J., J. Gao, and Y. Zhang, 2015, Combination of artificial neural network and clustering techniques for predicting phytoplankton biomass of lake Poyang, China, Limnology, 16, 179-191, doi: 10.1007/s10201-015-0454-7.

Kim, M., S. Baek, M. Ligaray, J. Pyo, M. Park, and K. Cho, 2015, Comparative studies of different imputation methods for recovering streamflow observation, water, 7(12), 6847-6860, doi:10.3390/w7126663

The clustering SOM maps of yearly partial dataset were found to similar to the pattern of SOM map with total dataset.

What important is not the number of events but whether the dataset was made with similar portions of 7 clusters of total data set. For example, if we made a dataset deliberately leaving out events for cluster 6 out of total dataset (371 events), then obviously the classification can be different to total dataset. In other words, the important matter is whether the sample well represents the population or not. As shown in following figure, if we made identical percentage of 7 clusters in samples, the classification of 7 clusters were identical even in 50 number of events.

Figure for SOM with different numbers of samples

Additionally, we also check the robustness in the patterns of component plane depending on the length of dataset. There are no significant pattern changes in duration, intensity, amount, antecedent soil moisture, soil moisture variation at 30cm and 60cm on down- and up- slope area even in 50 samples.

Figure for patterns of component planes with 50 samples

Therefore, if the events in sample is composed of identical portion of 7 cluster events to those of population, the machine learning clustering of different samples (50<n<350) provide identical results to those of population.

We added in the context “Further validation practices for various number of events, SOM with small event numbers (n=100) provided very similar classification results to the that of 396 events, if the portions of 7 clusters in dataset are identical. ."

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?

Equation 1 is used for calculation of all sensors in all depths. Of course, more data will be better (Ran et al., 2017). We actually have more data but they cannot be used simply we leave out other data because reliable data quality control for 10 years for only 30 points. Actually, there were several studies for representative point with similar numbers of data point (n=10) (Wei et al., 2017; Jia et al., 2013).

Wei, L., J. Dong, M. Gao, and X. Chen, 2017, Factors controlling temporal stability of surface soil moisture: a watershed-scale modelling study, Vadose zone, 16(10), doi:10.2136/vzj2016.12.0132.

Jia, Y., M. Shao, and X. Jia, 2013, Spatial pattern of soil moisture and its temporal stability within profiles on a loessial slope in northwestern china, Journal of Hydrology, 495, 150-161, doi: 10.1016/j.jhydrol.2013.05.001.

Ran, Y., X. Li, R. Jin, J. Kang, and M.H. Cosh, 2017, Strengths and weaknesses of temporal stability analysis for monitoring and estimating grid-mean soil moisture in a high-intensity irrigated agricultural landscape, Water Resources Research, 53, 283-301, doi:10.1002/2015WR018182.

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.

As reviewer point out, ensemble of soil moisture for temporal stability depend on depth and seasonal feature. Actually, we had selected the representative points for three different depths and the representative points for antecedent moisture were delineated and their component planes for ASM-10, ASM-30, and ASM-60 were evaluated (see Figure below). As you can see in this figure, no systematic difference between different depths of ASM can be found. In other words, implementation of 3 ASM for machine learning algorithm can be redundant. Considering the purpose of ASM for machine learning application, one ASM with least RMSE and lowest ITS is more appropriate for ASM consideration. Again, this is not for finding mean soil moisture but to delineate the most appropriate point for ASM evaluation

Figure for component planes for three different depths.

Following table showed RMSEs to mean soil moisture for all points. The point UP3-10 showed the minimum RMSE and ITS. We used UP3-10 as a representative point for ASM in machine algorithm in the paper. .

Table for RMSE for all points

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.

We used the MiniTRASE system(TDR) (probably one of most delicate and expensive TDR machines in market) for soil moisture observation having 2% error bound. We did not move the waveguide and the measurement seems consistent during measurement period. Of course, we did some experimental validation in the beginning of measurement campaign. There are more than 10 publications about soil moisture measurements and analysis for this hillslope (hydrological processes, journal of hydrology, water resources research). Soil moisture equipment corporation, 2005, TRASE operating instruction. Santa Barbara, CA.

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.

It could be zero in your case. But it is rarely happens in temporal stability processes with long time series datasets. The ITS with MRD and SDRD had been used for representing temporal stability in soil moisture measuring systems(Zhao et al., 2010). The contribution of MRD and SDRD in ITS can be changed depending on soil moisture measuring environmental characteristics. There seemed not many studies for only use MRD. That's why we select ITS as temporal stability.

Jia et al., 2013 indicated that "The value of the mean relative difference(MRD) for a point at a particular depth quantified whether that point was wetter or drier than the areal mean at the same depth. The standard deviation of the relative difference(SDRD) characterized the variability of relative difference at that point within the experimental period"

Jia, Y, M. Shao, X. Jia, 2013, Spatial pattern of soil moisture and its temporal stability within profiles on a loessial slope in northwestern China, Journal of Hydrology, 495, 150-161, http://dx.doi.org/10.1016/j.jhydrol.2013.05.001.

Following table showed MRD, SDRD and ITS for our dataset.

Table for MRD SDRD ITS for all points

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.

As far as we understand about the temporal stability from existing studies, it is not about local temporal variation but the relative variability to the spatial mean of soil moisture(Penna et al., 2013).

We agree your opinion that the temporal stability depends on seasonal feature and soil condition. However, the purpose of this paper is not about prediction of mean soil moisture but the most appropriate point for ASM.

The main reason to apply temporal stability to our sampling site is selection of representative point. Because there were some missing measurements in the time series of soil moisture, the average of antecedent soil moisture was changed depending on the missing measurements in our sampling sites. Since the representative point based on temporal stability have shown the point can follow the trend of average soil moisture and higher estimation ability about soil moisture average in previous researches, we used the UP3-10 for as an indicator for antecedent wetness without the missing data impacts. RMSE between mean of soil moisture and UP3-10 was minimum out of all points. Therefore, the reason why we select UP3-10 as the representative value of antecedent wetness condition. Table(above) showed RMSE for all points.

In order to clarify the paper on the context of ASM, the Figure 2 in previous version of paper is removed. Thanks for your understanding

Regrading to the raw data, we are welcome to share our data for possible co-work.

Penna, D., L. Brocca, M. Borga, and G. Dalla Fontana, 2013, Soil moisture temporal stability at different depths on two alpine hillslopes during wet and dry periods, Journal of Hydrology, 477(16), 55-71, doi:10.1016/j.jhydrol.2012.10.052.

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.

Reusser et al., gave us insight of applying the self-organizing map(SOM) to hydrological time series data in diagnostics and evaluation of model prediction. This method could help in classifying the rainfall events with small amounts. However, it is difficult to identify whether the slight increase in soil moisture is due to rainfall input or measurement error.

The measurement error before and after the rainfall event cannot be different. We interested the variation of soil moisture through the soil moisture different index and the measurement error will not be influence equation (4). In equation (4), we make difference between maximum soil moisture and antecedent moisture. Assuming the measurement error for these two terms is identical (from same machine), the impact of measurement error can be neglected. The measurement error of machine is less than 2 % in Vol. soil moisture. We attached the typical events of different cluster in supplementary material.

Technical points: - Line 64: "can be differently appeared"- please reformulate.

It is corrected as "can be different".

- 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.

It is corrected as "the generation of hillslope flow paths such as vertical flow and lateral flow"

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

We changed into 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).

We added ", hydrological model performance diagnostics (Herbst and Casper, 2008; Reusser et al., 2009" to line 80.

Reusser, D.E., T. Blume, B. Schaefli, and E. Zehe, 2009, Analysing the temporal dynamics of model performance for hydrological models, Hydrol Earth Syst. Sci., 13, 999-1018, https://doi.org/10.5194/hess-13-999-2009. Herbst, M., and M.C. Casper, 2008, Towards model evaluation and identification using self-organizing maps, Hydrol Earth Syst. Sci., 12, 657-667, https://doi.org/10.5194/hess-12-657-2008.

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

Thanks. we corrected captions as follows;

Figure 2. Soil moisture measurements for representative points (UP3-10) and the least temporally stable point (DO5-30) for the average soil moisture from 30 measurement points. RMSE and R2 were 1.72 and 0.90 for UP3-10 and 9.0 and 0.45 for DO5-60, respectively.

Figure 3. Structure of the (a) dendrogram for seven clusters using Ward's method and (b) SOM classifications in 96 neurons through a 16 $\times$ 6 matrix.

Figures 4. (a)–(j) Component planes of variable weightings with SOM clustering for the rainfall amount (a), antecedent soil moisture (b), soil moisture difference indices for the upslope at depths of 10 cm(c), 30 cm(d), and 60 cm(e), rainfall duration(f), rainfall intensity(g), and soil moisture difference indices for the downslope at depths of 10 cm(h), 30 cm(i), and 60 cm(j).

Figure 5. 2D array expression of k-fold ciess validation (k=10 years) by event projections for the summation of 10 SOMs of partial datasets to the SOM of the complete dataset.

Figure 6. A C4.5 decision tree of soil moisture difference indices (numbers in boxes) for hydrologic event classification. Depending on the threshold of soil moisture at a specific point noted in each box, the decision tree either go to left (Y) and right (N).

Figure 7. Statistical distributions in box plot of component vectors for (a) rainfall characteristics and antecedent soil moisture, (b) upslope volumetric soil moisture, and (c) downslope volumetric soil moisture for the SOM classification.

- Line 272: typo higher Thanks. we corrected it.

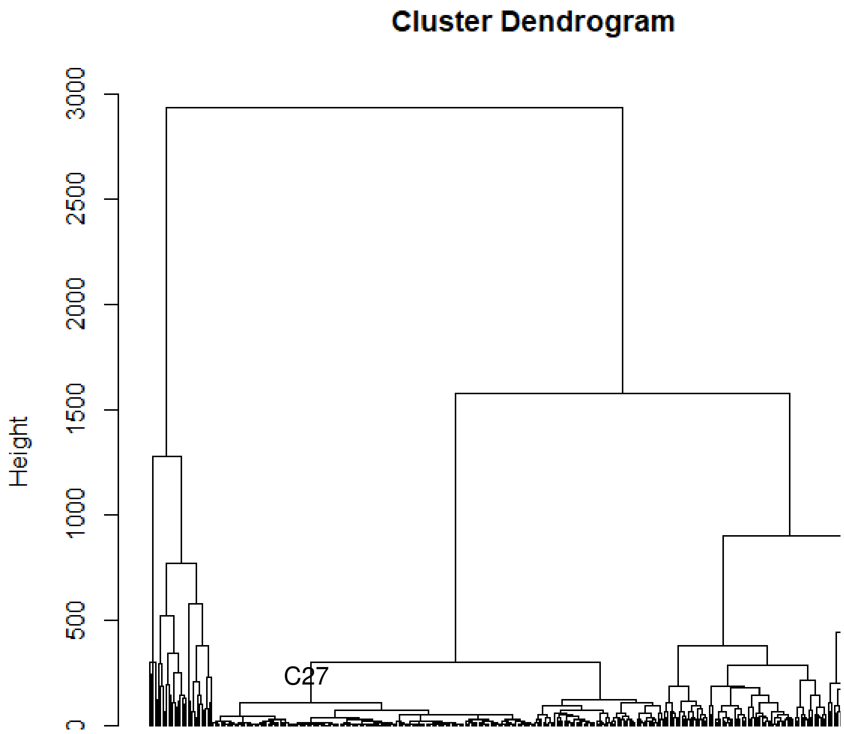
- Figure 3: DO5 30 has soil moisture values larger than 50%, what is the porosity of the soil? The soil moisture time series were measured using a multiplex based time domain reflectometer for 10 locations upslope (UP1âĂŠUP5) and downslope (DO1âĂŠDO5). At each location, three TDR sensors were inserted parallel to the surface at depths of 10, 30, and 60 cm at the upslope side of an installation trench. Average of soil porosities in the study area were 48.85% for upslope and 47.87% for downslope. Also the bulk density is 1.375 g/cm3 for upslope and 1.371 g/cm3 for downslope. Soil moisture greater than 50 % was rarely found under extreme events in rainy season. We believe this may be associated with development of substantial pipeflow in downslope part in extreme events.

We sincerely express our deep appreciation to two reviewers and editorial members of HESS.

Sincerely,

Sanghyun Kim, Prof. Dept. of Environmental Engineering Pusan National University kimsangh@pusan.ac.kr

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



# of clusters	Sum of square	Coefficient delta
1	8767491	
2	4455593	4311897.58
3	3207137	1248456.58
4	2395312	811824.39
5	1988096	407216.49
6	1688206	299889.83
7	1521255	166950.81
8	1383726	137529.06
9	1285079	98647.41
10	1212425	72653.42

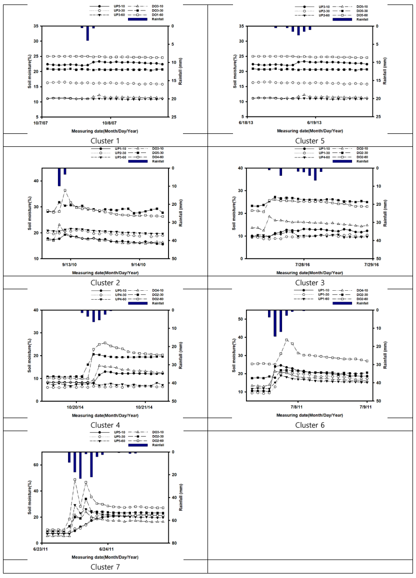
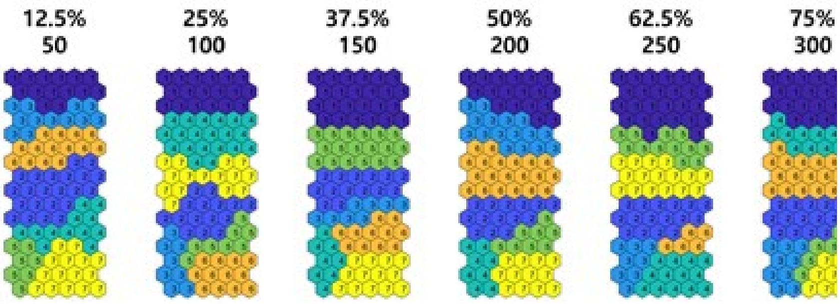
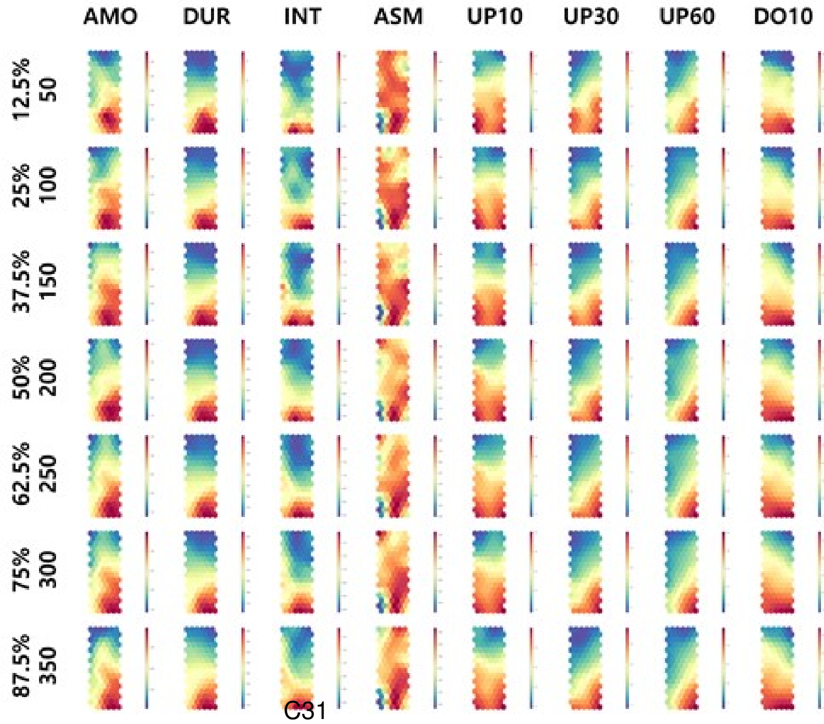
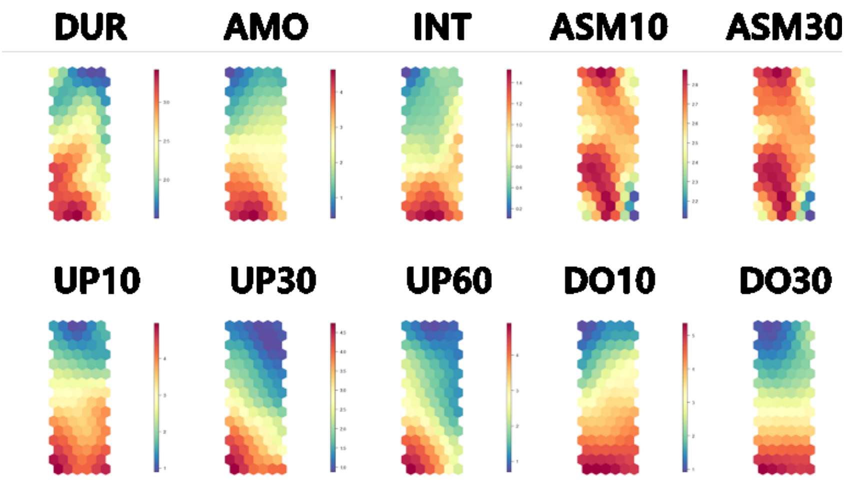


Fig. 3. Figure for representative recharge events for 7 clusters




AMO
DUR
INT
ASM
UP10
UP30
UP60
DO10
12.5%
50
25%
100
37.5%
150
50%
200
62.5%
250
75%
300
87.5%
350

C31


DUR
AMO
INT
ASM10
ASM30
UP10
UP30
UP60
DO10
DO30

	UP1-10	UP1-30	UP1-60	UP2-10	UP2-30	UP2-60	UP3-10	UP3-30	UP3-60	UP4-10	UP4-30
RMSE	5.93	6.19	4.99	3.38	3.56	4.67	1.72	3.18	4.28	7.33	4.61
	DO1-10	DO1-30	DO1-60	DO2-10	DO2-30	DO2-60	DO3-10	DO3-30	DO3-60	DO4-10	DO4-30
RMSE	6.39	3.21	7.88	3.58	4.96	5.77	4.95	4.49	3.14	1.98	1.82

	UP1-10	UP1-30	UP1-60	UP2-10	UP2-30	UP2-60	UP3-10	UP3-30	UP3-60	UP4-10	UP4-30
MRD(%)	-37	-40	-30	-17	-25	-31	4	5	-28	-49	-29
SDRD(%)	12	10	18	17	12	18	13	20	15	15	13
ITS(%)	39	42	34	24	28	35	13	21	32	51	32
	DO1-10	DO1-30	DO1-60	DO2-10	DO2-30	DO2-60	DO3-10	DO3-30	DO3-60	DO4-10	DO4-30
MRD(%)	39	10	45	-12	29	42	-17	27	15	-9	-4
SDRD(%)	22	18	21	18	20	25	26	15	16	12	14
ITS(%)	45	21	50	22	35	48	31	31	22	15	14