Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-574-AC3, 2019

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Soil moisture: variable in space but redundant in time" *by* Mirko Mälicke et al.

Mirko Mälicke et al.

mirko.maelicke@kit.edu

Received and published: 9 December 2019

We would like to thank Anonymous Referee 2 for the very constructive and insightful review of our work.

In the following, we will respond to the Referee's comments in the order of appearance (Referee comments in italics).

General comment:

I am not a hydrologist, so I cannot say anything about the level of novelty of the current work with respect to the published literature of which I am not well aware. On the

other hand, the proposed methodology seems very reasonable and effective to me. I liked the paper and the interpretation of the results. On the other hand, I personally found that the text should be improved: some sentences are open to ambiguity or unclear (the meaning or idea to be conveyed is there intuitively, but the structure of the sentence leaves it open to misinterpretations), and there are some repetitions that could be cutout in order to make the paper easier to follow (sometimes is hard). I suppose that the confusion in some sentence stem from the young age of the first Author: put some more efforts in making the text clearer and more specific in order to honor your work.
Response: We are very happy to hear the reviewer liked our manuscript and we will make an effort to improve the writing to reduce ambiguities and repetitions.

Specific comments:

1) *pp 9, lines 26-27: 'This distance is the foundation of Mean shift on the one hand and the compression quality calculations presented in section 2.5.2 below on the other hand'. The relevance of the distance in (3) for the mean shift clustering algorithm is somehow hidden during the description of the method in Sec. 2.3.1 (i.e., 'We tested different bandwidth parameters at a few examples and set the bandwidth to the 30% percentile of all pairwise vector distances between the dispersion functions of one year and depth'). If possible, I would like to rephrase this aspect putting more focus on the relevance of (3) within the clustering algorithm.*
Response: We will rephrase the respective paragraph. Equation (3) is important for Mean shift, as distances between input features in Mean shift are also calculated using the Euclidean distance. There are other algorithms that are i.e. based on ranks. We will put more focus on the relevance of Euclidean distance for Mean shift.

2) *In the Section 2.5.2 Compression quality, there are some unclear aspects to me.*

'To assure comparability we use one binning for all calculations of H (across years and depths). To achieve this, all pairwise distances between all spatial dispersion functions of all four years in all three depths are calculated. The discrete frequency distribution is formed from 0 up to the global maximum distance (between two dispersion functions) calculated using equation (3). The bins are formed equidistant using a width of the maximum function distance that still lies within the error margins calculated using equation (9). Thus, the information content of the spatial heterogeneity is calculated with respect to the expected uncertainties. This way we can be sure to distinguish exclusively those spatial dispersion functions that lie outside the error margins.' It is my understanding that the binning scheme is grounded on the distance function in (3), which make me to think that subsequent calculation (e.g., entropy, KLdivergence) will involve the distance in (3), but then 'To calculate the mean information content of the compressed series each cluster member is substituted by the respective cluster centroid. This substitution is obviously not a compression in a technical sense, but necessary to calculate the Kullback-Leibler divergence. Then a frequency distribution for compressed series X and the uncompressed series Y can be calculated. The Kullback-Leibler divergence DKL of X,Y is given in equation (12)' which compare compressed and uncompressed dispersion functions and does not involve the distance in (3) in any way. It seems that the binning (size of the bin and edges of the diverse bins) entails the diverse distance according to (3) (and the associated uncertainty, according with (9)), but the KL is evaluated for the dispersion function themselves (and not their distance)? What am I missing? Could you please further clarify how the distance (3) is involved in the evaluation of (12)? I would also briefly describe the meaning of the KL-divergence which is just introduced, but not commented.

Response: We thank the referee for pointing this out. We agree that the Kullback-Leibler divergence (KL) should be further described in section 2.5.2 (p.11). The Referee is right, this part can lead to misunderstandings and we

will carefully clarify respective sections. The main reason for possible confusion is that we combine two methods that both involve a step called binning. These two binnings are not linked to each other. The first binning refers to the dispersion function, here we pool observation points into lag distance classes for calculating the dispersion function. In the revised manuscript we will strictly introduce this binning as lag classes.

The second binning is necessary for calculating the Shannon entropy (equation 11) and therefore also the KL, which is using Shannon entropy. To this end we have to treat the set of distances between all dispersion functions as a discrete pdf. Thus "all pairwise distances between all spatial dispersion functions of all four years in all three depths are calculated" need to be pooled into meaningful distance classes (p.10 L22-23). Following Loritz et al. 2018 we think a meaningful minimum distance should be larger than the error margin. So, yes we base the calculation of KL on the distances between dispersion functions and not the dispersion functions themselves. And, the Referee is right, the binning for equation 12 is not the same as for calculating the dispersion function. We will clarify this in the revised manuscript.

3) *pp.12 lines 6-7: 'Dispersion declines with separating distance, as small values correspond to observations which have similar values while large values suggest the opposite' looking at Fig. 4 the dispersion function increases with the spatial lag, such that dispersion increases with separating distance. What am I missing here?*

Response: The Referee is right, this is a mistake. The dispersion is increasing with distance. We will replace 'decline' with 'increases'.

4) *pp.12 lines 18-23: 'As the spatial dispersion functions in the presented example are redundant in time, we compressed the information by replacing the dispersion function within one cluster by the cluster centroid. All four representative functions shown in Figure 4 c) exhibit increasing dispersion with separating dis-*

tance. For the blue and green cluster this happens step-wise at a characteristic distance of 500 m. That reminds us of a Gaussian variogram, which can also show a step-wise characteristic. The small grey cluster shows an increase at 500 and another one at 1000 m separating distance. In contrast the orange cluster, however, shows only a gentle increase with distance.' Are there any reasons for these stepwise increases or some physical related explanations for these behaviors?

Response: In geostatistics we may fit the Gaussian variogram function to experimental variograms which have strongly changing slopes. In close proximity to an observation it is very similar and the semivariance is rather constant, this is followed by a strong increase over a rather short increase in the lag distance. An experimental variogram of a variable that is closely linked to land use might look like this, if land use changes significantly within short distances. In my opinion, as this can only be seen in winter and spring, this originates from heterogeneous rainfall/throughfall input, whose spatial structure is still present in the dispersion function. Within the vegetation period these structures get evened out by the dominating effect of transpiring vegetation taking up considerable amounts of water before this water can shape soil moisture patterns. However, we cannot provide evidence that this explanation is correct.

5) *with reference to Fig. 4 and its discussion in Sec. 3.1 vegetation period is mentioned several times, would it be nice to have on Fig. 4.c this period highlighted (also in other figures where vegetation period is of relevance), for example as a light green bar along the x-axe or similarly, in order to understand when this vegetation period is 'on/off'.*

Response: Thank you for this good idea. We will add the suggested bar after defining a means for inferring vegetation period from temperature-sum curves.

6) *pp. 12, lines 24-26: 'In the vegetation period observations are similar even at large separating distances. As the orange cluster (Figure 4 c and d) covers significant*

parts of the vegetation period, the influence of vegetation on spatial soil water dynamics is considered to be dominant'. The latter sentence is misleading: during the so referred vegetation period which are the concurrent factors along with the vegetation-related influence that could possibly influence the soil moisture? How is it possible to discern the impacts of other factors in order to say that vegetation influence is the dominant one? Or, if the vegetation-related influence is the only factor, no surprise that it is the dominant one. Please consider revise the sentence or better support it.

Response: We agree with the Referee. We derived the 'vegetation period' only from the temperature-sum curve, following an approach equivalent to e.g Solantie (2004) or Seibert et al. (2017), and therefore cannot discern impacts of other factors. We will revise the paragraph to better express that dispersion functions are fundamentally different in summer/autumn (and how they are different). That the vegetation is responsible, is one possible explanation and we will move this part to the discussion.

7) *pp.13, lines 7-8: 'At the same time the observations get spatially more homogeneous in summer, particularly when the blue cluster emerges, because the dispersion at large lags decreases significantly.' I would substitute 'because' with 'i.e.', the decrease of the dispersion function is a consequence not a cause of the more homogeneous nature of the data during summer.*

Response: The Referee is right and we will change the sentence as suggested.

8) *pp. 13 lines 13-15: 'The green clusters emerge with strong rainfall events after longer previous dry spells (Fig. 5). We would have expected a third occurrence at the beginning of August, but the soil may already be too dry to bear a detectable dependency on separating distance'. It is not clear at which green cluster the Authors are referring to in this sentence, please clarify.*

Response: Thank you for pointing this out. We will reference Fig. 5 a,d instead of just Fig. 5.

9) *pp. 13 lines 17-18: 'The 50 cm dispersion functions (Fig. 5 f) show a clear spatial dependence and are similar'. I can see the trends of the two dispersion functions with respect to the lag and I can see that these trends are consistent among each other, but I don't see the similarity between the two dispersion functions that are characterized by fairly different values especially at larger lag. Please revise the sentence having care of the specificity of the wording.*
Response: We were intending to point out that the dispersion value at large lags is the only difference and will rephrase the sentence accordingly.

10) *pp. 13 lines 18-19: 'Not only the soil moisture observations have become much more homogeneous with depth, also the dispersion functions are more similar in shape.' The first part of the sentence is vague, more homogeneous with respect to what? In time or space? Looking at Fig. 5c, at fixed time (e.g., 2016-02), I can see a great level of heterogeneity across the moisture data, even larger than that recorded in 5.a-b. Furthermore, the dispersion function orange in Fig. 5f reaches values comparable to that of the dispersion functions in Fig. 5d (orange) and Fig. 5e (grey and green clusters).*
Response: We agree that our description of figure 5 needs to be specified. We will put more emphasis on the fact that in Fig. 5f the two functions are of similar shape, and only differ in the value of the last few lag classes. In contrast, 5d and 5e show dispersion functions of different shape. In Fig. 5d we find increasing functions vs. a flat function and in 5e two step-wise functions vs an increasing function. We agree that this has to be explained more properly. Furthermore we realized that the statement 'soil moisture observations have become much more homogeneous' does not hold like this will revise it.

11) *pp. 14 lines 1-2: 'We find dispersion functions with characteristic length of 500 m and the blue cluster persists throughout most of the year.' How is the characteristic length of the dispersion function defined? Please clarify.*
Response: With 'characteristic length' we are referring to the correlation length

of the dispersion function. We will add a definition to the respective section in the methods, where dispersion functions are introduced (2.2, p. 5 -6).

12) *pp. 16 lines 4: 'water water dynamics', water is repeated.*
Response: Thank you.

13) *pp. 18 line 27: 'In line with H4 spatial patterns of soil moisture were found to be persistent over longer time periods' Longer than what?*
Response: We did not compare the period lengths with anything specific and will therefore change the wording to '. . . were found to be persistent over weeks, if not months'.

14) *pp. 19 lines 17-18: 'We thus conclude that there is dependence of the dispersion on the rainfall pattern, which is reflected in their shape and characteristic lengths.' The sentence, as written, means that the shape and characteristic lengths are referred to that of the rainfall pattern, while I am imagining that are the shape and characteristic lengths of the dispersion function the ones that changes. Please revise the sentence.*
Response: Thank you for pointing this out. The Referee is right, we are referring to the shape and characteristic lengths of the dispersion function and will therefore revise the sentence.

15) *pp 20 lines 16-18: 'This Euclidean distance does, however, not provide information on the underlying cause of dissimilarity and thus a simple shift along the y-axis can result in the same level of dissimilarity as a change in the shape of the dispersion function.' Despite being clear from an intuitive point of view, this sentence can be sloppy to the most rigorous reader: the y-axe of what? Please revise, like 'a minor difference in the values of the dispersion functions, even though characterized by the a very similar shape, could results in . . ..'.*
Response: We thank the Referee and will revise the sentence as suggested.

16) *pp 22 line 1: 'A drying and then dry soil exhibits dispersion functions without spatial structure. Interestingly, these functions flatten out by minimizing the dispersion on large distance lags and we can thus see how the soil acts as a low pass filter.' The first sentence is obscure, especially when linked with the second one. Why the Authors claim that there is no spatial structure during drying and dry periods, when the associated dispersion functions clearly show a flat behavior for the majority of the spatial lags? As far as I have understood, the latter behavior is a sign of homogeneity in the soil moisture across space, which is a clear sign of a structure in space (maybe not that interesting, though) to me.*
Response: We agree with the Referee. Homogeneity across space is also to our understanding a sign of structure. We will revise the sentence and highlight the difference to the dispersion functions in wet periods.

Finally we would like to thank the Referee for the detailed, constructive and insightful comments.

References

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., and Zehe, E. (2018). On the dynamic nature of hydrological similarity. *Hydrology and Earth System Sciences, 22*, 3663–3684.

Seibert, S. P., Jackisch, C., Ehret, U., Pfister, L., Zehe, E. (2017). Unravelling abiotic and biotic controls on the seasonal water balance using data-driven dimensionless diagnostics. *Hydrology and Earth System Sciences, 21*(6), 2817-2841.

Solantie, R. (2004). Daytime temperature sum-a new thermal variable describing growing season characteristics and explaining evapotranspiration. *Boreal environment research, 9*(4), 319-334.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-574, 2019.