Replies to Reviewer 3

General comments to the authors

The proposed approach to identify the optimal number and location of a limited number of soil moisture sensors is sound and leverages well-known statistical techniques such as Principal Component Analysis and K-means clustering analysis. Clearly the research community is active in this topic and approaches of this type will be valuable to the scientific community, state and federal agencies, and watershed managers. While the approach is sound, it has two important cons in my opinion:

Reply: We thank the reviewer in acknowledging the proposed method as sound and valuable to the scientific community.

1) The approach requires a substantial amount of available information about the variable in question to decide the architecture of the network. This somewhat conflicts with the idea of deploying a new network, which is trying to resolve the problem of no having soil moisture data available.

Reply: Unlike methodologies proposed by other literature such as (Chaney et al., 2015) where a large number of in-situ characteristics datasets are required to build-up the soil moisture network (topography, land cover, soil properties), the methodology proposed in this study only requires soil moisture information from the WRF model (this will be emphasized in the discussion section). Since the WRF model can be run via globally freely available reanalysis datasets (e.g., ERA-5), it can be applied to any part of the world, even for areas where in-situ datasets are extremely sparse. Moreover, although WRF estimated soil moisture cannot represent the ground truth, they are ideal datasets to provide catchment and hydrometeorological characteristics, such as local climate, land cover, soil properties, topographies, which are the main drivers of local soil moisture heterogeneity. Therefore, we believe the proposed approach has the advantage over existing literature in soil moisture network design.

Chaney, N. W., Roundy, J. K., Herrera-Estrada, J. E., and Wood, E. F.: High-resolution modeling of the spatial heterogeneity of soil moisture: Applications in network design, Water Resour. Res., 51, 619-638, 2015.

2) The second drawback is that watershed managers will likely not spend substantial amount of time running complex simulations to generate a dataset to inform the proposed approach unless they are assisted by a group of scientists.

Reply: In the future, a decision support tool based on cloud computing could be developed, where WRF and the relevant data and code are ready to be applied anywhere in the world. In such ways, the watershed managers could focus on making optional choices instead of worrying about the background computational work.

I think the manuscript deserves publication. Below are some suggestions and questions to allow the authors improve the manuscript. Specific comments to the authors

Line 49: Remove comma between “probe” and “and Time..”

Reply: Agreed. This will be updated.

Line 50: Compared to other technologies such as neutron scattering and the gravimetric method, sensors relying on electromagnetic principles are probably not that old, particularly from the point of view of automated systems. They are about 30 years old, but not sure whether it classifies as one of the oldest.

Reply: Agreed. This will be updated.

Line 51: When calibrated, these sensors can be accurate. Otherwise measurements can have substantial bias.
Reply: Agreed. This will be updated.

Line 54: Economic considerations about what aspect? Please clarify.
Reply: This will be clarified.

Line 95: I suggest being more explicit and write “Po river plain and surrounding hilly areas”
Reply: Agreed. This will be updated as suggested.

Line 97-98: Please define acronym “ESA CCI”
Reply: The full name will be added.

Line 152: Please, consider a modification along these lines “5-minute angle soil database” or “5-minutes geographic resolution soil database”
Reply: It will be updated.

Line 157: Please use square brackets when nesting parentheses.
Reply: It will be updated.

Line 197: Remove “the” from “since the eigenvectors of the X” What is the depth of the soil moisture sensing by the current stations? How will this approach handle multiple sensing depths? It seems that the final network configuration may be different for different soil layers.
Reply: ‘the’ will be removed.

The soil moisture sensing by the current stations provides observations at multiple depths, and are maily centred at 10 cm, 25 cm, 45 cm and 70 cm. The case study presented in the paper is based on the surface soil moisture from the WRF output (at 10 cm). The reviewer is right. The final network configuration could be different for different soil layers. We will, therefore, include the analysis of root zone soil moisture in the updated manuscript.

In Figure 9 it is unclear what is the number of stations considered and the method (CA-max, CA-median) employed to create the “designed network”. Please provide more details in the figure caption. Do the current soil moisture monitoring stations only measure soil moisture? Do they observe other hydrological or meteorological variables that need to be taken into account at the time of designing a new network? To account for the network designed in probabilistic terms, have the authors considered using a probabilistic clustering method such as the Fuzzy C-means approach? Does the proposed approach account for the layout of existing monitoring stations?
Reply: The designed network is based on CA-Med and 90% variance contribution rate. This will be added in Figure 9 caption.

The in-situ soil moisture sensors only provide soil moisture information. There are separate rain gauges in the catchment.

We thank the reviewer on suggesting of using probabilistic clustering method such as the Fuzzy C-means approach. This study intends to explore the usefulness of the PCA/CA method which is relatively simple. For future studies, other methodologies will also be explored to examine

their effectiveness. However, we will add the probabilistic clusing method along with other potential approaches as a discussion in the manuscript.

No, the proposed scheme does not account for the layout of existing monitoring stations.

By looking at Figure 10, it seems that a complete re-arrangement of stations is required, which may conflict with existing agency resources and manpower. Judging from Figure 11 it seems that the whole effort of re-designing the network will remove some existing bias, but may not lead to a substantial improvement. Can the authors provide more details on the logic behind the location of existing stations?
Reply: For the current layout of the soil moisture stations, all of the sensors are located in the plain area, which can lead to a bias of the overall catchment soil moisture. However, we can see from Figure 10, some of the existing sensors are located near some of the designed sensors, which could be kept if located within the same cluster. But a lot more sensors are indeed required in the hill zone, where currently no sensors are installed. We will add this discussion in the updated manuscript.

Although in this case study the improvement is not substantial because the soil moisture spatial variations over the catchment are relatively small (see the figure below, from our previous study). However, for areas with large spatial soil moisture variations, the bias could be much noticeable.

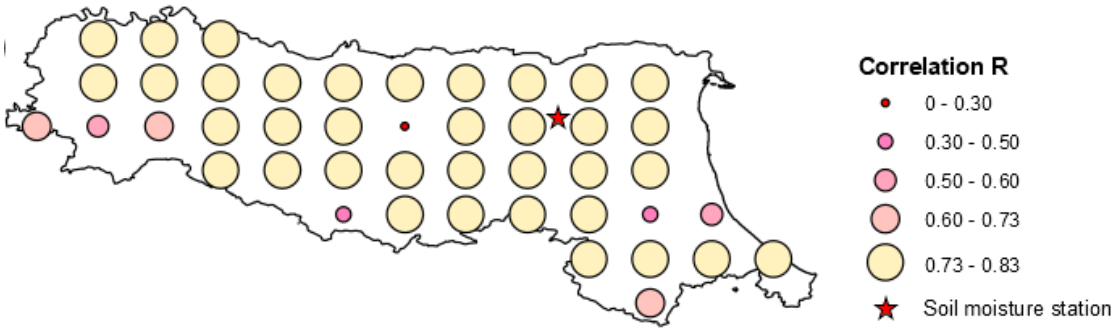

Figure 1. The cross-validation of spatially distributed WRF soil moisture against the in-situ soil moisture observation.

The existing stations could be initially installed for irrigation purpose, which hence is mainly located in the plain area.

It seems that the CA-max proposed clustering technique tends to initially place stations at or near the edges of the watershed. While this is the result of the proposed objective method, there are reasons why this might not be ideal. The CA-median seems to be less sensitive to this. How is the PCA superior to defining the number of clusters using something like the Silhouette method? How is the k-means part of this method different than variance minimization methods based on correlograms or semi variograms? How does this approach include or handle the changing land use/land cover?
Reply: CA-Max selects the maximum averaged soil moisture of a cluster. In the case study area, it is clear to see since the southern boundary of the catchment is mainly covered by densely covered trees which generally has higher soil moisture contents than the rest of the catchment, the selected locations tend to distribute near the southern boundary (70% in Figure 6 is a good example). For the CA-Med, as it selects the median averaged soil moisture of a cluster, the resultant locations are more homogeneously distributed. The CA-Max is focused

on the extreme soil moisture condition, whilst the CA-Med is more on the mean condition. Since they provide results in two aspects, it is useful to explore both in the study. The reasoning will be included in the updated manuscript.

Again, we thank the reviewer on suggesting alternative statistical approaches to solve the problem. We will include them in the discussion for future explorations. This paper focuses on assessing the effectiveness of the PCA and K-mean methodologies, and we can see that they result in a very good performance. Moreover, they are simple and can deal with a large number of datasets very efficiently, therefore have the potential to be widely adopted. Since we use the soil moisture temporal variations (10-year daily soil moisture datasets) for the PCA/CA analysis, the local characteristics such as the land use/land cover have already been considered by the analysis and represented by the CA results.

How can this method be applied to smaller catchment areas where a single pixel from remote sensing sources or distributed models is equal or even larger than the entire watershed?
Reply: The method can be applied to smaller catchment areas because the WRF model can simulate high-resolution soil moisture datasets (e.g., 1km, and for small catchments, the computational cost should be low for even higher resolution simulations).