Reviewer 1

Thanks for tacklig an important question in hydrogeophysics with your manuscript. I went through with pleasure. It is written straight forward, thus easy to follow. Figures are well presented.

Nevertheless I have some comments basically on the methodology. I hope they help to clarify/sharpen the main purpose of the manuscript, solve some unclear interpretation and maybe extending the research context.

To me the main purpose of the manuscript is the (more objective and formalized) cluster analyses enabled by a larger number of SNMR sounding compared to "just" manual interpretation of multiparameter geophysical data. However, I think, this part needs sharpening and clarifications.

Author response: Thank you for your comments. We have followed your recommendations and believe the manuscript have been significantly improved. Especially the comments on the clustering formulation has been clarified as well as adding some discussion on joint inversion vs clustering.

--- Komedal:

-Figure 3 gives the silhuette index for the clusters and the text introduces the terms "well defined", "fairly defined" and "poorly defined" that are used to decide for the 3-cluster case. First, the definitions of the terms is vague. For instance, "well defined" is defined by "most of the data ...". What does "most" mean? Give better definitions.

Author response: We have added a more thorough description of the silhouette. We have added the average silhouette index to the figure for each of the cluster sequences. We are interested in having both well defined clusters, while maintaining a high average silhouette index, which is the case for the Kompedal 3 cluster case. We have added the average silhouette index to the silhouette figures to better highlight this.

P7.L21-L26:
"The preferred number of clusters is chosen based on two criteria. Firstly, the highest average silhouette index as indicates that datapoints in general have the highest membership score with the given number of clusters. Secondly, we look at each cluster and their silhouette index. If more than 50% of the cluster is above the average silhouette index, the cluster is well-defined, between 30-50% the cluster is fairly-defined, and below 30% it is poorly-defined."

-Further, to me it is unclear why the 3-cluster case is taken. That case contains a "poor defined" cluster! Taking the siluette analyses I would go for two clusters. However, it is argued with lithological knowledge one would take 3 cluster and indeed when looking at Figure 4 it very much looks like 3 cluster. But this is not the outcome of the siluette analyses.

Author response: It is true that when looking only at the silhouette values for single clusters the two-cluster case looks better. We have added more description on how we take the total average silhouette index into account. In the Kompedal case, we have added more description on the clustering results.

P9L10-P10L2

"For our combined analysis, we begin by selecting the number of clusters, K, using the silhouette index. ***Error! Reference source not found.*** *shows results from four different clustering analyses with two to five clusters for the Kompedal data set. Each cluster is labeled with its index and the number of data points within each cluster. In each cluster, the silhouette indices are sorted to give a higher index when moving up the y-axis. We use the distinction of well-defined, fairly-defined and poorly-defined, subdivided as mentioned in the results section. In the two-cluster analysis in* ***Error! Reference source not found.****a, we see that both clusters are well-defined with more than 300 members in each and could be a well-suited number of clusters. With three clusters (****Error! Reference source not found.****b), cluster 1 is fairly-defined, while cluster 2 is poorly defined with many data points having a below-average silhouette index, and cluster 3 is well-defined. In the four-cluster analysis, two clusters, cluster 1 and cluster 4, become poorly defined, as seen in* ***Error! Reference source not found.****c. Lastly, five clusters yield three poorly-defined clusters (2, 3 and 4) with only few data points having a high membership score. The total average silhouette scores indicated by the grey line highlight that either two or three clusters should be used."*

-Finally it is agrued that cluster 3 is interpreted differently when being shallow or deep (page 10 lines 7 to 10). To me this "additional" interpretation to the cluster analyses is really weakening the more objective and automatic approach.

Author response: Thank you for the comment. This is indeed less objective. From the parameters presented, it is not possible to distinguish these different lithological units, unless more information is added, such as the water table to distinguish the unsaturated from saturated material. A more appropriate interpretation would be that the red cluster is a low-free water unit with high resistivity. For non-experts this might be too specific, which is why we opted for two interpretations. Another option would be to include the depth in the clustering. But as we tried only to include inverted parameters in this study we left out fixed parameters.

-It is argued that the cluster analyses is somehow a "brutal" approach for a smooth inversion. But then, why not using a layered inversion?

Author response: The smooth inversion is used as we do not have the flexibility to use a layered inversion with the current modelling setup for steady-state sequences used for SNMR. Here, we are bound to a fixed discretization stemming from the kernel calculations to limit computational time.

-In particular I am also suprised by the missing contrast of resistivities between saturated and unsaturated. Water saturation is a driver of the bulk resistivity. If saturation changes, resistivity changes! Is that maybe because of missing sensitivity of "inductive electromagnetic methods" to changes in "higher" resistivities, thus parameter uncertainties for rather resistive layers? It is surely not because of inability in the physical parameter as written for instance at page 12 line 10 or page 21 line 18-19. Electrical methods are used to monitor the vadose zone and water infiltration. Please be more clear. And not at least, TEM is used to monitor groundwater table! (Zamora-Luria et al 2024, Long-term monitoring of water table and saltwater intrusion ..., Near Surface Geophysics, Vol. 22). However, the cluster analyses solves that "problem" nicely into 3 clusters (because of the SNMR sensitvity to water content), thus, this seems to be the advantage of the cluster analyses. Please elaborate on this.

Author response: Yes, indeed the water saturation is a driver of the bulk resistivity. But as the saturated and unsaturated material has resistivities over 150ohmm, it is difficult to separate these in TEM as most sensitivity is below this. Electrical methods could perhaps be a more suited to separating these two in this case. Yes, TEM has been used for groundwater table monitoring but relies on very fixed inversion setups to catch these minimal changes as well as longer stacking as the system is stationary. Especially the discretization has to be matched accordingly, otherwise the water table might be smoothed out lying not at a layer bound but within a layer.

We have rephrased the sentences in the paper to more accurately describe, that it is the lack of sensitivity to changes in high resistivity of TEM, and not the lack of change in resistivity between the saturated and unsaturated media.

P16L5-L7

"*The lack of structure in the ρ indicates that the TEM is not sensitive enough to track this saturation change, whereas a lithological change would generally be expected to coincide with a larger ρ contrast, visible in the TEM data.*"

P21L18-19

"Similarly, relying solely on TEM data may make it difficult to detect the water table due to limited sensitivity to high-to-high resistivity contrasts."

--- Endelave:

As for Komedal, please show the siluette plot for different numbers of cluster.

Author response: The silhouette plot for Endelave has been added and described.

P19L3-P19L19

"As before, we start by selecting the appropriate number of clusters, through silhouette index analyses, shown in **Error! Reference source not found.**. Considering we expect a more heterogeneous geology, three to six clusters are used in the analysis. In **Error! Reference source not found.**, three clusters are used to partition the data, and result in one well-defined, one fairly-defined and one poorly-defined cluster, whereas the yellow has low and even negative silhouette values, indicating wrongly assigned data points. The average silhouette index is the highest found with the assigned clusters. By using four clusters in **Error! Reference source not found.**b, two are well-defined, one fairly and one poorly clustered. We see less negative silhouette index data here, while still maintaining a high average silhouette index. Further increasing the number of clusters to five reveals similar silhouette indexes but has two fairly-defined clusters, however the average silhouette index drops, see **Error! Reference source not found.**c. Using six clusters is similar with a few well-defined and fairly-defined, and with a lower average silhouette index. The silhouette analyses show that the number of clusters should either be three or four as they have well-partitioned clusters, with the highest silhouette index. Prior information from the area indicates that we have four distinct geological units: tills, sand aquifers, Paleogene clay, and possible saline intrusion into sand. The blue cluster in **Error! Reference source not found.**b was found to have important hydrogeological information, regardless of its low silhouette index and, as such, we used four clusters for further results."

I found the parameters and hydrological interpretation in the clusters somewhat suprising. Figure 8 indicates 4 cluster, no doubt, but especially the T2* times of sandy aquifer are quite low and the range between "clay", "till" and "sand" similar while for the saltwater saturated sand (saline sand is not a good term) T2* is higher? Why is that? Is that related to "impacts" (iron) on the T2' relaxation times? Explain! However taking a look at the spatial distribution the clusters are reasonable, so why is that? Apparently the cluster analyses works well and help identifying units. So this could be a real benefit. Please elaborate on this!

Author response: We agree that saline sand is not a good term and it has been changed throughout the manuscript.

The increase in T2* in the saltwater saturated sand cluster is surprising. Likely, this effect stems from the lithology in which the saltwater is intruding into is mostly different from the sand aquifer found in other soundings. This less compacted sand could be associated with larger pores and thereby an increase in T2* not directly linked to the saltwater.

Introduction/Methodology -> Context of research:

- Clearly there are other papers already dealing with joint interpretation of electrical methods and SNMR to solve the ambiguities in hydrogeophysical tasks (saltwater - clay, or vadose zone - clay). I think those should be mentioned. For instance to name just one - there are others as well:

Guenther and Müller-Petke, 2012, Borkum ...

- Furthermore it is also necessary to point out joint inversion approaches that are in particular used to get joint layer boundaries for the interpretation. This is also necessary to point out that SNMR demands a resistivity distribution. There are different approaches around, for instance you may refer to:

Behroozmand

Skibbe

Author response: We have rephrased parts of the introduction and included more papers in this subject.

P4L9-L12

"Others have used a joint-inversion approach where layer boundaries are set using multiple geophysical methods (Günther and Müller-Petke, 2012; Behroozmand et al., 2012). The joint approaches have the ability to delineate layer boundaries, not seen when inverted separately."

- SNMR may also be able to detect water in partly saturated sands with fast relaxation. The is research of SNMR in the vadose zone and soils. For instance:

Flinchum, B. A., Holbrook, W. S., Parsekian, A. D., & Carr, B. J. (2019). Characterizing the critical zone using borehole and surface nuclear magnetic resonance. Vadose Zone Journal, 18(1), 1-18.

Walsh, D., & Grunewald, E. (2012, January). Application of surface NMR measurements to characterize vadose zone hydrology. In Symposium on the Application of Geophysics to Engineering and Environmental Problems 2012 (pp. 229-229). Society of Exploration Geophysicists.

Hiller, T., Costabel, S., Radić, T., Dlugosch, R., & Müller-Petke, M. (2021). Feasibility study on prepolarized surface nuclear magnetic resonance for soil moisture measurements. Vadose Zone Journal, 20(5), e20138.

Discussion:

Appears more a like a repetition of already mentioned statements.

From a methodlogical point of view I would expect discussion on pro/cons compared to existing approaches of joint inversion/interpretation especially having the comments above on clustering in mind.

Also discuss why not or how to combine a joint inversion with clustering, for instance using a layered joint inversion.

Author response: We have added a paragraph on joint inversions and how we opted for the separate inversions to use the spatial distribution of TEM to try and track boundaries more spatially.

P28L25-L35
"Another way of exploiting collocated datasets is the use of joint inversion for layer boundary picking. Studies identifying layers from SNMR and TEM implementing various regularization techniques has shown promise in reducing the ambiguity found when interpreting each separately (Behroozmand et al., 2012); Skibbe et al., 2018). These approaches focus mostly on the collocated datasets and invert these jointly. In our study, the tTEM data is inverted separately with the full survey of more than 23000 datasets. As such, we have the ability to track the changes in resistivity in places where the SNMR is not present. Additionally, the framework for using joint inversion in steady-state SNMR is not established as kernels are calculated before the inversion, fixing the discretization. Further investigations could focus on implementing clustering in a joint inversion framework with a large spatial extent. This could alleviate some of the interpretational load when dealing with large datasets."

It seems as only T2' is used here and might results in similar values even for different units. Why not using T2 or T1 as possible with the APSU device?

Author response: Yes, only T2* has been used in this manuscript for the clustering. At the moment, T1 is fixed and set equal to T2 in the inversion framework. We decided to use T2* as at the time of writing the resolution capability of T2 in steady state had not been demonstrated. Additionally, the T2 does at these sites not add a lot of different information for the clustering.

Some minor text comments:

page 2 line 23: replace inconsistent by ambigious
Author response: Fixed

page 3 line 1: cannot distinguish: this is a bit harsh. there are papers around who deal with partly saturated NMR stuff (see above)

Author response: Sentence has been rephrased to say that it is difficult to distinguish these scenarios.

P3L4-L7
"As such, SNMR has difficulties distinguishing unconfined aquifers from semi-confined or confined aquifers without supplemental data, as the increase in water content cannot be established to be a saturation or a lithological transition (Behroozmand et al., 2015), Fig. 1."

page 3, second paragraph and figure 1: this is exactly what is described in (Günther, T., & Müller-Petke, M. (2012). Hydraulic properties at the North Sea island of Borkum derived from joint inversion of magnetic resonance and electrical resistivity soundings. Hydrology and earth system sciences, 16(9), 3279-3291.). Please cite or refer to it. It is a kind of perfect paper that lays ground for your paper as you go beyond this manual interpretation by your cluster analyses.

page 5 line 13: give numbers to deadtime

Author response: Fixed

page 5 line 14: define partly or fully decay

Author response: Added a description of this.
P6L1-L3

"Signals from very small pores can therefore partially or fully decay, i.e. lose its amplitude and coherency, before the instrument has begun recording data."

page 5 paragraph 2.3: how is resistivity for MRS inversion handled. Any coupling there to the TEM?

Author response: The resistivity structure from the nearest TEM sounding is used for kernel calculations for the MRS inversion and is fixed. (Stated in P5L30)

page 11 table 1: as above, 1000 ohmm for saturated sand appears too high (cluster 1) and the overlap to cluser 3 in terms of resistivities is also high. this needs to be explained.

Author response: The very coarse material does have quite a high resistivity. But it is hard for the TEM to map the difference between a 300 ohmm layer and a 800 ohmm layer. Therefore, the high resistivities could be a product of the lack of sensitivity in the coarse unit.

We have added a description of this before the table.

P13L11-L13

"The very high ρ (above 300 Ωm) is a product of very coarse material and that the TEM method can have limited sensitivity to determine resistivity above 150 Ωm."

page 16, table 2: t2* times of the saltwater sand and the freshwater sand are quite different. why?

Author response: As mentioned above, the newly deposited material in which the saltwater is intruding into, probably has a larger pore size than parts of the deeper aquifer which is mostly deposited by glaciers, likely has some remnant clay content.

page 16 line 17: saline sand -> replace by sand saturated with saline water (or something similar but saline sand is not correct)

Author response: This has been corrected in the manuscript.