Review of

Alleviating interpretational ambiguity in Hydrogeology through clustering-based analysis of transient electromagnetic and surface nuclear magnetic resonance data

The work that is presented in this manuscript is an important step towards an automated and thus fast hydrogeological interpretation of combined TEM and SNMR measurements. This combination has been applied in previous studies and has been proven to be powerful, especially within the frame of groundwater characterization in coastal areas, where the risk of relevant saltwater intrusion appears and frequently generates problems of freshwater production. In this environment, advanced knowledge of the aquifer situation, hydrogeologic modelling and temporal observation of the groundwater quality is desired and can most oftenly not be realized by drilling boreholes alone.

Recent improvements of TEM and SNMR have enabled options to produce big datasets that need to be interpreted in time to gain real social, economic and ecologic relevance in supporting freshwater production. Consequently, automated approaches such as the one presented in this study are the next logical step to deal with such data. The paper represents a significant contribution within the field of hydrogeophysics and its application and consequently fits in very well with the HESS journal.

The manuscript is well structured and the principle path to the results and conclusions is easy to follow. The figures are well presented and helpful to understand the key messages. However, some important technical information and discussions are missing, especially on details of the implementation and performance of the K-means approach. As I understand the study, the focus is not on the verification of the approach, e.g. by concrete ground truthing which is difficult given the large spatial extent covered by the surveys, but rather the introduction of the principal strategy that has still potential to be further developed. Please read my detailed concerns and recommendations below.

In the presented case studies, the outcome of the clustering analysis is basically tested against the classical manual interpretation at the basis of the original data with only a few observation boreholes. These are not sufficiently distributed over the profiles to enable reliable validation to the full extent. Consequently, the discussion section should honestly admit that there are still significant uncertainties, which cannot be resolved for several reasons. Please find details below.

However, the focus of this work and the value of automated clustering approaches in general is not on interpreting local situations as accurately as possible, even if the data at the specific position possibly allows this, but on providing a reliable overview at a larger scale. In this sense, I want to encourage the authors to include such reasoning in their motivation and discussion. Again, please find details below.

Finally, I assess the necessary revisions of the manuscript into the cathegory moderate-to-major.

Details

P2L12-13:

… electrical properties of the subsurface, specifically the resistivity, are used …

P2L22:

A limitation of…

P2L23:

The sentence starting with „An implication of …“ is overloaded with redundancy and word repetitions. Please reformulate.

P3L1:

„…making it immeasurable with…“ => „…which makes the clay-bound water not detectable with…“.

P3L3:

„established“ => „determined“

P3-Fig1:

I like the idea of this scheme. However, I do not see the necessity of providing the axes as colorbars. It is rather confusing – I tried without any success to relate the font color inside the scheme to the colorbars of the axes. If you insist of using them, please clarify in the figure caption how they should be understood. Otherwise, I suggest avoiding them.

Please change the label of y-axis: „SNMR water content“ instead of „water content“ to avoid confusion or misconception. As you correctly describe in the main text, clay normally exhibits high water contents, but apparently appears as low or zero water content in SNMR.

P5L5:

Incorrect statement: the increased stacking rate leads to higher singal-to-noise ratios, not to higher signal amplitudes.

P5L11:

Clarify with an additional brief sentence and a corresponding reference, why resistivity information is necessary for SNMR inversion.

P6L13:

I am not familiar with K-means approaches. However, as far as I know, the performance of any optimization algorithm strongly depends on the chosen termination criteria. Please explain in more detail, here or later in the main text, how your approach actually performes. How many iterations were chosen exactly? Did you reach its maximum number for the majority of the runs or did the algorithm converge properly to the predeterimed minimum distance? Can you quantify the minimum distance? Are there any criteria how to choose the minimum distance or is it rather a try-and-error process to find a proper threshold?

P7L8 (subsection 2.5)

This subsection, in my opinion, should appear at the very beginning of the methods section. Explain first what you did and where, before going into detail on each measurement method and processing approaches.

P8L2

Reading about „TEM soundings“ here is rather confusing – it indicates that you conducted the „traditional“ way of applying TEM. As I understand it, tTEM is rather a continuous sequence of TEM measurements while driving through the area. Please clarify.

Subsections 2.5.1 and 2.5.2:

You already provide results, their interpretation and some conclusions here, while introducing the investigation areas. This is Okay in principle, if all this information has already been puplished. In this case, please specify the corresponding references. Otherwise, these statements belong to Section 3.

Subsection 3.1.1 (on the description of the silhouette index analysis and interpretation of Fig. 3)

Unfortunately, some explanations in this passage remain unclear to me.

What exactly is the average silhouette index? Following the main text, it seems to be calculated for each cluster separately. But this is in contrast to the gray dashed lines in Fig.3 that represent a threshold for the whole data points in each subfigure. Moreover, it seems to be the very same value for each subfigure, no matter how many clusters are chosen - Is this plausible? Please clarify.

Furthermore, when having a closer look at Fig.3, I cannot follow the classification of „well-, fairly and poorly defined" that is made in the main text. It states for instance that cluster 1 in Fig.3b was well-defined (P8L22-24) but more than half of the data points in that cluster are below the average value, some points even appear with a silhouette index below 0. Following the definition in the main text this should be classified as a poorly defined cluster (P8L20: „…many data points below the average…"). Similar confusion appears for the interpretation of the other subfigures. However, maybe I had missed something relevant and completely misunderstood this analysis. Please clarify and reformulate the text accordingly.

P10-Fig4

Please specify the appreviations of the legend entries in the figure caption or provide a reference to table 1.

P11-Fig5 (and the other figures with profile data)

Please consider indicating <N>orth and <S>outh directions at beginning and end of the profile(s) to make it easier to follow the interpretation(s) in the main text.

P12L11

Following Archie's law, a change in water saturation should also lead to a contrast in rho. The fact that the change of water content measured by SNMR is not indicated by the TEM data is most likely due to the fact that TEM is not very sensitive to rho contrasts above 100 Ohm*m. This information should be given beforehand together with a suitable reference, e.g. when describing the scheme in Fig.1.

P12L13

I do not doubt this statement, but you should provide a reference to verify it. There are some papers around that study the relationship between water saturation and relaxation time for loose sediments.

P13-Fig6 (same for Fig.10 and 11)

I recommand focusing on T2* alone in the lower subfigure(s) to avoid confusion. My first reaction was wondering what additional information the ratio of T2*and T2 could provide. I guess there is no additional information in the T2 parameter compared to T2*, right?

P13L14

Please describe more in detail how the error bars in Fig.6 must be understood and how the uncertainty in assigning the data to the clusters affects the uncertainty of the water table estimation. What means „the uncertainty bars are based on the layer thickness“? Is it identical with the layer thickness? In this case, one could assume that, in general, we just have to decrease the layer thicknesses in the inversion model to increase the accuracy of the water table estimation. This would be nonsense, of course!

Sorry for stressing this aspect, but the confidence of non-invasive geophysical water table estimates is, in my experience, a frequent question in technical discussions with hydrogeologists. Your Figure 7 could be a key figure for the combined TEM/SNMR approach in this regard. However, it is not yet clear how to interpret it exactly.

And there is another aspect to consider when interpreting Fig.7. The water table is never identical with the interface between saturated and unsaturated zones. The saturated zone is always above the water table due to capillary forces. Depending on the lithology the difference between the two varies between a few cm (coarse sand) to several meters (till). Is such an interpretation of the descrepancy between the data points and the 1:1 line also reliable, given the actual geology?

P13L21 and Fig.7

I strongly recomment erasing the two yellow points from the crossplot. The attempt to verify the recent water table estimates with completely outdated information has no scientific value, even if it matched the 1:1 line by chance - which it obviously does not!

P13L12

Why do you not show the silhouette index plot for the Endelave site as well? Even if it was more complicated and difficult to interpret than for the case in Fig.3, the reader could learn a lot from it! How objective can the choice of the number of clusters even be? If an objective approach fails, we have to discuss criteria for subjective choices. As the number of clusters is crucial for this kind of analysis, I would expect a more detailed discussion on it for the two cases presented in this paper.

P13L13

I recomment to erase the sentence „Consider Figure 8,…“. The wording is technical incorrect, because the clustering results cannot be represented by two parameters. It is always three parameters regardless of how they are depicted. However, the sentence is not necessary at all.

P14L17

The red cluster might also include unsaturated sand or can you exclude this possibility for some reason?

P14L19

saturated sand aquifers

P14L22

„could indicate“ => „indicates“

P15L2

„would indicate“ => „indicates“

P15-Fig8

Please specify the appreviations in the legend.

P16L5

Okay, we learn later that unsaturated sand at this position is somewhat unreliable because of the borehole information. However, I miss a discussion in the paper about the resolution properties of the tTEM at shallow depths < 10 m. This is relevant for the interpretation of the near-surface till layer. TEM does obviously not resolve it, although one would also expect low resistivities here.

P17L8

Figure 9a

P17L15

What does „due to shifts in geological deposits“ mean in this context? Please explain more in detail.

P17L20

A few boreholes? There is only one borehole in the vicinity of sounding 1 and 2. Furthermore, I would expect a decrease in rho at the SA-TI interface. Again, this contrast can obviously not be resolved by the tTEM method in this specific case. However, in Fig. 6 we see that it resolves a rho contrast between SA and TI even at a similar depth. Why not here?

P18-Fig10

The reddish colors of Silt and Clay/till are difficult to distinguish in the figure. Please use another color for one of them.

P19L4

Please be precise! The organic matter does surely not lead to a drop in WC (the opposite is the case!), which is very obvious when inspecting the cited paper of Mashhadi et al. (2024). Similar to clay - also with gyttja it is actually not a „real“ but an apparent drop in WC caused by the fact that short T2*/T2 times are not detectable with SNMR.

P19L18

„might have“ => „might cause“

P20L8 – P21L6

This passage is redundant with the descriptions in the introduction and can be removed from the discussion section completely. Please only discuss here what is new and implied by the direct results of your study.

P21L20

As already mentioned (comment on **P12L13**), there are references to verify this statements. This is not a conclusion or finding of your study.

P21L23

„confidence in“ => „confidence of“

P2L26

„reduce“ => „reducing“

P2L28

„describe the most variance“ => „describe most of the variance“

P21L32

„informed“ => „qualified“

P22L5

provides

P22L29

I strongly doubt that the suggested approach is strictly „non-subjective“, because of the crucial predefinition of the number of clusters (please see my comments on the silhouette index approach). Use „less-subjective“ or a formulation such as „towards a non-subjective interpretation“.

P23L3

I disagree with this statement! Only for one of the site, the silhouette index approach was introduced and described, at all, and even this is hardly comprehensible (see my comments on section 3.1.1). For the second site, the choice of the number of clusters seems rather arbitrary to me. Finally, no concluding remarks can be made about the robustness of the suggested approach. This is also true for the whole K-means algorithm as it is used here and remains an objective for future research. The robustness could, for instance, be analysed in a pure synthetic parameter study.