https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

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

Mathias Vang[1], Jakob J. Larsen[2], Anders Vest Christiansen[1], Denys Grombacher[1]

[1]Department of Geoscience, Aarhus University

[2]Department of Electrical and Computer Engineering, Aarhus University

Correspondence: Mathias Vang (mva@geo.au.dk)

Abstract

Local characterization of groundwater systems is critical for managing and protecting vulnerable resources. Geophysical methods can provide dense imaging of subsurface parameters to delineate lithological boundaries and water tables for hydrogeological investigation. Though, using a single geophysical method for determining lithologies can yield erroneous interpretations as different lithologies can have similar properties. By using several geophysical methods, it is possible to reduce this risk and better assign likely lithologies to subsurface units. We present two case studies where transient electromagnetic (TEM) and surface nuclear magnetic resonance (SNMR) are used in combination to delineate hydrogeological structures. Novel spatially constrained inversion in SNMR was used to provide horizontal consistency between soundings. Three coincident parameters, resistivity from the TEM measurements and water content and relaxation time from the SNMR measurements were used in a K-means clustering scheme to resolve subsurface structures. The K-means clustering was evaluated with a silhouette index to pick the number of clusters. After clustering, each cluster was assigned a hydrogeological description based on the distinct features in the three parameters, e.g. a low resistivity, high water content, and high T_2^* is assigned as saltwater saturated sand. In the first case study, the clusters enabled improved resolution of a regional water table in an unconfined aquifer setting by the multi-geophysical approach. The water table estimates were positively evaluated against multiple boreholes within 500 m of coincident geophysical models. The second case study illustrates how clustering, of SNMR and TEM models, can delineate saltwater intrusion in an island coastal aquifer, which would not be possible with any of these methods individually. Additionally, the clustering resolved the main shallow aquifer on the island. Our work illustrates how the combination of geophysical data can be used to improve resolution of hydrogeological layers and reduce interpretational bias.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

1 Introduction

Climate-resilient groundwater management hinges on the need for detailed characterization of local groundwater systems (Dragoni & Sukhija, 2008). Historically, lithological descriptions of wells have been used to establish geological models to forecast local groundwater behavior and inform conceptual models of local systems (van Roosmalen et al., 2007). The high cost associated with drilling yields geological maps that are generally based on sparse point coverage, with long-distance interpolation, and simplicity assumptions between observations where structures may actually be complex. To address these data sparsity issues, geophysics can be used to delineate structures non-invasively, giving high resolution imaging of the subsurface to complement direct borehole observations (Binley et al., 2015). Methods based on imaging of subsurface electrical properties are used extensively in hydrological investigations, where spatial variations in the electrical properties, specifically the resistivity, of the subsurface are used to study pollution, explore groundwater resources, and delineate saltwater interfaces, among many other applications (Binley et al., 2015). Within methods imaging electrical properties, electromagnetic (EM) methods are widely used. They operate inductively by creating a varying magnetic field inducing eddy-currents in the ground (Nabighian & Macnae, 1991). The secondary magnetic field produced by the decaying eddy-currents is measured inductively at the surface. The measurements are rapid, which leads to high data acquisition rates that enable mapping of large areas using towed or airborne platforms (e.g. Auken et al., 2019; Sørensen and Auken, 2004). The EM data are translated into 1D models of resistivities by inversion (Christiansen et al., 2006), providing valuable insights into local (hydro-)geology. A limitation in these methods is that they rely on inconsistent links between lithology and resistivity. An implication of this is that assigning lithology to a specific electrical resistivity requires local knowledge of the link between resistivity and the associated lithology or geological unit (Dickinson et al., 2010). A common challenge is that different geological units have overlapping resistivity ranges making unique identification based on resistivity alone difficult or sometimes impossible.

Surface nuclear magnetic resonance (SNMR) provides direct sensitivity to water residing in large pores (Hertrich et al., 2007; Legchenko et al., 2002). By transmitting an excitation pulse oscillating at a specific frequency proportional to the Earth's magnetic field strength, the magnetic moment of hydrogen nuclei is shifted from its equilibrium state (Yaramanci et al., 1999). After terminating the pulse, the buildup magnetization decays and is related to the subsurface water quantity and pore parameters. This allows SNMR to track changes in water content across lithological boundaries and can provide valuable information on pore sizes. A limitation in SNMR is the inability to distinguish unsaturated sand from clay, as both will be seen with low WC, in the clays caused by the magnetization decaying extremely rapidly in

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

small pores making it immeasurable with the SNMR. As such, SNMR cannot distinguish 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. However, the combined interpretation of SNMR and TEM data, sensitive to different properties, may alleviate ambiguities in distinguishing between for instance unsaturated sand and clays (SNMR ambiguity) or clays and saline saturated sands (TEM ambiguity), which is highly relevant for coastal studies of unconsolidated settings (Costabel et al., 2017).

Consider the example of an unconfined/confined system, where SNMR cannot determine whether a transition from low to high water content marks the water table or a lithological shift from clays to sands. TEM can address this as it would resolve the conductive clay layer if present and delineate the lithological change to sand as seen in Figure 1. If it was an unconfined system, the TEM would image high resistivities in both layers while the saturation change is tracked by SNMR. Another example involves saline intrusion, where TEM cannot differentiate between saline sand and clay. If it is indeed a transition only in salinity, not water content, SNMR would reveal continuous high water content across the salinity boundary. SNMR alone would not be able to distinguish freshwater sand from saline sand, as it is only sensitive to the abundance of water and not salinity.

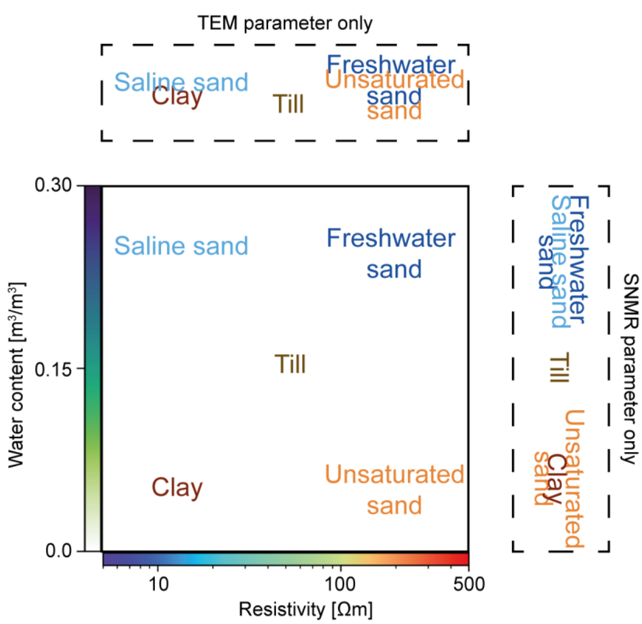

Figure 1 Different hydrogeological units resolved with TEM and SNMR. In dashed boxes, only one method is used, and the overlapping units show the ambiguities found. T_2^* can be implemented to further separate units.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

A multiple data type approach requires forming interpretations consistent with multiple geophysical model types simultaneously, which can be achieved through manual inspection of disparate data types. This enables one to distinguish hydrogeological layers through combined interpretation of all data types, but requires subjective choices regarding boundary delineation. An alternative approach employs statistical correlations across separate parameters to partition these into different clusters. One such approach is K-means clustering, which enables the subdivision of datasets based on multiple parameters (Kodinariya & Makwana, 2013). Different clustering approaches have also previously been applied to geophysical data and focus primarily on single source datasets, such as large EM datasets (Dumont et al., 2018) or large electrical resistivity datasets (Song et al., 2010). Some studies investigate clustering on derived parameters such as clay fraction and resistivity, both linked to EM surveys (Foged et al., 2014). Clustering across disparate data types, such as Bouguer anomaly data and magnetic data has been shown to improve the resolution of mineral deposits (Sun and Li, 2016). A study focused on delineating structures in urban settings by clustering on multichannel analysis of surface waves (MASW) and electrical methods to evaluate soil foundation structure (Le et al., 2022) and found the K-means clustering to resolve important structures in the shallow subsurface.

In this study, we demonstrate the benefits of combined SNMR and TEM data collection, where K-means clustering based on coincident models in two survey areas is shown to enhance interpretations and address ambiguities that persist if only a single data type is considered. The first example includes mapping of the water table in an unconfined meltwater plain aquifer, where a combined approach is used to address ambiguity as to the upper aquifer being confined/unconfined/or semi-confined across the investigated region. A second example taken from a small island shows how the method can delineate salt-water intrusion from clay-rich regions through a combined interpretation. We demonstrate a workflow for handling interpretations of SNMR and TEM simultaneously reducing possible interpretational bias.

2 Methods

2.1 Transient electromagnetic

In this study we use Transient Electromagnetics (TEM) to resolve subsurface resistivities. The tTEM instrument (Auken et al., 2019) was used in both field areas and can resolve the resistivity structure of the top 70m, however, here only the top 25m of the full model domain are used in the analyses. The induced voltages recorded by the tTEM are translated to 1D resistivity models by Spatially Constrained Inversion (SCI) using Aarhusinv (Auken et al., 2015; Viezzoli et al., 2009). The model is discretized into 30 layers with thicknesses varying

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

from 1m shallowly, to 10m at depth following resolution limitations at depth. The resulting resistivity models will be used for subsequent clustering.

2.2 Surface nuclear magnetic resonance

In this survey we use a recently developed technique for SNMR called steady-state. The steady-state has an increased stacking rate leading to a higher signal amplitude and a decrease in acquisition times (Grombacher et al., 2021). A set of transmit pulses, optimized to resolve the top 25m, was employed in both studies with the Apsu instrument with an acquisition time of 25min per site (Larsen et al., 2020). The resolved water content and the relaxation parameter, T_2^*, are used in the subsequent clustering. The SNMR models are discretized into 31 layers down to 50m, increasing in thickness at depth from 0.5 m to 4 m. The resistivity structure from the nearest TEM sounding is used for the inversion.

One limitation in SNMR is to detect water residing in very small pores. Because of instrument dead times associated with transmitting the excitation pulse, receiving data immediately after pulse termination is not possible. Signals from very small pores can therefore partially or fully decay before the instrument has begun recording data. As such, the magnetization from water residing in very small pores decay prior to data recording, which prevents observation of small pore water in SNMR. As such, SNMR water contents can be interpreted as a measure of "free" water or an effective porosity. T_2^* relaxation time is linked to pore sizes with low values occurring in small pores, while large pores have large values. This can be used to differentiate high water content units by their pore sizes.

2.3 Inversion considerations

Traditionally, 1D SNMR inversions are most commonly treated separately as limited measurements are carried out. However, recent acquisition speed-ups enabled by steady-state approaches have significantly enhanced spatial data density, which enables the use of horizontal constraints linking inversions of nearby measurement sites (Grombacher et al., 2021). One such example is the use of laterally constrained inversion (LCI) for SNMR as proposed by Behroozmand et al. (2012) where neighboring sites in a transect can be connected. Here, we add a dimension to the constraints using a spatially constrained inversion (SCI) framework, not only to bind models in line, but all neighboring models. Delauney triangulation is used to find the relevant neighbors as in Viezzoli et al. (2009). The strength of lateral bounds is scaled by the distance between models, with a maximum strength defined when models are closer than a threshold distance. This threshold distance is typically set to the nominal or average distance between neighboring soundings (Vang et al., 2024).

The computational load increases immensely when implementing SCI with many layers and parameters. To reduce the number of iterations, the SCI starting models are defined by single

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

site inversion results. This allows the SCI to converge within a few iterations. The TEM data are inverted separately with an SCI for the entire survey.

2.4 Clustering

Large datasets enable statistical approaches to inform on significant hydrogeological units. In the following examples, datasets are composed of 50 and 51 coincident SNMR/TEM soundings where a K-means clustering is employed (Kanungo et al., 2002) on their model parameters. The first step in this type of clustering is to select the number of clusters, K, into which the data sets will be clustered (Kodinariya & Makwana, 2013). After selecting the number of clusters, the algorithm makes an initial guess for the position of each cluster center in the parameter domain. The Euclidean distance from each data point to the cluster center is calculated, and each data point is assigned to the nearest cluster. The total distance from all data to their assigned clusters is then iteratively minimized through updating cluster center locations until either a minimum distance or a maximum number of iterations is reached.

To improve clustering of datasets for parameters exhibiting different sensitivities and spanning different ranges, normalization was used to ensure that each parameter has the same weight in the clustering algorithm. Here, we use a Z-score for normalization, where x is either resistivity (ρ), water content (WC), or relaxation parameter (T_2^*):

$$x_{i,norm} = \frac{x_i - \frac{1}{n}\sum_{i=1}^{n} x_i}{\sigma} \quad (1)$$

where σ is the standard deviation on the cloud of parameters from the inversion, x_i is the parameter value for the i^{th} data point and n number of data points. Following the normalization, we use the Scikit learn package in Python for the clustering and silhouette analysis (Pedregosa et al, 2011).

In this study, the number of clusters is chosen based on the Silhouette index, which calculates the membership S_i of each data point, i:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)}, S_i \in [-1, 1] \quad (2)$$

where a_i is average distance from data point i to other data points in the same assigned cluster, b_i is the minimum average distance of the i^{th} data point to all other data points in other clusters. The resulting index, or membership score, is a measure of how well a data point is associated with the assigned cluster. If the score of a given data point is 1 it infers that the data point is correctly assigned, while a score of -1 indicates that the data is wrongly assigned (Kodinariya & Makwana, 2013; Shutaywi & Kachouie., 2021). By evaluating these results, we can qualify the preferred number of clusters. In some cases, prior information can be used to fix the number of clusters, such as prior geological knowledge of the area (Dumont et al.,2018).

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

In this study we clustered on three parameters: WC, T_2^*, and ρ. The two geophysical methods used in this study have different sensitive volumes. SNMR inversion is discretized finely with 30 layers down to 50 m and the TEM has 30 layers in 120 m. To cluster on coincident values, a projection and averaging of the TEM ρ models onto the SNMR discretization is used. All TEM soundings within 60 m of an SNMR sounding are included. If there is no ρ model (TEM sounding) within 60 m, the nearest is used and mapped onto the SNMR discretization. This allows all SNMR points to be matched and prevents a reduction in data points.

2.5 Field site description

Two field surveys were conducted in different geologies to evaluate the use of clustering as a tool for alleviating interpretational ambiguity. Both sites were examined thoroughly with SNMR and TEM to provide the basis for the subsequent clustering analysis.

2.5.1 Kompedal

The first field site is Kompedal, a national forest in the Central Region, Denmark. The local geology consists of meltwater sand and glacial tills with varying clay contents. The sparse borehole coverage finds sand shallowly, and the water table varies from 5 m to 12 m in depth. Two geophysical surveys have been conducted here using TEM and SNMR, respectively. The scope of the surveys was to delineate the water table on a regional scale and assess whether the shallow aquifer can be considered unconfined or semi-confined across the region. The TEM data were collected with the tTEM instrument (Auken et al., 2019) while driving along the gravel roads within the forest, as seen in Figure 2a in blue. ρ in the area are generally high, above 200 Ωm, with some layers of lower ρ found at depth. There is little to no contrast between the unsaturated and saturated part of the meltwater sand in ρ. The SNMR survey consists of 50 soundings acquired over five days in June 2021, spread across the forest as seen in Figure 2a (Vang et al., 2023). The SNMR survey found low WC (~ 5 %) and low T_2^* values (~ 0.1 s) shallowly, with a sharp increase to higher WC (~ 25 %) at 6 m to 10 m depth. Layers with low WC and T_2^* can be associated with both unsaturated sands, and clay-rich material. The section indicated in Figure 2a will be used to show the results of the combined cluster analysis.

2.5.2 Endelave

The second location is a small 13 km^2 island, Endelave, in Kattegat, Denmark with a maximum elevation of 8 m. The island's geology consists of glacial till, meltwater sands, and post-glacial sands, while boreholes intercept Paleogene clay at depth throughout the island. Generally, the glacial tills are found in the west part of the island, where the post-glacial sands are found to the north. TEM and SNMR surveys shown in Figure 2b were conducted at this more

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

geologically heterogeneous location to resolve possible saltwater intrusion and delineate the shallow aquifer found in the meltwater sands and tills. The TEM soundings were acquired in April 2022 and cover the majority of the island and show ρ below 150 Ωm for the entire area. The ρ resolve buried valley structures and a very conductive basement. By TEM alone it is not possible to distinguish Paleogene clay from the saltwater saturated sand. The SNMR survey consists of 51 soundings over eight days in July and October 2023 and finds high WC shallowly in the east and north part of the island, where the west part shows low WC and T_2^*.

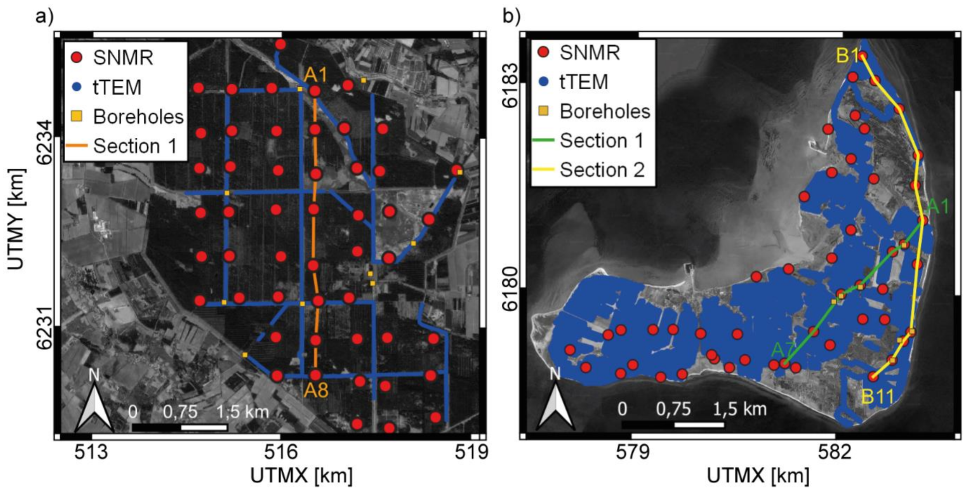

Figure 2 a) The Kompedal survey area. SNMR (red) together with TEM (blue) was collected in the area. b) Map of Endelave survey with SNMR (red) and TEM (blue). Map data: © Google Maps 2024, UTM Zone: 32N.

3 Results

3.1 Kompedal case study

3.1.1 Clustering analysis

For our combined analysis, we begin by selecting the number of clusters, K, using the silhouette index. Figure 3 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. If most of the data within a given cluster have values above the average silhouette index, it is considered as *well* defined, while clusters with some data below are considered *fairly* defined, and with many data points below the average silhouette index they are *poorly* defined. In the two-cluster analysis in Figure 3a, we see that both clusters are well-defined with more than 300 members in each. With three clusters (Figure 3b), two are well-defined, while cluster 2 is poorly defined with many data points having a below-average silhouette index. In the four-cluster analysis, two clusters, cluster 1 and

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

cluster 4, become poorly defined, as seen in Figure 3c. Lastly, five clusters yield three poorly defined clusters (2, 3 and 4) with only few data points having a high membership score. Given the silhouette indices, either choosing two or three clusters is appropriate. Prior hydrogeological information can be used to further qualify the choice between these (Dumont et al., 2018) and in Kompedal, we expect three distinct hydrogeological units, unsaturated sand, saturated sand and underlying till. The low silhouette indices in cluster 2 in Figure 3b, is a product of large variation within the cluster, which can be expected in glacial environments as mixing occured during deposition. Finally, three clusters were chosen to subdivide the data into meaningful and decently determined clusters.

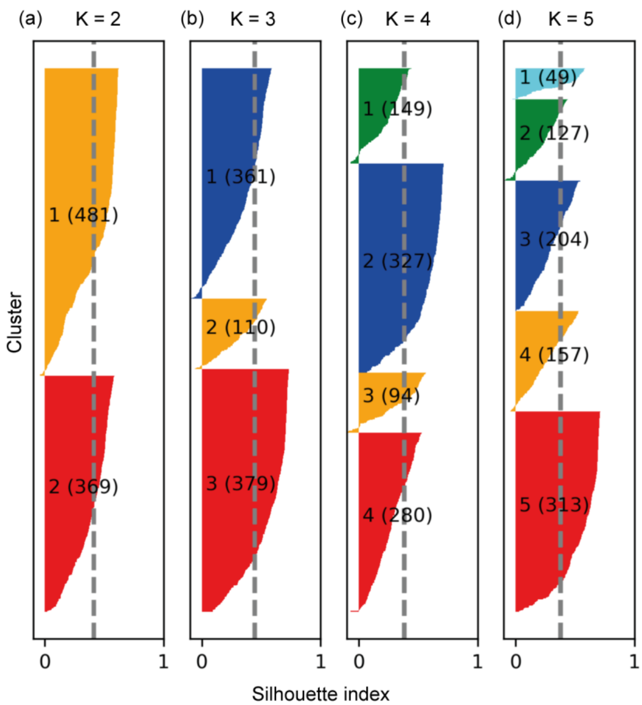

Figure 3 Silhouette index analysis for the Kompedal dataset. Four clustering routines were run with different number of clusters, K, (a) two, (b) three, (c) four and (d) five. The sorted silhouette values are shown for each cluster with the average value indicated by the grey dashed line.

Given three clusters, K-means clustering is used to partition the model parameters, WC, T_2^*, and ρ. In Figure 4a, the three model parameters are shown in a scatter plot where the color of a point reflects the assigned cluster. The other three 2D scatter plots, Figure 4b, c, and d, show the clustering results projected onto a plane that reveals correlations between two of the three. Cluster 1 in blue, is characterized by a high WC and high T_2^* value, and a high ρ. Table 1 shows that large variation occurs within this unit in the SNMR parameters as seen in Figure 4b. The unit is interpreted as a sandy aquifer given its high WC, high T_2^*, and high ρ. The

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

yellow cluster 2 has the largest variation in ρ, hence the low average silhouette index, but generally with lower ρ values than the other two clusters seen in Figure 4c, and with a large range in WC. A layer with these signatures is consistent with a saturated sandy till to a more clay rich till, with low ρ. The overlap with cluster 1 in WC and T_2^* in Figure 4b is interpreted as a gradual mixing of till and sands. Cluster 3 in red, has low T_2^* and a low WC and high ρ, which corresponds to unsaturated sand. However, low SNMR parameters in high ρ could indicate a silty deposit with smaller pore sizes, but with a similar conductivity. In places where the red cluster is found shallowly, it is interpreted as a unsaturated sand, and at depth under the water table, it is intrepreted as a saturated silt.

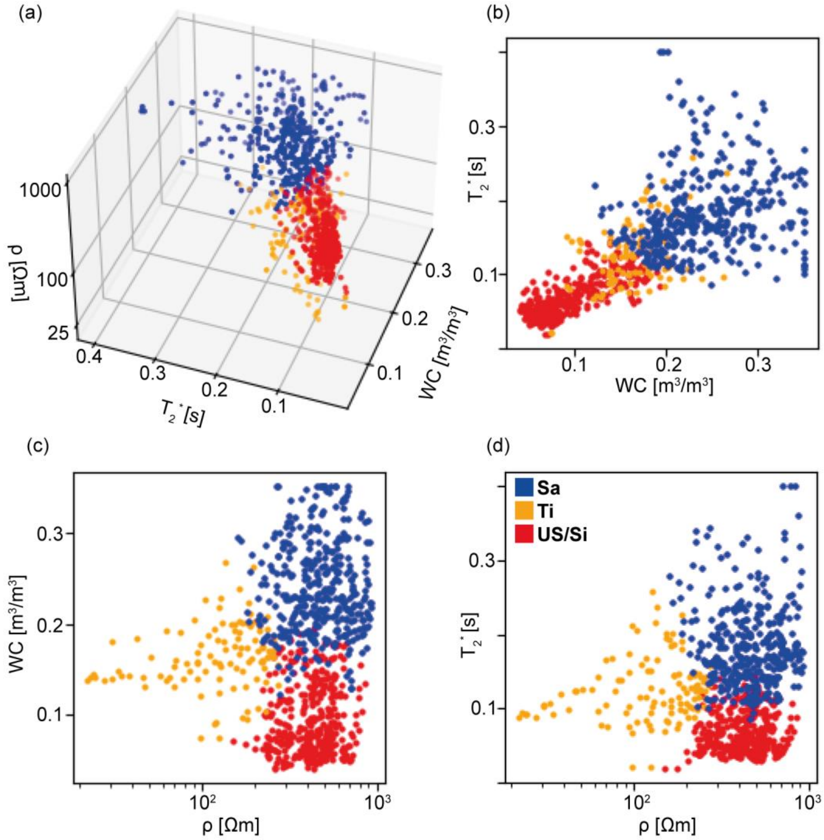

Figure 4 Clustering results from the Kompedal survey on three parameters: WC, T_2^* and ρ. (a) all three parameters in a scatter, (b) WC and T_2^*, (c) ρ and WC, and (d) ρ and T_2^*. The color of each datapoint defines the assigned cluster; 1-blue, 2-yellow, and 3-red.

Table 1: Cluster parameter bounds and interpreted geology for Kompedal.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

Cluster	**WC [m³/m³]**	**T_2^* [s]**	**ρ [Ωm]**	**Interpreted geology**	**Label**
1 (blue)	High [0.1-0.4]	High [0.1-0.4]	High [130-1000]	Saturated sand aquifer	**SA** (Sand Aquifer)
2 (yellow)	Medium [0.07-0.26]	Medium [0.03-0.26]	Low [20-300]	Saturated till	**Ti** (Till)
3 (red)	Low [0.04-0.18]	Low [0.03-0.14]	High [130-900]	Unsaturated sand	**US** (Unsaturated sand) or **Si** (Saturated silt)

3.1.2 Spatial interpretation

The three clusters are described in Table 1 and will be referred to by their labels, which are used in the figures to highlight their spatial extent. After assigning interpreted geologies to each cluster, we focus on their spatial position illustrated by a cross-section (location shown in Figure 2a). Consider section 1 in Figure 5, where the coincident data used in the clustering are shown as bars with colors associated with the assigned cluster. The US/Si cluster is situated mostly in the shallow subsurface extending from the surface down to depths of 5 m to 10 m. The grey lines track selected cluster boundaries at the sounding locations. The upper grey line in Figure 5 tracks the bottom of the US cluster and is interpreted to be a change from low-to-high saturation, since the US-cluster is defined by low WC and the underlying clusters have a higher WC. The SA-cluster is found in most soundings and has a variable thickness from 2 m to 17 m. The transition at sounding location 8, is from US to Ti-cluster likely due to lower ρ in this area. A second deeper grey line tracks the transition below the SA-cluster to the underlying Ti cluster.

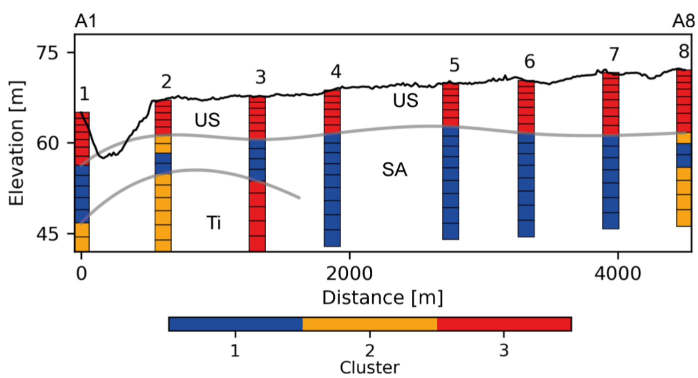

Figure 5 Clustering section from Kompedal where the partitioning of data is shown at every sounding. The grey lines track selected cluster boundaries. See Table 1 for cluster descriptions.

To evaluate possible variations within the boundaries estimated from the clustering, the profile shown in Figure 5 is reproduced in Figure 6 with ρ values and WC and T_2^*. Since clustering is

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

a discrete and often brutal partitioning of smoothly varying parameters, it is important to return to the original parameters for evaluation. The SNMR WC are shown as bars in Figure 6a and both T_2^*(left part of bar) and T_2(right part of bar) are shown side by side with the same color scale in Figure 6b and will be referred to as T_2^*/ T_2 profiles. The grey lines from the clustering are superimposed on this section to track variations within each cluster unit. Figure 6a displays the first section where shallow low WC and high ρ coinciding with the US-cluster where the T_2^*/ T_2 profiles show low values. Boreholes identify this unit as sand near sounding location 3 and 6, which match the interpreted geology as an unsaturated sand. The upper grey line is tracking an increase in WC from ~ 15 % to ~30 % in Figure 6a and from ~ below 0.1 s to above 0.15 s in T_2^* in Figure 6b, while there is no contrast in ρ. The lack of structure in the ρ indicates that this is likely a saturation change, as a lithological change would generally be expected to coincide with a ρ contrast. The elevated T_2^* is caused by less interaction with the grain surfaces because of increased saturation in the sand. Additionally, a borehole water table measurement coincides with this transition line at sounding location 6. The SA-cluster unit contains a range in WC from 20 % to 30 % and in T_2^* from 0.15 s to 0.3 s, indicating slight variations within the cluster. The SA-Ti transition coincides with a decrease in WC and ρ, interpreted as a similar reduction in pore size, a product of an increase in fines content. The T_2^* found in the Ti-cluster,

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

while quite varying, are generally lower than in the SA-cluster, consistent with the interpretation of increasing fines content at depth, as in a till.

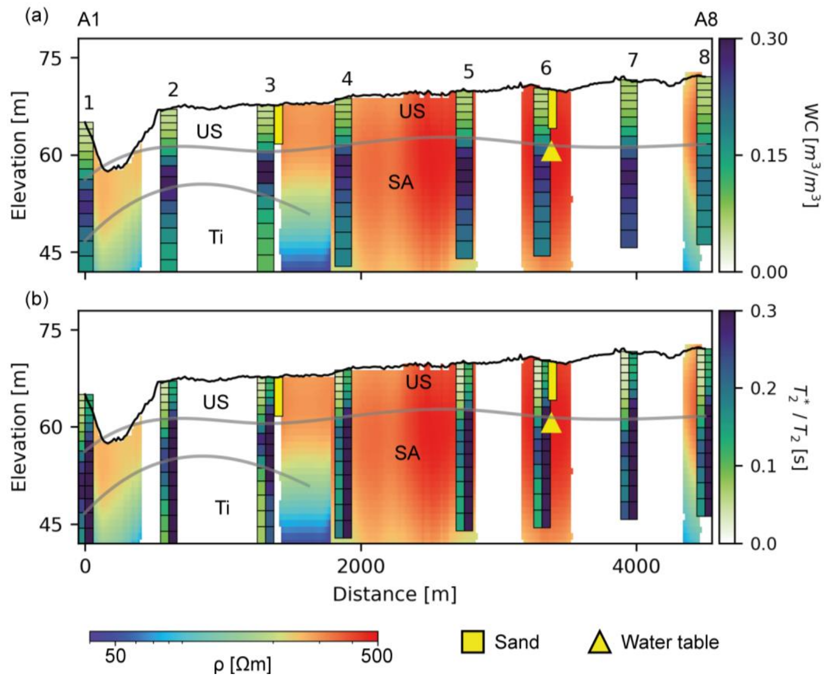

Figure 6 Profile of 8 SNMR soundings (bars) and TEM profile(background). Section 1 in Figure 2a. (a) SNMR WC, (b) a split bar with T_2^* (left) and T_2 (right). Boreholes at sounding location 3 and 6 are shifted ~40 m to avoid overlapping with the bars. Grey lines are tracking transitions between clusters in Figure 5.

To further evaluate the accuracy of the ability to track water tables by the upper cluster transition, consider Figure 7, where water tables from clustering are compared to available borehole-measured water tables within 500 m of SNMR sites. The clustering water tables are picked as the transition from the low WC US-cluster to any underlying cluster, SA or Ti, as both have a high WC compared to the US-cluster. The red line has a slope of 1 and the uncertainty bars are based on the inversion layer thickness, as the clustering method is ternary (i.e., it has three options) and consequently, some layers found at cluster transitions could be assigned to either cluster. We see that clustering tends to overestimate the water table elevation in many cases. This is a product of clustering being a brute thresholding in the parameter space. In this geology, the threshold from the clustering occurs at slightly lower WC than those coinciding with the water table and produces too shallow estimates. The trend, however, is similar to a slope of 1, indicating that a higher threshold could provide a better resolution of the regional water table. Additionally, the distance between borehole and coincident SNMR and TEM models could add uncertainty for the comparison, but this uncertainty is expected to follow the slope of 1. The two data points with yellow outline, far from the middle axis, stem

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

from the north of the area where the water table was measured in 1980, yielding some uncertainty due to possible long-term temporal or seasonal changes. Overall, the clustering captures the water table trend within an unconfined aquifer at a regional scale in an automated manner.

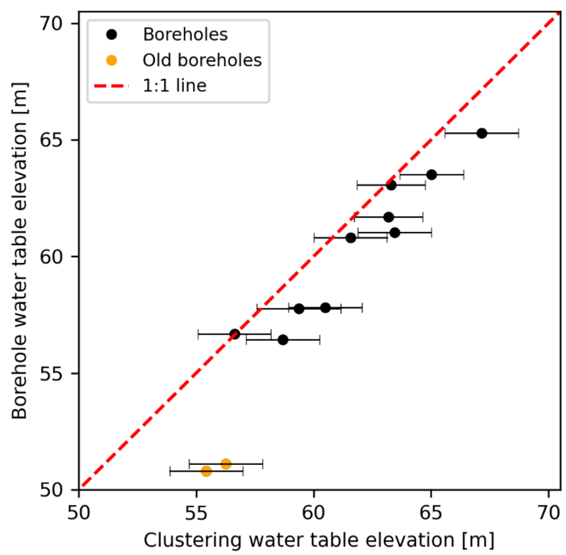

Figure 7 Borehole water table compared to the clustering water table at 12 SNMR locations with boreholes within 500 m. The red line has a slope of 1 and the error bars on the clustering estimates are based on the inversion layer thickness to provide a type of uncertainty. Yellow points have water table measurements over 40 years old.

3.2 Endelave case study

3.2.1 Clustering analysis

A similar methodology was used to examine the appropriate number of clusters for Endelave. Due to a more heterogeneous geology, four clusters were used to properly partition the data. Next, the partitioning of WC, T_2^* and ρ is inspected. Consider Figure 8, where the clustering results are shown, represented by the three clustered parameters (Figure 8a) or two (Figure 8b, c, d). First, the red cluster (1) is defined by quite low WC and T_2^* values (Figure 8b), while the ρ varies from 10 Ωm to 120 Ωm. This cluster exhibits properties consistent with till containing varying sand content and affecting ρ (Figure 8c). The green cluster (2) has mainly high ρ, high WC, and medium T_2^* values in Figure 8a. The high WC and ρ are properties associated with sand saturated aquifers. The yellow cluster (3) has similar SNMR attributes to the red cluster, with low WC and T_2^*, but has a lower ρ range illustrated in Figure 8d. This unit is interpreted to be of Paleogene clay due to the very low ρ found in this cluster. The range of WC found within the yellow cluster could indicate that layers with low to medium sand contents, but with low ρ are grouped here. The last cluster, blue (4) has a distinct T_2^* range in

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

Figure 8b and a large range of WC with ρ situated around 10 Ωm. The WC and T_2^* values would indicate that this layer has aquifer properties usually associated with sand, while the ρ indicates this as a conductive material. This is interpreted as saltwater saturated sand. In general, the clusters are not as distinct within the Endelave dataset, as the glacial interaction with the deposited sediment has caused a mixing of lithologies. This is evident from the ρ values where none exceed 130 Ωm, whereas the Kompedal survey consisted of ρ from 50 Ωm to 1000 Ωm. All the descriptions and interpreted geologies are found in Table 2.

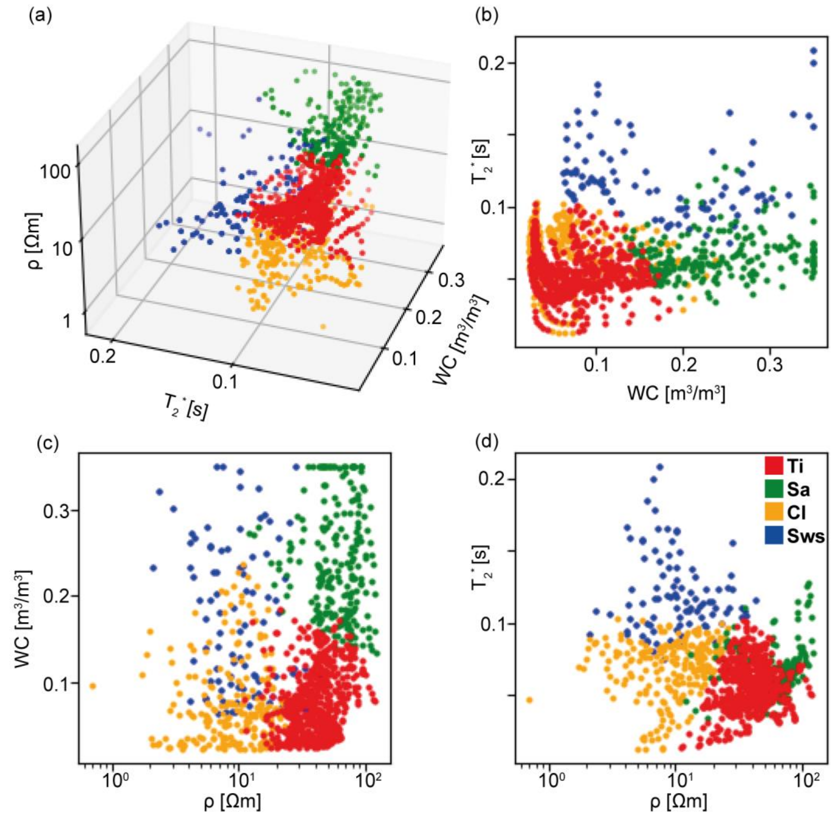

Figure 8 Clustering results from the Endelave survey on WC, T_2^* and ρ. (a) all three in a scatter, (b) WC and T_2^*, (c) ρ and WC, and (d) ρ and T_2^*. The color of each datapoint defines the assigned cluster.

Table 2: Cluster parameter bounds and interpreted geology for Kompedal.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

Cluster	WC [m^3/m^3]	T_2^* [s]	ρ [Ωm]	Interpreted geology	Label
1 (red)	Low-medium [0.03-0.18]	High [0.02-0.1]	High [10-120]	Till	**Ti** (Till)
2 (green)	High [0.15-0.40]	Medium [0.04-0.13]	High [15-120]	Sandy aquifer	**SA** (Sand aquifer)
3 (yellow)	Low-medium [0.03-0.18]	Low [0.02-0.1]	Low [1-25]	Paleogene clay	**Cl** (Clay)
4 (blue)	Low-High [0.07-0.40]	High [0.07-0.21]	Low [2-35]	Saltwater sands	**Sws** (Saltwater sand)

3.2.2 Spatial interpretation

Following the clustering we will examine their spatial extent on Endelave. We will show the results of two sections (Figure 2b). to see how the clustering performs in a more heterogenous setting. Consider first the section across the main shallow aquifer in Figure 9a, where we see a shallow Ti-unit corresponding to either a till or unsaturated sand. The SA-cluster unit has a thickness from 5 m to 12 m and is found below the Ti-cluster at sounding locations 3 to 6. Sounding location 1 is located 30 m from the coast, which aligns with the presence of the Sws-cluster. The Ti-cluster at depth is interpreted as a decrease in pore size from an increased clay or silt content. At sounding 7, all layers are grouped as the Ti-cluster, a sign of low SNMR parameters throughout the entire sounding location. The deepest discretized layers at most sounding locations are grouped in the Cl-cluster, tracked by a grey line, indicating a drop in ρ, as expected from the Paleogene clay.

To highlight possible saltwater intrusion, a section intersecting sounding locations at the coast is shown. The section in Figure 9b is quite complicated as it transects different geological regions. We consider three main points in this section, the Sws-cluster, the SA-cluster and the south end of the profile. In Figure 9b we see the Sws-cluster at sounding locations 1 to 3, defining a shallow and deep layer, while at sounding locations 6, 8 and 9 the cluster is seen shallowly at low elevations following the coast. The transition from the Sws-cluster to the underlying clusters is tracked by a grey line at sounding locations 1 to 3. Below the grey line at sounding locations 1 and 2, the layers are grouped with the Cl-cluster representing low ρ, lower T_2^* and WCs. It is important to note that even with combined SNMR and TEM, it will be hard to distinguish between saltwater and freshwater clays as both will be conductive and have a low free water content and T_2^* signatures in SNMR.

From sounding locations 3 to 8, the SA-cluster is found with a varying thickness from 2m to 10m. This unit, interpreted as the aquifer, is outlined in grey to compare with original parameter

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

values and borehole information later. At the south end of the profile, the clustering divides layers into Cl- and Ti-clusters, associated with clay and till by their low WC and T_2^*.

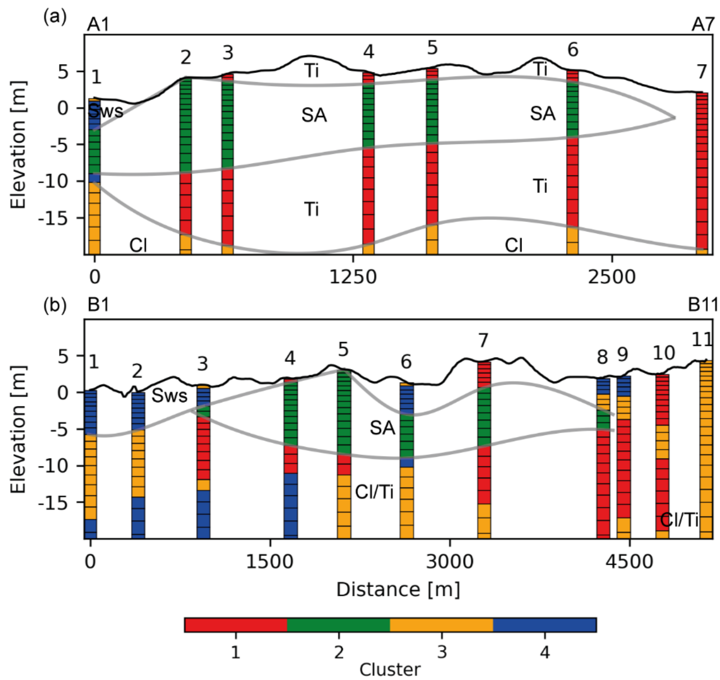

Figure 9 Clustering sections from Endelave where the partitioning of data is shown at every sounding location. (a) Section 1 (b) Section 2 in Figure 2a. The grey lines track selected cluster boundaries.

The discrete boundaries from the clusters are now used in the original parameter space to evaluate possible variations within the clusters and the estimated boundaries. In Figure 10, we consider the main shallow aquifer found on Endelave. The grey lines from Figure are used to delineate cluster extents and each unit is assigned a cluster label. Shallowly, the Ti-unit coincides with low WC and a high ρ in Figure 10a. T_2^*/T_2 are low in this unit and boreholes reveal either till or unsaturated sand here, matching the clustering interpretation. The upper Ti-SA grey line tracks an increase in WC at four sounding locations and coincides with a lithological change from clay to sand in two boreholes and coincides with a water table measurement in a separate borehole. This is interpreted as a semi-confined system with a water table at the layer boundary due to shifts in geological deposits. The SA unit here consists of high WC and low to medium T_2^* within a resistive unit. The boreholes identify this unit as sand or a mixture of sand and silt, which explains the range of WC grouped within this unit. The lower SA/TI transition tracks a decrease in WC, still with low to medium T_2^* seen in Figure 10b. The transition coincides with a decrease in ρ at sounding locations 1 and 2, and with a lithological boundary from sand to clay in a few boreholes. Furthermore, two boreholes

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

terminate exactly at this interface, which could indicate that the drillers hit something harder or more clay rich, prompting them to stop drilling. The Ti/Cl transition at depth tracks a decrease in ρ, which in the deep borehole is identified as a lithological boundary from clays and sand to Paleogene clay, agreeing with the clustering interpretations. This deep boundary is not seen in the SNMR-parameters as the Ti and Cl-clusters are only distinguishable by their ρ.

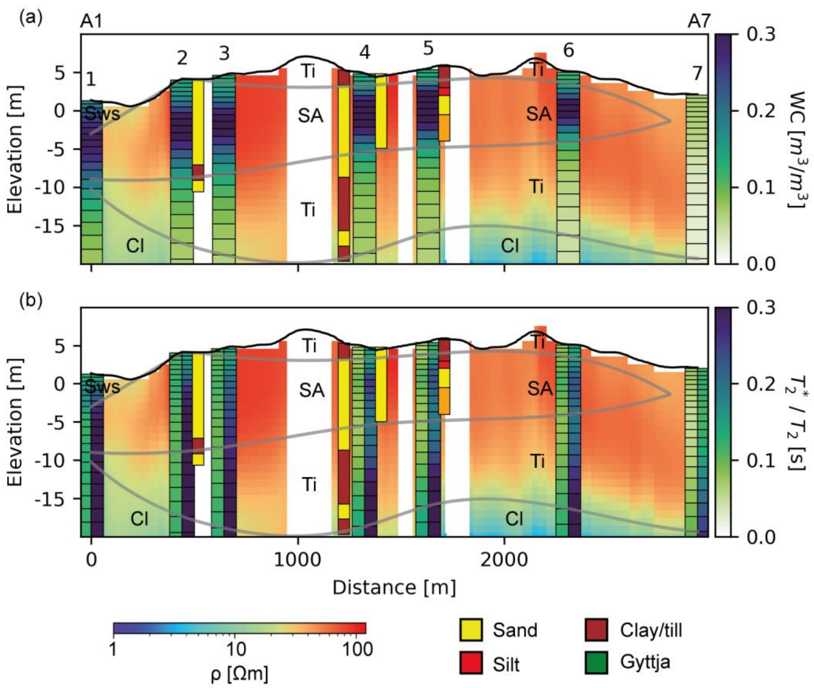

Figure 10 Profile of 7 SNMR soundings (bars) and TEM ρ (background). Section 1 in Figure 2b. (a) SNMR WC (b) a split bar with T_2^* (left) and T_2 (right). Grey lines are tracking transitions between clusters in Figure 9.

After reviewing the section through the main shallow aquifer in Figure 10, we will assess a second, more complex section. The grey lines from Figure will be used to delineate the cluster units and illustrate differences within the units. Consider now Figure 11, where the ρ and SNMR parameters are shown with lines following cluster transitions. The Sws unit is seen mainly at location 1 to 3 and is defined by high WC, very high T_2^* and low ρ. At sounding location 8, a borehole finds sand coinciding with the Sws-cluster in agreement with the saline sand interpretation. The high T_2^*/T_2 associated with the Sws-cluster in Figure 11b is a product of limited compaction within the newly deposited saline sand in the coastal environments. Below the Sws-unit, the grey line tracks a transition to lower WC and T_2^*, but maintaining the low ρ, which is defined by the Cl cluster. At sounding location 6, this transition is different with an increase in ρ tracking the border to the SA unit. Low WC at sounding locations 10 and 11

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

coincide with clay in a local borehole for the first 15 m where all layers are grouped within the Ti or Cl-clusters. The low water content and T_2^* signature at these locations prevent them from being clustered with the saline sands in Sws, highlighting the value of SNMR to distinguish these conductive units. The gyttja layer found in the borehole coincides with a drop in WC due to the increases in organic matter and was grouped with the Cl-cluster (Mashhadi et al., 2024).

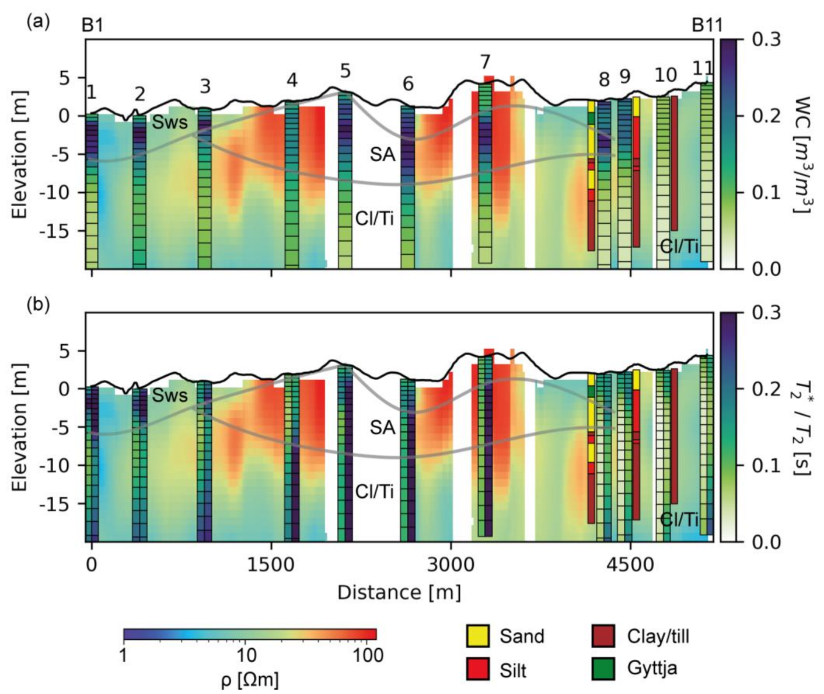

Figure 11 Profile of 11 SNMR soundings (bars) and TEM ρ (background). Section 2 in Figure 2b. (a) SNMR WC (b) a split bar with T_2^* (left) and T_2 (right). Grey lines are tracking transitions between clusters in Figure 9.

By clustering on this dataset, we have proven the ability to identify regions of possible saltwater intrusion. Figure 12 shows which sounding locations have layers that cluster within the saltwater aquifer, the freshwater aquifer, that have layers of both clusters, or only have the till and clay clusters. The saltwater cluster is observed mostly at the northern sounding locations where the post-glacial sands are located, but also along the east coast. The main aquifer unit, SA, is found in the east and north parts of the island, while the west part is dominated by the low water content clusters, shown in yellow and red. One sounding location with both saline and freshwater clusters far from the coast, is observed in the north of the island. The closest TEM sounding was acquired in a lowland south of this sounding, with elevation almost at sea level, which might have issues with saltwater intrusion. There is also a wetland close to this location, which might have a higher clay content with low ρ. If the TEM

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

and SNMR are not exactly coincident, some differences and anomalies in the clustering might occur. But in general, the K-means clustering is able to map this possible saltwater intrusion, which is a valuable asset in aquifer management.

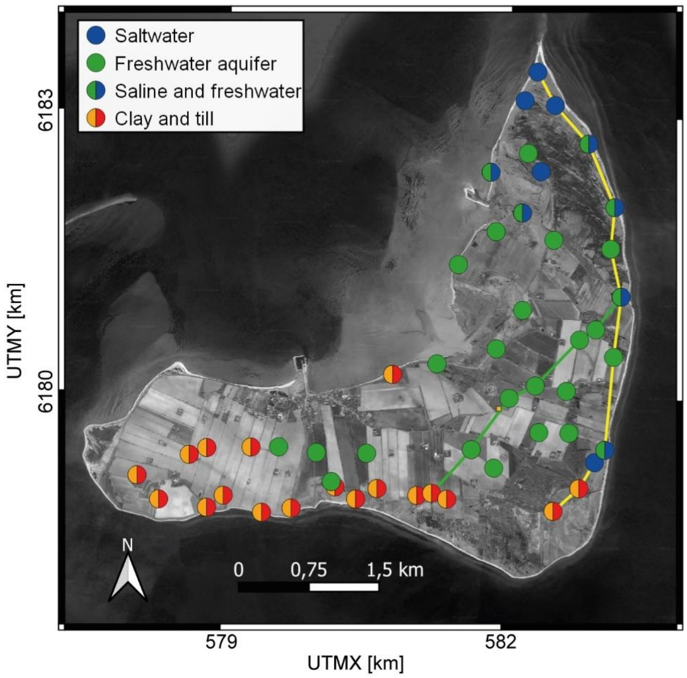

Figure 12 Sounding locations where salt water or fresh water has been identified. Locations with only clay and till clusters are shown with red and yellow. Map data: © Google Maps 2024, UTM Zone: 32N.

4 Discussion

In this study we investigated the use of clustering to combine the analysis of two geophysical methods, SNMR and TEM. Previously, the two methods were typically used together in either a joint-inversion or manual joint-interpretation approach. The limited availability of data has constrained the use of SNMR for clustering. Recent developments in SNMR have allowed for rapid acquisitions leading to much higher data densities (Grombacher et al., 2021) requiring a less subjective and fast interpretation scheme, such as clustering approaches. To track lithological boundaries using geophysics, it is necessary to have a contrast in layer parameters between the geological units. An example of this limitation in TEM occurs when saline sand lies above clay, as both have low ρ, creating little to no ρ contrast. In SNMR, a similar limitation occurs when distinguishing unsaturated sand and clay/till layers. SNMR can detect an increase in water content but cannot define whether the aquifer is unconfined or confined. Using these methods together can reduce ambiguities encountered when interpreting them separately, as they have complementary characteristics and different sensitivities. The TEM

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

ambiguity in the saline sands/clays example will be resolved by SNMR as a decrease in WC at the lithological boundary as SNMR is sensitive only to the quantity of water, not its salinity. In the SNMR unsaturated limit, TEM would resolve a resistive unit in the unconfined case and a conductive unit in the semi-confined or confined example. The ability to distinguish these layers and track them spatially between boreholes is why SNMR and TEM were chosen to find discrete boundaries using K-means clustering.

K-means clustering on geophysical models offers a simple, automated approach to identifying lithological transitions. It allows for reproduceable boundary definitions without subjective interpretations of the geophysical models. Discretizing smoothly varying parameters into predefined clusters is, however, brutal and there will be variability within the unit definitions. The ability to return to the original parameter space with cluster boundaries is crucial in addressing subtle variations within units and can be used to evaluate cluster transitions.

K-means clustering applied to geophysical models is not limited to SNMR and TEM parameters; it can also be extended to other collocated datasets with distinct sensitivities. For example, in areas where a deep water table is expected within a sand layer, seismic methods may be appropriate. However, because the seismic velocity of saturated sands can be similar to that of clays or tills, incorporating collocated TEM models can help reduce interpretational bias. Similarly, relying solely on TEM data may make it difficult to detect the water table due to limited resistivity contrast.

In SNMR, correlations between WC and T_2^* may exist. For example, in unsaturated sands, the low water content residing in the pores will be in close contact with the grain surfaces, resulting in interactions leading to low T_2^*/T_2. Since water content is proportional to signal amplitude, in low WC environments, low signal amplitude results in reduced confidence in the T_2^* estimates. When such parameters are linked, it might be of interest to simplify the approach by clustering on the product of water content and T_2^*. Thus, combining these two parameters may help reduce the influence of low-confidence T_2^* values in low water content environments. A similar option is to use a principal component analysis to reduce the basis to two parameters that describe the most variance, which in the Kompedal case would be resistivity and the product of the SNMR parameters. However, in more complex geologies such as Endelave, a decrease in basis dimension may reduce the ability to distinguish layers of high WC and low T_2^* from layers with low WC and high T_2^*. Through examining the data's variation and correlation, we can make informed decisions about whether to decrease the parameter space.

In this study, we focused on interpreting two survey areas using K-means clustering, which proved sufficient in meaningfully partitioning data and identifying lithological boundaries. One feature of the employed K-means clustering approach is the need to specify clusters

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

beforehand. In this study we based the choice of clusters on the silhouette index and prior geological information about the area. One alternative study uses agglomerative hierarchical clustering on SkyTEM data, which avoids selecting the number of clusters by starting with one cluster and subdividing until each data point has its own cluster (Dumont et al., 2018). This can alleviate some of the choices made for the silhouette index analyses and provide a better understanding of how clusters are further subdivided. A second challenge is to attribute uncertainties to the layer boundaries picked by the discrete K-means clustering. Here, others use fuzzy C-means where data points are assigned a membership score and can be partial members of more than one cluster (Paasche et al., 2007). Applying the fuzzy C-means can give an estimate of uncertainty for the picked cluster boundaries, i.e., if a data point could be a member of several clusters, it is less certain. This could apply to the Endelave data where the saline intrusion cluster in places could overlap with the freshwater cluster.

Since clustering is performed on coincident values, we are limited by the lowest dimension dataset, which in this case is the SNMR, e.g. on Endelave the survey consists of 51 soundings while there are over 23000 TEM soundings in the same area. This reduction in data space disregards large amounts of TEM data, which of course have valuable regional information, but lack coincident SNMR parameters. Additionally, lower data quantity can lead to clusters not representable for the area. If SNMR information could be extrapolated to the full TEM domain through appropriate spatially variable measures, it would allow for clustering on a much larger data set. Future research will focus on extrapolating SNMR parameters across the full TEM domain. This would enable a subdivision of the full TEM domain based on the coincident data clusters, and it will be possible better to delineate areas of potential saline intrusion spatially.

5 Conclusion

Through two field studies we demonstrated the automated spatial identification and separation of hydrogeological units in large scale geophysical campaigns. Recent improvements in the data acquisition rates of SNMR now offer data volumes sufficient to exploit clustering approaches when combining these data with other geophysical data. K-means clustering of complementary SNMR and TEM models is shown to provide a non-subjective approach, where enhanced hydrogeological interpretations can be formed by exploiting the complementary nature of two data types. To detect lithological boundaries, they must correspond to a contrast in geophysical properties. SNMR is shown to provide value when discriminating clay-rich sediments from saline saturated sand conditions, a challenging task based on only TEM models. Similarly, TEM is able to separate low-water content conditions from clay-rich conditions, which is impossible with SNMR alone. This is key to discriminating

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

between unconfined and semi-confined conditions. A silhouette index-based approach, combined with the a priori knowledge of the likely number of lithological units present, is shown to be a robust measure for selecting the number of clusters.

In the examples, clustering of NMR and TEM data provides a more complete characterization of local hydrogeological conditions than what can be achieved by each data set separately.

Data availability

The data shown in this study are available upon request from the corresponding author.

Author contributions

MV wrote the manuscript, gathered data and did the analysis. JJL helped with writing the manuscript. AVC helped with the analysis and correcting the writing. DG assisted in figure development, analysis and writing the manuscript.

Competing interest

The corresponding author declares that none of the authors have any competing interests.

References

Auken, E., Foged, N., Larsen, J.J., Lassen, K.V.T., Maurya, P.K., Dath, S.M. and Eiskjær, T.T.: tTEM—A towed transient electromagnetic system for detailed 3D imaging of the top 70 m of the subsurface. *Geophysics*, *84*(1), pp.E13-E22, https://doi.org/10.1190/geo2018-0355.1, 2019.

Auken, E., Christiansen, A.V., Kirkegaard, C., Fiandaca, G., Schamper, C., Behroozmand, A.A., Binley, A., Nielsen, E., Effersø, F., Christensen, N.B. and Sørensen, K.: An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data. *Exploration Geophysics*, *46*(3), pp.223-235, https://doi.org/10.1071/EG13097, 2015.

Behroozmand, A.A., Keating, K. and Auken, E.: A review of the principles and applications of the NMR technique for near-surface characterization. *Surveys in geophysics*, *36*, pp.27-85, https://doi.org/10.1007/s10712-014-9304-0, 2015.

Binley, A., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., Singha, K. and Slater, L.D.: The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. *Water resources research*, *51*(6), pp.3837-3866, https://doi.org/10.1002/2015WR017016, 2015.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

Christiansen, A.V., Auken, E. and Sørensen, K.: The transient electromagnetic method. In *Groundwater geophysics: a tool for hydrogeology* (pp. 179-225). Berlin, Heidelberg: Springer Berlin Heidelberg, https://doi.org/10.1007/3-540-29387-6_6, 2006.

Costabel, S., Siemon, B., Houben, G. and Günther, T.: Geophysical investigation of a freshwater lens on the island of Langeoog, Germany–Insights from combined HEM, TEM and MRS data. *Journal of Applied Geophysics*, *136*, pp.231-245, https://doi.org/10.1016/j.jappgeo.2016.11.007, 2017.

Dickinson, J.E., Pool, D.R., Groom, R.W. and Davis, L.J.: Inference of lithologic distributions in an alluvial aquifer using airborne transient electromagnetic surveys. *Geophysics*, *75*(4), pp.WA149-WA161, https://doi.org/10.1190/1.3464325, 2010.

Dragoni, W. and Sukhija, B.S.: *Climate change and groundwater: a short review* (Vol. 288, No. 1, pp. 1-12). London: The Geological Society of London, https://doi.org/10.1144/SP288.1, 2008.

Dumont, M., Reninger, P.A., Pryet, A., Martelet, G., Aunay, B. and Join, J.L.: Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies. *Journal of Applied Geophysics*, *157*, pp.1-9, https://doi.org/10.1016/j.jappgeo.2018.06.020, 2018.

Foged, N., Marker, P.A., Christansen, A.V., Bauer-Gottwein, P., Jørgensen, F., Høyer, A.S. and Auken, E.: Large-scale 3-D modeling by integration of resistivity models and borehole data through inversion. *Hydrology and Earth System Sciences*, *18*(11), pp.4349-4362, https://doi.org/10.5194/hess-18-4349-2014, 2014.

Grombacher, D., Liu, L., Griffiths, M.P., Vang, M.Ø. and Larsen, J.J.: Steady-State Surface NMR for Mapping of Groundwater. *Geophysical Research Letters*, *48*(23), p.e2021GL095381, https://doi.org/10.1029/2021GL095381, 2021.

Hertrich, M., Braun, M., Gunther, T., Green, A.G. and Yaramanci, U.: Surface nuclear magnetic resonance tomography. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(11), pp.3752-3759, https://doi.org/10.1109/TGRS.2007.903829, 2007.

Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, *24*(7), pp.881-892, https://doi.org/10.1109/TPAMI.2002.1017616, 2002.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

Kodinariya, T.M. and Makwana, P.R.: Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, *1*(6), pp.90-95, 2013.

Larsen, J.J., Liu, L., Grombacher, D., Osterman, G. and Auken, E., 2020. Apsu—A new compact surface nuclear magnetic resonance system for groundwater investigation. *Geophysics*, *85*(2), pp.JM1-JM11.

Le, C.V.A., Nguyen, N.N.K. and Nguyen, T.V.: Application of Clustering Method in Different Geophysical Parameters for Researching Subsurface Environment. Inżynieria Mineralna, https://doi.org/10.29227/IM-2022-02-05, 2022.

Legchenko, A., Baltassat, J.M., Beauce, A. and Bernard, J.: Nuclear magnetic resonance as a geophysical tool for hydrogeologists. *Journal of Applied Geophysics*, *50*(1-2), pp.21-46, https://doi.org/10.1016/S0926-9851(02)00128-3, 2002.

Mashhadi, S.R., Grombacher, D., Zak, D., Lærke, P.E., Andersen, H.E., Hoffmann, C.C. and Petersen, R.J.: Borehole nuclear magnetic resonance as a promising 3D mapping tool in peatland studies. *Geoderma*, *443*, p.116814, https://doi.org/10.1016/j.geoderma.2024.116814, 2024.

Mohnke, O. and Yaramanci, U.: Forward modeling and inversion of MRS relaxation signals using multi-exponential decomposition. *Near Surface Geophysics*, *3*(3), pp.165-185, https://doi.org/10.3997/1873-0604.2005012, 2005.

Nabighian, M.N. and Macnae, J.C.: Time domain electromagnetic prospecting methods, https://doi.org/10.1190/1.9781560802686.ch6, 1991.

Paasche, H. and Tronicke, J.: Cooperative inversion of 2D geophysical data sets: A zonal approach based on fuzzy c-means cluster analysis. *Geophysics*, *72*(3), pp.A35-A39, https://doi.org/10.1190/1.2670341, 2007.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J.: Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, pp.2825-2830, 2011.

Shutaywi, M. and Kachouie, N.N.: Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, *23*(6), p.759, https://doi.org/10.3390/e23060759, 2021.

https://doi.org/10.5194/egusphere-2025-406
Preprint. Discussion started: 30 April 2025

Song, Y.C., Meng, H.D., O'Grady, M.J. and O'Hare, G.M.: The application of cluster analysis in geophysical data interpretation. *Computational Geosciences*, *14*, pp.263-271, https://doi.org/10.1007/s10596-009-9150-1, 2010.

Sørensen, K.I. and Auken, E.: SkyTEM-A new high-resolution helicopter transient electromagnetic system. *Exploration Geophysics*, *35*(3), pp.191-199, 2004.

Sun, J. and Li, Y.: Joint inversion of multiple geophysical and petrophysical data using generalized fuzzy clustering algorithms. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, *208*(2), pp.1201-1216, https://doi.org/10.1093/gji/ggw442, 2016.

van Roosmalen, L., Christensen, B.S. and Sonnenborg, T.O.: Regional differences in climate change impacts on groundwater and stream discharge in Denmark. *Vadose Zone Journal*, *6*(3), pp.554-571, https://doi.org/10.2136/vzj2006.0093, 2007.

Vang, M., Grombacher, D., Griffiths, M.P., Liu, L. and Larsen, J.J.: High-density mapping of regional groundwater tables with steady-state surface nuclear magnetic resonance–three Danish case studies. *Hydrology and Earth System Sciences*, *27*(16), pp.3115-3124, https://doi.org/10.5194/hess-27-3115-2023, 2023.

Vang, M., Grombacher, D., Larsen, J.J. and Wison, S.: Efficient mapping of complex groundwater systems associated with braided rivers using small coil surface nuclear magnetic resonance. *Geophysics, **90(3)***, pp. 1-45, https://doi.org/10.1190/geo2023-0765.1, 2025.

Yaramanci, U., Lange, G. and Knödel, K.: Surface NMR within a geophysical study of an aquifer at Haldensleben (Germany). *Geophysical prospecting*, *47*(6), pp.923-943, https://doi.org/10.1046/j.1365-2478.1999.00161.x, 1999.