Journal of Applied Geophysics 184 (2021) 104203

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Time-series clustering approaches for subsurface zonation and hydrofacies detection using a real time-lapse electrical resistivity dataset

Damien Delforge [a,b,*], Arnaud Watlet [c], Olivier Kaufmann [d], Michel Van Camp [b], Marnik Vanclooster [a]

[a] Earth and Life Institute, Université catholique de Louvain, Belgium
[b] Royal Observatory of Belgium, Belgium
[c] British Geological Survey, United Kingdom
[d] University of Mons, Belgium

ARTICLE INFO

Article history:
Received 25 October 2019
Received in revised form 30 April 2020
Accepted 22 October 2020
Available online 01 November 2020

ABSTRACT

One main application of electrical resistivity tomography (ERT) is the non-invasive detection of geological or hydrological structures in the shallow subsurface. This paper investigates the capability of time-series clustering to retrieve such features on real time-lapse ERT datasets considering three aspects: (1) the comparison between three clustering algorithms k-means, hierarchical agglomerative clustering (HAC), and Gaussian Mixture Model (GMM), including the question of the optimal choice of cluster number and the identification of resistivity series whose classification is uncertain, (2) the effect of adding a spatial constraint in clustering, and (3) the robustness of the approaches to various representations of resistivity values and the number of time-steps involved in the clustering. The real time-lapse ERT dataset is obtained from dipole-dipole arrays on a 48 electrodes profile installed on the top of the Rochefort cave in Belgium. It consists of resistivity time-series defined over 465 days and associated with 1558 cells of the 2D ERT models derived from a time-lapse inversion. The clustering results are appreciated using clustering validation indices and further confronted with the expert-based structural model of the site. Results show that the three clustering algorithms provide similar spatial patterns on the standardized data and reveal correlated resistivity time-series. Some clusters are, however, spatially split and regroup time-series with a wide range of mean resistivity, suggesting different geological units within these groups. Clustering on the raw resistivity time-series may also appear inconsistent as the averaged resistivity series per cluster are highly correlated, thus missing the hydrological and functional traits of the subsurface elements. Applying a spatial constraint to the clustering of standardized data increases the number of clusters in order to retrieve spatially tied clusters. The grouped series are more homogeneous in terms of mean resistivity due to their spatial proximity, but some inconsistencies may remain due to synchronous hydrological forcing. Applying the clustering to various time-series representations allows us to gain confidence about the redundant spatial patterns. However, the patterns obtained from the clustering of the full standardized dataset cannot be reproduced from continuous sub-samples up to 100 days, but well from less than 20 samples picked randomly over the 465 days. Accordingly, our study highlights the importance of time-variable parameters in the identification of structural facies and hydrofacies with ERT while demonstrating the strength of long-term monitoring.

1. Introduction

The electrical resistivity of surface soil varies with the mineralogical composition of soil and rocks, temperature, the water content, and its solute composition. Electrical resistivity tomography (ERT) is a technique commonly used in geosciences that aims to capture these variations. ERT relies on electrodes, a current injection scheme, and the inversion of an associated resistivity model to map the resistivity of the shallow subsurface, either in two or three dimensions, and derive geological and hydrological interpretations (Banton et al., 1997; Samouëlian et al., 2005). Time-lapse ERT extends to an additional dimension by using repeated current injections over time, allowing the retrieval of temporal variation of resistivities. For their capabilities of generating a large amount of spatialized data at low cost, time-lapse ERT has been increasingly used in the near-surface geophysics community to investigate subsurface geology or hydrogeological processes (Barker and Moore, 1998; Kuras et al., 2009; Singha et al., 2014).

The visual or computer-assisted interpretation of inverted resistivity models remains challenging as the inversion procedure often relies on smoothness constraints, producing fuzzy patterns rather than a clear representation of subsoil heterogeneities (e.g., Günther et al., 2006; Loke and Barker, 1996). Besides, the resolution decreases, or the

* Corresponding author at: Earth and Life Institute, Université catholique de Louvain, Belgium.
E-mail address: damien.delforge@uclouvain.be (D. Delforge).

https://doi.org/10.1016/j.jappgeo.2020.104203
0926-9851/

uncertainties of the inversion image increase, as a function of the distance to the electrodes (e.g., Hermans and Irving, 2017). Accordingly, the improvement of the ERT models is approached from different angles. The first one focuses on the inversion procedure itself, for instance, by considering adaptable constraints to produce sharper results (Fiandaca et al., 2015; Nguyen et al., 2016). A second option, detailed in the next paragraph, is to apply post-inversion processing to enhance the interpretability of the model outputs. Other refinements may come from crossing strategies and datasets: the joint inversion of multivariate geophysical data (Doetsch et al., 2010; Infante et al., 2010) or the definition of an ensemble model either from a distribution of inversion parameters (Audebert et al., 2014) or from multiple electrode configurations (Ishola et al., 2014). Similarly to Paasche and Tronicke (2007), post-inversion approaches can be coupled with the inversion strategies by an iterative procedure (Doetsch et al., 2010; Elwaseif and Slater, 2012; Infante et al., 2010; Singh et al., 2017; Zhou et al., 2014) and be part in a fully automated way of an integrated ERT monitoring and modeling environment (e.g., Wilkinson et al., 2019).

In particular, post-inversion approaches can be defined in a mutually non-exclusive way, according to different aspects. Some papers target the detection and zonation of static features such as geological boundaries or structures (Caterina et al., 2013; Chambers et al., 2012, 2013; de Pasquale et al., 2019; Doetsch et al., 2010; Hsu and Yanites, 2010; Kutbay and Hardalaç, 2017; Xu et al., 2017), defects in covered landfill (Genelle et al., 2012), buried archeological objects or cavities (Elwaseif and Slater, 2010, 2012). Feature detection can also be improved on multivariate models or datasets (Giuseppe et al., 2014, 2018; see also Paasche et al., 2006). Other applications cover dynamic processes: mapping of the water or leachate infiltration front (Audebert et al., 2014; Scaini et al., 2017), the tracking of tracer's motion (Ward et al., 2016), or groundwater level monitoring (Chambers et al., 2015).

In any of these applications, the underlying algorithms can be summarized into the three following types: (1) gradient edge detection, (2) object segmentation into two groups (binarization), or more through (3) unsupervised classification, i.e., clustering. Clustering algorithms have some merits compared to other techniques. Ward et al. (2014) mentioned that gradient edge detection methods are limited since the steepest gradients are not always concurrent with geological interfaces, especially given the smoothness-constrained inversion and the lack of resolution at depth in ERT images. Also, gradient edge detection is applied in contexts where the substrate is typically organized in successive horizontal layers. As such, the applicability of this algorithm is challenged for anisotropic heterogeneous environments such as karst systems. Compared to segmentation algorithms that divide models into two subgroups, clustering algorithms have the advantage of not limiting the number of groups that can be defined according to their distinct resistive behavior, which on the other hand, raises the problem of the optimal choice of the number of clusters.

Overall, a few distinct clustering algorithms have been applied: fuzzy c-means (Chambers et al., 2015; Kutbay and Hardalaç, 2017; Paasche et al., 2006; Paasche and Tronicke, 2007; Singh et al., 2017, Ward et al., 2014;), k-means (Audebert et al., 2014; Giuseppe et al., 2014, 2018; Ishola et al., 2014; Scaini et al., 2017), Gaussian Mixture Models, GMM (Doetsch et al., 2010), and Hierarchical Agglomerative Clustering, HAC (Genelle et al., 2012; Xu et al., 2017). The fuzzy c-means, k-means, and GMM algorithms belong to the family of iterative relocation clustering algorithms. The fuzzy c-means and GMM are similar in the sense that they yield to probabilistic clusters, also termed soft or fuzzy clusters, meaning that an item may be assigned to several clusters with a given probability. On the contrary, k-means is a hard or crisp clustering algorithm according to which each item is assigned to a single group. HAC is another hard clustering algorithm based on a nested structure represented by a dendrogram.

Notwithstanding the availability of these advanced clustering techniques, methodological issues remain when applying clustering to real-world ERT datasets. Fuzzy c-means, for instance, was applied as a fuzzy algorithm to deal with the uncertainties brought by the smoothness of non-time-lapse ERT models (Ward et al., 2014). It is, however, not clear how such algorithms would work with time-lapse datasets. Further, Genelle et al. (2012) and Xu et al. (2017) used the HAC method to cluster for the first time ERT time-series of a time-lapse 2D dataset made of respectively 6 and 20 time-steps. Nevertheless, their study did not allow addressing critical issues such as the impact of alternative clustering algorithms on clustering results, the selection of the optimal number of clusters, or the evaluation of either the robustness or the uncertainties in clustering results. Also, Ward et al. (2014) suggested considering the local neighborhood and spatial constraints in clustering processes, an issue, which still needs to be further analyzed.

This paper focuses on the post-inversion clustering of ERT time-lapse datasets to extract and delineate spatially homogeneous features based on their resistivity patterns and aims at addressing the abovementioned concerns. We will refer to the term hydrofacies to denote spatial zones of similar patterns in their mean inverted resistivity, standard deviation, and correlation, assuming that they encompass common lithology and synchronous hydrological response at a daily time resolution. In particular, we discuss (1) the comparison and parametrization of three candidate clustering algorithms (k-means, GMM, and HAC) while addressing the question of the optimal number of clusters and the evaluation of the clustering results, and (2) the pertinence of including spatially explicit information in the clustering task. Finally, we discuss (3) the robustness of the clustering outputs to various representations of the resistivity data, whether or not log-scaled, normalized, differenced, decomposed, as well as the impact of the time span of the ERT model on the clustering outputs. Our analysis is based on a 2D-ERT dataset collected over a 465-days time-domain (Watlet et al., 2018a, 2018b). The long time span is particularly suitable to allow an exploratory analysis and to answer the aforementioned research questions. The geological interpretation of the study site and the ERT model are particularly detailed in these open-access references. In this issue, we focus mainly on introducing and investigating the characteristics and capabilities of the clustering methods for the zonation of time-lapse ERT models. To further encourage reproducibility and reusability, programming aspects exclusively relies on Scikit-learn (Pedregosa et al., 2011), an open-source Python package for machine learning.

2. Theoretical background

2.1. Time series clustering (TSC)

Clustering consists of grouping high dimensional data into fewer classes based on groups' inner similarities and groups' outer dissimilarities. In particular, time-series clustering (TSC) aims at grouping individual time-series together (Liao, 2005). TSC can be challenging due to the high dimensionality of time datasets: (M,N) where M is the number of time-series (samples) and N the number of time steps (features). Averaging clusters reduces the dimensionality to (k,N) where k is the final number of clusters. A clustering algorithm defines clusters and their members based on criteria involving distance or similarity measures. In machine learning, clustering is also defined as unsupervised classification since there are no predefined labeled groups that could serve as a basis for training.

Due to the combined effect of data structure and dimensionality, the diversity of fields of application, the different clustering purposes, and the nature of hunted patterns, a wide variety of TSC approaches are found in the literature (Liao, 2005; Aghabozorgi et al., 2015). These are, for the most part, declined under three aspects:

1. A time-series representation, which denotes any transformation of the time-series reducing the dimension of the dataset before the clustering;

2. A clustering algorithm relying on a distance measure;
3. An evaluation technique.

A prior reduction of the dataset dimensionality has several advantages: diminution of the memory consumption and speed-up of the clustering algorithm, noise reduction, and the harmonization of time-series data of unequal length or resolution into a dataset where an equal number of features characterize each sample time-series. If no prior reduction is applied, we refer to the raw-data-based approach (Liao, 2005). As far as TSC is concerned in this study, ERT series are univariate, real-valued, uniformly sampled, and smoothed (due to inversion smoothing constraint), of equal length, and relatively short. For these reasons, we mainly considered a raw-data-based approach until Section 4.3, where dimension reduction is tested. Still, raw TSC usually involves a scale transformation. The z-standardization is used in the vast majority of cases and applied to each of the M samples, i.e., the individual time-series X_i:

$$X_{z,i} = \frac{X_i - \mu(X_i)}{\sigma(X_i)}, i \in [1, M] \quad (1)$$

where $\mu(X_i)$ and $\sigma(X_i)$ stand for the mean and standard deviation estimates for X_i.

In Section 4.3, we refer to four time-series representations, each of them either applied to the resistivity or the log-resistivity, that are: the raw data without transformation, the z-standardized data (Eq. (1)), the differenced data ($X_i(t) - X_i(t-1)$) followed by z-standardization (Eq. (1)), and the decomposed data using principal component analysis (PCA) on the z-standardized data (Eq. (1)). Differencing removes the seasonal variation of the mean resistivity and will most likely result in a clustering that is more sensitive to the synchronous response of daily variations. By decomposing the covariance matrix of the dataset, PCA reduces its dimension from N to several orthogonal components that explain most of the variance of the dataset (see Section 3.2). Since PCA is applied to the z-standardized data, the covariance matrix is equivalent to the correlation matrix, and the PCA reveals correlation patterns in the time-series across space.

2.2. Clustering algorithms

There are no strict restrictions on the use of conventional clustering algorithms for the specific case of TSC. However, it is common to have distance functions modified according to the purpose of clustering. Two cases arise depending on whether the aim is to group synchronous and linearly correlated series (similarity in time), or whether the procedure must rely on elastic measures of distance tolerant to some distortions or asynchronies (similarity in shape). This paper focuses on the first case, i.e., the similarity in time compliant with our definition of hydrofacies, and that is usually addressed using Euclidean distances, squared Euclidean distances, or correlation-based distance. On a z-standardized dataset (Eq. (1)), the correlation coefficient $R_{X_iX_j}$ between two time-series is related to their squared Euclidean distance ${d_{X_{z,i}X_{z,j}}}^2$ such that $R_{X_iX_j} = 1 - {d_{X_{z,i}X_{z,j}}}^2/2N$. Despite the introduction of new distance metrics, Euclidean-based distances remain the simplest and one of the most competitive options (Keogh and Kasetty, 2003).

There is an extensive and non-exclusive taxonomy dedicated to the description of clustering algorithms (Tan et al., 2019). An important distinction is based on the clustering structure. If the algorithm produces an independent partition of k clusters, it is called a partitional algorithm. On the other hand, if clustering produces a tangled structure of groups and subgroups, it is referred to as hierarchical, although it is possible to retrieve a partition of k clusters based on a cut-off distance. Another dichotomy is based on the hard (or crisp) or probabilistic (or soft, fuzzy) nature of the partition. Hard clustering labels each object i to one unique cluster, while probabilistic clustering defines a probability of membership. Another type of clustering algorithms is prototype-based or center-based clustering. These algorithms partition objects based on their distance from the centroid of the cluster, and, for these reasons, tend to produce convex clusters centered on the mean. In the 2D example of Fig. 1, clusters A and C are convex since they could be averaged to a characteristic element, the centroid that belongs to the cluster. Cluster A has a spherical covariance matrix, while C has an anisotropic covariance matrix. On the reverse, cluster B is concave, and the centroid is no longer a reliable prototype. In general, concave clusters are extracted using methods that consider the local neighborhoods and densities around each sample (e.g., Section 2.2.2). It allows extracting dense clusters regardless of their structural arrangement. However, these approaches are challenged in the case of an ERT model given smoothness constraints and the subsequent lack of sharp variations in resistivity.

This study relies on three prototype-based clustering algorithms so that the resistivity series can be averaged into a mean representative series per cluster. These are k-means, hierarchical agglomerative clustering (HAC), and Gaussian Mixture Models (GMM). The k-means algorithm is partitional, hard, and tends to produce convex spherical clusters. HAC is hierarchical and hard. The covariance structure of the clusters depends on the distance metric and the constraints applied to the agglomeration. Finally, GMM is partitional, probabilistic, and not tight to a spherical covariance. The Python Scikit-learn library (Pedregosa et al., 2011) provides all the clustering algorithm used in this study.

2.2.1. k-means

The k-means algorithm is the most used clustering method (Berkhin, 2006). It is a partitioning relocation clustering algorithm based on the principle of finding a partition C of k clusters by minimizing the sum of squared Euclidean distances between each object i belonging to a cluster $c \in C$ with respect to the cluster centroid μ_c. The objective function to minimize is then:

$$\phi = \sum_{i \in M} \min_{c \in C} \|i - \mu_c\|^2 \quad (2)$$

The original k-means algorithm is referred to as Lloyd's algorithm and consists of a simple series of repeated steps:

1. k clusters centers $\mu_c$are randomly sampled given a uniform probability;
2. Each object i is assigned to the cluster closest center;

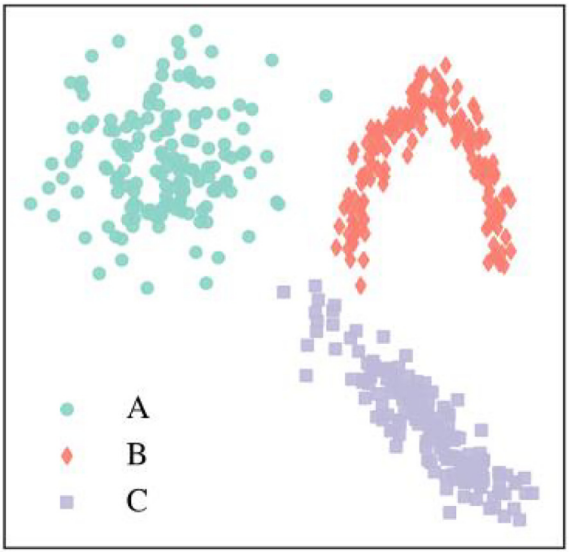

Fig. 1. Example of three different cluster distributions in two dimensions. *Cluster A is convex and spherical. Cluster B is concave. Cluster C is convex and anisotropic.*

3. ϕ is computed with respect to μ_c;
4. A new μ_c is obtained by averaging cluster members.

The steps 2 to 4 are repeated until ϕ is stable. The k-means algorithm tends to produce convex clusters of equal variances across the feature space, i.e., spherical clusters as cluster A in Fig. 1. The cluster prototype is the centroid μ_c. Due to the random initialization of cluster centers (step 1), a few repetitions of the full process (steps 1 to 4) are usually required to avoid convergence to suboptimal results. The best clustering partition, i.e., with minimal ϕ, is kept. Depending on the dataset, k-means may remain unstable and yields non-deterministic outputs. This study relies on Scikit-learn's implementation of the k-means++ algorithm (Arthur and Vassilvitskii, 2007). The k-means++ implementation improves the speed and accuracy of the original k-means by modifying the randomized initialization scheme (step 1). The idea is to spread initial centers allocation. The first center is still sampled given uniform probability distribution, while the subsequent centers are sampled given probability densities inversely proportional to the distance to previously defined centers.

2.2.2. Hierarchical agglomerative clustering (HAC)

Hierarchical Agglomerative Clustering (HAC) differs from k-means as it provides a nested structure of the clustering through a dendrogram. HAC uses a bottom-up approach: it starts from individual samples i as leaves and merges them into branches based on their proximity until one cluster remains. Clusters are progressively merged based on their relative proximity. The proximity is defined by linkage methods defining the distance between clusters. This study focuses on the Ward linkage method (Ward, 1963) that minimizes the sum of squared differences within all clusters. Hence, the output is quite similar to that of k-means and tends to produce convex clusters of equal covariances as well (Fig. 1, cluster A), which could be averaged into a cluster prototype, i.e., the centroid. Unlike k-means, which relies on random initializations of cluster centers, HAC's outputs are stable and do not require several iterations.

A particularity of the Scikit-learn's implementation lies in the opportunity of constraining the merging of branches by providing a connectivity matrix (Abraham et al., 2014). Such a matrix is binary of square shape (M,M), where M is the number of samples and distinguishes connected objects from disconnected objects so that two branches can be merged only if spatially connected objects exist between them. This functionality could be used to retrieve non-spherical clusters in the N-dimensional feature space such as Cluster B and C in Fig. 1. In that case, the connectivity matrix is computed using a nearest-neighbor approach. In this case, the connectivity matrix is computed from the mesh of the ERT model so that two cells are connected if they share an edge. This capability is used in Section 4.2 as a spatial constraint in order to retrieve spatially homogeneous clusters.

2.2.3. Gaussian mixture model (GMM)

Gaussian Mixture Models (GMMs) aim at modeling a dataset as a linear mixture of k Gaussian distributions defined in the N-dimensional feature space (Berkhin, 2006). Multivariate Gaussian models are related to the Mahalanobis distance that evaluates the distance of samples to a given distribution (Gallego et al., 2013). As a probabilistic algorithm, the clustering is soft so that each object i has a probability of belonging to each cluster. For a given object, these probabilities sum up to one. GMM requires as input the number of clusters k and relies on the expectation-maximization algorithm (Dempster et al., 1977) to find an optimal clustering. Expectation-maximization is closely related to the k-means algorithm as it involves iterative relocations: the starting point is a random initialization of k Gaussian distributions that are iteratively reallocated by updating the GMM parameters, i.e., the mixture k weights, the k mean vector of dimension N, and the k NxN covariance matrix. Doing so, GMM maximizes the overall likelihood L that each object belongs to the Gaussian mixture.

By default, the Scikit-learn implementation of GMM uses the same initialization strategy as k-means++ (Section 2.2.1) and automatically assign each sample to the most likely group. GMM is non-deterministic and different realizations may give different outcomes due to the random initialization. Different types of covariance matrix exist. Choosing a spherical type will add the constraint that the variance in each of the N dimensions should be approximately equal. As a result, GMM would yield probabilistic convex spherical clusters, similarly to what would be expected from the k-means and Ward-HAC methods. Here, we did not add constraints on the covariance so that each cluster may have its specific covariance matrix, and GMM may retrieve anisotropic convex clusters such as Cluster C in Fig. 1.

2.3. Clustering evaluation

The evaluation of clustering is a difficult task as it is an unsupervised classification meaning that ground-truth labels are usually not available. However, the literature suggests different clustering validation indices aiming at providing both (1) a statistical evaluation of the clusters to measure how well their members are tight and separated from the other clusters and (2) comparing two different partitions in terms of similarity. We used both kinds of indices in this study. They are all implemented within the Scikit-learn library.

2.3.1. Silhouette index (SI)

Silhouettes were introduced to measure how well an object belongs to its own cluster and as a tool to objectify the choice of the number of clusters k in partitioning algorithms such as k-means (Rousseeuw, 1987). For an object i part of the M samples, a Silhouette value $S(i)$ relies on the mean intra-cluster distance a_i and the mean nearest-cluster distance b_i:

$$S(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}}, i \in [1, M] \tag{3}$$

$S(i)$ ranges from -1 to 1 and renders the degree of membership of the object to its cluster. A negative value suggests that the object is assigned to the wrong cluster and stand for an outlier. By averaging Silhouette values, an overall Silhouette index (SI) can be computed to render the clustering quality:

$$SI = \frac{1}{M}\sum_i S(i), i \in [1, M] \tag{4}$$

Typically, k is chosen in such a way that SI is maximum. A comparative analysis of 30 validation indices reports SI as the best index on various synthetic datasets and in the top tier on real datasets (Arbelaitz et al., 2013). However, SI tends to endorse clustering that produces convex clusters, such as k-means or HAC, and may be inappropriate if the algorithm allows the retrieval of anisotropic or concave clusters. This case may happen with GMM or the HAC algorithm if constrained with a connectivity matrix.

2.3.2. Information criteria

The number of components using GMMs are usually not optimized using the Silhouette index but based on information criteria relying on the log-likelihood of the GMM and accounting for the number of free parameters in the model. For this purpose, the Akaike Information Criterion AIC (Akaike, 1974) and the Bayesian Information Criterion BIC (Schwarz, 1978) are usual:

$$AIC = -2\log(L) + 2d \tag{5}$$

$$BIC = -2\log(L) + d\log(M) \tag{6}$$

where $\log(L)$ is the log-likelihood, M the number of samples, and d is the number of degrees of freedom related to the model. The degrees of

freedom d is given by summing the covariance, mean, and mixing weights free parameters. With no constraint on the covariance, the number of covariance free parameters are given by the half of the number off-diagonal elements and the number of diagonal elements, i.e., $kN(N+1)/2$, where N is the number of time-steps in the case of TSC. The number of mean parameters is given by kN since the mean vector is of dimension N. At last, the number of weight parameters is given by $k-1$ since $k-1$ parameters are sufficient to describe the mixture weights as they sum up to one. Unlike SI, AIC or BIC should be minimal for an optimal k.

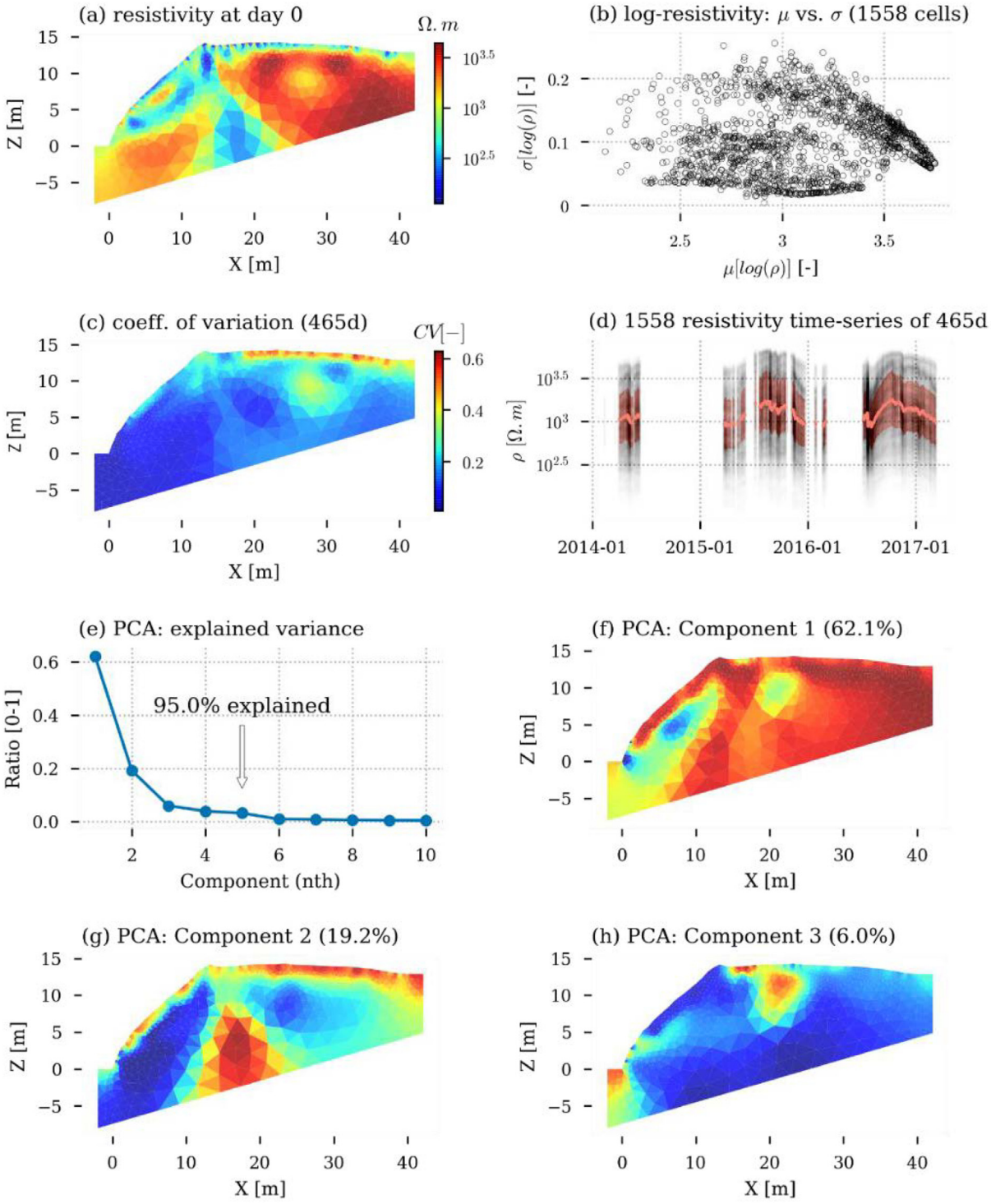

Fig. 2. Time-lapse ERT model of the Rochefort cave subsurface. (a) Inverted resistivity ρ for the reference model at day 0. (b) Scatterplot of the mean log-resistivity $\mu[log(\rho)]$ and standard deviation of log-resistivity $\sigma[log(\rho)]$ computed over the 465 days for the 1558 spatial cells. (c) Coefficient of variation ($CV = \sigma(\rho)/\mu(\rho)$) of the resistivity computed over the 465 days. (d) Inverted daily resistivity time-series (log-scale). The red line is the mean time-series, with the shaded red areas representing the interquartile range. Missing data are left blank. (e) Variance explained by the principal component analysis (PCA) of the z-standardized resistivity dataset (Eq. (1)). (f to h) First, second, and third principal components associated with (e). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

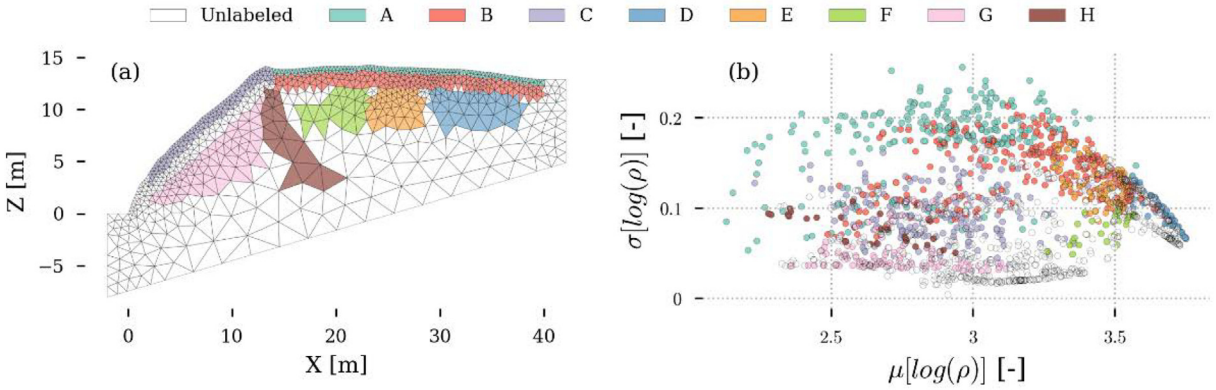

Fig. 3. Expert-based classification of the time-lapse ERT model from Watlet et al., 2018a. (a) Spatial zonation of the groups. (b) Scatterplot of the mean log-resistivity $\mu[log(\rho)]$ and the standard deviation of log-resistivity $\sigma[log(\rho)]$ for the 1558 resistivity time-series. Colors correspond to the groups identified in (a).

2.3.3. Adjusted mutual information (AMI)

Based on the information theory, the adjusted mutual information (AMI) is used to measure the similarity between two partitions, or the classification performance if one partition is considered as ground truth data. Another application is consensus clustering, which aims at identifying a more robust partition from an ensemble of different clustering algorithms' outputs based on their degree of agreement (Monti et al., 2003; Vinh and Epps, 2009). In this paper, AMI is used to compare the similarity of two clustering partitions with and without spatial connectivity constraint (Section 4.2), the outcomes of different time-series representations, or the robustness of the clustering derived from subsamples instead from the whole dataset (Section 4.3). AMI is an adjusted measure of similarity. Adjustment in clustering comparison is needed to account for the expected similarity score of randomness, which may vary according to the number of clusters k. It allows having a similarity score ranging from 0 to 1, with 0 corresponding to the score of random labeling and 1 reflecting a perfect agreement between two clustering outputs. Scikit-learn's implementation of AMI relies on Vinh et al. (2010). Considering two clustering partitions vector **U** and **V**:

$$AMI(\mathbf{U},\mathbf{V}) = \frac{I(\mathbf{U},\mathbf{V}) - \mathbb{E}\{I(\mathbf{U},\mathbf{V})\}}{\max\{H(\mathbf{U}),H(\mathbf{V}))\} - \mathbb{E}\{I(\mathbf{U},\mathbf{V})\}} \quad (7)$$

where $H(\mathbf{U})$ and $H(\mathbf{V})$ are the information entropy of the given partition, and $I(\mathbf{U},\mathbf{V})$ is the mutual information between both partitions. The expected mutual information for randomness is $\mathbb{E}\{I(\mathbf{U},\mathbf{V})\}$ based on random partitions preserving the number of clusters k and the number of members in each cluster. In general, AMI has the advantage that its score remains unchanged in case of permutations of the cluster labels. It is particularly useful for comparing agreement between two partitions since one object may belong to two respective clusters that are similar, but most likely labeled differently.

3. Study site and data

3.1. Rochefort cave laboratory

The clustering approaches are applied to a real dataset collected at the Rochefort Cave Laboratory (Watlet et al., 2018b). This site (Camelbeeck et al., 2012) is located near the town of Rochefort, in one

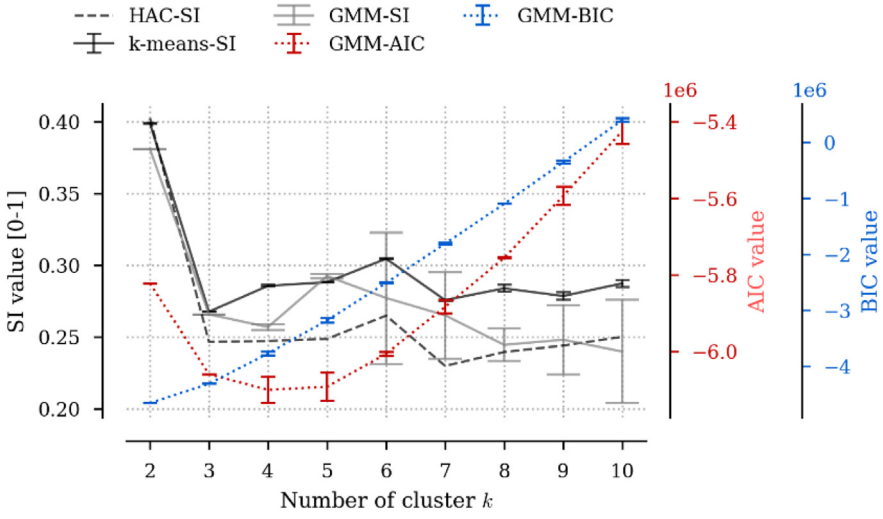

Fig. 4. Comparison of clustering validation indices for different numbers of clusters and clustering algorithms. *The Silhouette index (SI) is reported for the k-means, HAC, and GMM models in black or grey with respect to the left axis. The blue and red lines report on both the right axes the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) obtained for the GMM models. Regarding k-means and GMM, the error bars represent 2 standard deviations across 20 runs, each of them including 20 random initializations.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the largest karst systems of Belgium. It hosts several kinds of sensors and instruments, which have been progressively installed in natural caves or at the surface since the early '90s to monitor several types of natural processes, from active movements along geological fractures (Camelbeeck et al., 2012) to water infiltration patterns in both the saturated and unsaturated zones of the karst system (Poulain et al., 2018). It is in the latter context that the ERT dataset used in this paper has been acquired.

The experiment, detailed in Watlet et al., 2018a, allowed collecting ERT datasets daily between 2014 and 2017, which still represents, to the best of our knowledge, the longest, high-resolution ERT monitoring experiment conducted in a karst environment. The electrodes are permanently installed along a line of 48 electrodes at 1-m intervals. The line starts at the bottom of a doline and goes all the way to the top of a flat limestone plateau. Most of the electrodes are permanently buried at shallow depth, while the first six electrodes are directly attached to the outcropping limestone. Measurements were carried out first via an ALERT system (Kuras et al., 2009) and then with a Syscal Pro (Iris Instruments) and include dipole-dipole and gradients protocols. Data quality was assessed via reciprocal measurements. Resistivity models were processed using BERT (Rücker et al., 2006; Günther et al., 2006) using a time-lapse inversion scheme with a reference model. For a detailed presentation of the measurements and the inversion aspect, see Watlet et al., 2018a.

3.2. Time-lapse ERT model

The time-lapse ERT dataset (Fig. 2) is obtained from dipole-dipole arrays. The spatial grid consists of 1558 cells corresponding to the number of samples *M* to be clustered. Each of them is assigned to a resistivity time-series defined on 465 daily time steps defining the number of dimensions *N* of the dataset. Due to the particularly challenging context of using ERT on a karst system (Watlet et al., 2018a), several gaps occur throughout the dataset (Fig. 2d). Although more continuous datasets could be used for this study, such gaps are inherent to field measurements and should be accounted for when searching for semi-automated tools to support the interpretation of time-lapse ERT results. Moreover, this dataset has two main strengths: (i) it images a complex fractured limestone area and therefore shows a vast range of resistivity patterns both spatially and temporally, and (ii) it is a 2D case study, which is an advantage when testing several clustering approaches. These two aspects seem ideal to explore different clustering methods in the context of identifying geological features with distinct hydrological patterns, i.e., hydrofacies.

The structural interpretation of the ERT model is described into details in Watlet et al., 2018a using a series of external information form an in-situ borehole, geological observations, and a 3D model from a UAV-based photoscan (Triantafyllou et al., 2019) performed in the cave. The interpretation resulted in a segmented classification of the model into the eight zones shown in Fig. 3. The highly resistive zones under the plateau (Fig. 2a) were interpreted as low porosity limestone (Fig. 3, zones D & F). More conductive patterns were attributed to either the soil (Fig. 3, zones A & C), the karstified limestone areas (Fig. 3, zones B & E), or a zone of increased fracture intensity with strong dip in the middle of the image (Fig. 3, zone H). Lastly, zone G presents a low and relatively constant resistivity (see Fig. 2) related to the presence of clayey limestone. The classification was limited to the upper model because the experts took into account the loss of resolution in the lower part. Based on the PCA first component (Fig. 2f), the dynamic high resistive limestone zone F is correlated with the clayey limestone (Fig. 3, zone G). The other massive limestone zone D (Fig. 3) is correlated with the rest of the model (Fig. 2f to h). On the second component (Fig. 2g), the superficial zones A to C appear more clearly. The porous limestone area E is also identifiable in blue tones, as well as a spot of higher conductivity on the reference model (Fig. 2a), or using the coefficient of variation (Fig. 2c). The patterns of the third component

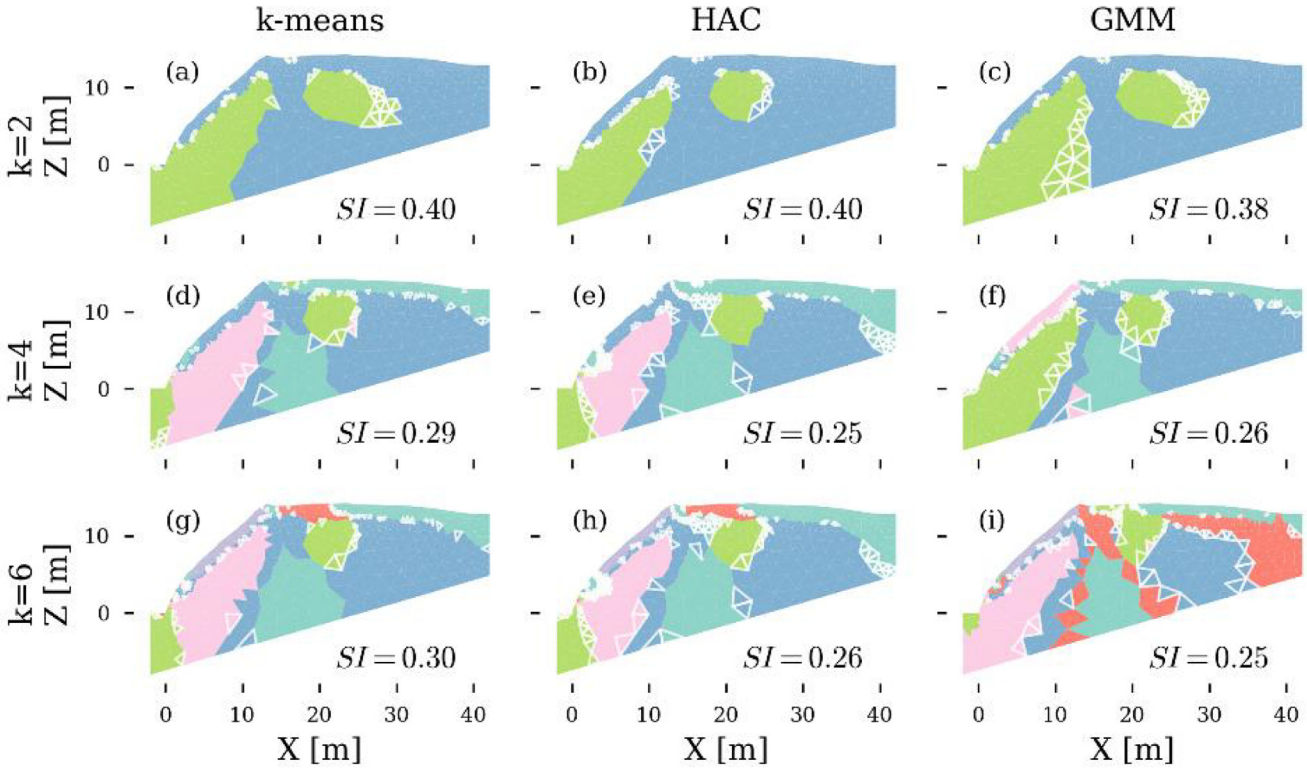

Fig. 5. Comparison of the spatial clustering patterns for k = 2, 4, and 6. *The figure columns refer to the three clustering algorithms applied to the time-lapse ERT dataset: k-means, HAC, and GMM. The rows represent different choices regarding the number of cluster k. SI is the corresponding average Silhouette index associated with the partition. The white edges correspond to negative SI values indicating potentially misclassified cells.*

(Fig. 2h) are mostly redundant with the first one, except for the lower left part of the model. However, that area was not considered in Fig. 3a since it consists of extrapolated resistivity values given that the first electrode is located at the origin of the X and Z axis.

4. Results

4.1. Comparison of clustering algorithms

This first section compares k-means, HAC with Ward's linkage method, and GMM. The clustering algorithms are applied to the z-standardized (Eq. (1)) resistivity time-series (Fig. 2d). The appropriate number of clusters k is studied by relying on the Silhouette Index (SI, Eq. (4)) for k-means, HAC, and GMM. Regarding GMM, the optimal number of clusters is also appreciated using the AIC and the BIC criteria (Eq. (5) and (6)). The results are reported in Fig. 4. The higher SI yields the preferred k. On the contrary, the lowers AIC or BIC indicate the preferred k for GMM. Since GMM and k-means presents a risk of non-deterministic outputs due to random initialization, their related curves are represented with error bars relative to 2 standard deviations resulting from 20 runs of the clustering algorithm. For both algorithms, each run involves 20 random initializations allowing to select the best model (see Sections 2.2.1 and 2.2.3). As suggested by the small error bars, the k-means clustering appears stable. In contrast, the GMM model is relatively unstable for k=6 and above with respect to the SI value. This is not the case regarding the AIC or BIC. Hence, the GMM is likely to generate different patterns, although they have a similar log-likelihood L (Eq. (4) and (5)).

SI values are relatively low (<0.4), indicating weak compactness and low separability, as one would have expected from a smooth dataset. Still, all indices agree that the optimal number of clusters k is 2, except GMM-AIC (red), suggesting k between 3 and 5. As a second-best, a k value of 6 appears for the HAC and the k-means method, which may be geologically relevant given the different lithologies described in

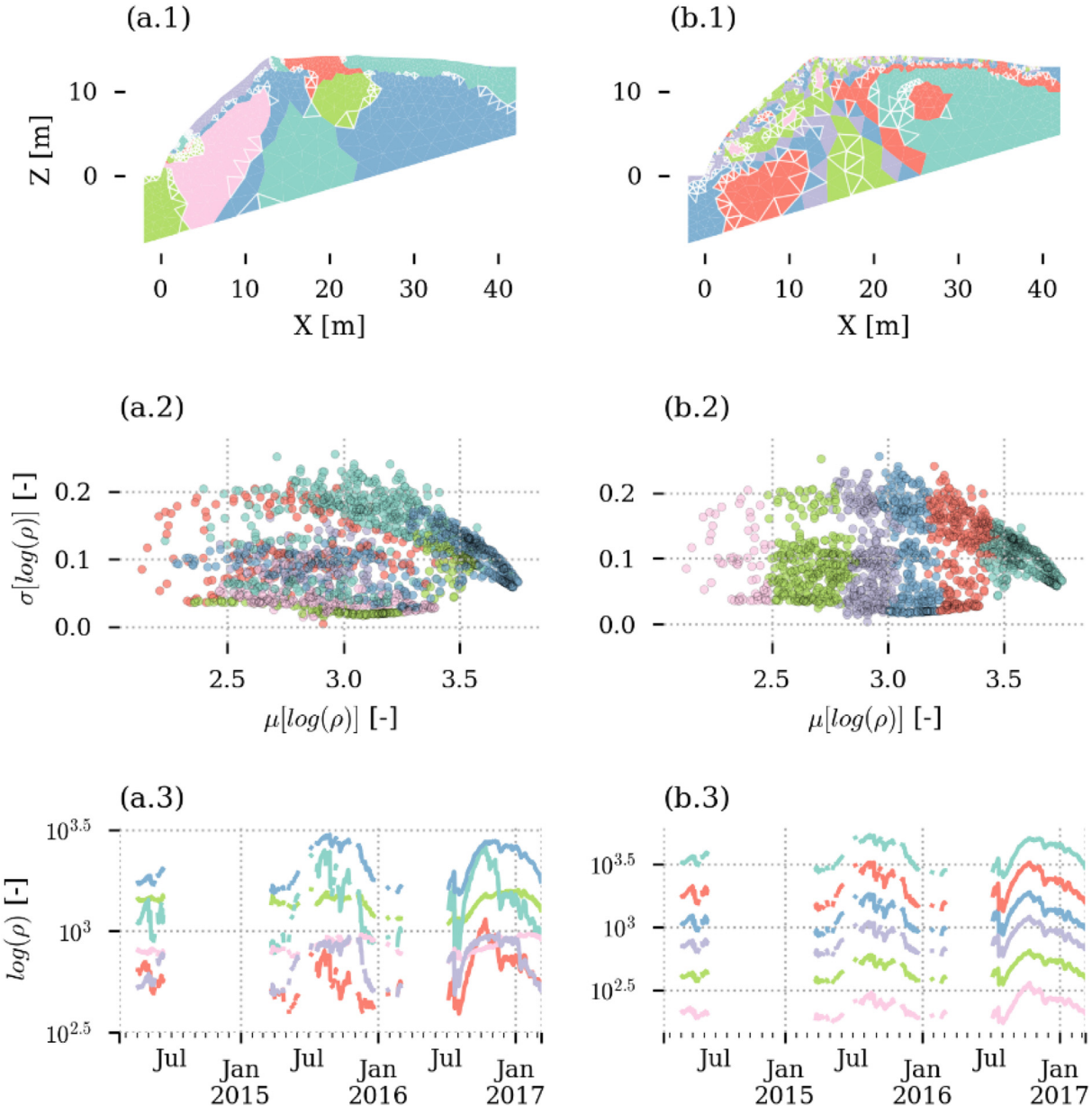

Fig. 6. Diagnostic plot for HAC clustering (k=6). (a) On the z-standardized inverted log-resistivity data. (b) On the inverted log-resistivity data. The first row shows the spatial representation of the clustering partition with the cells having negative SI values displayed with white edges; the second one their distribution in the scatterplot of the mean log-resistivity $\mu[log(\rho)]$ versus its standard deviation $\sigma[log(\rho)]$ for each of the 1558 ERT series; the third one shows the averaged log-resistivity time-series per clusters.

Fig. 3. The k-means algorithm is stable since almost no deviation in the SI is observed. For comparison, the clustering for k values of 2, 4, and 6 are visualized spatially in Fig. 5. Cells with white edges are those having a negative Silhouette value. Since GMM is unstable (Fig. 4), the spatial patterns represented in the figure are the product of one single realization and is subject to changes across runs. This is particularly the case of GMM with k = 6 (Fig. 5i), for which the displayed patterns were intentionally selected to depict a pattern that differs from the one retrieved by k-means and HAC (Fig. 5g and h).

Even if some group attribution may differ with GMM (e.g., Fig. 5f), the spatial patterns of zonation are generally similar regardless of the method. In general, the patterns of Fig. 5 match well with the one highlighted by the PCA decomposition (Fig. 2f to h). With $k = 2$, the green cluster is representative of the slope's subsurface (mostly clayey limestone, Fig. 3, zone G) plus an additional inclusion below the plateau matching roughly the dense limestone group F in Fig. 3. The green cluster is divided into two parts once $k = 4$, except with GMM that instead identified the top part of the slope (Fig. 5f, Fig. 3, zone C). Another split occurs with the corresponding blue cluster: the top surface appears as being dynamically related to the deeper low resistivity area in the fractured zone (see Figs. 2a and 3). With $k = 6$, the slope's surface appears as a cluster on its own (violet) in all cases. As an additional comparison with Fig. 3, the k-means and HAC outputs (Fig. 5g and h) present a horizontal division of the plateau into two clusters. The identified red cluster suggests that different dynamics occur at the surface of the fractured area and above the dense limestone area (Fig. 3, zone F), which was mainly visible on the PCA first and third components (Fig. 2f and h). The red cluster of the GMM clustering (Fig. 5i) is more in phase with Fig. 3 as it separates the soil surface from the underlying bedrock. In the next sections, HAC will be exclusively considered as it is similar but computationally faster than k-means, and does not have stability issues such as GMM.

In contrast with Fig. 5, Fig. 6 reports the $k = 6$ HAC clustering applied on the log-resistivity inverted data: the z-standardized log-resistivity for the first column (Fig. 6a.x) and the raw log-resistivity for the second one (Fig. 6b.x). Applying HAC on the z-standardized resistivity (Fig. 5h) or the z-standardized log-resistivity (Fig. 6a.1) provides spatially similar clusters on this long term dataset. Fig. 6a.3 reports the averaged raw log-resistivity time-series and their distinct dynamical patterns, especially for the delay and magnitude of resistivity declines that occur during fall. The pink and the green cluster (Fig. 6a.3) are less responsive to variation in resistivity over time and were regrouped together in the $k = 2$ partition in Fig. 5. Fig. 6a.2 shows that some clusters (e.g., lime green, blue, or turquoise) are spread over the entire statistical space defined by the mean log-resistivity and its standard deviation. Hence, if such a cluster gathers correlated series, it most likely groups different geological materials together, thus, different hydrofacies, which encourage to consider raw log-resistivity as well for more consistency.

Nevertheless, clustering on the raw log-resistivity alone yields the quantization of the ERT models into iso log-resistivity clusters. This is most visible on the statistical scatterplot of Fig. 6b.2. In other words, the clustering of the full dataset of 465 days is roughly equivalent to the clustering of the mean of the 1558 log-resistivity series. No information about the dynamical nature of resistivity is leveraged to define the clusters. Consequently, the clustering produces averaged time-series (Fig. 6b.3) that are highly correlated, hence, poorly representative of the hydrological states of the subsurface system. Still, some spatial zones are of interest such as the spatial red node within the turquoise cluster below the plateau that map to the porous limestone area of lower resistivity compared to the surrounding and denser limestone (Fig. 2a, Fig. 3., zone E). Besides, if some spatially organized patterns are consistent in both approaches, this is not the case of the pink cluster of Fig. 6a.1 that should rather be eventually broken up into a lower and upper part (Fig. 6b.1). Indeed, while spatially tied, the pink cluster presents a wide range of mean log-resistivity.

Ideally, an adequate clustering method for the recovery of hydrofacies should leverage both information about the correlated dynamics from the z-standardized data and the raw resistivity. For

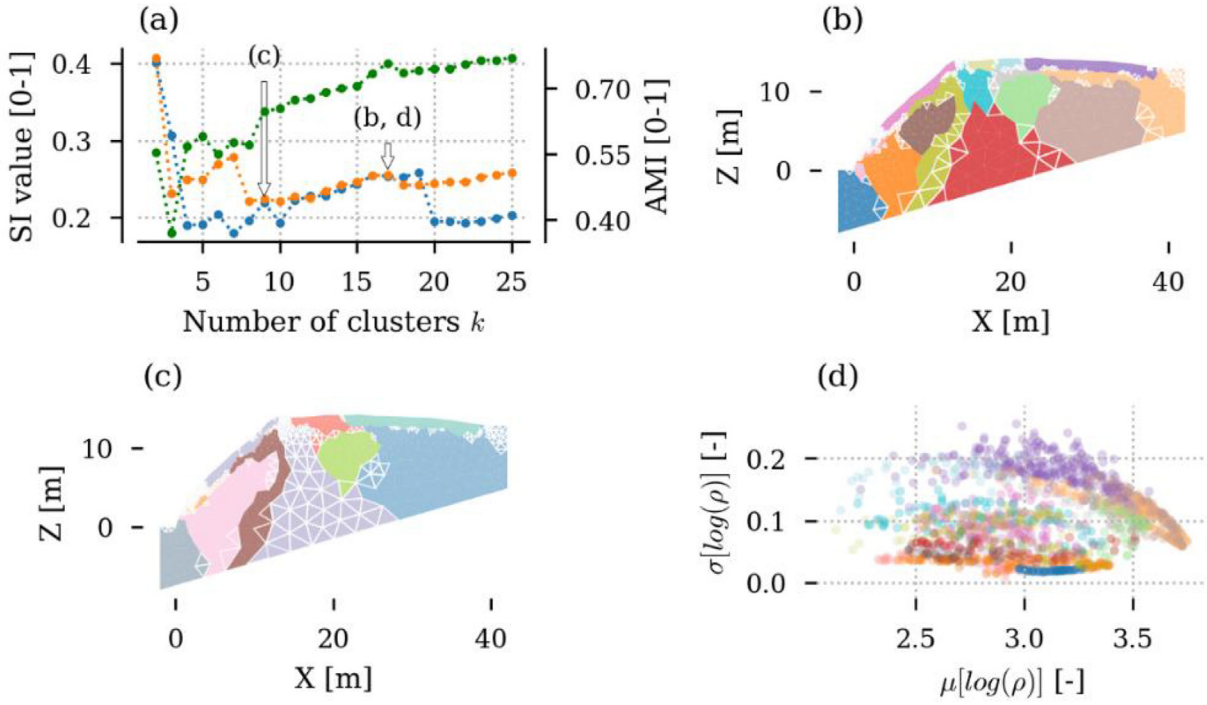

Fig. 7. Selection of the number of clusters for the HAC method with connectivity constraint. (a) Silhouette Index (SI, Eq. (4)) for the HAC with connectivity constraint (blue), without it (orange), and their similarity (green) given by the Adjusted Mutual Information (AMI, Eq. (7)). (b) HAC with connectivity constraint and k = 17 clusters. (c) HAC with connectivity constraint and k = 9. (d) Scatterplot of the mean log-resistivity $\mu[log(\rho)]$ versus its standard deviation $\sigma[log(\rho)]$ for the clustering presented in (b). Cells with negative Silhouette values in (b) and (c) are showed with white edges. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

instance, this could be done either by considering the raw log-resistivity in the clustering processes with some weighting scheme or by defining a posteriori a consensus clustering (see Monti et al., 2003) between a.1 and b.1 in Fig. 6. We did not develop these methods because of our inability to validate or identify the number of clusters with such a dual approach that involves a wide range of potential compromises. This opportunity would rather be investigated using virtual experiments, which falls beyond the scope of our study. Notwithstanding, Fig. 6 portrays the clustering results in a way that allows a fine diagnostic of the outcome and, eventually, a supervised reclassification of the groups, when intra-cluster resistivity ranges are too broad.

4.2. Spatially constrained clustering

To mitigate the inconsistencies brought by the wide ranges of mean log-resistivity and standard deviation (Fig. 6a.2) within clusters, a first possibility is to disjoint those that are spatially (Fig. 6a.1) or statistically (Fig. 6a.2) split. Another possibility is to spatially constrain the clustering by providing a spatial connectivity matrix to the HAC algorithm (see Section 2.2.2). The constraint will increase the number of the cluster over six, up to the point that the partition is both spatially and dynamically consistent in terms of correlation. The process of selecting the appropriate number of clusters k with the Silhouette Index (SI,

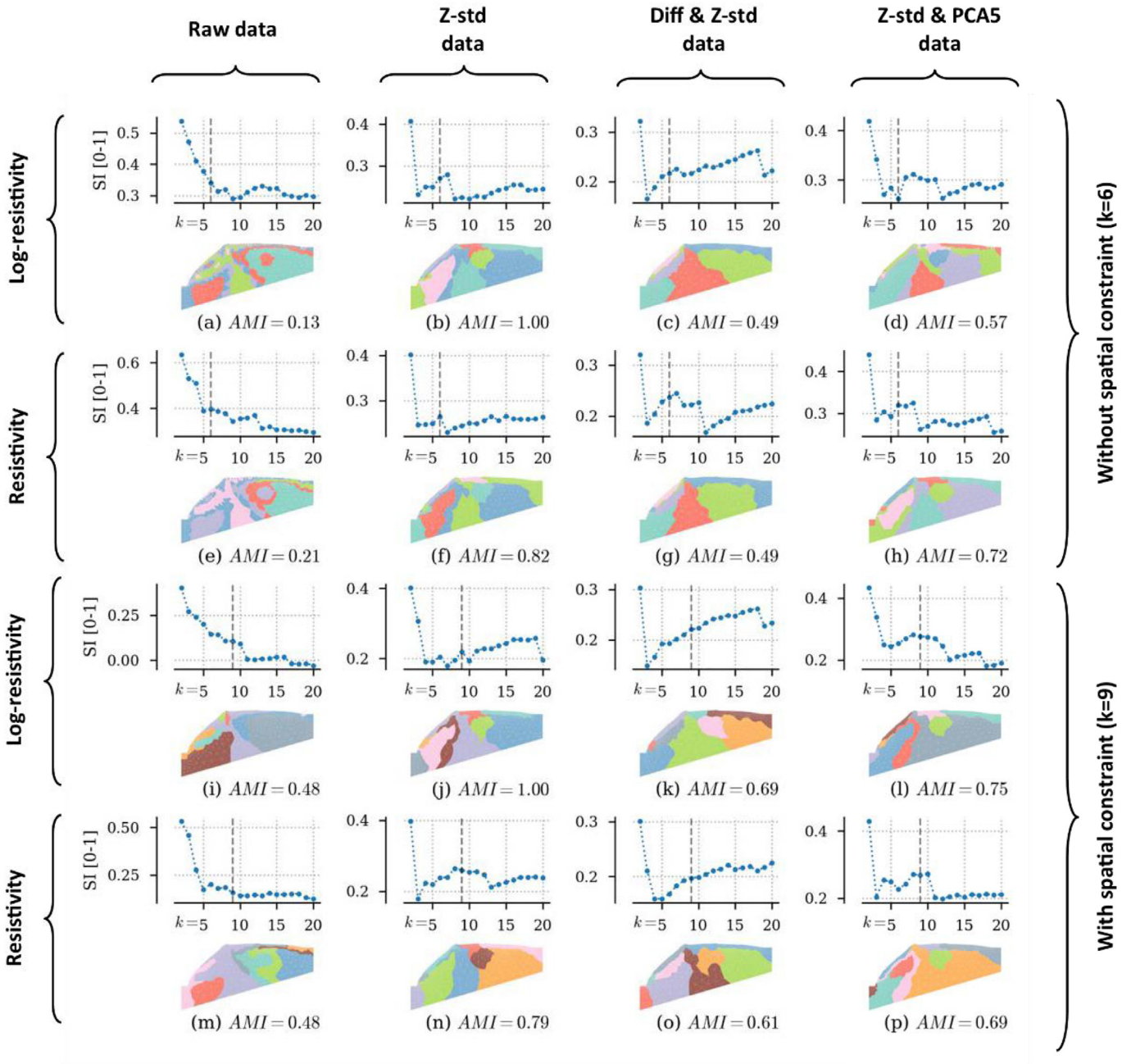

Fig. 8. HAC clustering applied to various time-series representation. (a to h) With $k = 6$ and (i to p) with $k = 9$ clusters and a spatial connectivity constraint. Time-series representations considered for the clustering are either the resistivity or the log-resistivity as raw data (column one), z-standardized data (column two), differenced and z-standardized data (column 3), and decomposed z-standardized data into 5 PCA components. Each label (a to p) shows the Silhouette variation according to the number of cluster k and the spatial patterns of clusters for the k indicated by the vertical dashed line in the Silhouette plot.

Eq. (4)) is repeated in Fig. 7a using HAC on the z-standardized log-resistivity data. Respectively, the blue and the orange curves report the SI with and without the use of the spatial connectivity constraint. The AMI similarity (Eq. (7)) between the two approaches is given by the green curve. Above k=6, a first local optimum appears at 9 clusters. With k=9, the spatial organization of patterns (Fig. 7c) is comparable to what is seen in Figs. 5 or 6, but the top slope, here in violet (zone C in Fig.3), has merged with a wider fractured area (Fig.3, zone H). The latter is poorly defined as being fully characterized by negative SI values (Fig. 7.c. white edges). This means that the dynamics found in this area are more similar to those of other clusters, mainly the turquoise one in Fig. 7c, if compared to Figs. 5 or 6.

Further apart, a better optimum is found around 17 clusters (Fig. 7a), which coincide with a small peak in the AMI, and equivalent SI in both clustering approaches. Thus, we have interpreted this point as a methodologically consistent number of clusters. Above 19 clusters, the SI with connectivity constraint drops to 0.2. With k=17 (Fig. 7b), the main former spatial patterns remain recognizable with the notable difference that a second horizon appears in the plateau, as in Fig. 3 (zone B) or Fig. 5i. Another difference is that the pink cluster of Fig. 7c has been split into two parts. Besides, the spatial constraint has the effect of restoring more consistent groups in terms of average resistivity and standard deviation (Fig. 7d). A part of this consistency is, however, explained by spatial smoothness constraint in the inversion scheme. Finally, k=17 results in the presence of many small clusters, mainly in uncertain areas (see Figs. 5h and 6a.1) at the bottom of the slope or above the fractured area. However, if the smaller groups and those located relatively far from the surface electrodes are ignored, the partition would provide about ten spatially distinct groups that are geophysically interpretable, similarly to Fig. 3. Still, it appears that the more conductive porous limestone area (Fig. 3E) does not appear even with many clusters as high as 17.

4.3. Sensitivity and robustness of clustering partitions

In Fig. 8, the clustering task is applied to various time-series representations that are or not log-scaled, normalized, differenced, or decomposed (Section 2.1). The first block (a to h) applies HAC with k set to 6 clusters while the second one (i to p) considers 9 clusters with a spatial connectivity constraint. Within each block, the two rows represent the choice to work either on the resistivity (Ω. m) or its log-transformation. Then, the clustering is applied, respectively to the columns of Fig. 8, on these raw datasets (Raw data), the z-standardized (Eq. (1)) ones (Z-std data), their first order differences followed by a z-standardization (Diff & Z-std data), and finally on the five first components of the PCA decomposition of the z-standardized data. Each labeled pair of figures represent the Silhouette index (SI, Eq. (4)) as a function of the number of clusters k, and below it, the spatial patterns on the

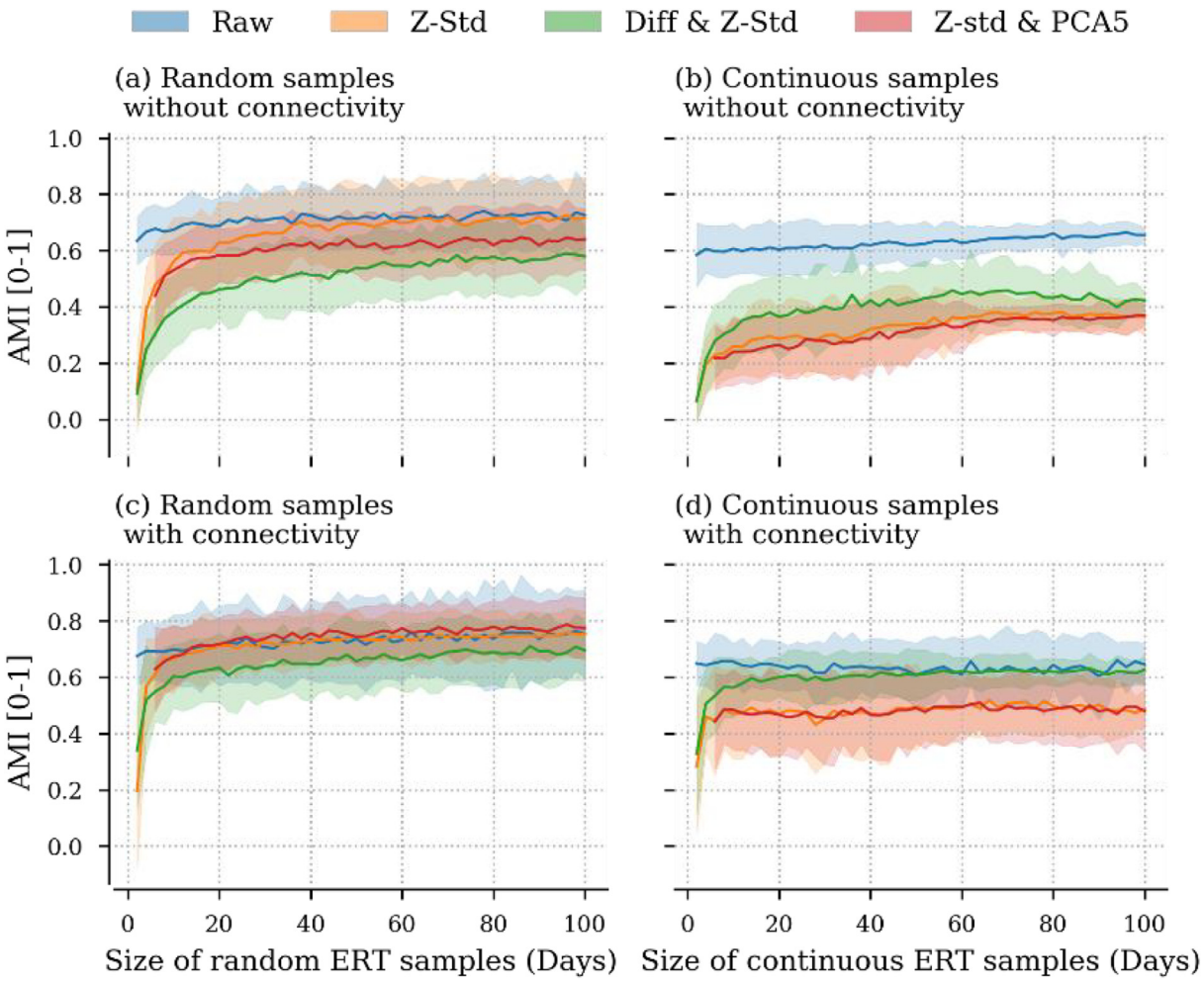

Fig. 9. Convergence of HAC clustering partitions for various representations of log-resistivity data with increasing size of the sample sets. AMI (Eq. (7)) is computed between every 50 runs of clustering on the sample sets and the partition retrieved on the full dataset of 465 days: (a) with random samples (k = 6); (b) with continuous samples (k = 6); (c) with random samples and with connectivity constraint (k = 9); (d) with continuous samples and with connectivity constraint (k = 9). The curve presents the mean and the 2 standard deviation bands for each representation of Fig. 8 (Raw, Z-std, Diff & Z-std, Z-std & PCA5).

bottom related to k=6 or 9, whether a spatial connectivity constraint is considered or not, in phase with Figs. 6a and 7c. Although the selected k is not always a local optimum, arbitrarily fixing the number of clusters allows comparing the similarity of partitions with the Adjusted Mutual Information (AMI, Eq. (7)). We chose as references the clustering applied on the z-standardized log-resistivity (b and j), which therefore have an AMI of 1. An AMI of zero reflects the score of two random partitions. In Fig. 8, an AMI reaching 0.7 shows comparable spatial patterns with the reference. Of course, visual differences in the top of the model and closer to the electrodes have much more impact on the AMI score due to the variable resolution of the grid.

In terms of spatial patterns, the new representation based on differencing (Diff & Z-std) produces interesting cluster distributions. Without spatial constraint, the clusters are nevertheless mostly continuous. The massive limestone area on the right of the fractured area (Fig. 3F) does not appear. The differencing removes the seasonal variation of the resistivity (see Fig. 6a.3), making this spot more synchronous with the rest of the limestone area below the plateau. The zone is, however, identified when the connectivity constraint is applied. Another interesting cluster is the banana-shaped one below the slope surface corresponding to the area of clayey limestone (G in Fig. 3). The shape maps well with the ones retrieved from raw resistivity data (a, e, i, m), indicating a consistent cluster. Regarding the application of the PCA (Z-std & PCA5), the clustering produces similar clusters compared to the reference, except for (d), but the value of 6 clusters does not seem appropriate given its sub-optimal SI. Otherwise, not much information is lost from the decomposition, and this option could be considered for reducing the computational requirement of the clustering task. In general, the selection of the number of clusters based on the SI is dependent on the time-series representation.

Yet, all clustering tasks shown in Fig. 8 are applied to the full time-span of the dataset, which is 465 days. Another aspect of sensitivity is related to the question: how much information (i.e., days) is necessary to retrieve the clustering partitions of Fig. 8? The question is addressed in Fig. 9 with HAC clustering applied with and without connectivity constraint on the log-resistivity data and its four representations shown in Fig. 8. The selected days are sampled according to two strategies. The first one (a and c) picks random but different (without replacement) ERT samples meaning that the samples could be spread over the full time-span of 465 days. The other strategy (b and d) picks continuous samples (i.e., consecutive days). For each given size, the sampling is repeated 50 times. The AMI is computed between each of the 50 clustering outputs and the partition obtained with the same time-series representation on the full time-span of 465 days (Fig. 8 a to d and i to l).

According to the random sampling strategy (Fig. 9a and c), the clustering applied on samples of the raw log-resistivity (blue) provides stable AMI across the range of sampling sizes (2 to 100 days). Compared to the full dataset, similar clustering partitions with high AMI ≅ 0.7 are obtained even on small sample sets, with or without connectivity constraints. It means that there is not much added-value of a long time-span when the clustering of raw resistivity is performed. Therefore, it does not matter much if the sample sets are continuous or not. Regarding, the clustering on the z-standardized data (orange) and decomposed data (red), the convergence of the AMI mostly occurs with samples sets below 20 days with $k = 6$ and without connectivity constraints. The decomposed data (red) has a lower convergence limit (~ 0.6) compared to the z-standardized data (orange). This is because of the unstable 465 days patterns retrieved with PCA (Fig. 8d), which was the reference for computing the AMI. Convergence is faster and occurs mostly between 10 days when the clustering is applied with k=9 and a connectivity constraint (Fig. 9c). However, there is a drop in the AMI limit when the clustering is applied on continuous sample sets (Fig. 9b and d), and AMI does not exceed 0.5 even with a time-span as significant as three months. Since it is not the case with random sampling (Fig. 9a and c), one may conclude that the clustering applied on the full dataset is mostly based on seasonal variation, as shown in Fig. 6a.3. This behavior is different for the differenced dataset (green)

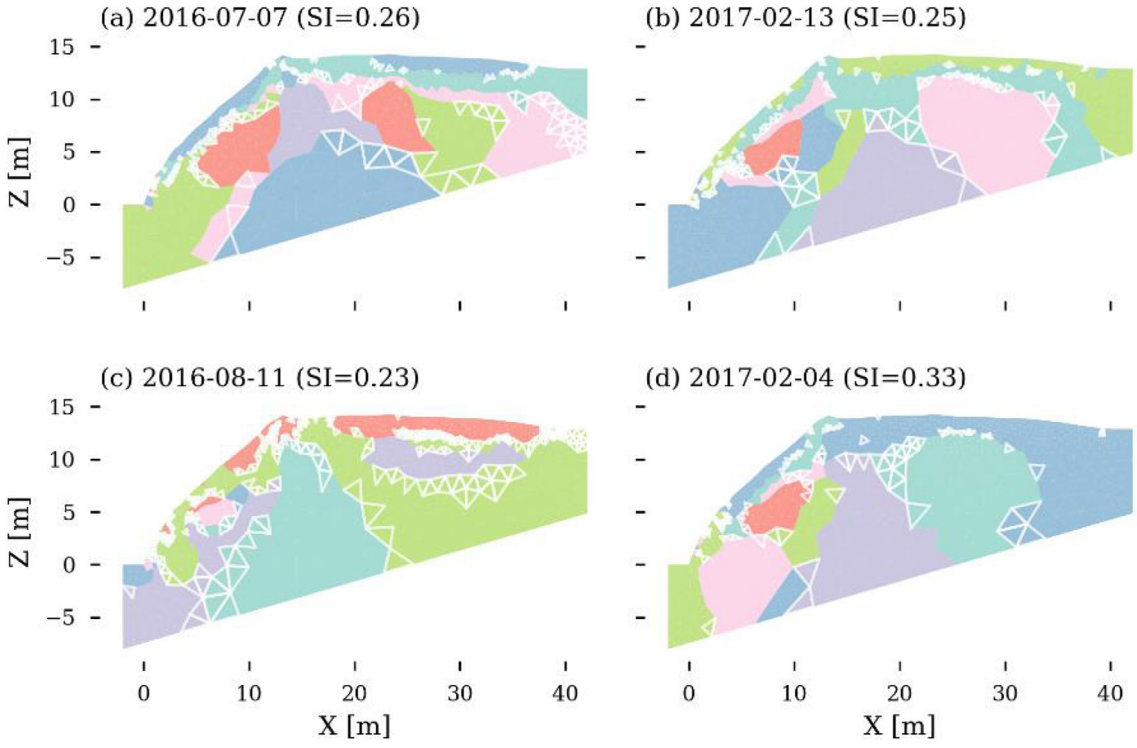

Fig. 10. HAC clustering ($k = 6$) applied to four continuous samples of 20 days of log-resistivity. (a to d) shows the results from four different starting days with the Silhouette index (SI, Eq. (4)) reported in parenthesis. The negative Silhouette values (Eq. (3)) are displayed in with white edges.

as the seasonal variation is removed by the differencing. Consequently, with random sampling (Fig. 9a and c), it converges less rapidly and to a lower AMI compared to the z-standardized (orange) and the decomposed (red) dataset. The loss of AMI is also lower when it comes to continuous sampling (Fig. 9b and d).

Regarding the continuous sampling strategy, the drop of AMI for the z-standardized dataset (orange) and the decomposed ones (red) may raise several concerns related to the geophysical investigation and the methodology for recovering hydrofacies from correlated dynamics. Indeed, Fig. 10 shows the different clustering partitions obtained from four different continuous periods of 20 days. The spatial patterns substantially deviate from the usual one retrieved on the full dataset. On the one hand, it may suggest that the patterns retrieved on the z-standardized are not robust unless applied on a long term ERT dataset covering at least a year, given the daily measurement strategy and the seasonal patterns shown in the data (Fig. 6a.3). On the other hand, the results simply reflect changes through time in dynamically correlated features across space, which may include changes in the optimal number of clusters. From that point of view, hydrofacies may change over time according to the hydrological states of the systems, water distribution, and the patterns of hydrological connectivity. The result of Fig. 10 may portray some of these changes; however, as the focus of this paper is on clustering methods, we will not attempt a premature hydrological interpretation. The point is instead to underline the sensitivity of the method and the need to develop more robust methods for the clustering of hydrofacies, for instance, by considering the raw resistivity or other geophysical data in the process. Finally, it is worth recalling that the difficulty of the clustering task in this particular case is linked to the complexity of the karstic site. The retrieval of coherent groups may be easier in a less heterogeneous environment.

5. Conclusion

Nowadays, computer-assisted vision is increasingly used to extract and delineate geological and hydrological features, sometimes referred to as litho or hydrofacies, from ERT models. While early studies provided applications for non-time-lapse ERT models, applications to time-lapse models are still underrepresented. On short time-lapse models (< 20 days), Genelle et al. (2012) and Xu et al. (2017) developed the first applications based on time-series clustering (TSC), assuming that these structures can be extracted based on the similarities observed in the time dynamics of resistivity. We have introduced the basic principle of clustering, together with three clustering evaluation metrics and one clustering similarity metric. Using a 465 days time-lapse ERT model of 1558 cells acquired from the surface of a heterogeneous karstic environment (Watlet et al., 2018a, 2018b), this paper studies: (1) the comparison between the three clustering algorithms k-means, hierarchical agglomerative clustering (HAC), and Gaussian Mixture Model (GMM), including the question of the optimal choice of cluster number and the identification of potentially misclassified spatial cells, (2) the effect of adding a spatial constraint in clustering, and (3) the robustness of the clustering outputs to various representations of the resistivity data as well as the impact of the number of days considered in the ERT model for the clustering task.

Specifically, applied to 1558 z-standardized resistivity series of 465 days, the three candidate algorithms produce similar spatial patterns that highlight temporarily correlated area across space. We considered 6 clusters based on our clustering evaluation metrics. However, such clusters may be spatially split and may include cells with substantial differences in their mean raw resistivity or standard deviation. Hence, clustering based on the correlation of resistivity series obtained from z-standardized data may retrieve geologically inconsistent groups. On the other hand, clustering on the raw resistivity time-series is dominated by their mean resistivity. Accordingly, the retrieved clusters depict isoresistivity areas, but their averaged temporal dynamics are all correlated, and nothing is learned about the specific dynamic property of the subsurface elements. This therefore encourages to work on the standardized resistivity while checking the raw resistivity distribution within clusters.

In the second part, we consider the HAC specificity of adding a spatial constraint such that clusters are spatially tied into one feature. The constraint had the expected effect of increasing the suggested number of clusters to 9 or 17. With 9, one of the clusters would have needed to be separated for consistency. With 17, the expected patterns are well represented overall by about ten clusters, while the remaining clusters were either relatively small or distant from the electrodes, thus deserving less consideration. The results were indeed more consistent in their raw resistivity and standard deviation while applied on the z-standardized data due to the spatial proximity of the cells, but some raw resistivity patterns are not always revealed from correlation patterns.

In the last section, HAC with and without connectivity constraint was applied to 8 different time-series representations where the resistivity is, or not, logarithmically scaled, standardized, differenced, or dimensionally reduced with principal component analysis. The major differences in spatial patterns remained between the raw resistivity and the other representations revealing correlated areas. The redundancy of patterns across the different representations creates confidence in the patterns that are restituted. However, our sensitivity analysis based on smaller sample sets showed that these patterns are associated with the seasonal dynamics of resistivity and cannot be retrieved from the standardized data even with continuous sample sets of 100 days. It also shows how much interpretation can vary between a single ERT survey and time-lapse experiments, as well as from one short-term time-lapse survey to another. Still, less than 20 days are necessary to retrieve the long-term patterns if they are not continuous but randomly picked in the model. If this last result may encourage long-term ERT monitoring of at least one year for the retrieval of robust clusters. It may also depict the temporal variability of water distribution, hydrological processes, and so hydrofacies if they are identified from short-term correlated resistivity.

In general, our results encourage to practice the clustering of time-lapse ERT models with various numbers of clusters, various time-series representations, and various sample sets to gain confidence from redundancies between the resulting patterns. Shortly, more robust clustering methods for the identification and zonation of hydrofacies and lithofacies will benefit from integrating both the information about raw resistivity and temporal dynamic similarity, and eventually other geophysical datasets (e.g., Giuseppe et al., 2014, 2018; Paasche et al., 2006). Regarding the algorithms, HAC was particularly interesting for its versatility. Besides its ability to constraint spatially the clustering, HAC can be applied with any distance metrics. If HAC does not account directly for the uncertainty in the clustering such as GMM, such uncertainty may be computed from bootstrap samples in the particular case of time-lapse ERT datasets. In that spirit, HAC is used to generate consensus clustering that may already be helpful to create a final clustering from several clustering partitions (Monti et al., 2003). Further guidelines will most likely be fruitfully developed in combination with synthetic experiments combining resistivity and hydrological modeling since the main difficulty of clustering is its unsupervised nature and the difficulty of appreciating the validity of the outcomes.

Acknowledgment, samples, and data

This work is part of a Ph.D. supported by a FRIA grant from the Belgian Fund for Scientific Research (FSR-FNRS). The original ERT Data is publicly available from Watlet et al. (2018b). Code recipes for clustering the ERT dataset using the Scikit-learn library are available from a Mendeley Data repository: http://dx.doi.org/10.17632/zh5b88vn78.2.

A.Watlet publishes with the permission of the Executive Director, British Geological Survey (UKRI-NERC).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. Front. Neuroinform. 8. https://doi.org/10.3389/fninf.2014.00014.

Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y., 2015. Time-series clustering a decade review. Inf. Syst. 53, 16–38. https://doi.org/10.1016/j.is.2015.04.007.

Akaike, H., 1974. A new look at the statistical model identification. Springer Series in Statistics. Springer, New York, pp. 215–222. https://doi.org/10.1007/978-1-4612-1694-0_16.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recogn. 46, 243–256. https://doi.org/10.1016/j.patcog.2012.07.021.

Arthur, D., Vassilvitskii, S., 2007. K-means++: the advantages of careful seeding. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics.

Audebert, M., Clément, R., Touze-Foltz, N., Günther, T., Moreau, S., Duquennoi, C., 2014. Time-lapse ERT interpretation methodology for leachate injection monitoring based on multiple inversions and a clustering strategy (MICS). J. Appl. Geophys. 111, 320–333. https://doi.org/10.1016/j.jappgeo.2014.09.024.

Banton, O., Cimon, M.-A., Seguin, M.-K., 1997. Mapping Field-Scale Physical Properties of Soil with Electrical Resistivity. Soil Sci. Soc. Am. J. 61, 1010. https://doi.org/10.2136/sssaj1997.03615995006100040003x.

Barker, R., Moore, J., 1998. The application of time-lapse electrical tomography in groundwater studies. Lead. Edge 17, 1454–1458. https://doi.org/10.1190/1.1437878.

Berkhin, P., 2006. A survey of clustering data mining techniques. Grouping Multidimensional Data. Springer-Verlag, pp. 25–71. https://doi.org/10.1007/3-540-28349-8_2.

Camelbeeck, T., van Ruymbeke, M., Quinif, Y., Vandycke, S., de Kerchove, E., Ping, Z., 2012. Observation and interpretation of fault activity in the Rochefort cave (Belgium). Tectonophysics 581, 48–61. https://doi.org/10.1016/j.tecto.2011.09.027.

Caterina, D., Beaujean, J., Robert, T., Nguyen, F., 2013. A comparison study of different image appraisal tools for electrical resistivity tomography. Near Surf. Geophys. 11 (6), 639–657. https://doi.org/10.3997/1873-0604.2013022.

Chambers, J.E., Wilkinson, P.B., Wardrop, D., Hameed, A., Hill, I., Jeffrey, C., Loke, M.H., Meldrum, P.I., Kuras, O., Cave, M., Gunn, D.A., 2012. Bedrock detection beneath river terrace deposits using three-dimensional electrical resistivity tomography. Geomorphology 177-178, 17–25. https://doi.org/10.1016/j.geomorph.2012.03.034.

Chambers, J.E., Wilkinson, P.B., Penn, S., Meldrum, P.I., Kuras, O., Loke, M.H., Gunn, D.A., 2013. River terrace sand and gravel deposit reserve estimation using three-dimensional electrical resistivity tomography for bedrock surface detection. J. Appl. Geophys. 93, 25–32. https://doi.org/10.1016/j.jappgeo.2013.03.002.

Chambers, J.E., Meldrum, P.I., Wilkinson, P.B., Ward, W., Jackson, C., Matthews, B., Joel, P., Kuras, O., Bai, L., Uhlemann, S., Gunn, D., 2015. Spatial monitoring of groundwater drawdown and rebound associated with quarry dewatering using automated time-lapse electrical resistivity tomography and distribution guided clustering. Eng. Geol. 193, 412–420. https://doi.org/10.1016/j.enggeo.2015.05.015.

de Pasquale, G., Linde, N., Doetsch, J., Holbrook, W.S., 2019. Probabilistic inference of subsurface heterogeneity and interface geometry using geophysical data. Geophys. J. Int. https://doi.org/10.1093/gji/ggz055.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Methodol. 39, 1–22.

Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S.A., Green, A.G., 2010. Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. Geophysics 75, G53–G64. https://doi.org/10.1190/1.3496476.

Elwaseif, M., Slater, L., 2010. Quantifying tomb geometries in resistivity images using watershed algorithms. J. Archaeol. Sci. 37, 1424–1436. https://doi.org/10.1016/j.jas.2010.01.002.

Elwaseif, M., Slater, L., 2012. Improved resistivity imaging of targets with sharp boundaries using an iterative disconnect procedure. J. Environ. Eng. Geophys. 17, 89–101. https://doi.org/10.2113/jeeg17.2.89.

Fiandaca, G., Doetsch, J., Vignoli, G., Auken, E., 2015. Generalized focusing of time-lapse changes with applications to direct current and time-domain induced polarization inversions. Geophys. J. Int. 203 (2), 1101–1112. https://doi.org/10.1093/gji/ggv350.

Gallego, G., Cuevas, C., Mohedano, R., Garcia, N., 2013. On the mahalanobis distance classification criterion for multidimensional normal distributions. IEEE Trans. Signal Process. 61 (17), 4387–4396. https://doi.org/10.1109/TSP.2013.2269047.

Genelle, F., Sirieix, C., Riss, J., Naudet, V., 2012. Monitoring landfill cover by electrical resistivity tomography on an experimental site. Eng. Geol. 145-146, 18–29. https://doi.org/10.1016/j.enggeo.2012.06.002.

Giuseppe, M.G.D., Troiano, A., Troise, C., Natale, G.D., 2014. K-Means clustering as tool for multivariate geophysical data analysis. An application to shallow fault zone imaging. J. Appl. Geophys. 101, 108–115. https://doi.org/10.1016/j.jappgeo.2013.12.004.

Giuseppe, M.G.D., Troiano, A., Patella, D., Piochi, M., Carlino, S., 2018. A geophysical k -means cluster analysis of the Solfatara-Pisciarelli volcano-geothermal system Campi Flegrei (Naples, Italy). J. Appl. Geophys. 156, 44–54. https://doi.org/10.1016/j.jappgeo.2017.06.001.

Günther, T., Rücker, C., Spitzer, K., 2006. Three-dimensional modelling and inversion of dc resistivity data incorporating topography - II. Inversion. Geophys. J. Int. 166, 506–517. https://doi.org/10.1111/j.1365-246x.2006.03011.x.

Hermans, T., Irving, J., 2017. Facies discrimination with electrical resistivity tomography using a probabilistic methodology: effect of sensitivity and regularisation. Near Surf. Geophys. 15 (1), 13–25. https://doi.org/10.3997/1873-0604.2016047.

Hsu, H.-L., Yanites, B.J., Chen, C. Chih, Chen, Y.-G., 2010. Bedrock detection using 2D electrical resistivity imaging along the Peikang River Central Taiwan. Geomorphology 114, 406–414. https://doi.org/10.1016/j.geomorph.2009.08.004.

Infante, V., Gallardo, L.A., Montalvo-Arrieta, J.C., de León, I.N., 2010. Lithological classification assisted by the joint inversion of electrical and seismic data at a control site in Northeast Mexico. J. Appl. Geophys. 70, 93–102. https://doi.org/10.1016/j.jappgeo.2009.11.003.

Ishola, K.S., Nawawi, M.N.M., Abdullah, K., 2014. Combining multiple electrode arrays for two-dimensional electrical resistivity imaging using the unsupervised classification technique. Pure Appl. Geophys. 172, 1615–1642. https://doi.org/10.1007/s00024-014-1007-4.

Keogh, E., Kasetty, S., 2003. On the need for time series data mining benchmarks: a survey and empirical demonstration. Data Min. Knowl. Disc. 7, 349–371. https://doi.org/10.1023/a:1024988512476.

Kuras, O., Pritchard, J.D., Meldrum, P.I., Chambers, J.E., Wilkinson, P.B., Ogilvy, R.D., Wealthall, G.P., 2009. Monitoring hydraulic processes with automated time-lapse electrical resistivity tomography (ALERT). Compt. Rendus Geosci. 341, 868–885. https://doi.org/10.1016/j.crte.2009.07.010.

Kutbay, U., Hardalaç, F., 2017. Development of a multiprobe electrical resistivity tomography prototype system and robust underground clustering. Expert. Syst. 34, e12206. https://doi.org/10.1111/exsy.12206.

Liao, T.W., 2005. Clustering of time series data: a survey. Pattern Recogn. 38, 1857–1874. https://doi.org/10.1016/j.patcog.2005.01.025.

Loke, M.H., Barker, R.D., 1996. Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method1. Geophys. Prospect. 44 (1), 131–152. https://doi.org/10.1111/j.1365-2478.1996.tb00142.x.

Monti, S., Tamayo, P., Mesirov, J., Golub, T., 2003. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach. Learn. 52 (1), 91–118. https://doi.org/10.1023/A:1023949509487.

Nguyen, F., Kemna, A., Robert, T., Hermans, T., 2016. Data-driven selection of the minimum-gradient support parameter in time-lapse focused electric imaging. Geophysics 81 (1), A1–A5. https://doi.org/10.1190/geo2015-0226.1.

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

Paasche, H., Tronicke, J., Holliger, K., Green, A.G., Maurer, H., 2006. Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on Fuzzy c -means cluster analyses. Geophysics 71 (3), H33–H44. https://doi.org/10.1190/1.2192927.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Poulain, A., Watlet, A., Kaufmann, O., Camp, M.V., Jourde, H., Mazzilli, N., Rochez, G., Deleu, R., Quinif, Y., Hallet, V., 2018. Assessment of groundwater recharge processes through karst vadose zone by cave percolation monitoring. Hydrol. Process. 32, 2069–2083. https://doi.org/10.1002/hyp.13138.

Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Rücker, C., Günther, T., Spitzer, K., 2006. Three-dimensional modelling and inversion of dc resistivity data incorporating topography - I. Modelling. Geophys. J. Int. 166, 495–505. https://doi.org/10.1111/j.1365-246x.2006.03010.x.

Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A., Richard, G., 2005. Electrical resistivity survey in soil science: a review. Soil Tillage Res. 83, 173–193. https://doi.org/10.1016/j.still.2004.10.004.

Scaini, A., Audebert, M., Hissler, C., Fenicia, F., Gourdol, L., Pfister, L., Beven, K.J., 2017. Velocity and celerity dynamics at plot scale inferred from artificial tracing experiments and time-lapse ERT. J. Hydrol. 546, 28–43. https://doi.org/10.1016/j.jhydrol.2016.12.035.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464. https://doi.org/10.1214/aos/1176344136.

Singh, A., Sharma, S.P., Akca, İrfan, Baranwal, V.C., 2017. Fuzzy constrained Lp-norm inversion of direct current resistivity data. Geophysics 83, E11–E24. https://doi.org/10.1190/geo2017-0040.1.

Singha, K., Day-Lewis, F.D., Johnson, T., Slater, L.D., 2014. Advances in interpretation of subsurface processes with time-lapse electrical imaging. Hydrol. Process. 29, 1549–1576. https://doi.org/10.1002/hyp.10280.

Tan, P.-N., Steinbach, M., Kumar, V., Karpatne, A., 2019. Introduction to Data Mining. Global edition. Pearson Education Limited, Harlow, United Kingdom.

Triantafyllou, A., Watlet, A., Mouélic, S.L., Camelbeeck, T., Civet, F., Kaufmann, O., Quinif, Y., Vandycke, S., 2019. 3-D digital outcrop model for analysis of brittle deformation and lithological mapping (Lorette cave Belgium). J. Struct. Geol. 120, 55–66. https://doi.org/10.1016/j.jsg.2019.01.001.

Vinh, N.X., Epps, J., 2009. A novel approach for automatic number of clusters detection in microarray data based on consensus clustering. 2009 Ninth IEEE International

Conference on Bioinformatics and BioEngineering. IEEE. https://doi.org/10.1109/bibe.2009.19.
Vinh, N.X., Epps, J., Bailey, J., 2010. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. J. Mach. Learn. Res. 11, 2837–2854.
Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244. https://doi.org/10.1080/01621459.1963.10500845.
Ward, W.O.C., Wilkinson, P.B., Chambers, J.E., Oxby, L.S., Bai, L., 2014. Distribution-based fuzzy clustering of electrical resistivity tomography images for interface detection. Geophys. J. Int. 197, 310–321. https://doi.org/10.1093/gji/ggu006.
Ward, W.O.C., Wilkinson, P.B., Chambers, J.E., Nilsson, H., Kuras, O., Bai, L., 2016. Tracking tracer motion in a 4-D electrical resistivity tomography experiment. Water Resour. Res. 52, 4078–4094. https://doi.org/10.1002/2015wr017958.
Watlet, A., Kaufmann, O., Triantafyllou, A., Poulain, A., Chambers, J.E., Meldrum, P.I., Wilkinson, P.B., Hallet, V., Quinif, Y., Van Ruymbeke, M., Van Camp, M., 2018a. Imaging groundwater infiltration dynamics in the karst vadose zone with long-term ERT monitoring. Hydrol. Earth Syst. Sci. 22, 1563–1592. https://doi.org/10.5194/hess-22-1563-2018.
Watlet, A., Kaufmann, O., Triantafyllou, A., Poulain, A., Chambers, J.E., Meldrum, P.I., Wilkinson, P.B., Hallet, V., Quinif, Y., Van Ruymbeke, M., Van Camp, M., 2018b. Data and Results for Manuscript Imaging Groundwater Infiltration Dynamics in Karst Vadose Zone with Long-Term Ert Monitoring. https://doi.org/10.5281/zenodo.1158631.
Wilkinson, P., Chambers, J., Meldrum, P., Watson, C., Inauen, C., Swift, R., Curioni, G., 2019. The automated geoelectrical data processing workflow of the PRIME infrastructure monitoring system. 2019 1st Conference on Geophysics for Infrastructure Planning, Monitoring and BIM. EAGE. https://doi.org/10.3997/2214-4609.201902562.
Xu, S., Sirieix, C., Riss, J., Malaurent, P., 2017. A clustering approach applied to time-lapse ERT interpretation case study of Lascaux cave. J. Appl. Geophys. 144, 115–124. https://doi.org/10.1016/j.jappgeo.2017.07.006.
Zhou, J., Revil, A., Karaoulis, M., Hale, D., Doetsch, J., Cuttler, S., 2014. Image-guided inversion of electrical resistivity data. Geophys. J. Int. 197, 292–309. https://doi.org/10.1093/gji/ggu001.