Waste Management 195 (2025) 22–31

Contents lists available at ScienceDirect

Waste Management

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

An integrated study of hard and soft cluster analyses for detecting leachate in a MSW landfill site using geoelectrical data

Giorgio De Donno [a], Davide Melegari [a], Valeria Paoletti [b,c], Ester Piegari [b,*]

[a] Dipartimento di Ingegneria Civile Edile e Ambientale (DICEA), "Sapienza" Università di Roma, Rome, Italy
[b] Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse (DiSTAR), Università degli Studi di Napoli "Federico II", Naples, Italy
[c] Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli Osservatorio Vesuviano, Via Diocleziano, 328, 80124 Naples, Italy

ARTICLE INFO

Keywords:
Leachate
Cluster analysis
K-means
Fuzzy C-means
Electrical resistivity tomography
Induced polarization tomography

ABSTRACT

An accurate assessment of leachate levels necessitates the integration of various parameters. Traditional geophysical prospecting methods often lack measurable accuracy because they focus on individual parameters rather than effectively integrating data. This may lead to inconsistent estimates of leachate depth and make the evaluation of prediction reliability challenging. In this study, we exploit hard and soft cluster analyses to improve the effectiveness of geoelectrical methods in identifying the extent of leachate accumulation zones. A machine learning-based approach employing hard clustering on resistivity and induced polarization data was recently proposed to obtain integrated model sections that highlight leachate accumulation zones in municipal solid waste landfills. In those models, areas with different colours represent areas characterized by specific ranges of values of the considered physical quantities and have strictly defined boundaries. This is an intrinsic limitation of hard clustering that carries out cluster assignments without providing an assessment of the reliability of the reconstruction. In contrast, soft clustering approaches provide estimation of the cluster membership that allows a refinement of cluster boundaries, improving the identification of groups in the data. We apply hard and soft cluster analyses to geoelectrical data for detecting leachate accumulation zones in a landfill located on a steep slope in Central Italy. There, leachate may not only contaminate groundwater but also trigger instability phenomena. Among the different clustering algorithms, we selected K-means due to its simplicity of implementation, its ability to identify clusters that are both compact and distinct, and its faster performance compared to Fuzzy C-means. The clusters associated with the leachate accumulation zones represent approximately the 11% of total investigated subsoil and are characterized by values of resistivity, chargeability and normalized chargeability in the ranges of about 1.5–5 Ωm, 10–70 mV/V and 4.5–37 mS/m, respectively. Then, we applied the Fuzzy C-means algorithm to obtain the degree of membership of points belonging to such areas and better outline their boundaries. By considering a fuzzy membership greater than 0.5, we achieve an accuracy exceeding 90% in identifying leachate in wells. Furthermore, identifying zones with lower membership we delineate the boundaries of less saturated regions as well as those that are more saturated, providing a reconstruction of potential preferential leachate flows within the waste mass. These findings have important practical implications as they contribute to cost reductions for future drilling and monitoring processes.

1. Introduction

Landfill leachate is a toxic liquid produced within landfills that can pose serious risks to public health if it percolates to the groundwater. Traditional non-destructing testing for detecting leachate in municipal solid waste (MSW) landfills often involve geophysical field surveys, among which electrical and electromagnetic methods have been undoubtedly the most used methods for this purpose in the last decades (see e.g., Soupios et al., 2017; De Donno et al., 2024 for recent reviews). Geophysical techniques are indeed rapid and non-invasive as they can be employed for retrieving spatially continuous models of the landfill with limited costs (e.g., Di Maio et al., 2018; Juarez et al., 2023; Mary et al., 2023). Despite these advantages, geophysical models should always be validated through direct information (i.e. borehole or well data), since they derive from measurements not directly sensing hydrogeological or contaminant properties (Day-Lewis et al., 2017). Recently, the

* Corresponding author.
E-mail address: ester.piegari@unina.it (E. Piegari).

https://doi.org/10.1016/j.wasman.2025.01.034
Received 10 October 2024; Received in revised form 20 January 2025; Accepted 22 January 2025
Available online 29 January 2025

combined use of different geoelectrical methods, such as electrical resistivity tomography (ERT) and induced polarization (IP), has been growing for leachate detection (e.g., Xia et al., 2023) and MSW landfill characterization (e.g., De Donno and Cardarelli, 2017b; Flores-Orozco et al., 2020; Ma et al., 2024). In fact, low resistivity values can be mainly correlated with an increase on bulk (electrolytic) conduction, due to the leachate produced within the landfill. Furthermore, surface conduction taking place in the electrical double layer (EDL), and triggered by biogeochemical activity occurring in the waste mass, can be sensed by measuring the voltage decay when a current is switch-off (Binley and Slater, 2020). However, the combined results of those techniques were almost only visually inspected, leaving room for ambiguous interpretation of the leachate-saturated zones (Martinho, 2023). Conversely, an accurate assessment of the fully saturated zones is pivotal to ensure that the drainage interventions will be effective for landfill management (Mukherjee et al., 2015).

In the light of these challenges, there has been a growing interest in the use of Machine Learning (ML) techniques for integrating multiple datasets in the field of Geosciences (Yu and Ma, 2021) and specifically for landfill monitoring (Sharma and Sood, 2024). The combined use of satellite images and ML has enabled for example the identification of illegal landfills (Torres and Fraternali, 2021). Various deep learning architectures, such as Fully Convolutional Networks (FCN) and U-Net, have been employed for semantic segmentation of landfill areas in satellite images (Fraternali et al., 2024). These models can accurately classify and outline landfill boundaries, significantly improving waste detection rates compared to manual inspection methods (Niu et al., 2023). Recurrent neural networks trained in the learning phase on daily precipitation, daily average temperature and the accumulated amount of landfilled waste has been proposed to predict quantity and quality of landfill leachate (Ishii et al., 2022). Majchrowska et al. (2022) proposed a deep-learning method for waste detection and classification in natural and urban environments based on open-source datasets covering different waste categories (bio, glass, metal and plastic, non-recyclable, other, paper, and unknown). Ershadi et al. (2023) demonstrated the suitability of using Machine Learning models to forecast the long-term leaching from Construction and Demolition waste (CDW). Liang et al. (2023) combined the chemical insights of the machine learning models with infrared spectroscopy for a quick characterization of biomass and waste. Gaur et al. (2024) suggested the application of Artificial Neural Networks, Genetic Algorithms, and Support Vector Machines to enhance leachate treatment, addressing the challenges posed by the heterogeneous composition of solid waste. Niu et al. (2024) presented an approach adopting effective location-routing strategies to balance the total cost, carbon emissions and distance between residential areas/waste disposal centers.

Several review studies over the last five years have provided a thorough analysis of the different ML models and techniques applied in waste management (e.g., Abdallah et al., 2020; Guo et al., 2021; Xu et al., 2021; Lu and Chen, 2022, Xia et al., 2022; Majchrowska et al., 2022; Munir et al., 2023; Fraternali et al., 2024, Huang et al., 2024). Recently, we have proposed a first ML-based approach to facilitate the identification of leachate accumulation zones starting from individual model sections of geoelectrical data (Piegari et al., 2023). In parallel, Sun et al. (2023) used a deep network for multi-view ERT fusion during the salt tracing experiments in a hazardous waste landfill, while Zhang et al. (2024) integrated geotechnical, geochemical and geophysical data also using a cluster analysis for offshore landfill characterization. Starting from these results, in this work, we aim to go beyond the state-of-the art clustering approaches for leachate detection, by integrating hard and soft clustering procedures applied to geophysical ERT and IP data acquired in MSW landfills. The benefit of this approach compared to traditional methods is potentially twofold: on one hand we can provide an integrated and quantitative model of the landfill, on the others we can assess the uncertainty associated with the reconstruction. We first use hard cluster analysis as a tool for locating the leachate accumulation areas, characterized by the lowest electrical resistivity and highest chargeability values. Secondly, we assess the boundaries of the leachate accumulations through the application of a soft clustering algorithm, using the membership function as a proxy for estimating the robustness of our procedure. We apply this approach to a MSW landfill located, in Central Italy, on a steep slope, where the detection of leachate levels is crucial to prevent possible instability phenomena triggered by its accumulation.

2. Material and methods

2.1. Site description

The investigated site is a MSW landfill built in the ‘80 s at the service of a medium-size city located in Central Italy (Fig. 1). The landfill site is located on a steep slope (average slope percent around 40%) and was provided with a bottom liner (geo-membrane) and with an embankment built at the bottom to prevent slope instability phenomena. Owing to the original design of the landfill, the volume of buried waste is greater on the southern side (higher elevation, around 510–520 m a.s.l.), progressively decreasing toward the bottom of the landfill on the northern side (elevation: 460–450 m a.s.l.). The bedrock is a marly-arenaceous flysch, which is not expected to be sensed by the geophysical survey, since the bottom liner acts as an insulator preventing the current flow outside the landfill.

The landfill is equipped with a monitoring network of wells, among which five representative Casagrande-type piezometers were chosen (P1-P5 in Fig. 1). We selected these wells since they are all the same type, they were still active at the time of measurements and leachate levels were logged during the same days of the geophysical campaign. Piezometers P1-P3 were almost dry (leachate levels of only few centimetres), while P4 and P5 displayed leachate levels of about 8 and 11 m from the well bottom, respectively (corresponding to elevations around 460 and 477 m a.s.l.).

2.2. Geophysical data acquisition and processing

We acquired 4 ERT/IP curvilinear profiles (L1-L4) spaced approximately 40–50 m apart (Fig. 1) and having a length ranging from 325 m (L1) to 500 m (L4), located along the road tracks built for site management. We used the IRIS Instruments Syscal Pro resistivity-meter with a dipole–dipole array and electrodes spaced 5 m apart. For IP data acquisition we used a 50% duty cycle with a pulse length of 2 s and 20 logarithmic spaced gates for sampling the IP decay curve (first gate centred at 0.03 s, last gate at 1.75 s).

Raw data were filtered for outliers, negative voltage values and decay curves showing increasing voltage. Geophysical data were then inverted for resistivity ρ [Ωm] and integral chargeability M [mV/V], that describe the resistive and capacitive effect of the waste mass (Binley and Slater, 2020). Specifically, chargeability is proportional to the ratio of the polarization strength of the medium divided by its electromigration strength, which reflects both changes in saturation and salinity of the leachate (electromigration), as well as an increase of the biogeochemical activity (polarization).

The contribution of the surface conduction can be isolated by computing the so-called normalized chargeability MN [mS/m] at the end of the inversion process by (e.g., Slater and Lesmes 2002):

$$MN = \frac{M}{\rho} \tag{1}$$

In this study, the ERT and IP inversion procedures were computed using the 2.5D approximation for forward modelling (Dey and Morrison, 1979), by linearizing each curvilinear profile (Fig. 1). We note that a 3D inversion of acquired data would not be feasible in this specific case, due to the large spacing between the profiles. Therefore, before performing

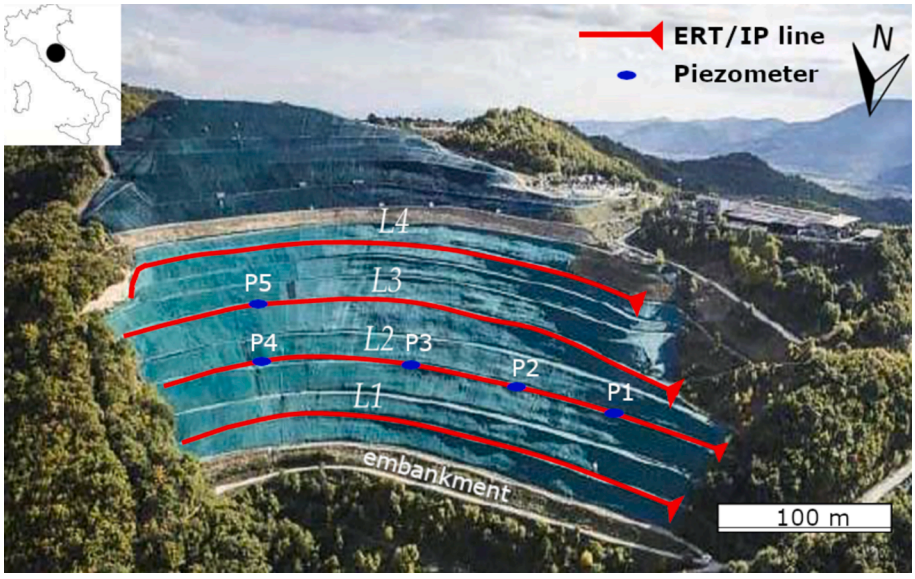

Fig. 1. View of the studied MSW landfill in Central Italy, showing the locations of the four electrical profiles (red lines, L1-L4). Red arrows indicate the direction of the survey along the lines (from the first to the last electrode). Blue ellipses mark the locations of the five piezometers (P1-P5).

the inversion, the error related to curvilinear *vs.* linear acquisition was evaluated for each quadrupole in terms of the percentage absolute error (AE) between 2D and 3D geometric factors K (Everett, 2013), as follows:

$$AE_i(\%) = 100\left|\frac{K_i^{3D} - K_i^{2D}}{K_i^{2D}}\right| \text{ with } i = 1, 2, \cdots, M, \quad (2)$$

being M the number of measurements (quadrupoles).

The three-dimensional visualizations shown in the following sections were obtained by using a newly developed MATLAB code (https://github.com/GiorgioDeDonno/ERT_IP_3D_view).

The 3D view of the AE for each geophysical measurement (Fig. 2) shows low AE values (almost everywhere below 3%), except for some isolated zones located at the boundaries of the landfill or in the bottom areas on L2 and L3, where the error increases, due to an increase in curvature at the ends of profiles. Overall, the reliability of our 2.5D approximation is confirmed by the mean absolute error for each line (sum of the AEs divided by the number of quadrupoles), which ranges from a minimum value of 0.5% (L1) to a maximum of 1.5% (L4), with L2 and L3 showing intermediate values (1.1% and 0.8%, respectively). Therefore, the impact of the linearization of curvilinear acquisition on the geophysical model is marginal almost everywhere, and we can use the linearization of the profiles in the inversion procedure.

We carried out data inversion using the VEMI software (De Donno and Cardarelli 2017a). This code implements a Gauss-Newton iterative

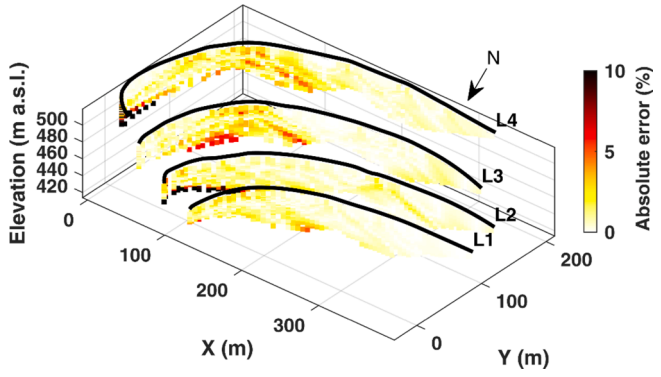

Fig. 2. Absolute error committed for each measurement due to the linearization of the curvilinear profiles over the studied landfill.

formulation (Loke and Barker 1996), where the chargeability dataset is inverted following the linear approximation proposed by Oldenburg and Li (1994). We set inequality constraints on chargeability ($M \geq 0.1$ mV/V) to avoid negative IP values in the inverted models. The inversion was conducted using the 2.5D approximation. Nevertheless, the geophysical models will be presented in this work through 3D views to ease the interpretation of the leachate accumulation across the entire landfill.

2.3. Proposed approach integrating hard and soft clustering

Fig. 3 illustrates the proposed procedure for localizing the leachate accumulation zones integrating hard and soft clustering. Our approach is based on the following steps: after ERT/IP data acquisition (i) and 2.5D inversion (ii), we combine all the geophysical information in a proper joint space (iii), then we apply a hard cluster analysis to the unique 3D dataset using the K-means algorithm(iv). We get integrated models (v) which help us to detect the presence of leachate by selecting the cluster characterized by the lowest resistivity values and the highest chargeability and normalized chargeability values. Finally, after applying a soft clustering procedure through the Fuzzy C-means (vi), we restrict the analysis on the identified cluster for assessing the boundary of the leachate accumulation areas based on the values of the membership function (vii). It is worth noting that our choice of clustering algorithms fell on the combined use of K-means and Fuzzy C-Means for several reasons. The first reason is linked to the characteristics of the dataset. The distribution of resistivity, chargeability, and normalized chargeability values occupies a relatively dense area of the parameter space, making density-based and hierarchical algorithms ineffective.

Among partitioning algorithms, K-means is commonly recognized as the most effective in identifying evenly sized clusters. Its combination with Fuzzy C-Means allows evaluating the percentage of overlap between different clusters. This percentage, particularly in relation to the cluster associated with the presence of leachate, facilitates a more reliable assessment of the extent of the leachate area. The use of Fuzzy C-means is preferred over other commonly used soft clustering methods, such as Gaussian Mixture Models, due to its computational efficiency and the fact that it does not require the underlying data to follow a specific distribution.

In the following section, we briefly summarize the principles of the K-means and Fuzzy C-means algorithms, highlighting their similarities and differences.

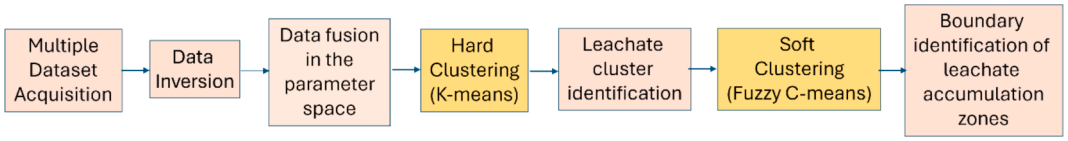

Fig. 3. Sketch of the proposed procedure to identify leachate accumulation zones through hard and soft clustering starting from geoelectrical tomographic data.

2.3.1. K-means

The K-means algorithm is an unsupervised ML partitioning algorithm subdividing a dataset into K non-overlapping groups (clusters). The iterative process of K-means starts assigning a set of K centroids. It then computes the distance between each data point and the K centroids, assigning each point to the cluster with the closest centroid. Next, the position of each centroid is updated to minimize the sum of the squared Euclidean distances between the data points and the cluster centroids, as well as the total within-cluster sum of squares (WCSS), defined as the sum of within-cluster variations for all K clusters (Bhattacharya, 2021). These steps are repeated until convergence is reached, which occurs when there are not significant changes in the assignments of the clusters or when a maximum number of iterations is reached.

K-means advantages are basically its simplicity and computational efficiency, and the convergence guarantee, even though not necessarily to the global optimum (Bhattacharya, 2021). In fact, K-means finds local minimum of WCSS, and different initial positions of centroids could be reflected in different cluster solutions. To overcome this problem, we performed a sensitivity analysis by varying the number of initial centroid configurations starting from 50. The stability of solution was found with a number of initial configurations greater than 500, which is the value used in this work. We performed the cluster analysis using the MATLAB function *kmeans*, employing the default option for centroid initialization. This means that the k-means++ algorithm was used to initialize the cluster centroids, which has been demonstrated to result in improved clustering results and faster convergence (David and Vassilvitskii, 2007). Another limitation of K-means is the requirement to predefine the number of clusters, as well as the need to select one or more criteria for determining the optimal number of clusters. In this work, we build upon the results of our previous study that found an optimal number of clusters using the elbow method (Piegari et al., 2023). Finally, since it is a hard clustering procedure, the K-means only returns a clustered section (or volume in 3D) without an evaluation of the reliability of the partitioning procedure, as the assignment to the cluster is sharp (each pixel is only assigned to one cluster).

2.3.2. Fuzzy C-means

Fuzzy C-Means (FCM) is a partitioning clustering algorithm that allows data points to belong to multiple clusters with varying degrees of membership (Bezdek et al., 1984). Unlike hard clustering algorithms such as K-means, where each data point is assigned to a single cluster, FCM assigns membership coefficients between 0 and 1 for each data point to every cluster. Similarly to K-means, the iterative process of FCM starts assigning a set of C centroids randomly. It then computes the data point membership coefficients for each cluster based on the distance between the data point and the cluster centroids. Next, the position of each cluster centroid is computed to be the mean of all data points, weighted by their membership coefficients for that cluster. The iterative process that updates the cluster centroids and membership coefficients continues until convergence is reached, i.e. when the maximum change in cluster centroids between two iterations is less than a specified threshold or when a maximum number of iterations is reached. The membership coefficients are calculated based on the inverse of the distance between the data point and each cluster centroid, raised to the power of $2/(m\text{-}1)$, where m is the parameter that controls how fuzzy the clusters will be. We set $m = 2$ in this work, which is a commonly accepted choice, as using squared distances in calculations typically ensures that the membership degrees are meaningful without excessively blurring the boundaries between clusters (Mendel et al., 2014).

In summary, compared to hard clustering algorithms, FCM provides a more flexible way of assigning data points to clusters, allowing for overlapping clusters and accounting for uncertainty in cluster assignments.

In our study, we carried out both hard and soft clustering procedures on $\log_{10}$-transformed and rescaled variables X between [0 1], such as i.e. for the resistivity:

$$X_\rho = \frac{\log_{10}\rho - \min(\log_{10}\rho)}{\max(\log_{10}\rho) - \min(\log_{10}\rho)} \tag{3}$$

with analogous formulations for M and MN.

The cluster analyses were performed by using software packages available in the Statistics and Machine Learning Toolbox of MATLAB.

3. Results

3.1. Geophysical models

A 3D view of the inverted geophysical models is shown in Fig. 4a, 4b and 4c for resistivity, chargeability and normalized chargeability, respectively.

Our results show that the landfill can be sub-divided into 3 main layers from the top to the bottom, based on its electrical properties: i) covering layer with high resistivity ($>$ 10 Ωm) and low chargeability ($<$ 10 mV/V), laterally heterogeneous, related to the covering soil and the unsaturated waste; ii) waste mass with different saturation degrees from leachate, having low resistivity ($<$ 10 Ωm) and high chargeability ($>$ 10 mV/V) and iii) bottom layer, with an increasing resistivity and a decreasing chargeability due to the presence of the bottom liner preventing the current flow outside the landfill. The resistivity increase (and chargeability decrease) related to the bottom liner is not as sharp as expected (being the liner nearly an insulator); this may be due to the sensitivity decrease of electrical methods with depth, which may cause the position and the magnitude of the anomalies at larger depths to be biased (De Donno and Cardarelli, 2017b). Although the waste mass heterogeneity locally affects the lateral variation of electrical parameters in the middle layer, electrical parameters characteristic of leachate ($\rho \sim$ 2–3 Ωm and $M \sim$ 30–40 mV/V) are located in the eastern part of the landfill. Here, the bottom of the landfill is deeper, likely due to an impluvium of the old topographic surface (before the construction of the landfill). We note that the correspondence between electrical parameters of the geophysical models and leachate level is not straightforward, as similar values of the electrical parameters are recorded both in correspondence with dry wells (i.e. P2 and P3), and wells with significant leachate levels (P4 and P5).

The MN model (Fig. 4c), also displaying a three-layer geometry, helps us discerning the surface conduction mechanism that is affected by the presence of the waste mass (microbial activity) more than the changes in saturation. However, in this case, visual inspection alone makes it difficult to establish an effective threshold for isolating the area that is fully saturated by leachate, as the imaging is dependent on the colour scale. For instance, high MN values (around 10 mS/m) are observed near P2 and P3 (which are dry), as well as in the vicinity of P4 and P5 that show significant leachate levels. To overcome this limitation and improve the interpretability of geophysical sections, in the

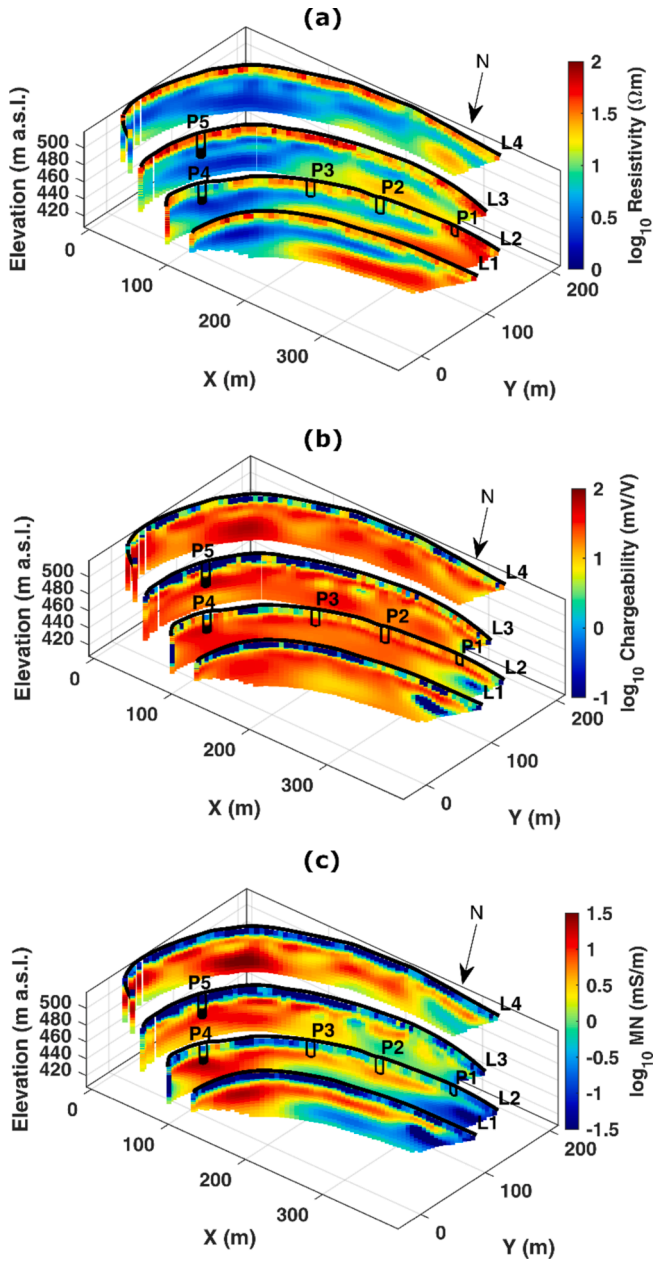

Fig. 4. 3D view of the resistivity (a), chargeability (b) and normalized chargeability (c) inverted models for the four profiles over the studied MSW landfill. The piezometric levels logged in wells (black filled rectangles) are superposed to the models and their diameters are exaggerated for the sake of clarity.

following we apply an integrated clustering approach to get a comprehensive model and its associated reliability to be used for the assessment of the leachate-saturated areas.

3.2. Cluster analysis

First, we applied the *K*-means algorithm to get a clustered model of the investigated sections (Fig. 5). The number of clusters was set equal to 10, following the results of our previous study on the same dataset (Piegari et al., 2023). This optimal number resulted from the Elbow method with an explained variance equal to 95% (Fig. 5c in Piegari et al., 2023). This allowed for maximizing information while avoid overfitting, as too many clusters would lead to non-interpretable regions regarding the ranges of physical properties. The clustering algorithm was initialized using 500 different configurations of centroids, then we selected the configuration minimizing distortion. The resulting clustering in the transformed space is shown in Fig. 5a, while a landfill 3D view is depicted in Fig. 5b. This model incorporates the three investigated geophysical parameters (ρ, *M*, *MN*) in one comprehensive reconstruction of the four profiles, using an appropriate colour scale ranging from dark green (cluster no.1) to dark red (cluster no. 10). The former displays the highest resistivity and the lowest chargeability (and *MN* values), so that the presence of leachate for these pixels is extremely unlikely. Conversely, the latter highlights the lowest values of ρ and moderate to highest of *M* and *MN*, likely associated to the leachate accumulation.

The agreement between the dark red cluster (n. 10) and the leachate levels logged in P4 and P5 is good with a precision of 1 and 0.8, respectively. The main accumulation zone is almost only localized in the eastern part of the landfill (presence of an old impluvium) and the extension of the saturated area seems to be reduced with respect to that resulting only from a visual inspection of the geophysical models (see Fig. 4). However, the localization of the leachate accumulation zones suffers from the intrinsic limitations of *K*-Means, which assigns each data exclusively to one cluster, making the identification of the edges of the accumulation areas challenging.

To this end, we applied the Fuzzy C-means algorithm to the investigated dataset and we extracted, from the reconstruction, only the cluster no.10 (Fig. 6), likely related to leachate accumulation. The data point cloud and the associated memberships for cluster no.10 are shown in Fig. 6a and 6b, where we can isolate the extreme values (lowest resistivity and highest chargeability) for the four profiles (filled circles in Fig. 6a), which are located in the middle of the accumulation zone (Fig. 6c, yellow squares). We were also able to select the boundary points (Fig. 6a, filled squares) characterized by low memberships ($<$ 0.5), located close to the cluster no.9, mainly representing the boundary of the accumulation areas (Fig. 6c, purple squares).

The 3D views of this dark red cluster and its fuzzy membership, shown in Fig. 6c and 6d respectively, help us to recognize a preferential pathway of the leachate from the top (L4) to the bottom (L1) of the landfill (blue line in Fig. 6c and 6d). As shown in Fig. 6d, high membership degrees belonging to cluster n.10 are primarily found at the boundaries of the red zones in Fig. 6c, with the exception of L3, while lower membership values are visible in the middle of these zones. Looking at the saturated zones (light blue filled wells), there is a very good agreement with the clustered models (dark red pixels) and leachate levels logged in wells.

4. Discussion

A correct definition of the boundaries of the fully-saturated areas is crucial for landfills located on steep slopes, as a continuous leachate accumulation can increase pore pressure causing slope failure (Jianguo et al., 2010; Feng et al., 2018). The results presented in Fig. 6 clearly show that the computation of membership can be a good proxy in this context, as the highest membership values for L1, L2 and L4 surround the cluster area, while L3 shows an anomalous behaviour, likely because the cluster covers a deeper zone at the bottom of the landfill that is poorly sensed by geophysical methods.

We notice in Fig. 6a that, despite the resistivity values being restricted within a very narrow range ($\sim$ 1.5–5 Ωm), both *M* and *MN* span a wider range (*M* $\sim$ 10 – 70 mV/V; *MN* $\sim$ 4.5 – 37 mS/m). This is not unexpected, because even a small variation in conduction (due change in saturation) can greatly influence the induced polarization response, since chargeability tends to decrease in saturated zones, as also recently shown for similar environments (Melegari and De Donno, 2024). These results are well aligned with those presented in a recent brief review of case studies about characterization of MSW landfills through ERT and IP methods (Abdulrahman et al., 2016). They reported that a small difference in conductivity between fully saturated and partially saturated waste can result in a decrease of chargeability around

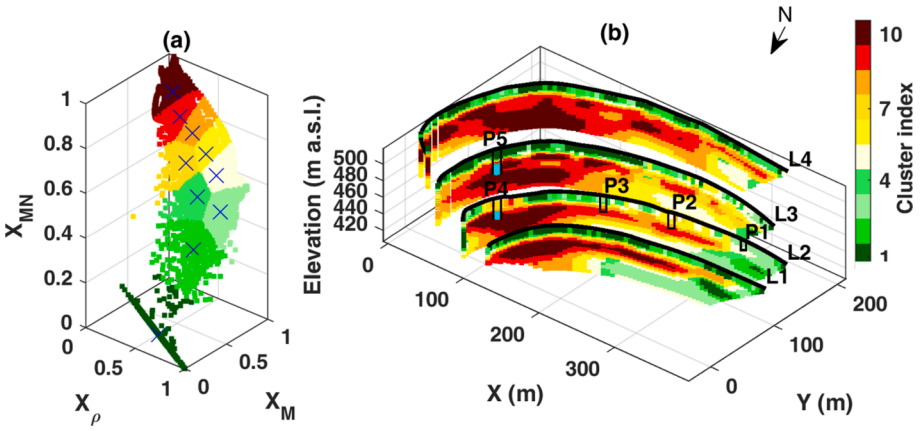

Fig. 5. (a) K-means clustering in the normalized log10-space, where position of centroids is marked with blue cross points. The dimension of the *MN*axis is doubled with respect to the other two axes for the sake of clarity. (b) 3D view of the clustered model of the landfill. The piezometric levels logged in wells (light blue filled rectangles) are superposed to the models.

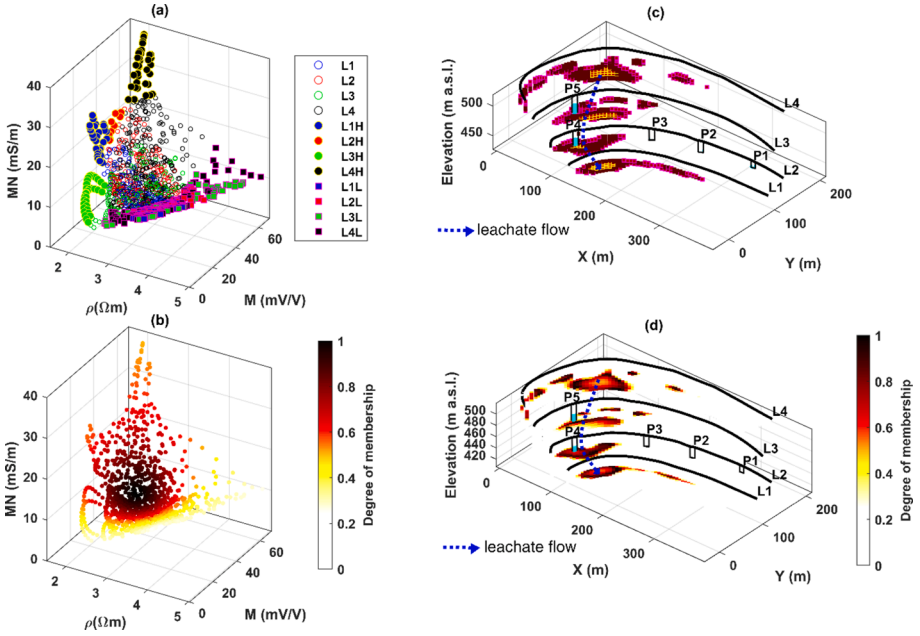

Fig. 6. Results of soft cluster analysis: (a) Scatterplot of cluster n.10 (dark red cluster in Fig. 5a) in the unscaled parameter space, for the four profiles (L1-L4). The filled circles are the extreme values having the lowest resistivity and highest normalized chargeability values. The filled squares show the boundary values having a membership lower than 0.5. (b) Degree of membership related to cluster n.10. (**c**) 3D view of the cluster n. 10 (dark red in Fig. 5a), where the yellow points are the extreme values of Fig. 6a (with the lowest resistivity and highest *MN*) and the purple points are the low-membership boundary values of Fig. 6a. (d) 3D view of the degree of membership belonging to cluster n.10. The piezometric levels logged in wells (light-blue filled rectangles) are superposed to the models and their diameters are exaggerated for the sake of clarity. The dotted blue line represents the inferred leachate preferential pathway.

50%, as in our case. Furthermore, *M* and *MN* are linked to surface polarization that is directly proportional to the biogeochemical activity that can be highly variable in the waste mass (Flores-Orozco et al., 2020). Conversely, DC resistivity is mainly related to the ion concentration and water content (Ma et al., 2024) that are not expected to significantly vary in the saturated area. Looking at the distribution of points belonging to the different profiles in parameter space (Fig. 6a), we note how the peaks in the point cloud (filled circles in Fig. 6a) clearly display different ranges of the geophysical parameters. The points belonging to L3 (green points) are located at the bottom of the section where they are affected by both a strong regularization and a very low sensitivity of the IP method (Piegari et al. 2023): the former effect leads to the arc-shaped points while the latter to the fact that chargeability is quite low despite the low resistivity values.

For the other three lines (L1, L2 and L4), there is a clear difference between the extreme points located upstream (L4, having the highest values of *M* and *MN* and low ρ values) and downstream (L1, showing low values of *M* and *MN* and the lowest ρ values), with L2 having an intermediate range of values. The very high IP response recorded for L4 likely corresponds to less saturated waste and biogeochemically active zones (high organic matter), which is responsible for high rates of microbial activity, resulting in high polarization that surpasses the high salinity and conductivity (Flores-Orozco et al., 2020). In contrast, for L1, ρ reaches the lowest values (< 2 Ωm) associated with relatively lower and constant values of *M*. Increasing concentrations of leachate result in the increase of fluid and surface conductivity, thus yielding lower IP responses (Flores-Orozco et al., 2020). These features suggest the presence of the highest degree of leachate saturation in the lowest profile L1, as expected, due to the steep slope of the landfill that leads leachate pathways downhill and confirms that peak values of *MN* cannot be

always associated with the highest degree of leachate saturation.

Since FCM such as K-Means tends to find clusters with spherical shapes, the lowest membership values in Fig. 6b are attributed to data points that are farthest from the centroid and exhibit asymmetry (Han et al., 2012). The results confirm that the membership degrees of Fuzzy C-means cannot be interpreted as higher or lower probabilities of representing areas with leachate, but simply as areas that deviate more or less from the values of electrical properties associated with the centroid. Therefore, the data corresponding to the extreme points associated with the most saturated areas (filled circles in Fig. 6a), are associated with the lowest membership values, as well as the boundary zone between cluster no. 9 and 10 (filled squares and bottom zones in Fig. 6a and 6c).

To better understand how the points belonging to the red cluster are distributed across the four profiles, we show histograms of the three geoelectrical properties (Fig. 7a, b, c), along with the distribution of fuzzy membership for each data point belonging to the dark red cluster (Fig. 7d).

The shape of the resistivity distributions (blue histograms, Fig. 7a) evolves from right-skewed (uphill, profile L4) to left-skewed (downhill, profile L1), providing evidence that the lowest resistivity values are found in the areas at the bottom of the impluvium on which the landfill is located, following the natural percolation of fluids by gravity. This effect is likely due to the peculiar shape of the bottom of the landfill (impluvium) which narrows downstream thus favouring the leachate accumulation, and decreasing the resistivity. Regarding the chargeability values, it is interesting to note that as we move from uphill to downhill (i.e., from L4 to L1), their distributions (orange histograms, Fig. 7b) progressively lose their right-skewed shape and become more focused around the mean value, while the normalized chargeability distributions (yellow histograms, Fig. 7c) remain right-skewed in all profiles except for L3. On one hand, the narrow chargeability distribution for L1, displaying an overall low average chargeability value (Fig. 7b-top), is due to the decrease of resistivity, i.e. increase of surface and fluid conductivity at the bottom of the landfill (Fig. 7a-top), counterbalancing the high *MN* values (Fig. 7c-top). On the other hand, the right-skewed *M*distribution for L4 (Fig. 7b-bottom) is mainly dominated by the increase of surface polarization clearly visible in the respective *MN* distribution (Fig. 7c-bottom), likely due to a higher

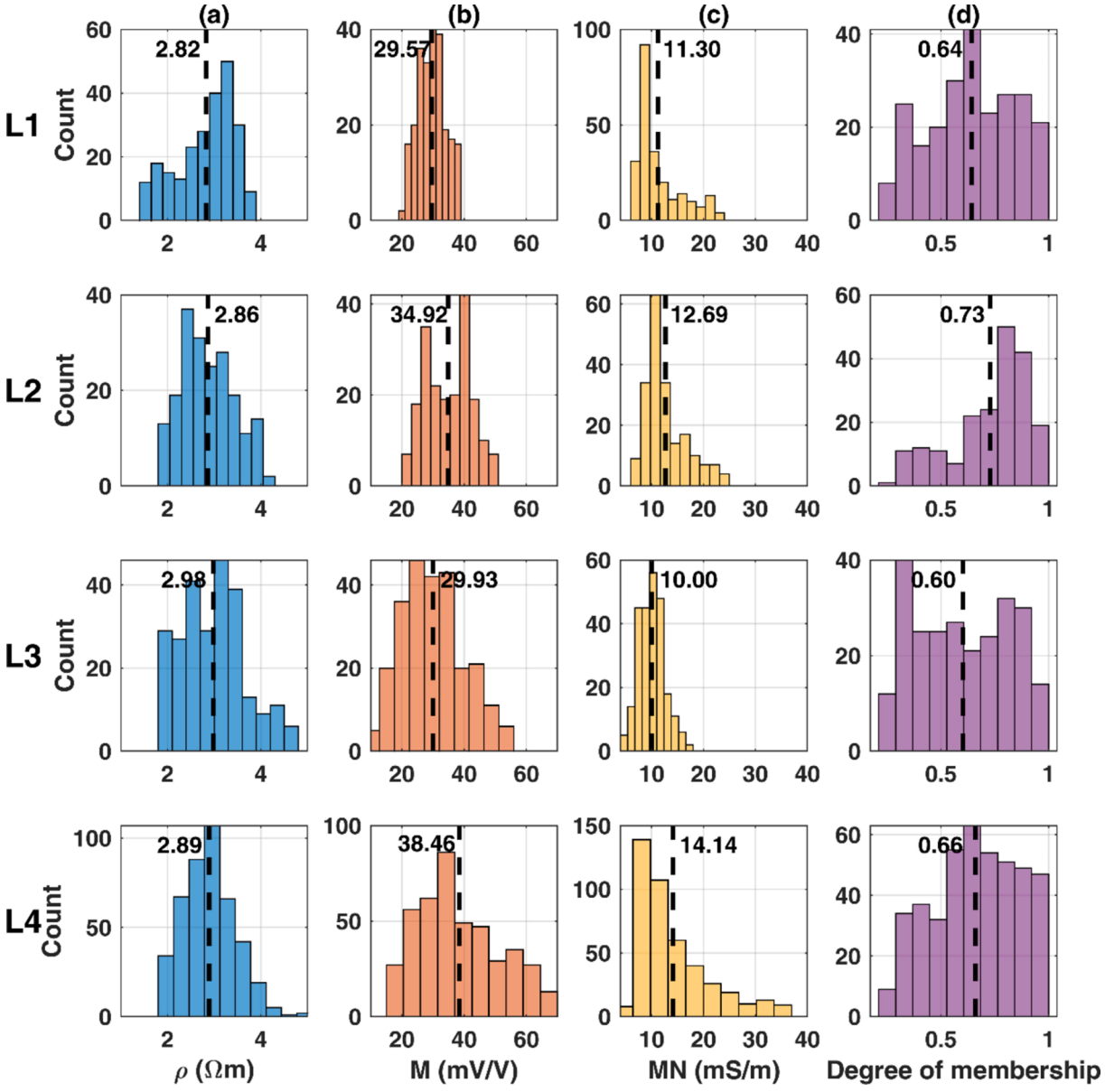

Fig. 7. Histograms of the three geophysical parameters and fuzzy membership for data in the individual profiles (cluster no. 10)**:** (a) resistivity, (b) chargeability, (c) normalized chargeability, (d) fuzzy membership. The dashed lines indicate the mean values, which are also superimposed to the plots.

biogeochemically activity at the top of the landfill (fresh waste). Finally, the distributions of the membership values (purple histograms in Fig. 7) clearly show that the greatest number of points with the lowest membership values belong to L3 and are localized at the bottom of the model section (Fig. 6d). However, for this profile, shallow pockets of leachate are identified by our cluster analyses, and the strong agreement found between one of these pockets and the leachate level measured in P5 supports our findings.

To quantify the reliability of our models, we reported in Fig. 8 the comparison between the levels logged in P4 and P5 wells and both the geophysical (ρ and M) (Fig. 8a and 8c for P4 and P5, respectively) and the clustering (index and membership) parameters (Fig. 8b and 8d for P4 and P5, respectively). On one hand, the standalone geophysical reconstruction leaves room for ambiguity in defining the threshold for the saturated zones, changes being very smooth and somewhere conflicting. On the other hand, the ML-based reconstruction is much more accurate for leachate detection, with levels well in agreement with the cluster analysis. Considering the widespread metrics used for evaluating ML-based models (see Powers, 2011 for the definitions), we achieved for the couple of wells (P4, P5) an accuracy of (0.9, 0.92), a precision of (1, 0.8), a true positive rate (TPR) of (0.75, 1) and a false positive rate (FPR) of (0, 0.13). Overall, the values of the metrics are very good, except for TPR of P4, where the leachate level is found at a slightly lower elevation compared to the well data. This value is also affected by the limited depth of the well, leading to only 4 pixels belonging to the saturated area (3 true positive, 1 false negative). In fact, for the calculations we only considered the pixels lying within the wells, even if it is straightforward that the saturation areas are also extended well below the maximum depth of the wells (with a consequent further improvement of the metrics). Additionally, the value of the membership (Fig. 8b, grey line) of the only biased point (located just below the level at an elevation of 459 m a.s.l.) is quite high (around 0.25) even if it was assigned to another cluster, thus testifying that the assignment was conflicting between clusters no. 9 and 10. Moreover, considering also the P1-P3 dry wells (0% of false negatives and true and false positives and 100% of true negatives), the overall accuracy raises to 0.95.

Finally, we quantify the extension of leachate accumulation zones as the percentage of points belonging to cluster no. 10, which is about 10.68% of the total investigated subsoil. Assuming the extreme points (characterized by the highest ρ and lowest M and MN values), located in the middle of the accumulation areas (1.8% of the total pixels, yellow points in Fig. 6c), as saturated regardless of their memberships, this percentage is almost unchanged at 10.64% when considering points characterized by fuzzy membership higher than 0.25. However, it decreases to 8.34% for fuzzy membership higher than 0.5 (yellow areas in Fig. 6c) and further drops to 5.10% when considering only membership values higher than 0.75.

It is worth noting that the use of a fuzzy membership threshold may introduce potential limitations or biases, as it could lead to an underestimation or overestimation of the leachate presence. However, adding constrains on data (i.e. considering points with a membership lower than 0.5 and with ρ, M and MN values respectively larger and smaller than those of the cluster centroid), helps defining a range of uncertainty (purple areas in Fig. 6c), which is useful for assessing the effectiveness of our procedure.

A full benefit-cost analysis of the presented procedure is beyond the

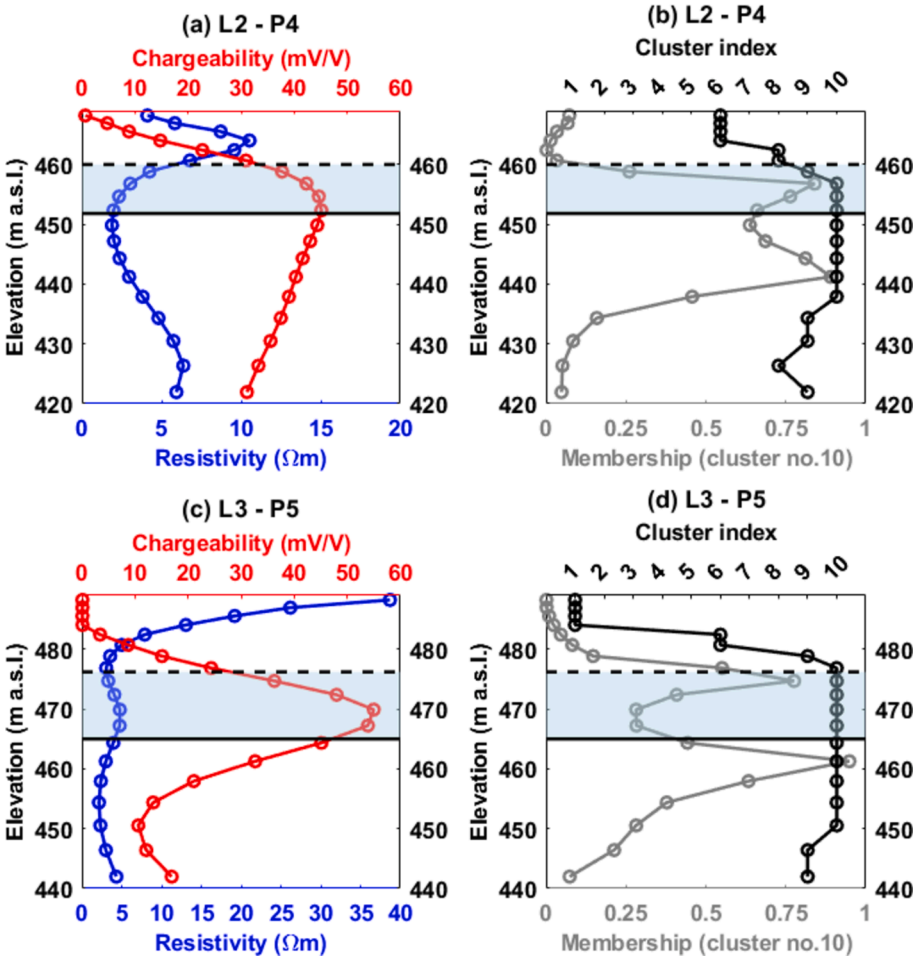

Fig. 8. Vertical profiles of geophysical parameters ρ and M (a, c), clustering index and membership (b, d), extracted at the locations of P4 and P5. The well bottom is marked by a black solid line, the leachate level by a dashed line and the saturated area is shown by the light blue filled area.

scope of this work, being it focused only on selected profiles and dependent on the peculiar geometry of the studied landfill. However, the integration of geophysical data can allow for a drastic reduction of the cost related to the drilling of wells, whose position could be biased when using only classical geophysical parameters, as demonstrated in Fig. 8. As far as the whole monitoring process is concerned, on one side the preventive application of this new method can lead to a guided choice of the monitoring network, on the other one a semi-automatic interpretation of geophysical data can be linked to alert systems during the landfill management. In addition, using this advanced data analytics, landfills can be efficiently monitored also for gas emissions, waste decomposition, and other critical factors. The ability to handle large amounts of data and provide real-time insights is essential in the complex and dynamic environment of landfill operations.

5. Conclusions

Landfills represent complex environments where the accuracy of their characterization improves with the inclusion of multiple parameters; as these parameters are interrelated, an integrated analysis enhances our understanding of the complex dynamics operating within these systems. Differently from traditional methods often relying on single-attribute analysis, we demonstrated that clustering techniques ease the integration of multiple geophysical parameters with a dual advantage. They provide model sections that simultaneously show the ranges of variation of multiple geophysical parameters in colors that are easier to interpret, even for non-specialist operators. Furthermore, they allow for automated processing of large datasets, reducing human bias and error while providing reproducible results.

Through a sequential application of K-Means and FCM algorithms we integrated geophysical data and delineated the extent of leachate accumulation zones within a MSW landfill. These zones were identified by points belonging to the cluster with the lowest ρ and the largest M and MN, each exhibiting a different degree of fuzzy membership. Specifically, they were characterized by resistivity values ranging from 1.5 to 5 Ωm, chargeability values between 10 and 70 mV/V and normalized chargeability values from 4.5 to 37 mS/m. The points with higher fuzzy memberships allowed us to identify the most saturated zones and point out preferential leachate flows. Those with lower memberships help quantifying uncertainties in boundaries of leachate zones, which were highlighted as purple areas in Fig. 6c.

The results of our approach were found in very good agreement with direct information. Using a fuzzy membership value greater than 0.5, we attained an accuracy of over 90% in detecting leachate in wells P4 and P5 and 100% in dry wells.

In conclusion, we demonstrated that the proposed approach based on a combined use of hard and soft clustering algorithms can be an effective tool for a quick and reliable identification of the leachate-saturated zones in MSW landfills, starting from geoelectrical data. Due to the effectiveness of such data in characterizing leachate properties, this method could be applied to various types of landfills and has important implications in waste management strategies~~,~~ Indeed, accurate estimates of leachate levels help sustainable management practices as the depth of leachate can influence treatment choices and operational practices (Wang and Qiao, 2024).

Future work should thoroughly explore the link between the geophysical and petrophysical parameters in such complex scenarios, in order to invert geophysical data for saturation directly.

CRediT authorship contribution statement

Giorgio De Donno: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation, Conceptualization. **Davide Melegari:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Valeria Paoletti:** Writing – review & editing, Methodology, Data curation. **Ester Piegari:** Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

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.

Acknowledgements

The authors wish to thank the Editors and three anonymous reviewers for their thorough revisions, which have definitely helped improve the quality of the manuscript. Open access publishing facilitated by Università degli Studi di Napoli Federico II, as part of the Elsevier - CRUI-CARE agreement.

Data availability

Data will be made available on request.

References

Abdallah, M., Talib, M.A., Feroz, S., Nasir, Q., Abdalla, H., Mahfood, B., 2020. Artificial intelligence applications in solid waste management: A systematic research review. Waste Manag. 109, 231–246.
Abdulrahman, A., Nawawi, M., Saad, R., Abu-Rizaiza, A.S., Yusoff, M.S., Khalil, A.E., Ishola, K.S., 2016. Characterization of active and closed landfill sites using 2D resistivity/IP imaging: case studies in Penang, Malaysia. *Environ. Earth Sci.* 75, 1–17.
Bezdek, J.C., Ehrlich, R., Full, W., 1984. FCM: the fuzzy c-means clustering algorithm. Computer & Geosciences 2, 191–203.
Bhattacharya, S., 2021. A Primer on Machine Learning in Subsurface Geosciences Vol. 1, 1–172.
Binley, A., Slater, L., 2020. Resistivity and induced polarization: Theory and applications to the near-surface earth. Cambridge University Press.
David, A., Vassilvitskii, S., 2007. K-means++: The Advantages of Careful Seeding. In: SODA '07 Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035.
Day-Lewis, F.D., Slater, L.D., Robinson, J., Johnson, C.D., Terry, N., Werkema, D., 2017. An overview of geophysical technologies appropriate for characterization and monitoring at fractured-rock sites. J. Environ. Manage. 204, 709–720.
De Donno, G., Cardarelli, E., 2017a. VEMI: a flexible interface for 3D tomographic inversion of time-and frequency-domain electrical data in EIDORS. Near Surf. Geophys. 15 (1), 43–58.
De Donno, G., Cardarelli, E., 2017b. Tomographic inversion of time-domain resistivity and chargeability data for the investigation of landfills using a priori information. Waste Manag. 59, 302–315.
De Donno, G., Melegari, D., Paoletti, V., Piegari, E., 2024. Electrical and Electromagnetic Prospecting for the Characterization of Municipal Waste Landfills: A Review, in: Anouzla, A., Souab, S., (Eds.), Technical Landfills and Waste Management: Volume 1: Landfill Impacts, Characterization and Valorisation, pp. 1–29.
Dey, A., Morrison, H.F., 1979. Resistivity modelling for arbitrarily shaped two-dimensional structures. Geophys. Prospect. 27 (1), 106–136.
Di Maio, R., Fais, S., Ligas, P., Piegari, E., Raga, R., Cossu, R., 2018. 3D geophysical imaging for site-specific characterization plan of an old landfill. Waste Manag. 76, 629–642.
Ershadi, A., Finkel, M., Susset, B., Grathwohl, P., 2023. Applicability of machine learning models for the assessment of long-term pollutant leaching from solid waste materials. Waste Manag. 171, 337–349.
Everett, M.E., 2013. Near-surface applied geophysics. Cambridge University Press.
Feng, S.J., Chen, Z.W., Chen, H.X., Zheng, Q.T., Liu, R., 2018. Slope stability of landfills considering leachate recirculation using vertical wells. Eng. Geol. 241, 76–85.
Flores-Orozco, A., Gallistl, J., Steiner, M., Brandstätter, C., Fellner, J., 2020. Mapping biogeochemically active zones in landfills with induced polarization imaging: The Heferlbach landfill. Waste Manag. 107, 121–132.
Fraternali, P., Morandini, L., Gonzalez, S.L.H., 2024. Solid Waste Detection in Remote Sensing Images: A Survey. Waste Manag. 189, 88–102.
Gaur, V.K., Gautam, K., Vishvakarma, R., Sharma, P., Pandey, U., Srivastava, J.K., Varjani, S., Chang, J., Jonathan, H.H.N., Wong, W.C., 2024. Integrating advanced techniques and machine learning for landfill leachate treatment: Addressing limitations and environmental concerns. Environ. Pollut. 354124134.
Guo, H., Wu, S., Tian, Y., Zhang, J., Liu, H., 2021. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. Bioresour. Technol. 319, 124114.
Han, J., Kamber, M., Pei, J., 2012. Data Mining: Concepts and Techniques, 3rd edn. Vol. 1, Elsevier.
Huang, L.T., Hou, J.Y., Liu, H.T., 2024. Machine-learning intervention progress in the field of organic waste composting: Simulation, prediction, optimization, and challenges. Waste Manag. 178, 155–167.

Ishii, K., Sato, M., Ochiai, S., 2022. Prediction of leachate quantity and quality from a landfill site by the long short-term memory model. *Journal of Environmental Management* 310, 114733.
Jianguo, J., Yong, Y., Shihui, Y., Bin, Y., Chang, Z., 2010. Effects of leachate accumulation on landfill stability in humid regions of China. Waste Manag. 30 (5), 848–855.
Juarez, M.B., Mondelli, G., Giacheti, H.L., 2023. An overview of in situ testing and geophysical methods to investigate municipal solid waste landfills. Environ. Sci. Pollut. Res. 30 (9), 24779–24789.
Liang, R., Chen, C., Sun, T., Tao, J., Hao, X., Gu, Y., Xu, Y., Yan, B., Chen, G., 2023. Interpretable machine learning assisted spectroscopy for fast characterization of biomass and waste. Waste Manag. 160, 90–100.
Loke, M.H., Barker, R.D., 1996. Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method. Geophysical Prospecting 44 (1), 131–152.
Lu, W., Chen, J., 2022. Computer vision for solid waste sorting: A critical review of academic research. Waste Manag. 142, 29–43.
Ma, X., Zhang, J., Schwartz, N., Li, J., Chao, C., Meng, J., Mao, D., 2024. Characterizing landfill extent, composition, and biogeochemical activity using electrical resistivity tomography and induced polarization under varying geomembrane coverage. Geophysics 89 (4), E151–E164.
Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykow, M.A., Kwasigroch, A., Majek, K., 2022. Deep learning-based waste detection in natural and urban environments. Waste Manag. 138, 274–284.
Martinho, E., 2023. Electrical resistivity and induced polarization methods for environmental investigations: an overview. Water Air Soil Pollut. 234(4):215, 1–53.
Mary, B., Sottani, A., Boaga, J., Camerin, I., Deiana, R., Cassiani, G., 2023. Non-invasive investigations of closed landfills: An example in a karstic area. Science of the Total Environment 905, 167083.
Melegari, D., De Donno, G., 2024. September. ERT and TDIP Survey for Mapping of Leachate Plumes: Application to a MSW Landfill in Italy Vol. 2024(1, 1–5.
Mendel, J. M., Hagras, H, Tan, W., Melek, W., Ying, H., 2014. Introduction To Type-2 Fuzzy Logic Control: Theory and Applications. 1-376, Wiley-IEEE Press.
Mukherjee, S., Mukhopadhyay, S., Hashim, M.A., Gupta, B.S., 2015. Contemporary Environmental Issues of Landfill Leachate: Assessment and Remedies. Crit. Rev. Environ. Sci. Technol. 45 (5), 472–590.
Munir, M.T., Li, B., Naqvi, M., 2023. Revolutionizing municipal solid waste management (MSWM) with machine learning as a clean resource: Opportunities, challenges and solutions. Fuel 348, 128548.
Niu, B., Feng, Q., Yang, J., Chen, B., Gao, B., Liu, J., Li, Y., Gong, J., 2023. Solid waste mapping based on very high resolution remote sensing imagery and a novel deep learning approach. Geocarto Int. 38 (1), 2164361.
Oldenburg, D.W., Li, Y., 1994. Inversion of induced polarization data. Geophysics 59 (9), 1327–1341.
Piegari, E., De Donno, G., Melegari, D., Paoletti, V., 2023. A machine learning-based approach for mapping leachate contamination using geoelectrical methods. Waste Manag. 157, 121–129.
Powers, D.M., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. International Journal of Machine Learning Technologies 2 (1), 37–63.
Sharma, K., Sood, M., 2024. Monitoring, classification and analysis of waste disposal sites using Machine Learning. Procedia Comput. Sci. 235, 1558–1567.
Slater, L.D., Lesmes, D., 2002. IP interpretation in environmental investigations. Geophysics 67 (1), 77–88.
Soupios, P., Ntarlagiannis, D., Sengupta, D., Agrahari, S., 2017. Characterization and monitoring of solid waste disposal sites using geophysical methods: current applications and novel trends. In: Sengupta, D., Agrahari, S. (Eds.), Modelling Trends in Solid and Hazardous Waste Management. Springer, pp. 75–103.
Sun, X., Qian, X., Nai, C., Xu, Y., Liu, Y., Yao, G., Dong, L., 2023. LDI-MVFNet: A Multi-view fusion deep network for leachate distribution imaging. Waste Manag. 157, 180–189.
Torres, R.N., Fraternali, P., 2021. Learning to Identify Illegal Landfills through Scene Classification in Aerial Images. Remote Sens. (Basel) 13 (22), 4520.
Wang, J., Qiao, Z., 2024. A comprehensive review of landfill leachate treatment technologies Front. Environ. Sci. 12. https://doi.org/10.3389/fenvs.2024.1439128.
Xia, W., Jiang, Y., Chen, X., Zhao, R., 2022. Application of machine learning algorithms in municipal solid waste management: A mini review. Waste Manag. Res. 40 (6), 609–624.
Xia, T., Meng, J., Ding, B., Chen, Z., Liu, S., Titov, K., Mao, D., 2023. Integration of hydrochemical and induced polarization analysis for leachate localization in a municipal landfill. Waste Manag. 157, 130–140.
Xu, A., Chang, H., Xu, Y., Li, R., Li, X., Zhao, Y., 2021. Applying artificial neural networks (ANNs) to solve solid waste-related issues: A critical review. Waste Management 124, 385–402.
Yu, S., Ma, J., 2021. Deep learning for geophysics: Current and future trends. Reviews of Geophysics 59 (3), e2021RG000742.
Zhang, Z., Yuan, Z., Hu, L., Coulon, F., Liu, H., Cheng, Z., Wu, S., Lui, Q., Pi, X., Wu, W., Fei, X., 2024. Comprehensive geophysical, geotechnical, and geochemical assessments of an offshore landfill in Singapore. J. Hazard. Mater. 480, 135908.