



SAPIENZA
UNIVERSITÀ DI ROMA

Advanced Geophysical Imaging of Landfills with ERT/TDIP Data: Full-Waveform Processing, Stochastic Inversion and Clustering Analysis

Ph.D. course in Environmental and Hydraulic Engineering (XXXVIII cycle)

Ph.D. candidate:
Davide Melegari

Tutor
Prof. Giorgio De Donno

Coordinator
Prof. Paolo Monti

Academic Year 2025/2026

**Advanced Geophysical Imaging of Landfills with ERT/TDIP Data:
Full-Waveform Processing, Stochastic Inversion and Clustering Analysis**

PhD thesis. Sapienza University of Rome
© 2025/2026 Davide Melegari. All rights reserved.

Author's affiliation:

“Sapienza” University of Rome – DICEA, via Eudossiana 18, Rome, Italy

Author's email: davide.melegari@uniroma1.it

Abstract

Municipal solid waste landfills are complex systems in which leachate accumulation can generate both environmental and geotechnical hazards. In this context, Electrical Resistivity Tomography (ERT) and Time-Domain Induced Polarization (TDIP) provide complementary sensitivity to pore-fluid conductivity and interfacial polarization processes, thus offering significant potential for non-invasive landfill monitoring. However, the practical use of these methods is still limited by three major issues: the poor quality of TDIP data, the lack of rigorous uncertainty quantification in local deterministic inversions, and the difficulty of integrating multiple geoelectrical outputs into a single, operationally meaningful interpretation.

This thesis addresses these limitations through three complementary research lines. The first focuses on the improvement of TDIP data quality by means of full-waveform acquisition and advanced signal processing specifically designed for 100% duty-cycle measurements, enabling a full-decay forward modelling and inversion. The second investigates uncertainty analysis through a global inversion framework for joint ERT/TDIP data, with the aim of moving beyond single deterministic models toward ensembles of plausible solutions and associated statistical indicators. The third develops machine-learning clustering approaches for the integrated interpretation of multiparameter tomographic models, progressively moving from hard clustering to target-oriented fuzzy mapping.

Within the broader framework of environmental engineering, these developments are relevant because they contribute to the identification, delineation, and monitoring of leachate-related targets in complex landfill systems, thus supporting more informed site characterization and management decisions. Rather than proposing a single unified workflow from the outset, the thesis develops three distinct but converging methodological directions, providing both advances in applied geophysics and practical tools for more reliable landfill monitoring and management.

Contents

Abstract	i
1 Introduction	1
2 Background	3
2.1 State of the art of electrical methods in landfills: strengths and sources of interpretative ambiguity	3
2.2 Electrical properties	4
2.3 Electrical resistivity tomography (ERT) and time-domain induced polarization (TDIP)	7
2.4 Paper #1: From full-waveform data processing to a full decay forward modelling and inversion	12
2.5 Paper #2: Global inversion strategies	14
2.6 Paper #3, #4 and #5: Machine-learning techniques for integrated interpretation	15
3 From full-waveform data processing to a full decay forward modelling and inversion	17
Paper #1: Full-waveform electrical data processing and inversion for imaging MSW landfills: from denoising to full-decay inversion	18
3.1 Abstract	18
3.2 Introduction	18
3.3 Acquisition and processing of full-waveform data	21
3.3.1 Site description and data acquisition	21
3.3.2 100% duty-cycle full-waveform data processing	23
3.3.2.1 Harmonic suppression and spike removal	23
3.3.2.2 Drift correction	24
3.3.2.3 Stacking and gating	26
3.4 Full-decay data inversion	27
3.4.1 Modelling and inversion algorithms	27
3.4.2 Inversion results	30
3.5 Discussion	33
3.6 Conclusions	35
3.7 Final remarks	36

4	Global inversion strategies	41
	Paper #2: Global joint inversion of ERT and TDIP data using very fast simulated annealing: uncertainty quantification and application to landfill characterization	42
4.1	Abstract	42
4.2	Introduction	42
4.3	Method	44
4.3.1	Forward modelling	44
4.3.2	Inversion procedure	45
4.3.2.1	VFSA inversion	45
4.3.2.2	Inversion parameters and <i>a priori</i> information	46
4.3.2.3	Uncertainty estimation	48
4.3.2.4	Petrophysical transformation	49
4.4	Results	49
4.4.1	Synthetic example	49
4.4.1.1	Model geometry and inversion parameters	49
4.4.1.2	Inversion results and uncertainty assessment	51
4.4.2	Field case study: municipal solid waste (MSW) landfill	51
4.4.2.1	Site description and data acquisition	51
4.4.2.2	Inversion results and uncertainty assessment for the 2019 dataset	53
4.4.2.3	Inversion results and uncertainty assessment for the 2025 dataset	54
4.4.2.4	Comparison of petrophysical models	57
4.5	Discussion	58
4.5.1	Uncertainty assessment and posterior analyses	58
4.5.2	Comparison with alternative stochastic inversion approaches	59
4.5.3	Implications for environmental geophysics	59
4.5.4	Petrophysical interpretation	60
4.6	Conclusions	61
4.7	Final remarks	62
5	Machine-learning techniques for integrated interpretation	65
5.1	Paper #3: A machine learning-based approach for mapping leachate contamination using geoelectrical methods	66
5.1.1	Abstract	66
5.1.2	Introduction	66
5.1.3	Methods	68
5.1.3.1	Geophysical imaging	69
5.1.3.1.1	Forward modelling	69
5.1.3.1.2	Data inversion	70
5.1.3.2	K-Means clustering	71
5.1.4	Site location and geophysical measurements	72
5.1.5	Inversion results	73
5.1.6	Cluster analysis results	74
5.1.7	Discussion	76
5.1.8	Concluding remarks	77

5.2	Paper #4: Fuzzy Mapping of Integrated Multi-Parameter Tomographic Data to Detect and Localize Geophysical Targets	79
5.2.1	Abstract	79
5.2.2	Introduction	79
5.2.3	Method	81
5.2.3.1	ERT and IP: field measurements and data inversion	81
5.2.3.2	FCM clustering	82
5.2.4	Results	84
5.2.4.1	Site description and field survey	84
5.2.4.2	Synthetic data	86
5.2.4.3	Field data	88
5.2.5	Discussion	88
5.2.6	Conclusion	92
5.3	Paper #5: An integrated study of hard and soft cluster analyses for detecting leachate in a MSW landfill site using geoelectrical data . .	94
5.3.1	Abstract	94
5.3.2	Introduction	95
5.3.3	Material and Methods	96
5.3.3.1	Site description	96
5.3.3.2	Geophysical data acquisition and processing	97
5.3.3.3	Proposed approach integrating hard and soft clustering	98
5.3.3.3.1	K-means	99
5.3.3.3.2	Fuzzy C-means	100
5.3.4	Results	101
5.3.4.1	Geophysical models	101
5.3.4.2	Cluster analysis	103
5.3.5	Discussion	104
5.3.6	Conclusions	110
6	Conclusion	113
	References	115

List of Figures

2.1	Main conduction pathways and EDL polarization in a non-metallic porous medium (Flores-Orozco et al., 2020).	7
2.2	Example of representation of observed data through the apparent resistivity pseudo-section (ERT tomography of a MSW landfill, Rennes FR).	8
2.3	Injected current and measured potential for the 100% duty cycle waveform (Olsson et al., 2015).	9
2.4	Voltage charge curve after current switch-on.	10
2.5	Example of representation of observed data through the apparent integral chargeability (TDIP tomography of a MSW landfill, Rennes FR).	11
3.1	Workflow of the proposed procedure from full-waveform TDIP acquisition (100% duty-cycle) to petrophysical interpretation.	21
3.2	Orthophoto of the study area showing the four ERT/TDIP profiles (G2, G4, G6 and G9). Yellow circles mark electrode positions; the subset used for the detailed results is highlighted by a black outline. Orange circles indicate the five piezometers (P1–P5). The inset shows a field photograph of the landfill terraces.	22
3.3	Examples of waveform denoising. (a) Harmonic suppression: raw (grey) and filtered (blue) potentials; dashed vertical lines indicate the time-gating boundaries. (b) Spike removal using a Hampel filter: outliers (squares) detected outside the adaptive bounds (lower/upper limits) are replaced to obtain the despiked trace (green); circles indicate the updated series after replacement.	23
3.4	Drift analysis and correction for a representative waveform. (a) Drift removal using a Cole–Cole drift model fitted to selected end-of-pulse subsets (blue markers); the reported slopes refer to the fitted drift before/after correction. (b) Residuals of the drift fit for the selected subsets, with an example of only accepted (good-fitting) cases; the resulting STD_{drift} is reported. (c) Final stacked and gated IP decay after drift correction, with total uncertainty (error bar) and STD_{gating}	24

3.5	Example of the processed dataset for profile G4. (a) Observed apparent resistivity pseudo-section ρ_a^{obs} . (b) Observed apparent chargeability for gate #10, M_{a10}^{obs} ; square markers highlight measurements extracted from negative-polarity transients (N). (c) Quality metric V_{STD} (%) used for data screening. (d) Set of processed decay curves (coloured), with discarded curves shown in light grey, after stacking and time gating.	28
3.6	Full-decay inversion results for profile G4. (a) DC resistivity (Ωm). (b) Intrinsic chargeability m_0 (mV/V). (c) Relaxation time constant τ (s). (d) Frequency exponent c (dimensionless). Model regions not shown correspond to low-coverage areas (according to the adopted coverage cut-off value). (e) Examples of observed (symbols) and predicted (solid lines) decay curves for four selected quadrupoles (IDs reported in the legend).	32
3.7	3D visualization of the inverted models for profiles G2, G4, G6 and G9. (a) DC resistivity ρ . (b) Intrinsic chargeability m_0 . (c) Relaxation time τ . (d) Frequency exponent c . Models are displayed only within the high-coverage region defined by the adopted cut-off value; axes are in UTM coordinates and elevation.	38
3.8	Comparison between a-posteriori integral chargeability obtained from (a–d) a conventional two-step inversion and (e–h) the full-decay inversion. Panels correspond to the four profiles (G2, G4, G6 and G9). Shaded regions indicate areas below the adopted coverage cut-off value.	39
3.9	Water content (a–d) and CEC (e–h) models retrieved with the relation of Revil et al. (2020) using inverted parameters from full-decay inversion of full-waveform data. Panels correspond to the four profiles (G2, G4, G6 and G9). The piezometric levels logged in wells are superposed with white filled rectangles.	40
4.1	Synthetic model: (a) true resistivity model, (b) mean resistivity model, (c) minimum misfit resistivity model, (d) true chargeability model, (e) mean chargeability model, (f) minimum misfit chargeability model. The true position of the bottom liner is marked with a black solid line, as in the case of the true position of the leachate level.	50
4.2	Synthetic model: uncertainty estimation in percentage values (a) and absolute values (b) for resistivity, uncertainty estimation in percentage values (c) and absolute values (d) for chargeability, (e) posterior correlation matrix (elements are numbered by rows downward starting from the upper-left corner), (f) error function vs. number of iterations for the 10 independent runs. The true position of the bottom liner is marked with a black solid line, while the true position of the leachate level with a black dashed line.	52
4.3	Study area.	53

4.4	2019 field dataset: (a) apparent resistivity pseudo-section, (b) mean resistivity model, (c) minimum misfit resistivity model, (d) apparent chargeability pseudo-section, (e) mean chargeability model, (f) minimum misfit chargeability model. The estimated liner position is marked with a white solid line. The P1 and P2 wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.	55
4.5	2019 field dataset: uncertainty estimation in percentage values (a) and absolute values (b) for resistivity, uncertainty estimation in percentage values (c) and absolute values (d) for chargeability, (e) posterior correlation matrix (elements are numbered by rows downward starting from the upper-left corner), (f) error function vs. number of iterations for the 10 independent runs. The estimated liner position is marked with a black solid line. The wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.	55
4.6	2025 field dataset: (a) apparent resistivity pseudo-section, (b) mean resistivity model, (c) minimum misfit resistivity model, (d) apparent chargeability pseudo-section, (e) mean chargeability model, (f) minimum misfit chargeability model. The estimated liner position is marked with a white solid line. The P1 and P2 wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.	56
4.7	2025 field dataset: uncertainty estimation in percentage values (a) and absolute values (b) for resistivity, uncertainty estimation in percentage values (c) and absolute values (d) for chargeability, (e) posterior correlation matrix (elements are numbered by rows downward starting from the upper-left corner), (f) error function vs. number of iterations for the 10 independent runs. The estimated liner position is marked with a black solid line. The wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.	57
4.8	Petrophysical models achieved at the landfill site using the relations after Revil et al. (2020): (a) 2019 water content, (b) 2019 CEC, (c) 2025 water content, (d) 2025 CEC. The estimated liner position is marked with a black solid line. The P1 and P2 wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.	58
5.1	Sketch of the traditional and ML based approaches. The red dotted line encloses the proposed procedure.	69
5.2	Map of the study area for the case study (landfill in central Italy).	72
5.3	Resistivity (a-d) and chargeability (e-h) models for the case study. Absolute errors are 2.2%, 4.2%, 7.4% and 4.5% for resistivity models and 1.1 mV/V, 2.1 mV/V, 3.7 mV/V, and 4.6 mV/V for chargeability models. The piezometric levels in wells (white filled areas) are superposed to the models.	73

5.4	Results of the cluster analysis. Panel (a) : explained variance as a function of the number of clusters. The arrows indicate the best choice for K. Panel (b): scatterplot of $\log_{10}$ -transformed geoelectrical data in the normalized space, where clusters are marked by different colours. In each cluster, the location of the centroid is shown by a blue cross.	75
5.5	Comparison between normalized chargeability models (a-d) and integrated sections after cluster analysis (e-h) for the case study. The piezometric levels are superposed to the models.	76
5.6	Aerial map of the investigated landfill with the ERT/IP lines and the selected line L2 used for the detailed analysis.	85
5.7	Synthetic simulation. The geophysical target (leachate accumulation zone) is marked by the white trapezoid. White pixels indicate no data coverage or unreliable values excluded from the clustering analysis. (a) True models. (b) Inverted models.	86
5.8	FCM clustering results for the synthetic model: (a) 3D $\log_{10}$ parameter space, (b) clustered section, and (c) membership section for the cluster associated with the geophysical target.	87
5.9	Field-data inversion results along line L2: (a) resistivity, (b) chargeability, and (c) normalized chargeability. The leachate levels logged in wells are shown by the white-filled portions of the black rectangles. White pixels indicate no data coverage or unreliable values excluded from clustering.	89
5.10	FCM clustering results for the field data: (a) 3D $\log_{10}$ parameter space, (b) clustered section, and (c) membership section of the cluster associated with the geophysical target. The well data are superposed to the sections.	90
5.11	Normalized CH and BIC indices used to evaluate the optimal number of clusters for the analyzed examples. The dashed vertical line marks the preferred solution for the field case.	90
5.12	Horizontal profile of the membership values extracted at 458 m a.s.l. from the membership section associated with the geophysical target. The profile is compared with the available well information.	92
5.13	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).	97
5.14	Absolute error committed for each measurement due to the linearization of the curvilinear profiles over the studied landfill.	99
5.15	Sketch of the proposed procedure to identify leachate accumulation zones through hard and soft clustering starting from geoelectrical tomographic data.	100

5.16	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 on the models and their diameters are exaggerated for clarity.	102
5.17	(a) K-means clustering in the normalized $\log_{10}$ space, where centroid positions are marked with blue crosses. The dimension of the M_N axis is doubled with respect to the other two axes for clarity. (b) 3D view of the clustered landfill model. The piezometric levels logged in wells (light blue filled rectangles) are superposed on the models. . . .	104
5.18	Results of the soft-cluster analysis: (a) scatterplot of cluster no. 10 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, while the filled squares show the boundary values with membership lower than 0.5. (b) Degree of membership related to cluster no. 10. (c) 3D view of cluster no. 10, where the yellow points are the extreme values and the purple points are the low-membership boundary values. (d) 3D view of the membership degree belonging to cluster no. 10. The light-blue filled rectangles indicate the piezometric levels logged in wells. The dotted blue line represents the inferred preferential leachate pathway.	105
5.19	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, and (d) fuzzy membership. The dashed lines indicate the mean values, which are also superposed on the plots.	107
5.20	Vertical profiles of geophysical parameters ρ and M (a, c), together with 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.	109

Chapter 1

Introduction

Waste management is one of the main challenges facing contemporary society and despite the progress of the circular economy, land disposal sites are currently the most common destination for waste globally. The main cause of contamination associated with Municipal Solid Waste (MSW) landfills is related to leachate, which is a liquid formed when rainwater filters through waste placed in a landfill. When this liquid comes into contact with buried waste, it draws out chemicals or constituents from them. This potentially highly polluting liquid can vary in composition over time and depending on the type of waste, but typically contains high concentrations of dissolved organic matter, inorganic chemicals and heavy metals. In addition to polluting the soil and groundwater systems due to leakage from the landfill, the leachate can also generate geotechnical instability in the case of waste disposal sites located on slopes. For these reasons, leachate must be properly managed, monitored and collected, in order to maintain its level at the bottom as low as possible.

Among near-surface geophysical techniques, electrical methods, such as electrical resistivity tomography (ERT) and time-domain induced polarization (TDIP), have repeatedly demonstrated potential for landfill investigations, even though the interpretation of electrical models in landfills is often ambiguous: the detected anomalies may have multiple competing explanations (e.g., leachate, moisture, clayey cover, metallic inclusion, heterogeneity of waste composition), and the inversion process itself can introduce non-uniqueness, resolution loss, and uncertainty that are rarely communicated in a form directly usable for decision-making (De Donno et al., 2024).

This thesis addresses the above challenges through three complementary research lines, each designed to overcome a specific limitation that currently affects geoelectrical investigations in landfill environments.

First, a methodological line is devoted to improving the quality and reliability of TDIP data, with particular emphasis on full-waveform acquisition and processing strategies aimed at mitigating noise, spikes, drift, and other waveform distortions. In this thesis, this line is not an end in itself: the processing stage is specifically developed to produce waveform data of sufficient quality for full-decay inversion, so that the TDIP data can be exploited more consistently than in conventional gated approaches.

A second research line focuses on inversion uncertainty, exploring global inversion strategies to move beyond single deterministic solutions and provide a more robust

characterization of plausible subsurface models.

Third, an interpretation line addresses the need for more objective and operationally usable outputs, through multiparameter machine-learning clustering methods that simplify the interpretation of ERT/TDIP-derived models and provide confidence-aware target localization.

Each research line is represented by a dedicated publication (or set of publications, in the case of the clustering topic). Accordingly, this thesis uses them as the core scientific contributions, while expanding on the broader theoretical framework, the links between research lines, and the methodological aspects that are only partially addressed in individual articles. For paper-specific details (e.g., full methodological derivations, parameter settings, and complete result sections), the reader is referred to the corresponding chapters/publications.

Rather than presenting these components as a fully unified workflow from the outset, this thesis develops and validates them as three distinct but converging directions. Their progressive integration is a key perspective for future work, with the long-term goal of building a complete processing–inversion–interpretation framework in which advanced polarization modelling (including full-decay and global inversion approaches) can be directly linked to environmental parameter estimation (e.g., saturation-related indicators) and machine-learning-assisted interpretation.

The remainder of the thesis is organized according to the three research lines introduced above, after this introduction (Chapter 1):

- Chapter 2 presents the state of the art of electrical methods for landfill characterization and provides the theoretical background that is not covered in the appended papers.
- Chapter 3, corresponding to the Paper #1, deals with the full-waveform processing and full-decay inversion framework.
- Chapter 4, corresponding to the Paper #2, focuses on global inversion of ERT and IP data, including motivation, principles, and implementation.
- Chapter 5 addresses machine-learning techniques for integrated interpretation: hard clustering (Paper #3), soft clustering with synthetic analysis and cluster-number selection (Paper #4), soft clustering on target-oriented field data (Paper #5).
- Chapter 6 summarizes the main contributions, discusses limitations and open challenges (e.g., single workflow, time-lapse monitoring), and outlines future work.

Part of the research activities underlying this PhD was conducted during a visiting period (March–June 2025) at Géosciences Rennes, CNRS – Université de Rennes (France). The visit focused on advanced TDIP characterization of an MSW landfill through full-decay forward modelling and inversion for leakage detection, and on the development of a dedicated pole–dipole-based investigation line aimed at improving sensitivity to leakage-related targets. The results of this visiting activity are not reported in this thesis; they are mentioned here to contextualize the broader research framework and collaborations.

Chapter 2

Background

2.1 State of the art of electrical methods in landfills: strengths and sources of interpretative ambiguity

Applied geophysics represents an effective tool for the accurate characterization of landfill properties. Its relevance is indirectly acknowledged in Article 13 of the D.Lgs. 36/2003 on landfills, which regulates leachate management and monitoring activities. At the European level, landfill management and monitoring are framed by Council Directive 1999/31/EC and its later amendment Directive (EU) 2018/850, in coordination with the general principles established by Directive 2008/98/EC on waste. The landfill operator is required to demonstrate compliance with regulatory requirements, including the assessment of leachate flows and volumes. Such an assessment may be supported by spatial screening carried out using geophysical techniques. A variety of geophysical methods can be employed for these purposes, mainly including electrical, electromagnetic (EM) and seismic methods (De Donno et al., 2024).

Among geophysical methods, electrical techniques have been increasingly used for landfill characterization (Soupios and Ntarlgiannis 2017; Nguyen et al. 2018; Juarez et al. 2023). These methods have proven to be powerful tools for this purpose, as the strong electrical contrasts between waste, leachate, and surrounding formations make the electrical properties an appropriate proxy for landfill studies. In addition, electrical prospections offer several advantages compared to other techniques and direct investigations, such as non-invasive survey, cost-effectiveness, rapid data collection, high resolution capabilities, spatial coverage. The acquisition of tomographic high-resolution data also reaches significant depths and enables the assessment of both resistive –with the so-called electrical resistivity tomography (ERT) method– and capacitive behaviour –using the time domain induced polarization (TDIP) method– of the waste mass and the surrounding formations. Overall, these methods cover the aspects and issues concerning landfill characterization and management such as leachate detection, waste characterization, characterization of cover material, settlement and erosion detection, long-term monitoring (De Donno et al., 2024).

In fact, ERT and TDIP methods have complementary sensitivity to landfill-relevant processes. On the one hand, ERT primarily reflects bulk electrical conduction controlled by fluid conductivity, saturation, porosity and temperature. On the other,

TDIP reflects charge storage and relaxation phenomena associated with electrical double layers, membrane polarization, and interfacial effects; in the anthropogenic environment, polarization can be influenced by metallic elements, organic matter, microbial activity, and saline leachate (Binley and Slater, 2020). IP information is therefore essential for landfill monitoring, as it helps reduce interpretative ambiguity. In fact, different materials may exhibit similar resistivity values, while displaying a distinct polarization behavior, or vice versa, thereby enhancing the reliability and reducing the ambiguity of the final geophysical reconstruction.

However, the joint interpretation of ERT–TDIP in landfills still leaves residual ambiguities. In particular, the relationship between highly conductive areas and chargeability responses remains difficult to interpret, and the optimal strategy for combining the two datasets is still an open issue. Furthermore, landfill environments amplify the intrinsic difficulties of geophysical interpretation. Waste bodies are heterogeneous, anisotropic, and often time-variable, while electrical responses may be dominated by high noise levels arising from localized conductive objects or by acquisition artefacts.

These residual ambiguities motivate the main three aspects addressed in this thesis: (i) TDIP data quality, tackled through full-waveform acquisition and processing; (ii) model uncertainty, obtained through global inversion; and (iii) interpretability of multiparameter outputs, achieved via clustering-based integration into a single model for target delineation.

2.2 Electrical properties

The electrical resistivity ρ [Ωm], as well as its reciprocal electrical conductivity σ [S/m], is a physical property of material and it is defined as the capability of a medium to resist to an electric current flow I . Therefore, by indirectly inferring from other measurements, it is possible to evaluate the capability of a medium to be a good conductor or not. For this reason, resistivity is a key parameter in the context of MSW landfills. Because leachate is typically rich in dissolved ions, it commonly exhibits low resistivity, which contrasts with materials typically located in landfills (i.e., unsaturated waste, geomembranes, such as high-density polyethylene - HDPE), thus it is possible to identify it on the basis of ERT models.

In porous media such as waste masses, resistivity reflects the combined effect of several conduction mechanisms. Specifically, conduction is mainly due to the motion of ions (electrolytic or ionic conduction, σ_f) or, more rarely, to the motion of electrons as charge carriers (electronic or ohmic conduction, σ_m). In applied geophysics, ρ is typically interpreted as bulk (electrolytic) conduction, i.e., the motion of free ions within the pore fluid that provides a low-resistance pathway for current flow. An important exception in landfills is the presence of metallic waste, which can introduce strong electronic conduction pathways that can locally dominate the measured response, potentially masking the ionic conduction of the porous matrix and complicating the interpretation of resistivity anomalies. Archie (1942) empirically demonstrated that the ratio between the bulk (volumetric) resistivity and the water resistivity (ρ_w) of which the rocks and soils are saturated remains constant:

$$\frac{\rho}{\rho_w} = \text{const} = F, \quad (2.1)$$

where the constant F , named formation factor, is an intrinsic characteristic of each rock or soil type, expressed as:

$$F = a\Phi^{-m}, \quad (2.2)$$

where Φ is the porosity (given by the ratio between pore volume V_v and bulk volume V_T), while a (tortuosity factor) and m (cementation exponent) are two constants to be determined experimentally as they vary depending on the type of rock or soil. In case of partially saturated porous media, eq. (2.1) is then modified as follows:

$$\rho = a\Phi^{-m}S^{-n}\rho_w, \quad (2.3)$$

where S is the degree of saturation (expressed as the ratio between water volume V_w and pore volume V_v), and n is an experimental constant referred to as saturation exponent. Archie's law relies on the assumption that conduction is mainly electrolytic through pore fluids and that the solid matrix is essentially non-conductive. These conditions are often only partially met in practice; clay-related surface conduction and strong heterogeneity can cause significant deviations. Archie-type relationships should therefore be regarded as first-order guidance and must be applied with caution in landfill environments.

In addition to the bulk conduction that has been described previously, soils may also show other conduction mechanisms. For environmental engineering applications, a relevant conduction mechanism emerging in a porous medium is the one associated with the so-called Electrical Double Layer (EDL) (Binley and Slater, 2020). For structural reasons, on the surfaces of discontinuity grain/fluid of particles, an accumulation of negative charges is usually observed. Hence, positive dissolved ions are adsorbed on surface in order to restore local neutrality, forming a first layer (called Fixed or Stern layer), characterized by a high concentration of cations. More externally, weaker chemical bonds are established, generating a second layer (named Diffuse or Gouy layer), much thicker than the Fixed layer. The sum of these two layers gives the EDL, which generates an additional surface conduction (or electromigration) described below.

Furthermore, under the action of an external electric field, soil may act as a capacitor, which means that it is able to be polarized. This effect, known as Induced Polarization (IP), is characterized by the displacement of charged elements in the subsoil in response to an external electric field. The polarization mechanisms differ depending on the type of medium and on the frequency of excitation (ω_0) in which the phenomenon is activated, or its inverse, named relaxation time (τ_0), which is the time taken by the medium to return to the equilibrium condition after the electric field has stopped (Binley and Slater, 2020). Low polarization phenomena ($\omega_0 < 10^6$ Hz) are associated with the polarization of the pore/grain system of a saturated porous

material; therefore it is related to the peculiar geometry or composition of the porous medium. Although different polarization mechanisms as membrane or interfacial polarization can also be non-negligible, in landfill environments, the measured polarization is often governed by EDL-controlled interfacial processes. They are caused by surface charging and counter-ion dynamics on charged interfaces, generated by the highly reactive surfaces of the waste matrix (mineral fines and organic matter), by biogeochemical processes such as microbial/biofilm activity and by ion-exchange/sorption reactions that modify EDL chemistry (e.g. Kessouri et al. 2019; Rahmani et al., 2025).

Therefore, in order to represent the IP contribution, the measured electrical conductivity generated by the EDL can be expressed as a complex property (σ_s^* , defined here as frequency-independent for the sake of simplicity), which can be split into a real part (σ_s'), that measures the surface conduction (or electromigration) strength generated by EDL, and an imaginary component (σ_s''), representing any temporary, reversible charge storage mechanisms (Binley and Slater, 2020):

$$\sigma_s^* = |\sigma_s| e^{i\varphi} = \sigma_s' + i\sigma_s'', \quad (2.4)$$

where $|\sigma_s|$ is the surface conductivity magnitude and φ is the phase, which represents the lag of the induced electric field with respect to the injected current. The magnitude is related to the real and imaginary parts of the complex conductivity as follows:

$$|\sigma_s| = \sqrt{(\sigma_s')^2 + (\sigma_s'')^2}, \quad (2.5)$$

On the other hand, since at low frequencies used in this application the charge storage is small relative to electromigration, the phase can be written as:

$$\varphi = \tan^{-1} \left[\frac{\sigma_s''}{\sigma_s'} \right] \approx \left[\frac{\sigma_s''}{\sigma_s'} \right], \quad (2.6)$$

which states that the phase angle is proportional to the polarization strength of the medium divided by the electromigration strength of the medium.

The whole conductivity (Fig. 2.1) is therefore a complex parameter, both including electrolytic conductivity (σ_f) and complex surface conductivity (σ_s^*) representing conduction and polarization occurring in the EDL:

$$\sigma^* = \sigma_f + \sigma_s^* = \sigma_f + \sigma_s' + i\sigma_s'', \quad (2.7)$$

since the solid-matrix contribution (σ_m) for low-frequency investigation and non-metallic materials is commonly treated as negligible compared with pore-fluid conduction and EDL effects. In the end, the real and imaginary components of the measured complex conductivity are related to the electrolytic and surface conductivity mechanisms as follows:

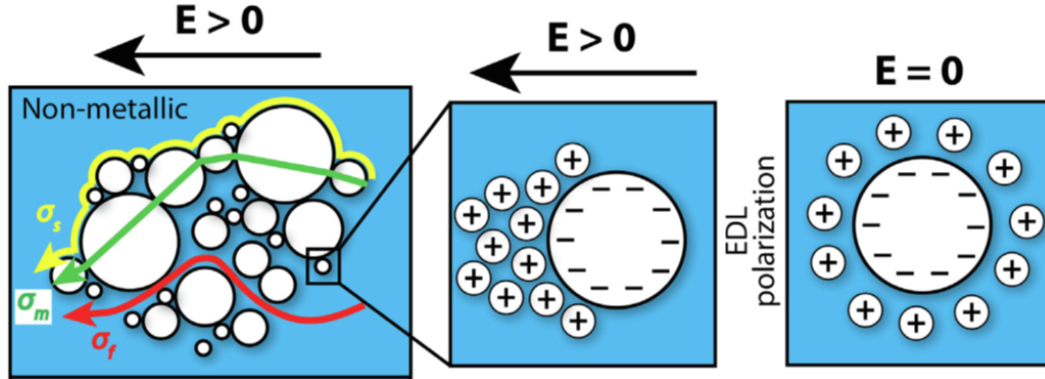


Figure 2.1. Main conduction pathways and EDL polarization in a non-metallic porous medium (Flores-Orozco et al., 2020).

$$\sigma' = \sigma_f + \sigma'_s, \quad (2.8a)$$

$$\sigma'' = \sigma''_s. \quad (2.8b)$$

Equation (2.8a) describes that real conductivities result from the parallel electrolytic and surface conduction pathways within the porous medium. At the same time, eq. (2.8b) highlights that the imaginary conductivity only senses the surface pathway and is therefore independent of the electrolytic conduction pathway. Moreover, generalizing the eq. (2.6), being the phase an index of the IP response, if σ' (i.e., volumetric and/or grain/fluid interface conductivity) increases because of a higher saturation of the medium, the IP effect decreases even if polarization is present. It should also be acknowledged that this conceptual trend, namely the possible decrease of the IP response in fully leachate-saturated zones, is not consistently applied, as simplified approaches were adopted for clustering (e.g., Papers #3 and #4) for the sake of simplicity.

2.3 Electrical resistivity tomography (ERT) and time-domain induced polarization (TDIP)

In field surveys, the complex conductivity $\sigma^*(\omega)$ is inferred from controlled-source geoelectrical measurements acquired either in the frequency-domain (FD), through sinusoidal excitation at discrete (low) frequencies, or in the time-domain (TD), through the analysis of the transient response following current injection. Therefore, while frequency-domain measurements aim at reconstructing $\sigma^*(\omega)$ explicitly across a range of frequencies, time-domain response must be transformed or interpreted to retrieve equivalent spectral information. Field-scale comparisons have shown that TD methods can offer improved robustness and signal-to-noise ratio under realistic survey conditions, often leading to more reliable parameter estimation than FD ones, particularly in electrically noisy environments and for low-chargeability targets (Maurya et al., 2018).

Among the TD techniques, the most popular is the ERT method, where data acquisition is often carried out grounding simultaneously along an alignment a large

number of electrodes, all connected through multi-conductor cables to a resistivity-meter (powered by an external battery). The resistivity-meter is a single device that combines together both the ammeter and the voltmeter, being capable of measuring, for each quadrupole i , both the current intensity (I_i) and the potential difference (ΔV_i^{OBS}). Consequently, the measured apparent resistivity $\rho_{a,i}^{OBS}$ is simply given by:

$$\rho_{a,i}^{OBS} = K_i \frac{\Delta V_i^{OBS}}{I_i} \quad \text{with } i = 1, 2, \dots, N_Q, \quad (2.9)$$

where K_i is the geometric factor and N_Q is the number of observations (quadrupoles) (Binley and Slater, 2020). The values in eq. (2.9) are simply provided by measuring the steady-state transmitted current (I_i) and the resulting voltage differences (ΔV_i^{OBS}) on electrode pairs at different distances from the current injections, both recorded by the resistivity-meter. A convention is generally adopted to represent the measurements, consisting of reporting the measured apparent resistivity values on a vertical cross-section, the so-called "pseudo-section" (Fig. 2.2). Such representation of apparent resistivity indicates only in a qualitative way the trend of resistivity variations in the subsoil both horizontally and vertically, since the measured resistivity is only an "apparent" value (Binley and Slater, 2020).

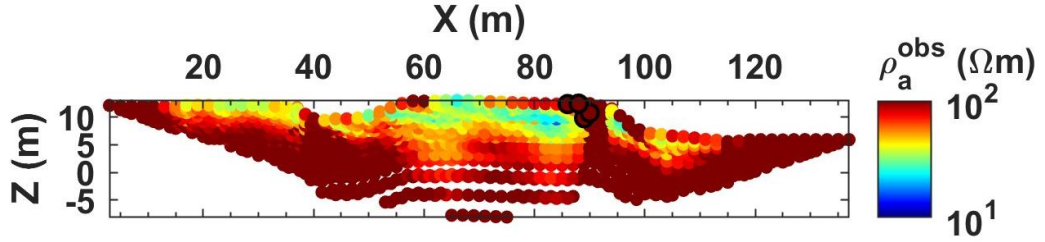


Figure 2.2. Example of representation of observed data through the apparent resistivity pseudo-section (ERT tomography of a MSW landfill, Rennes FR).

In fact, the investigated medium is not homogeneous; therefore, when a current I is injected through the electrodes, the observed potential difference ΔV^{OBS} is a function of not only the distance between the electrodes, but also the resistivity and relative position of the different media (layers, inclusions, etc.) with respect to the electrodes. Therefore, the resistivity value from equation (2.9) does not coincide with the "true" value of resistivity of the subsoil, which cannot be directly measured. Consequently, the correct approach for retrieving the real resistivity model of the area is to perform data inversion. The inversion procedure is an iterative mathematical process that involves comparing the apparent resistivity predicted by a given resistivity model with the corresponding observations. For this reason, the knowledge of the electrostatic field equation and of its solution for the potential is required (see Chapter 3).

Nevertheless, when polarization is present, ERT measurements alone are insufficient because the voltage response becomes time dependent. TDIP measurements explicitly record this transient behaviour, enabling the estimation of polarization-related parameters in addition to resistivity. Under the assumption of low-frequency

polarization, the IP phenomenon is often modelled through the chargeability m defined as:

$$\sigma_{\infty} = \sigma_0 (1 - m) , \quad (2.10)$$

where $0 \leq m < 1$, σ_{∞} is the DC (direct-current) electrical conductivity (the DC electrical resistivity is $\rho_{\infty} = 1/\sigma_{\infty}$) and σ_0 is the instantaneous conductivity. This formulation corresponds to the linear approximation of the polarization effect, valid for typical near-surface TDIP surveys and small chargeability values (Seigel, 1959). So far, the acquisitions of ERT and TDIP data have been performed sequentially, using the same equipment and multi-electrode configuration, but the current must be injected for a longer time than in a stand-alone ERT survey, so as to activate the polarization mechanisms. To understand it better, it is useful to look at the raw full wave-forms data rather than the sampled data.

Recent developments in geoelectrical equipment (e.g. ABEM Terrameter LS2, IRIS I-FullWaver, etc.) have given access to advanced acquisition systems, enabling a 100% duty cycle current waveform (Fig. 2.3). The instruments allow for faster acquisitions and a better Signal-to-Noise Ratio (SNR), which are critical parameters in landfill environments (Olsson et al., 2015).

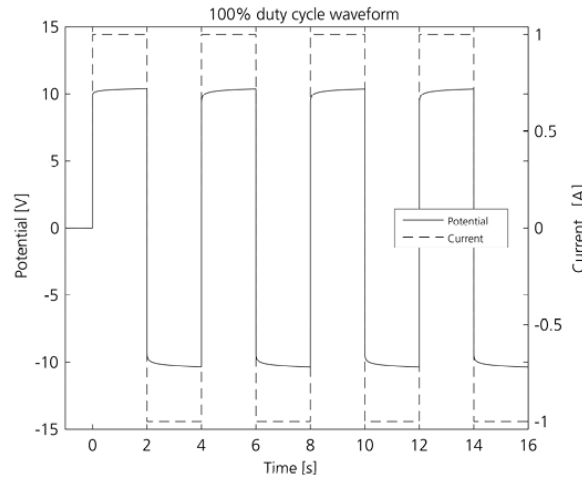


Figure 2.3. Injected current and measured potential for the 100% duty cycle waveform (Olsson et al., 2015).

When a current is activated in case of polarizable media, a potential increase due to polarization effects (charging phase) is observed (Fig. 2.4), starting from the instantaneous voltage immediately after the activation of the current (V_0) up to the so-called stationary voltage (V_{∞}) at the end of the pulse (for simplicity V is used instead of ΔV to define a voltage difference).

We can now define:

$$M_a = \frac{V_{\infty} - V_0}{V_{\infty}} , \quad (2.11)$$

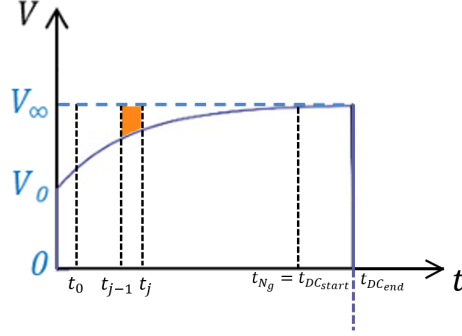


Figure 2.4. Voltage charge curve after current switch-on.

where M_a [dimensionless or mV/V], named (apparent) chargeability, is the model parameter of the IP effect as it expresses the contribution of the polarization effects into the recorded potential. In other words, M_a quantifies the capability of a material to store electrical charges when an injected current is applied. The apparent chargeability M_a typically ranges from a few mV/V to hundreds mV/V. Therefore, the capacitive effect is orders of magnitude weaker than the resistive one.

As well as for resistivity, only an apparent chargeability (M_a) can be experimentally observed (Fig. 2.5) and an inversion process is required for retrieving the "true" chargeability. The problem would be closed if both the potential V_∞ and V_0 are measurable, so that the apparent chargeability is obtained by applying (2.11), but this is not the case. In fact, even if the stationary voltage (V_∞) is actually measurable with good accuracy once the pulse duration has been set correctly, the measurement of the instantaneous voltage (V_0) is not practically feasible. This is due to the fact that the measurement is affected by the disturbances that occur immediately after the current switch-on (i.e., spikes) and by electromagnetic (EM) induction phenomena, arising when there is a variation of the electric field with time. However, induction phenomena generally decay quickly after the current switch-on (1 - 20 ms). As a consequence, measurements shall be performed with a certain delay time $t_0 = t_{dly}$, given by the time interval in which induction phenomena occur, in order to neglect them (Binley and Slater, 2020). Therefore, also to avoid the effects of spikes, an integral measurement of the chargeability is often computed, by appropriately sampling (generally in a logarithmic way) the IP curve through a finite number $j = 1, 2, \dots, N_g$ of time windows (gates) $\Delta t_j = t_j - t_{j-1}$ (Fig. 2.4). For every j -th gate, the area above the charge curve is calculated, normalizing it to the maximum value of potential calculated in the same time interval. This gives the observed partial chargeability:

$$M_j^{OBS} = \frac{\int_{t_{j-1}}^{t_j} (V_\infty - V_j) dt}{V_\infty(t_j - t_{j-1})}, \quad (2.12)$$

Therefore, in practice the observations are given, for each quadrupole $i = 1, 2, \dots, N_Q$, by the integration over the entire sampling range (from $t_0 = t_{dly}$ to t_{N_g}), yielding the so-called integral (total) apparent chargeability $M_{a,i}^{OBS}$, given by the weighted average of the partial chargeabilities (M_j^{OBS}) for their sampling

windows:

$$M_{a,i}^{OBS} = \frac{\sum_{j=1}^{N_g} M_j^{OBS} \Delta t_j}{\sum_{j=1}^{N_g} \Delta t_j}. \quad (2.13)$$

Thus, instead of the instantaneous apparent chargeability, given by (2.11), which cannot be measured practically, the $M_{a,i}^{OBS}$ value is used, even if it is not its actual reproduction. This implies that in the reconstruction of the inverted model, the chargeability values may not correspond to the absolute (real) ones, thus only relative comparisons can be made (Binley and Slater, 2020). In the following, the term chargeability will therefore be specified as theoretical, apparent, integral or intrinsic, depending on the context.

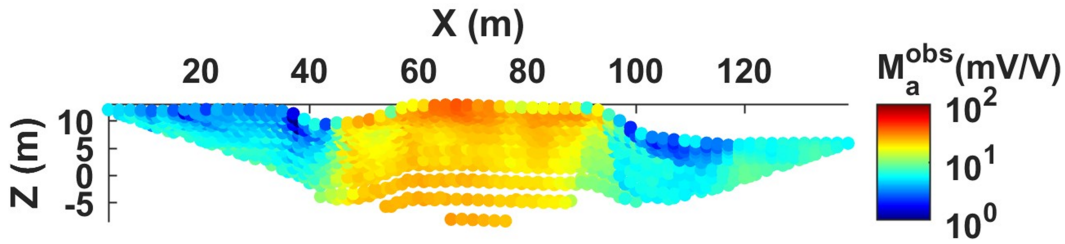


Figure 2.5. Example of representation of observed data through the apparent integral chargeability (TDIP tomography of a MSW landfill, Rennes FR).

As a final remark for landfills, other than being distinguished as a medium with high conductivity due to a great ions' concentration, the leachate may be also associated with a decreasing chargeability. In fact, as already mentioned, when the saturation increases a lot, the IP response decreases, but without collapsing to zero. Therefore, the leachate is characterised by a chargeability that tends to be lower with respect to zones where dry waste and high biochemical activity are present. This trend is not universal and must be interpreted in the context of waste composition, salinity, and local biogeochemical conditions.

Finally, in Papers #1 and #2, the inverted geoelectrical parameters are also transformed into water (leachate) content and cation exchange capacity (CEC) through petrophysical relationships derived from the joint use of DC conductivity and normalized chargeability (Revil et al., 2020). In this approach, the conductive term mainly reflects pore-fluid transport and saturation conditions, whereas the polarization term provides information on surface-related electrochemical properties of the waste matrix. As a result, water content is estimated as a saturation-dependent variable, while CEC is used to describe the ion-exchange capacity of the medium. These transformations extend the interpretation from geophysical observables to environmentally meaningful parameters, although they must be applied with caution in landfill environments because of strong heterogeneity and variable fluid chemistry.

2.4 Paper #1: From full-waveform data processing to a full decay forward modelling and inversion

As introduced before, to the new generation of resistivity-meters, the complete waveform of the injected current and measured potential signal can now be recorded. Full-waveform data are very useful because they allow to increase the reliability and quality of data through the possibility of performing precise signal processing, especially for TDIP measurements which are more sensitive to noise than resistivity measurements. Moreover, thanks to developments in time-domain measuring procedures, the use of a current injection via a 100% duty cycle waveform makes it possible to retrieve the IP information during the on-time (i.e., when the potential increases during the acquisition). Therefore, by acquiring TDIP measurements while the current is being transmitted, without off-time periods (i.e., using a 50% duty cycle), the acquisition time is effectively halved. It is also demonstrated that the use of a 100% duty cycle waveform generates a superposition of the IP charge-up and discharge. Consequently, another advantage of this acquisition mode is that the SNR can increase up to a factor of 2 with respect to a 50% duty cycle, resulting in an improved IP data quality (Olsson et al., 2015). While taking these significant developments into account, drawbacks still remain for TDIP measurements. Specifically, there are limitations that compromise the correct extraction of information from TDIP data in the field. In practice, the most critical issues are related to: (i) noise components (e.g., harmonics and transient spikes), (ii) slow time-dependent distortions such as electrode polarization drift, and (iii) waveform distortions related to coupling and superposition effects that complicate the separation of the true polarization response from acquisition artefacts.

The first research line of this thesis is documented in Paper #1, which develops an advanced processing-and-inversion workflow for 100% duty-cycle full-waveform TDIP data acquired in MSW landfills. The paper combines harmonic suppression, spike removal, Cole–Cole-based drift correction, stacking/gating, and automatic quality control to obtain reliable decays for full-decay spectral inversion, and compares the results with the classical two-step integral-chargeability approach. For this reason, full-waveform acquisition and processing are not only a technological upgrade, but also a methodological prerequisite: it enables waveform-level quality control and dedicated signal processing steps that are essential to obtain reliable IP decays. This is particularly important when the subsequent interpretation relies on full-decay inversion, which is inherently sensitive to the transient shape and therefore requires a processed dataset. Once the full waveforms have been quality-controlled and processed, the resulting decays can be used in the inversion framework. Before this step, it is necessary to establish a consistent forward model for dispersive electrical behaviour. This behaviour has been thoroughly explained previously in terms of complex conductivity (eq. 2.7). To model it through parametrization, the Cole-Cole model (Cole & Cole, 1941; Pelton et al., 1978) is one of the simplest and most used ones. It can be expressed in terms of complex resistivity, as a function of: ρ_0 , that denotes the direct-current (DC) resistivity [Ωm]; m_0 , which is the intrinsic chargeability [dimensionless or V/V] and is different from the integral chargeability previously introduced; τ_ρ , that represents the time constant [s] and

equal to the inverse of the angular frequency ω at which the imaginary resistivity $\rho''(\omega)$ reaches its minimum; c , that is the frequency exponent [dimensionless] and defines the sharpness/flatness of the spectrum. This parameterization is widely used because of its simplicity and physical interpretability, but in inversion it may suffer from parameter correlations (especially between m_0 and c , as demonstrated by Fiandaca et al. 2018), which can reduce model stability and increase non-uniqueness. Nevertheless, it provides a practical and well-established basis for forward modelling of TDIP data, and it is therefore adopted here as the starting spectral description of the dispersive electrical response.

Although the landfill-oriented Paper #1 reports the workflow and its application, a compact summary of the underlying full-decay forward/inversion formulation is provided here to make the thesis self-contained.

Since the IP data are acquired in TD the complex impedance must be transformed back into a transient (time-dependent) response. The TD step-off response is obtained through an inverse Fourier sine transform (De Donno and Cercato 2023):

$$V_q^{S-OF}(t) = Z_q(0) - \frac{2}{\pi} \int_0^\infty \Im m \left(-\frac{Z_q^*(\omega)}{i\omega} \right) \sin(\omega t) d\omega, \quad (2.14)$$

where $Z_q^*(\omega)$ is the complex electrical impedance and $Z_q(0)$ is the DC response for $\omega = 0$. Then, the TD step-off response is evaluated via a Fast Hankel transform for fixed log-spaced times (0.01–8 s), after interpolating the frequency response to 10 points per decade with cubic splines to improve transform accuracy.

Full-decay inversion is performed within a Gauss–Newton framework using the above-mentioned Cole-Cole parameterization (model $\mathbf{m}$) of the polarization response, with $\log_{10}$ -transformed model parameters ($\mathbf{x}$) and inequality constraints $[a, b]$ to improve numerical stability and keep the solution within physically plausible ranges (Kim and Kim 2011). The modified Gauss–Newton model update ($\delta \mathbf{x}^{(n)}$) is written in the transformed parameter space as:

$$\delta \mathbf{x}^{(n)} = \left[(\mathbf{P}^T \mathbf{J})^T \mathbf{W}^T \mathbf{W} (\mathbf{P}^T \mathbf{J}) + \beta^{(n)} \mathbf{R}^T \mathbf{R} + \lambda^{(n)} \mathbf{I} \right]^{-1} \times \left\{ (\mathbf{P}^T \mathbf{J})^T \mathbf{W}^T \mathbf{W} [\mathbf{d} - g(\mathbf{m}^{(n)})] - \beta^{(n)} \mathbf{R}^T \mathbf{R} (\mathbf{m}^{(n)} - \mathbf{m}^0) \right\}, \quad (2.15)$$

being $\mathbf{J}$ the Jacobian matrix, $\mathbf{P}^T = \frac{\partial \mathbf{m}}{\partial \mathbf{x}} = \frac{\log 10 (b-\mathbf{m})(\mathbf{m}-a)}{(b-a)}$, $\mathbf{W} = \text{diag}\left(\frac{1}{s^2}\right)$ the data weight matrix (s are the observed standard deviations), $\mathbf{R}$ the smoothness matrix (second-difference operator using the four neighbours of the k -th block), $\mathbf{I}$ the identity matrix, $\mathbf{d}$ the observed and $g(\mathbf{m}^{(n)})$ the predicted data vectors and β and λ regularization parameters at the n -th iteration; if the error increases, an inner loop explores alternative λ values to recover a decreasing-misfit update.

Further details regard the computation of the Jacobian in the frequency domain followed by the same time transform used for the forward response. In particular, the frequency-domain Jacobian is derived by matrix multiplication of the FD sensitivity with the partial derivative of the complex resistivity with respect to the CC parameters using the chain-rule (De Donno and Cercato, 2023) as:

$$J_{q,k}^{FD} = \frac{\partial Z_q^*(\omega)}{\partial \rho_k^*(\omega)} \cdot \frac{\partial \rho_k^*(\omega)}{\partial m_k}. \quad (2.16)$$

In the present thesis, these formulations provide the methodological basis just for the full-decay inversion workflow discussed in Paper #1, while later chapters focus on uncertainty and interpretability aspects through global inversion and machine-learning-based integration.

2.5 Paper #2: Global inversion strategies

As discussed in the first section, widely used geoelectrical investigations in landfill environments are affected not only by data-quality issues (particularly for TDIP), but also by a second major limitation: the lack of rigorous uncertainty quantification in inversion results. In fact, ERT/TDIP models are obtained through deterministic local inversion schemes, resulting in a single best-fitting model under given regularization assumptions. Local methods (such as the Gauss–Newton approach presented in Paper #1) iteratively update a starting model and converge toward a nearby minimum of the objective function, depending on initialization, regularization choice, and the local curvature of the misfit surface (Loke and Barker, 1996; Sen and Stoffa, 2013). Even if these methods are computationally efficient, they do not provide probabilistic information about the range of models compatible with the data.

In order to overcome these limitations, in Paper #2, a global inversion strategy for ERT/IP data is introduced. These methods differ in how they explore the model space, exploring a much broader portion of the parameter space through stochastic sampling. For this reason, they reduce dependence on subjective choices and enable a statistically meaningful assessment of uncertainty. Despite these advantages, global inversion has rarely been applied to 2D TDIP problems, due to their computational cost. In fact, this routine needs to repeat forward modelling over thousands to millions of trial models, becoming expensive when the inversion is performed in a joint ERT/TDIP framework. To reduce this computational cost, model-reduction strategies have been applied to probabilistic inversion of ERT data (e.g. Fernández-Martínez, 2015; Aleardi et al., 2022), but this solution may affect spatial resolution and can bias the resulting uncertainty estimates.

Considering a good balance between computational efficiency and global search capability, the Very Fast Simulated Annealing (VFSA) method (Ingber, 1989) is adopted as global inversion algorithm. This method is applied on an unreduced full 2D mesh, using parallel independent runs, logarithmic parameterization and inequality constraints to retrieve a probabilistic description of model variability (Cercato, 2011; Sen and Stoffa, 2013). The choices of inversion parameters and the operation of the algorithm itself are described in detail in the Paper #2.

Although this approach has previously been applied to ERT inversion (Sharma, 2012), to our knowledge this represents one of the first applications on an unreduced full 2D mesh in a joint ERT/TDIP framework. In this context, global inversion is particularly relevant because interpretative decisions (e.g., identifying leachate accumulation zones) can carry operational consequences. In fact, interpretations

are often based on anomalies whose geometry, depth, and intensity may be strongly influenced by smoothness constraints, poor sensitivity at depth, and limited data coverage. By contrast, the global inversion strategy proposed explicitly estimates model uncertainty from the ensemble of accepted solutions, thereby helping to distinguish robust features from weakly constrained regions and reducing the risk of over-interpretation in environmental and geotechnical decision-making.

In summary, while the state of the art confirms the diagnostic potential of ERT/TDIP for landfill monitoring, it also highlights a persistent methodological gap: uncertainty is rarely quantified in a rigorous statistical sense, especially in 2D landfill-focused applications. Paper #2 is intended as a direct response to this gap, by introducing a computationally demanding but methodologically robust global inversion approach that complements the deterministic workflows commonly used in practice and provides an explicit uncertainty layer for interpretation.

2.6 Paper #3, #4 and #5: Machine-learning techniques for integrated interpretation

Beyond data quality (Section 2.4) and uncertainty estimation (Section 2.5), a third limitation often encountered in landfill geophysics is interpretative complexity: ERT/TDIP inversion typically yields separate geophysical sections (such as ρ and m_0 inverted models) that are commonly interpreted by qualitative comparison, leaving residual ambiguity and making the results difficult to communicate in an operational way to other professionals involved in the landfill management. This gap is particularly relevant in such environments, where heterogeneous waste composition, clay-rich barriers, metallic debris, and spatially variable saturation can produce electrical signatures that are not uniquely attributable to leachate. As a result, even when anomalies are detected, translating multi-parameter geophysical outputs into a clear delineation of leachate-related zones remains difficult.

To address these interpretative ambiguities, Paper #3, #4 and #5 present a novel approach for data integration where geoelectrical models are processed and analyzed using machine learning (ML)-based algorithms. ML techniques have paved the way for advancements in many different fields, since it allows the analysis of large datasets and develops intelligent systems that are capable of adapting and improving over time (Zhang et al., 2022). It can be broadly classified into two main categories: unsupervised learning and supervised learning. In this Thesis, the focus is on unsupervised clustering, that makes it possible to identify areas with similar electrical properties, i.e., corresponding to a plausible accumulation of leachate, by assigning them to different groups (clusters), as well as to provide a confidence proxy via membership. This results in an integrated and quantitative representation of the subsurface that can be interpreted more consistently for leachate mapping.

A first step in this direction is provided by hard clustering, where each cell is assigned to a single cluster, yielding a unique "integrated model" that replaces separate sections with a classified image of the subsurface. The K-Means algorithm (Lloyd, 1982) is used, which has numerous advantages, including reduced computational complexity and very simple implementation. At the same time, the choice of the initial position of the centroids and the number of clusters with which to

divide the dataset can greatly influence the effectiveness of the analysis conducted. This approach is deeply described in Paper #3, which proposes distance-based color coding to facilitate the identification of leachate-accumulation zones in a less ambiguous manner than visual inspection of individual parameter maps.

However, hard clustering intrinsically produces sharp boundaries and does not provide an explicit measure of how reliable the assignment is at each location. For this reason, in this research line a second algorithm used is the Fuzzy C-Means algorithm (Bezdek, 1984), which allows the confidence of cluster assignment to be evaluated locally. In fact, this algorithm belongs to the soft clustering group, where each cell is associated with membership values (between 0 and 1) that quantify the degree of belonging to each cluster. In practice, membership values are used here as a practical proxy for classification confidence, not as posterior inversion uncertainty. This is the key methodological advance documented in Paper #4, where hard clustering is used to first localize the cluster(s) most compatible with leachate, and fuzzy C-means is then applied to evaluate membership within and around these clusters, improving the delineation of leachate boundaries and enabling comparisons with well observations. In addition, this work addresses a critical practical aspect that is often left implicit in clustering-based interpretation, namely the selection of an optimal number of clusters. This is investigated through objective cluster-validity criteria (Calinski–Harabasz and BIC) and synthetic tests, providing a reproducible basis to balance over-segmentation before transferring the methodology to field data.

Finally, Paper #5 generalizes the approach into a target-oriented fuzzy mapping strategy for integrated multiparameter tomographic datasets, emphasizing that in many environmental and engineering applications the goal is not a full stratigraphic reconstruction, but the robust detection and localization of a specific target (leachate accumulation in this case). Paper #5 formalizes the use of fuzzy membership as a reliability indicator, combines geoelectrical data with well data to validate the clustering models. In particular, the ultimate goal is to improve the ability of clustering to delineate the most saturated zones on the basis of chargeability-related information.

The present Thesis complements Papers #3–#5 by explicitly framing clustering as a response to a persistent state-of-the-art gap in landfill investigations: geoelectrical imaging often delivers valuable but fragmented information, whereas operators require a compact, quantitative, and communicable output.

Chapter 3

**From full-waveform data
processing to a full decay
forward modelling and inversion**

Paper #1: Full-waveform electrical data processing and inversion for imaging MSW landfills: from denoising to full-decay inversion

Davide Melegari^{1,*} and Giorgio De Donno¹

¹ Dipartimento di Ingegneria Civile Edile e Ambientale, “Sapienza” Università di Roma, Rome, Italy

* Corresponding author: davide.melegari@uniroma1.it

The manuscript presented in this chapter is currently under review for publication in the Journal of Applied Geophysics and is reported here in its complete form.

3.1 Abstract

Geophysical imaging and monitoring of municipal solid waste landfills are pivotal for assessing leachate accumulation, which can contaminate the surrounding environment and/or trigger instability for slope landfills. Electrical resistivity tomography and time-domain induced polarization have proven effective for these purposes, although they are often strongly affected by noise in such complex environments. In this study, we developed an advanced workflow for processing and inversion of full-waveform electrical data with a 100% duty cycle, designed to enhance the signal-to-noise ratio and facilitate the application of full-decay inversion. The processing workflow includes harmonic suppression, spike removal, drift correction based on Cole–Cole modelling, automatic gate filtering, and rigorous quality control. Filtered data are inverted using a spectral Cole–Cole formulation within a 2.5D framework and compared with results obtained from the classical two-step integral chargeability inversion. The spectral inversion enables the a-posteriori computation of an unbiased integral chargeability, which assists in estimating water content and cation exchange capacity through well-established petrophysical relationships. This approach is applied to a landfill located on a steep slope in Central Italy. The results reveal a three-layer landfill structure, with conductive zones associated with leachate accumulation and highly chargeable zones related to biologically active waste masses. The results are validated by leachate levels logged in two wells. The study demonstrates that the advanced processing of 100% duty-cycle TDIP data, combined with full-decay inversion, can significantly enhance the reliability of IP interpretation in landfill environments, supporting improved leachate monitoring and risk assessment.

Keywords: electrical resistivity tomography, time-domain induced polarization, full-waveform data, full-decay inversion, leachate.

3.2 Introduction

Waste management represents a critical issue of contemporary society, and despite the progress of the circular economy, land disposal sites are currently the most common destination for waste globally (Kaza et al., 2018). One of the main environmental problems in municipal solid waste (MSW) landfills is leachate, which may cause

contamination of the surrounding environmental compartments (groundwater and soil) in case of leakage. For imaging and monitoring of leachate accumulations, geophysical surveys offer the advantages of being non-invasive and cost-effective, allowing us to obtain a conceptual model of the investigated site (Soupios et al., 2017). Among the different geophysical techniques, direct-current (DC) electrical and electromagnetic (EM) methods are currently the most commonly used for imaging and monitoring MSW landfills (see De Donno et al., 2024 for a recent review). In particular, electrical resistivity tomography (ERT) has been widely used in the last decades because it is a high-resolution method able to take advantage of the strong contrast between leachate (highly conductive) and the surrounding environment in terms of electrical resistivity (e.g., De Carlo et al., 2013; Maurya et al., 2017). In some cases (e.g., Gazoty et al., 2012; De Donno et al., 2017), the ERT method has been coupled with the study of induced polarization (IP), which can provide complementary diagnostic information. In fact, the IP effect is not directly related to the presence of leachate but can be influenced by several factors due to different polarization phenomena, such as fluid content, interfacial processes, waste composition, biological activity, and the presence of metallic inclusions (e.g., Bückner et al., 2018, 2019; Binley & Slater, 2020; Rahmani et al., 2025). However, the IP signal is characterized by an intrinsic error considerably higher than that of resistivity, since the corresponding potential is of negligible magnitude compared to the resistive one (on the order of a few mV or hundreds of μV compared to a few V or hundreds of mV), and consequently appropriate strategies to maximize the signal-to-noise ratio (SNR) should be applied during both the acquisition and the processing phases (Dahlin et al., 2012). Traditionally, combined ERT and TDIP surveys are carried out using a 50% duty-cycle waveform for current injections, which has the theoretical advantage of resetting the potential for each pulse, but it is considerably slower due to the need to double the acquisition time. This limitation has been overcome recently by using a 100% duty-cycle waveform (Olsson et al., 2015), where the IP information is retrieved during the on-time period. Furthermore, recent developments in TDIP equipment have given access to full-waveform recordings of measured potential and transmitted current, which can enable advanced signal processing and increase the reliability and quality of the IP data; despite these improvements made in data acquisition, the correct extraction of the IP information may still be harmed by different types of noise (Olsson et al., 2016). Specifically, drift can generate a progressive variation in the signal recorded throughout the entire acquisition, thereby significantly altering resistivity and chargeability measurements. This noise is mainly due to electrochemical polarization, arising from the use of polarizable electrodes, and to current-induced polarization, since the electrodes are used both for potential readings and for current injections (Dahlin et al., 2002). Additionally, spikes caused by anthropogenic sources (i.e., electric fences) or by polarity switches, as well as harmonic noise due to power lines oscillating at a base frequency (e.g., 50 Hz), can also contribute to increasing the noise level. In any case, filtering out these noises is the only feasible option to achieve a good-quality IP dataset to be used for inversion in a complex landfill environment, since using non-polarizable electrodes or separating current and potential cables is highly time-consuming (Dahlin et al., 2012).

The filtered IP response is usually integrated within a time window extended

from tens of milliseconds (discarding the inductive effects) to some seconds (long enough to correctly sense the low-frequency capacitive effects) after the current switch of polarity, by sampling the voltage decay curve with a number of generally log-spaced gates. However, the resulting integral apparent chargeability is not strictly comparable to the theoretical apparent chargeability, since their mathematical formulations are different (Siegel, 1959), and consequently the final chargeability model will be biased. This is due to the fact that the theoretical chargeability is a ratio of voltage values taken immediately before and after the current switch-off, the latter being practically not measurable (Binley & Slater, 2020). Additionally, the IP response is compressed into a single parameter, and therefore the whole information contained in the voltage decay curve cannot be extracted. To overcome this limitation, a full-decay inversion for TDIP data has been recently developed for near-surface applications (Fiandaca et al., 2012), which requires the modelling of the entire potential decay curve and thus high-quality IP data. This inversion methodology leads to a more accurate description not only of the IP response but also of the full electrical response in terms of complex conductivity, ensuring a more correct distinction between saturated and unsaturated zones and a more accurate assessment of waste distribution (Fiandaca et al., 2013). Moreover, this approach enables a rigorous characterization of the leachate and waste-body spectral parameters, recovering a more direct link to interfacial polarization mechanisms correlated to waste biodegradation and active redox processes within leachate accumulation zones (Flores-Orozco et al., 2020). Furthermore, a-posteriori integral chargeability can be retrieved after full-decay inversion in order to apply the established petrophysical relations, yielding, for example, water content and cation exchange capacity (CEC) values (Revil et al., 2020). However, only a few examples of full-decay inversion of IP data have been applied to MSW landfills, often due to the low quality of the available IP datasets.

In light of this, the aim of this work is to develop and apply advanced geophysical processing and inversion procedures (Fig. 3.1) for landfill management, particularly oriented to:

1. develop a compact processing routine for 100% duty-cycle IP data to increase the SNR, thus enabling the application of advanced inversion approaches;
2. apply a full-decay inversion algorithm to the processed data and compare the results with the classical approach in terms of integral chargeability on the same lines;
3. achieve final petrophysical sections in terms of water content and CEC, which can be directly linked to leachate accumulation zones and waste-mass biochemical activities.

This approach is applied to the characterization of an MSW landfill located on a steep slope in Central Italy, where leachate accumulation can also lead to instability phenomena.

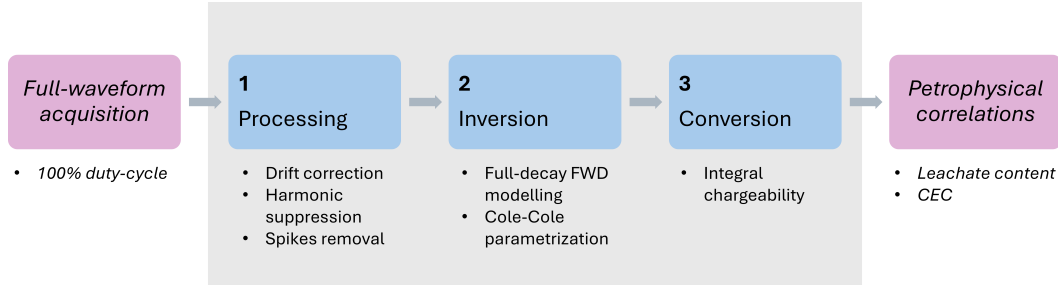


Figure 3.1. Workflow of the proposed procedure from full-waveform TDIP acquisition (100% duty-cycle) to petrophysical interpretation.

3.3 Acquisition and processing of full-waveform data

3.3.1 Site description and data acquisition

The field case is an MSW landfill built in several phases starting in the 1980s and active until ten years ago (end of waste disposal), located near a medium-sized city in Central Italy. The landfill is located on a steep slope and occupies a large portion of the hillside, approximately between 580 m (top of the slope) and 410 m (valley) a.s.l., with an average slope of 20° (Fig. 3.2). The disposal site is a controlled landfill, so it is equipped with a multi-barrier system (top and bottom liner, drainage system) to avoid leachate leakage pathways to the subsoil. For this reason, the HDPE bottom liner acts as an electrical insulator and the marly-arenaceous flysch bedrock is not expected to be detected by geophysical investigations.

A previous geophysical campaign was conducted on four profiles (G2, G4, G6, and G9 in Fig. 3.2) using a standard acquisition of resistivity and integral chargeability data with 48 electrodes and the SyscalPro (IRIS Instruments) resistivity-meter. The inverted models for resistivity and integral chargeability, presented in Piegari et al. (2023) and De Donno et al. (2025), highlighted significant leachate accumulations in the eastern part of the landfill. Although the resistivity models depicted the leachate distribution well, the IP reconstruction is affected by the simplified approach used for data acquisition, as well as the maximum depth of investigation reached.

Building on these results, a new campaign was conducted to estimate the current leachate levels after establishing a drainage system based on the findings of the previous survey. At the same time, we also aimed at improving the quality of the IP results with respect to the previous investigation, to enable the link between geophysical and petrophysical parameters. For this reason, data acquisition was carried out on the same four geoelectrical profiles as in the previous campaign. We used the ABEM Terrameter LS2 (GuidelineGeo) resistivity-meter, which supports a 100% duty-cycle acquisition of full-waveform data, providing a good signal-to-noise ratio and a faster acquisition, particularly important along such long and complex profiles. Data were acquired using a multiple-gradient protocol with 81 stainless steel electrodes spaced from 4 m to 6 m (from G2 to G9) through a 4×21 roll-along layout configuration. We acquired 1498 measurements (full-waveform current and potential signals) for each profile, with a satisfactory mean contact resistance of 749Ω . The current was injected with five pulses (four canonical pulses plus one



Figure 3.2. Orthophoto of the study area showing the four ERT/TDIP profiles (G2, G4, G6 and G9). Yellow circles mark electrode positions; the subset used for the detailed results is highlighted by a black outline. Orange circles indicate the five piezometers (P1–P5). The inset shows a field photograph of the landfill terraces.

additionally inserted by the manufacturer protocol at the beginning of the signal) having a length of $4 + 0.3$ s (4 s devoted to IP and 0.3 s to resistivity measurements, respectively) with a sampling rate of 1 kHz.

3.3.2 100% duty-cycle full-waveform data processing

Raw potential signals were processed by adapting for a 100% duty-cycle dataset the approach originally proposed by Olsson et al. (2016) for a 50% duty-cycle dataset. An example of a raw full-waveform potential signal (profile G4) affected by noise (drift, spikes and harmonic noise) is shown in Fig. 3.4a (grey line).

3.3.2.1 Harmonic suppression and spike removal

Firstly, we remove harmonic noise using a simple moving average over a 20 ms sliding window, matching the 50 Hz power-line period. Since this approach is intentional but may affect the earliest samples, we later exclude these samples from inversion. Figure 3.3a demonstrates that no additional compensation is necessary.

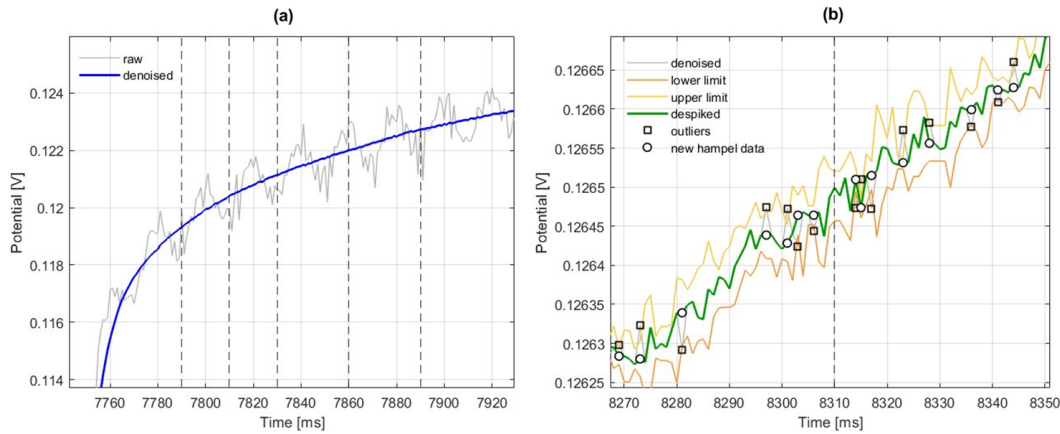


Figure 3.3. Examples of waveform denoising. (a) Harmonic suppression: raw (grey) and filtered (blue) potentials; dashed vertical lines indicate the time-gating boundaries. (b) Spike removal using a Hampel filter: outliers (squares) detected outside the adaptive bounds (lower/upper limits) are replaced to obtain the despiked trace (green); circles indicate the updated series after replacement.

Secondly, de-spiking of the data is carried out following Olsson et al. (2016) using the so-called Hampel filter (Liu et al., 2004), which is commonly used for the removal of outliers. Even in this case, spikes affect IP information more than resistivity measurements, as they can significantly contaminate the potential value obtained from the average of the samples of a short-duration gate. For each sample given as input, the Hampel function calculates the median relative to itself and its neighbours. It also calculates the standard deviation of each sample relative to the same interval in the form of the mean absolute deviation (MAD). If the sample in question differs from the median by a certain multiple of the MAD, then it is replaced by the median itself. For this application, the number of neighbours on both sides of the sample has been set to 2, and the spike-detection threshold value

to double the MAD value. An example of the identification and substitution of the spikes is shown in Fig. 3.3b.

3.3.2.2 Drift correction

To assess drift, we select three data subsets from the full-waveform pulses. This avoids harmonic-noise residuals after denoising. We compute subset samples (blue circles in Fig. 3.4a) as average values over a 20 ms fundamental period, centred every 125 ms at the end of each pulse, where the IP effect is lower, though not zero.

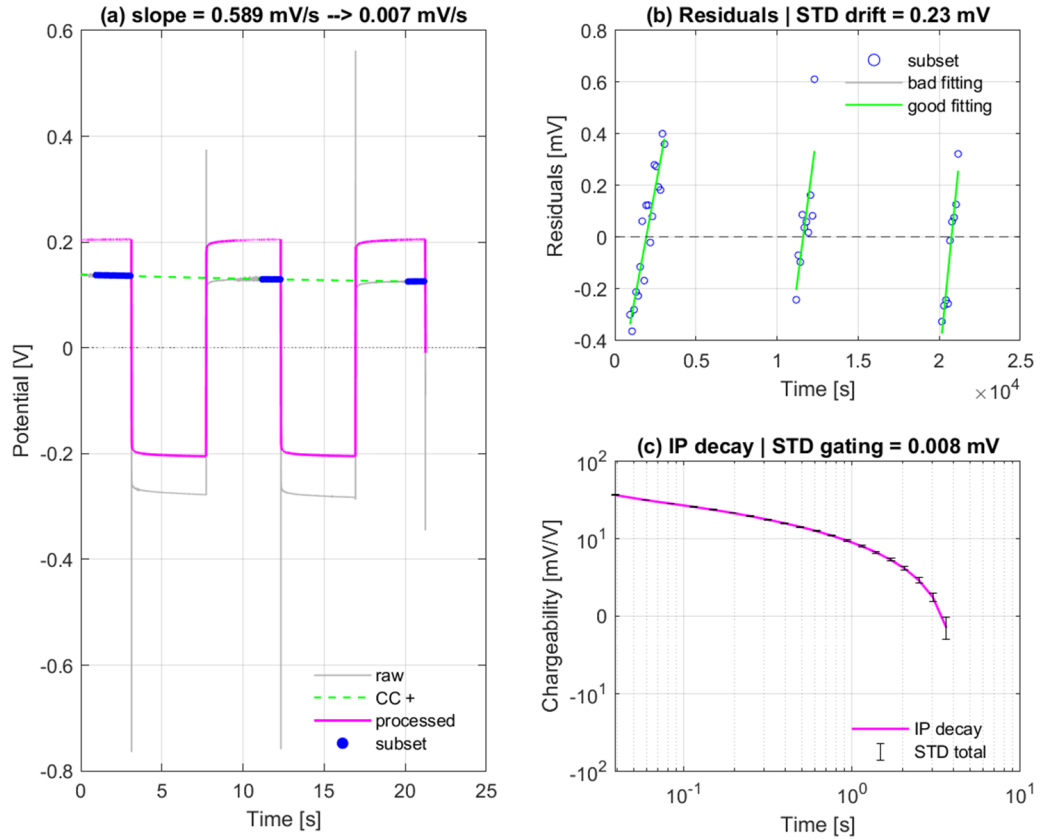


Figure 3.4. Drift analysis and correction for a representative waveform. (a) Drift removal using a Cole–Cole drift model fitted to selected end-of-pulse subsets (blue markers); the reported slopes refer to the fitted drift before/after correction. (b) Residuals of the drift fit for the selected subsets, with an example of only accepted (good-fitting) cases; the resulting STD_{drift} is reported. (c) Final stacked and gated IP decay after drift correction, with total uncertainty (error bar) and STD_{gating} .

For the first pulse, where the potential signal theoretically starts at zero, we sample the last 70% of the pulse. For other windows, we use the last 25%. We select subsets only from positive pulses (nos. 1, 3, and 5) to guarantee the same pulse polarity and apply linear interpolation to them. If the slope of the interpolated line does not exceed 0.05 mV/s (maximum drift of about 1 mV during acquisition), drift is negligible and correction is skipped. Otherwise, we fit the model proposed by Olsson et al. (2016), based on the Cole–Cole (CC) model (Pelton et al., 1978):

$$u_{\text{drift}}(n) = d + m_0 \sum_{j=0}^{100} (-1)^j \frac{\left(\frac{n}{\tau f_s}\right)^{jc}}{\Gamma(1 + jc)}. \quad (3.1)$$

where τ and c follow the standard CC notation (time constant and frequency exponent, respectively), m_0 represents the amplitude of the drift, d is added as an offset constant, f_s is the sampling frequency (1 kHz), n is the sample index and Γ is Euler’s Gamma function. Bounds reflect common ranges for the CC parameters, discarding unrealistic values: $d \in \mathbb{R}$; $m_0 \in [-1, 1]$; $\tau \in [10^{-6} \text{ s}, 10^3 \text{ s}]$; $c \in [10^{-5}, 1]$; while starting values are data-driven: $d = \text{mean}(u_{\pm})$, $m_0 = \max(u_{\pm}) - \min(u_{\pm})$, $\tau_0 = 1 \text{ s}$, and $c_0 = 0.1$.

The drift correction is then performed by fitting our subset to the CC model through three different methods: nonlinear least squares (Marquardt et al., 1963), interior-point (Byrd et al., 2000), and trust-region (Coleman et al., 1996). The best-fitting curve is that yielding the minimum root mean square error (RMSE), and we subtract it from the entire raw potential waveform (both polarities) to obtain the processed potential value. In fact, even if the drift model is estimated only from positive pulses, we assume it to be polarity-independent over the measurement window. Therefore, we obtain the processed signal (magenta solid line in Fig. 3.4a) as well as the respective residual values for each positive pulse, as differences between our fitting and potential recordings for each sample. Badly fitted curves were then removed, modifying the approach originally proposed by Fiandaca et al. (2022) for 50% duty-cycle waveforms. In our case, residuals were grouped for each pulse (Fig. 3.4b) and classified as regularly shaped if two conditions are met: (i) a polarity-consistent linear trend (positive slope correlated with the positive pulse), and (ii) a drift uncertainty in terms of standard deviation (STD_{drift} , after Olsson et al., 2016) below an acceptable threshold value of 1 mV. The motivation for these choices is that, unlike the case of 50% duty-cycle data, drift correction is performed during the charge period. Therefore, the zero baseline in our data does not reflect the true intrinsic value of the signal, but only the difference from the fitted curve. The residual slope of each accepted filtered record was also checked, and if the value exceeded the above-mentioned threshold (0.05 mV/s), the measurement was filtered out from our dataset (Table 3.1).

Table 3.1. Summary of the drift-correction workflow for each profile. “Raw data” is the number of acquired quadrupoles (excluding measurements affected by power drops). “W/o drift” indicates measurements for which drift is below the adopted threshold and therefore no drift correction is applied; percentages in this column are referred to the raw dataset.

Profile	Raw data	W/o drift	Not validated	Processed
G2	1497	280 (19%)	41 (3%)	1176 (97%)
G4	1497	291 (19%)	41 (3%)	1165 (97%)
G6	1489	293 (20%)	46 (4%)	1150 (96%)
G9	1484	355 (24%)	31 (3%)	1098 (97%)

3.3.2.3 Stacking and gating

After data processing, the IP stacked potentials are obtained from the processed signals V_{PROC} by

$$V_{\text{IP}}^{\text{STACK}} = \frac{1}{N_{\text{PULSES}}} \sum_{j=1}^{N_{\text{PULSES}}} (-1)^{j+1} [V_{\text{DC}}(j) - V_{\text{PROC}}(j)]. \quad (3.2)$$

where V_{DC} is the potential retrieved at the end of each pulse (the last 0.3 s) and $N_{\text{PULSES}} = 4$ in this work, starting from the second one. In fact, the first pulse is used only for drift estimation, and it is not included in the stacked data (same approach to compute predicted stacked data), since the potential signal starts from zero. Similarly, the DC stacked potential $V_{\text{DC}}^{\text{STACK}}$ is achieved by averaging the DC potential values in analogy to Eq. (3.2), and apparent resistivity can be derived from $V_{\text{DC}}^{\text{STACK}}$ by using the implemented approach of Fiandaca et al. (2012).

Then, the IP signals were gated using $N_{\text{GATES}} = 20$, with a log-increasing width, as reported in Table 3.2.

Table 3.2. Time-gating scheme used to extract the stacked IP decay curve.

Gate number	1	2	3	4	5	6	7	8	9	10
Gate width (s)	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.09	0.11
Gate number	11	12	13	14	15	16	17	18	19	20
Gate width (s)	0.13	0.16	0.19	0.23	0.27	0.33	0.40	0.48	0.57	0.68

We chose to start our gating with a delay of 30 ms to minimize the influence of the induction effects occurring at earlier times and the edge effects of the harmonic suppression. Following the approach proposed by Olsson et al. (2016), within each gate, we apply a Gaussian window coefficient (Harris, 1978) to highlight the local decay trend and suppress the high-frequency noise. Since the IP signal in each gate has an exponential trend, by convolving the stacked signal within the gate with the Gaussian window and fitting the resulting smoothed signal with an exponential, we are able to estimate the intra-gate noise. This applies to an ideal IP charge, whereas in the case of variations such as those found in heterodox transients (Fiandaca et al., 2022) it may also reveal a non-exponential trend. For this reason, when the exponential trend does not provide an adequate description (poor fit quality), a linear fit was employed. The gated potential $V_{\text{IP}}^{\text{GATE}}$ is then computed through the fitted curve at the gate centre and the gate noise is quantified as the RMSE between the convolution and the fit. This error is used to define the error associated with each gate ($V_{\text{STD}}^{\text{GATE}}$) as the ratio with the respective $V_{\text{IP}}^{\text{GATE}}$, achieving a measure used for quality control, which could not be estimated using the stacking procedure between pulses, given the effect of pulse superposition (Fiandaca et al., 2012). We fixed the required quality threshold equal to 5%, discarding every gate that exceeds that value. The IP dataset can also be expressed in terms of measured apparent chargeability m_a^{GATE} for each gate g by:

$$m_a^{\text{GATE}}(g) = 1000 \frac{V_{\text{IP}}^{\text{GATE}}(g)}{2V_{\text{DC}}^{\text{STACK}}}, \quad g = 1, 2, \dots, N_{\text{GATES}}, \quad (3.3)$$

where the factor 2 is justified by the structure of the 100% duty cycle, in which the polarization effect is due to the superposition of the discharge of the previous polarity and the charge of the corresponding polarity (Olsson et al., 2015). Together with the IP decay, the cumulative data standard deviation (STD) was finally calculated for each gate following the approach of Olsson et al. (2016) and displayed along the IP measurements (Fig. 3.4c).

Finally, after gating, we apply an automatic gate-filtering routine to remove the outliers (bad decay curves) in the dataset. Regarding the possible types of heterodox transient, we define as acceptable only the “P” (starting from positive values with a decreasing slope) and the “N” (starting from negative values with an increasing slope) curves (after Fiandaca et al., 2022). For P/N decays, accepted gates display neither an increase/decrease nor a steep drop (more than 25% with respect to the previous gate). Finally, we fit a single-exponential model to the saved gates, and we accept only curves displaying a good fit ($R^2 > 0.9$). If this criterion is not met, the procedure iteratively removes the first gate and recalculates the exponential model until the fitting is satisfactory. After this phase, the decay curves having more than 50% rejected gates are discarded (total number of discarded curves around 5% across the entire dataset). As an example, Fig. 3.5d reports the filtered dataset of line G4, in which it is possible to note the coherence of the IP decay curves obtained through this processing, represented together with the pseudo-sections of the observed data (Fig. 3.5a for apparent resistivity, Fig. 3.5b for apparent chargeability at gate no. 10, and Fig. 3.5c for the average of V_{STD}^{GATE} across all gates, defined as V_{STD}). After the processing phase, the data vector for inversion (**d**) is given by the DC stacked potential V_{DC}^{STACK} and the IP gated potential $V_{IP}^{GATE}(g)$ for each quadrupole measurement:

$$\mathbf{d}_q = \left[V_{DC,q}^{STACK}, V_{IP,q}^{GATE}(1), V_{IP,q}^{GATE}(2), \dots, V_{IP,q}^{GATE}(N_{GATES}) \right]. \quad (3.4)$$

3.4 Full-decay data inversion

3.4.1 Modelling and inversion algorithms

Modelling and inversion of processed data are carried out adopting the so-called 2.5D approximation, that is, assuming that conductivity is invariant in the y -direction (normal to the profile section x - z), while the electric potential is still distributed in three dimensions. In this case, assuming an external point source located at (x_S, z_S) , the resistive and capacitive response of the subsurface is described within a domain D by the Fourier-transformed Poisson complex equation (De Donno and Cercato, 2023):

$$-\nabla \cdot [\sigma^*(x, z, \omega) \nabla \phi^*(x, z, \lambda, \omega)] + \lambda^2 \sigma^*(x, z, \omega) \phi^*(x, z, \lambda, \omega) = I^* \delta(x_S) \delta(z_S), \quad \forall (x, z) \in D. \quad (3.5)$$

where σ^* is the complex electrical conductivity ($\rho^* = 1/\sigma^*$ being the complex resistivity), generally depending on the angular frequency $\omega = 2\pi f$, with f the

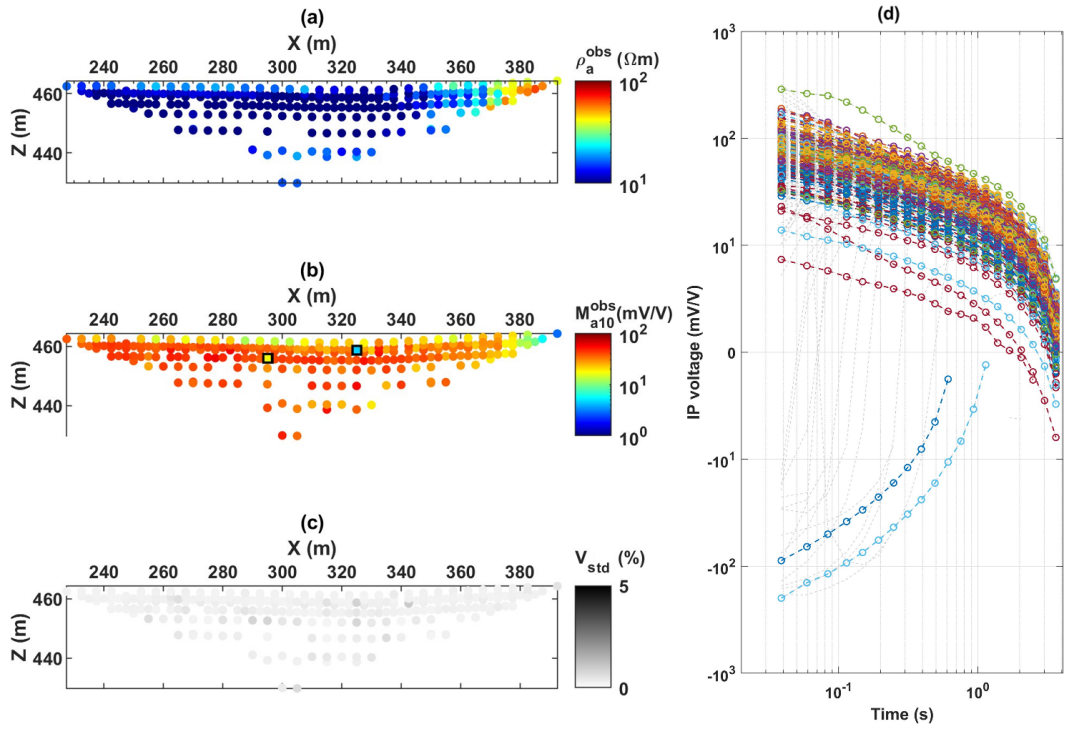


Figure 3.5. Example of the processed dataset for profile G4. (a) Observed apparent resistivity pseudo-section ρ_a^{obs} . (b) Observed apparent chargeability for gate #10, M_{a10}^{obs} ; square markers highlight measurements extracted from negative-polarity transients (N). (c) Quality metric V_{STD} (%) used for data screening. (d) Set of processed decay curves (coloured), with discarded curves shown in light grey, after stacking and time gating.

frequency, ϕ^* the complex-valued transformed electric potential, λ the transformed variable, I^* the injected current, and δ the Dirac delta function.

A complex electrical measurement of rock material provides information on both electrolytic (ohmic) conduction and polarization effects (surface conduction) and on its frequency dependence. In environmental studies conducted at low frequencies ($f < 10^3$ Hz), the dominant polarization mechanism is the polarization of ionic charges associated with the electrical double layer (EDL) at the mineral–fluid interface (Binley et al., 2005).

The frequency dependence of complex resistivity (or conductivity) is typically characterized using phenomenological models (e.g., Cole and Cole, 1941; Pelton et al., 1978). For geophysical applications, the CC model is among the most widely adopted, and it expresses the complex resistivity as

$$\rho^*(\omega) = \frac{1}{\sigma^*(\omega)} = \rho_0 \left[1 - m_0 \left(1 - \frac{1}{1 + (i\omega\tau)^c} \right) \right]. \quad (3.6)$$

where ρ_0 denotes the direct-current (DC) resistivity [Ω m], m_0 is the intrinsic chargeability [mV/V], τ represents the time constant [s], and c is the frequency exponent [dimensionless].

Equation (3.5) is solved for each frequency within the range [10^{-4} Hz, 10^{10} Hz] using logarithmic sampling with five points per decade and the finite-element method. Neumann and Dirichlet boundary conditions are applied respectively on the surface and on lateral and bottom boundaries placed sufficiently far from the source (De Donno and Cardarelli, 2017b). The complex electrical impedance, $Z_q^*(\omega_j)$, is then obtained, for each quadrupole $q = 1, \dots, N_q$ and angular frequency ω_j ($j = 1, \dots, N_f$), by computing the potential difference between the corresponding potential electrodes and normalizing it by the injected current, where N_q and N_f denote the total number of quadrupoles and frequencies, respectively. The DC response, $Z_q(0)$, is computed using the same approach for $\omega = 0$.

Since IP data were acquired in the time domain, the complex impedance is then back-transformed into a step response V_{STEP} using an inverse Fourier sine-transform of the frequency-domain response following the approach of Fiandaca et al. (2013), as implemented in MATLAB by De Donno and Cercato (2023). For a 100% duty cycle, the stack response is computed by

$$V_{\text{STACK}(j)}(t) = \sum_{m=1}^j \sum_{k=1}^2 (-1)^{m+k} V_{\text{STEP}}(t + (k-1)T + (j-m)T), \quad (3.7)$$

$$j = 1, 2, \dots, N_{\text{ST}} + 1.$$

Then, predicted potentials $V_{\text{DC}}^{\text{STACK}}$ and $V_{\text{IP}}^{\text{GATE}}(g)$ were computed following the same procedures described for stacking and gating of the observed data.

We used in this work our inversion algorithms based on a Gauss–Newton formulation and described extensively after De Donno and Cardarelli (2017b) for integral chargeability and De Donno and Cercato (2023) for full-decay inversion, respectively. The spectral inversion is performed using the CC model ($m_k = [\rho_{0k}; m_{0k}; \tau_k; c_k]$, $k = 1, 2, \dots, M$, with M the number of elements), setting inequality constraints such that $a_k < m_k < b_k$, with $a_k = (0.1 \Omega \text{ m}, 0.1 \text{ mV/V}, 0.001 \text{ s}, 0.1)$ and $b_k =$

($10^4 \Omega \text{ m}$, 500 mV/V , 10 s , 0.6). We discarded unrealistic values or values inconsistent with those expected in an MSW landfill site. Inversion is carried out using a $\log_{10}$ -transformation of the model parameters, which is particularly suitable for electrical models spanning many orders of magnitude (Kim and Kim, 2011):

$$x_k = \log_{10} \left(\frac{m_k - a_k}{b_k - m_k} \right). \quad (3.8)$$

We used the model resolution matrix (MRM) as a proxy for assessing data coverage (Menke, 2015), with different cut-off values [10^{-3} , 10^{-4} , $2 \cdot 10^{-6}$, 10^{-4}] for discarding pixels in the CC models, according to the difference in sensitivity between the four parameters (García-Fiscal and Flores, 2018).

The predicted data at the last iteration were then used to compute the a-posteriori integral chargeability values, following the approach of Fiandaca et al. (2012), in order to: (i) compare the integral chargeability derived from a rigorous spectral inversion with the one arising from a classical IP inversion approach (Oldenburg and Li, 1994); and (ii) use DC resistivity and integral chargeability as input for estimating water content θ_l and CEC through the petrophysical relations of Revil et al. (2020), as follows:

$$\theta_l = \left[\frac{1}{\sigma_l} \left(\sigma_0 - \frac{M_n}{R} \right) \right]^{1/m} \quad (3.9)$$

$$\text{CEC} = \frac{M_n}{\theta_l^{m-1} \rho_g \lambda}. \quad (3.10)$$

where M_n (in S/m) is the normalized chargeability (Slater and Lesmes, 2002), defined as the product of the chargeability and the direct-current conductivity σ_0 ; m denotes the cementation exponent, which we set equal to 2, as well as the saturation exponent n in the second Archie's law (modified after Winsauer et al., 1952); σ_l corresponds to leachate conductivity; ρ_g denotes the grain (waste) density; and $R = \lambda/B \approx 0.1 \pm 0.02$ after Revil et al. (2017), where λ is the apparent mobility of the counterions for polarization (in $\text{m}^{-2} \text{s}^{-1} \text{V}^{-1}$) and B is the apparent mobility of the counterions for surface conduction (in $\text{m}^{-2} \text{s}^{-1} \text{V}^{-1}$). The average leachate conductivity, directly logged in the wells, was equal to $1.0 \pm 0.2 \text{ S/m}$, while the waste density ρ_g was assumed constant and equal, in a first approximation, to 1500 kg m^{-3} (Steiner et al., 2022).

3.4.2 Inversion results

Although datasets were acquired on the whole G2–G9 profiles, we inverted only subsets of data pertaining to the eastern part of the landfill (see Fig. 3.2 for location), where leachate accumulation zones were already detected in the previous geophysical campaign (Piegari et al., 2023). Figures 3.6a–d present the inverted CC model for profile G4, alongside the fitting results for four selected quadrupoles (Fig. 3.6e). The inverted cross-sections are able to image the landfill site down to a maximum depth of investigation (DOI) of around 40 m, detecting a three-layer model:

1. the uppermost layer (thickness of a few metres), having high resistivity ($\rho_0 > 20 \Omega \text{ m}$) and relatively low chargeability (m_0 between 5 and 20 mV/V), related to the presence of the covering layer;
2. a conductive and chargeable middle layer (with variable thickness along the profiles due to the geometry of the landfill) related to the waste mass, where $m_0 > 100 \text{ mV/V}$ (even with significant spatial variability) and $\rho_0 < 10 \Omega \text{ m}$ (with large areas in the middle part of the sections having $\rho_0 < 5 \Omega \text{ m}$), likely due to leachate accumulation;
3. the bottom layer, which is visible as an increase of resistivity and a respective decrease of chargeability, although it falls within a low-coverage region. This electric behaviour is due to the presence of the low-permeability high-density polyethylene (HDPE) bottom liner, acting as an insulator ($\rho \sim 10^7 \Omega \text{ m}$).

In particular, the most conductive zones ($\rho_0 \sim 2 \Omega \text{ m}$), located between 250 and 350 m in the x -direction and at an elevation interval between 460 and 440 m a.s.l. (Fig. 3.6a), are associated with lower chargeability values with respect to the neighbouring zones (Fig. 3.6b).

Additionally, the CC inversion also provides τ and c models (Figs. 3.6c–d), even though they are not as well resolved as ρ_0 and m_0 . Time-constant values are mainly between 0.5 and 1.5 s, with three local anomalies where τ approaches or exceeds 2 s (Fig. 3.6c). The frequency exponent c (Fig. 3.6d) shows a spatial pattern similar to m_0 , with lower values (0.2) outside the landfill and higher values (0.6) within the waste mass, consistent with a more heterogeneous medium and more overlapping polarization mechanisms. The quality of the fitting is satisfactory, as demonstrated by the good agreement between predicted and observed decay curves, though the error increases for the weaker IP signals (i.e., green line in Fig. 3.6e) as expected.

The results of the G4 profile are confirmed by the overall 3D view of the CC models for the four profiles (Fig. 3.7), positioned to follow the potential leachate pathway from upstream (G9) to downstream (G2). The DC resistivity models (Fig. 3.7a) clearly show the transition between the various terraces of the waste-disposal site, within the three main layers that characterize the entire landfill. The conductive middle zone, located on the western side of the lines and associated with the leachate accumulation zone, narrows downstream due to the peculiar morphology of the bottom layer (impluvium), already highlighted by previous investigations (Piegari et al., 2023). Moreover, a high-resistivity area appears at the bottom of the landfill in the middle of profiles G6 and G9, likely related to an old embankment used to separate different zones during waste collection. Another anomaly is also visible in the western part of the G2 profile, where a low-conductivity shallow area appears, likely due to an inversion artefact caused by the vicinity to the embankment, where some electrodes of G2 grounded outside the landfill can interfere with the metallic reinforcements of the embankment.

The high-chargeability anomalies in the intermediate layer clearly delineate the boundaries of the waste mass, displaying a moderate biological activity (or, locally, possible metallic inclusions), with values ranging from 100 to 500 mV/V. Time constant and frequency exponent also follow the stratigraphy of the landfill, with

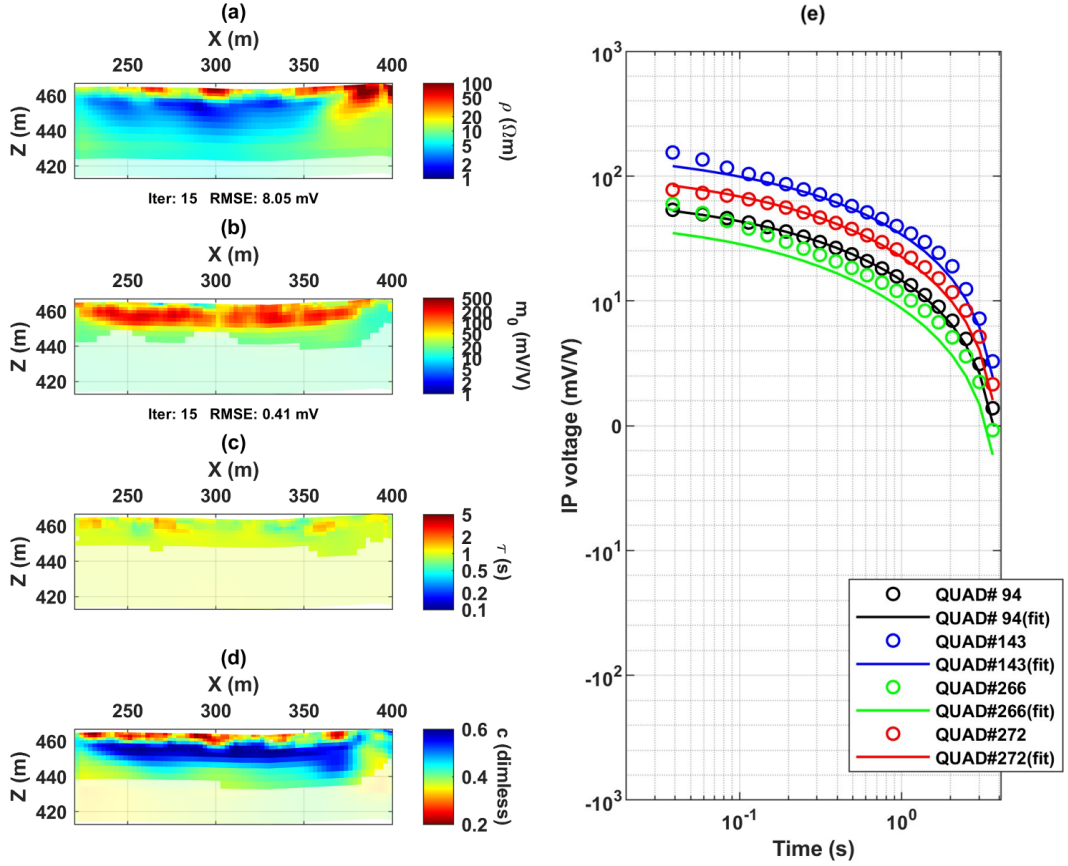


Figure 3.6. Full-decay inversion results for profile G4. (a) DC resistivity (Ωm). (b) Intrinsic chargeability m_0 (mV/V). (c) Relaxation time constant τ (s). (d) Frequency exponent c (dimensionless). Model regions not shown correspond to low-coverage areas (according to the adopted coverage cut-off value). (e) Examples of observed (symbols) and predicted (solid lines) decay curves for four selected quadrupoles (IDs reported in the legend).

the same limitations previously explained for G4, both in terms of coverage and interpretability.

After inversion we compare the integral chargeability obtained from the two-step inversion method (Figs. 3.8a–d) proposed by Oldenburg and Li (1994, III method – nonlinear approach) with the a-posteriori integral chargeability calculated after the full-decay inversion (Figs. 3.8e–h). Because the two inversion approaches employ different types of data (integral apparent chargeability and IP voltage, respectively), the RMSE values are not directly comparable. Although both methods reveal a similar layering, the chargeable intermediate zone (waste mass) appears smoother in the two-step inversion (Fig. 3.8a–d). Additionally, in the classical inversion the anomalous zone extends downward without significant variation, although deep areas fall within low-coverage regions. For the full-decay case, the waste mass is more sharply defined, as expected, and the high chargeability values associated with the waste mass are confined to regions having good coverage. In the shallow part of the landfill, the two reconstructions exhibit some common shapes and lateral variation of the IP response (i.e., at $x \approx 300$ m for line G4 in Figs. 3.8b and 3.8f), even though with higher heterogeneity for the full-decay models.

Finally, the integral chargeability derived from the full-decay inversion is used to apply the petrophysical relations described in Eqs. (3.9) and (3.10) for estimating θ_l (Figs. 3.9a–d) and CEC (Figs. 3.9e–h), respectively, linked to leachate accumulation zones and waste biodegradation activities. Regarding the CEC analysis, sections exhibit localised anomalies, which are likely associated with the strongest interfacial-polarization mechanisms caused by biogeochemical activity. At the same time, the models present higher levels of water content in the intermediate layer, as expected, which may be correlated with saturated waste zones. In fact, the high θ values (0.5–0.6) match the levels logged in wells P4 and P5 along the G4 and G6 profiles. However, the low leachate level logged in P3 (tens of centimetres) corresponds to an area with a slightly lower water-content value (0.4–0.5), though not close to zero as expected. This is likely due to a smaller leachate-accumulation zone and to the fact that the waste mass can exhibit a high moisture content due to its field holding capacity, still not contributing to a continuous flow path within the landfill (see Zhang et al., 2023, for a review).

3.5 Discussion

The presented results highlight that resistivity and chargeability convey different but complementary information in MSW landfills. Low resistivity reliably identifies leachate-accumulation zones, whereas high chargeability is primarily associated with waste presence, biogeochemical activity, and interfacial-polarization processes. The observed vertical misalignment between conductive and chargeable zones is consistent with landfill-ageing processes, leachate migration, and decreasing biological activity with depth. This behaviour is not unexpected, since the increase in volumetric conductivity observed in the presence of saturated leachate, as more ion exchange occurs, could lead to a decrease in the IP response (Abdulrahman et al., 2016). Hence, chargeability is not directly proportional to the presence of leachate, but rather to the presence of waste, as it can be influenced by several other features

such as waste composition, aggregation, biological activity, and the presence of metallic inclusions (e.g., Binley & Slater, 2020; Flores-Orozco et al., 2020). Another reason is that leachate, formed when compacted fresh-aged waste interacts with rainwater, seeps over time into the bottom layers due to gravity, while biological waste activity decreases. This reduces both leachate production and, notably, the IP signatures. As a result, areas of higher conductivity shift downward and areas of higher chargeability move upward. This effect creates the observed shift between the maxima of the two models. The observed decoupling likely reflects this, but in such complex environments many factors may locally increase m_0 independently of fluid content, as discussed above.

Conversely, the other CC parameters (τ and c) have less information content. On the one hand, the main drawback of the time constant is the lower sensitivity with respect to the other parameters, and its variations around 1 s may reflect couplings with m_0 and c (Tarasov and Titov, 2013). Because of this, τ anomalies should be viewed carefully. Taking this into consideration, these areas are in the well-covered shallow region, and they may reflect slower polarization processes associated with interfacial-polarization effects (Maxwell–Wagner/membrane) due to the presence of areas with biofilm or active chemical surfaces. On the other hand, the strong link between c and m_0 indicates that these parameters are correlated, as pointed out by Fiandaca et al. (2018). For this reason, c provides little extra information beyond what m_0 already shows in this dataset, because the problem becomes more ill-posed and the c values could change to compensate m_0 and τ (Macnae, 2025). For this reason, τ and c can be included as supporting parameters in landfill characterization, but discussed alongside ρ_0 and m_0 only in areas with sufficient coverage.

The results of the a-posteriori conversion to integral chargeability (Fig. 3.8) underline how the full-decay inversion can provide models with higher resolution compared to those reconstructed with a classical two-step inversion. The motivation is twofold: (i) thanks to the forward modelling of the entire decay curve the dataset is enlarged, enabling a more accurate description of waste distribution compared with a single integral value; and (ii) the estimation of the intrinsic chargeability is unbiased compared with the classical approach, where predicted and observed apparent (integral) chargeability formulations differ since it is unfeasible to reliably record the instantaneous value of the potential when current is shut off or changes sign (Binley & Slater, 2020). Anyway, an advanced processing workflow, such as the one presented in this paper, is crucial to obtain a high-quality 100% duty-cycle IP dataset that can enable this type of inversion.

The petrophysical models were then derived from well-established empirical relationships, whose effectiveness depends on the assumptions regarding grain density and leachate conductivity. The former parameter was not studied in this work, since it would require drilling boreholes and extracting waste samples for laboratory analysis. Nevertheless, an accurate description of grain density is almost always unfeasible for MSW landfills, even using direct inspections, due to the extreme spatial heterogeneity of waste composition. The latter parameter (leachate conductivity) is known only at scattered points (wells), but its variation throughout the site is not significant (± 0.2 S/m), so the assumption of an average constant value can be reasonable. Therefore, these models can be useful for defining the predominant spatial trends of leachate accumulation and waste distribution where the coverage

is good. In fact, good agreement was also found for CEC models, where the most biologically active areas are easily identifiable, as in the case of the two central anomalies of line G6 (Fig. 3.9g).

This work underlines the potential of implementing in the future a continuous time-lapse monitoring system of landfill sites, based on the proposed workflow, in order to automatically detect changes in saturation and waste-mass biochemical activities. This implementation could lead to the dual objective of checking leachate accumulation over time and, on the other hand, assessing the long-term effectiveness of the management procedures in use. Our investigation was guided by the aim of assessing the effectiveness of the drainage system developed after the results of the previous geophysical survey (Piegari et al., 2023). In this sense, an overall reduction of the low-resistivity areas can be seen, since the conductive anomalies are confined to a smaller portion of the landfill compared with the previous resistivity models. However, a thorough evaluation of the changes in leachate-accumulation areas is beyond the scope of this work, because it requires a time-lapse survey, considering at least four measurements per year to take into account the effect of seasonality. In addition to that, the integration of electrical data with seismic techniques (e.g., seismic tomography) could strengthen the petrophysical link, enabling the estimation of saturation given the porosity (Steiner et al., 2022).

3.6 Conclusions

This work demonstrates the effectiveness of the proposed approach encompassing processing of 100% duty-cycle full-waveform electrical data with full-decay inversion for the characterization of MSW landfills, oriented to the imaging of leachate-accumulation zones and the characterization of the waste mass.

The proposed processing workflow for 100% duty-cycle full-waveform data can help to enhance the signal-to-noise ratio of IP measurements, enabling the application of full-decay inversion for CC parameters. Full-decay inversion of TDIP data provides a more accurate and physically consistent description of the subsurface electrical response than classical two-step integral chargeability inversion. While both approaches identify similar large-scale layering, the full-decay method is more consistent and reliable, highlighting sharper variations of electrical parameters, more localised anomalies, and a more reliable delineation of biologically active waste zones, particularly where data coverage is adequate.

Although the spectral inversion retrieves τ and c models, these parameters are only weakly constrained and strongly correlated with intrinsic chargeability. As a result, they provide limited independent information in this dataset and should be interpreted cautiously, mainly as supporting indicators alongside resistivity and chargeability.

The full-decay inversion enables the estimation of an unbiased a-posteriori integral chargeability, which can retain higher resolution compared with the classical two-step inversion approach. This feature can assist the application of well-established petrophysical relationships to estimate water (leachate) content and CEC. While these results are still subject to the underlying assumptions and to parameter uncertainties, they provide valuable insight into leachate saturation and zones of en-

hanced biogeochemical activity, supporting landfill risk assessment and management strategies.

3.7 Final remarks

In closing this chapter, it is useful to recall the conceptual framework introduced in the Background. In porous media, when electrolytic conductivity increases because of high saturation and leachate salinity, the IP response may decrease even if polarization mechanisms are still present. In this sense, fully leachate-saturated zones are mainly associated with low resistivity and with a chargeability response that can be lower than in drier waste or in zones where biochemical activity is more pronounced, also because complete saturation limits the interfacial conditions required for a strong IP response. Accordingly, the association between low resistivity and high chargeability is not regarded afterwards as a universal signature of fully saturated leachate, but as an operational indicator of leachate-affected and electrochemically active waste domains. This also explains why the simplified synthetic models or clustering approaches discussed later use this association for practical target detection, without explicitly separating the possible decrease of IP response caused by full leachate saturation.

In this thesis, no dedicated synthetic benchmark was performed to quantify the improvement of full-decay inversion over classical integral-chargeability inversion. The advantage of full-decay inversion, as shown in previous studies cited in this chapter and in the field results comparison reported in Fig. 3.8 of the manuscript presented here, is that the temporal shape of the TDIP response is exploited instead of reducing the polarization behaviour to a single integral chargeability value. This allows the inversion to estimate the intrinsic chargeability m_0 , which better represents the overall polarization strength of the investigated medium than a chargeability value integrated over a fixed time window. A systematic synthetic comparison in terms of data misfit and model resolution remains a useful future development.

A separate point concerns the quantitative assessment of the full-waveform acquisition and processing gain. A direct comparison with conventional instrument-derived IP chargeability values was not performed in this thesis because the 2025 data were acquired in "RES" mode in the ABEM instrument options. This configuration was selected for practical reasons, to reduce acquisition time in the landfill environment and to allow the adopted current-pulse configuration for full-waveform processing. Under this acquisition mode, standard IP chargeability values were not provided as instrument output. Therefore, a paired comparison with classical gated IP data was not available, and remains an important future development for quantifying the gain in terms of signal-to-noise ratio or quality of the IP data.

The results also show that not all Cole–Cole parameters have the same interpretative value in this landfill context. Resistivity and intrinsic chargeability provide the most stable and environmentally meaningful information, whereas the relaxation time τ and the frequency exponent c are more difficult to relate unambiguously to waste properties. They are therefore considered complementary descriptors of the spectral response rather than primary interpretative parameters.

Finally, the reliability of the full-decay inversion depends on the quality of

waveform processing. In this thesis, the gating was started with a delay of 30 ms to minimize the influence of induction effects occurring at earlier times and the edge effects of harmonic suppression. The Hampel-based despiking was adopted following the workflow of Olsson et al. (2016), whose effectiveness for full-waveform TDIP processing had already been demonstrated. Its parameters were adapted to the characteristics of the recorded signal and checked against the raw and processed waveforms, while corrections of fractions of millivolts may still be relevant for weak IP responses at late times. Future developments will also consider an update of this processing scheme, including a more automated and adaptive selection of filtering and despiking parameters. The comparison with piezometric levels is also regarded as qualitative, because estimated water content is indirectly derived from geoelectrical parameters, whereas well levels provide local hydraulic information in a highly heterogeneous waste body.

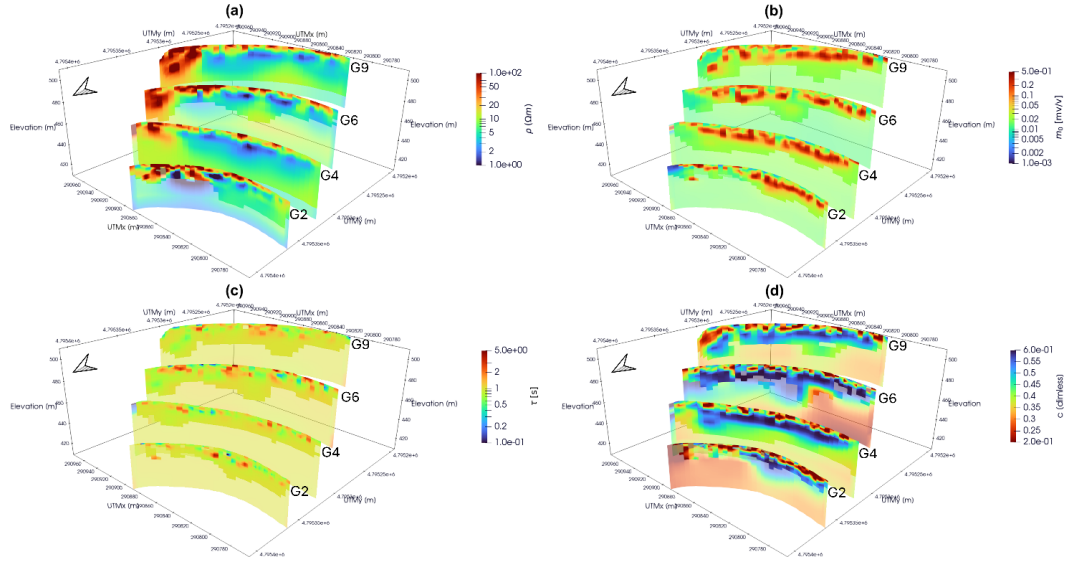


Figure 3.7. 3D visualization of the inverted models for profiles G2, G4, G6 and G9. (a) DC resistivity ρ . (b) Intrinsic chargeability m_0 . (c) Relaxation time τ . (d) Frequency exponent c . Models are displayed only within the high-coverage region defined by the adopted cut-off value; axes are in UTM coordinates and elevation.

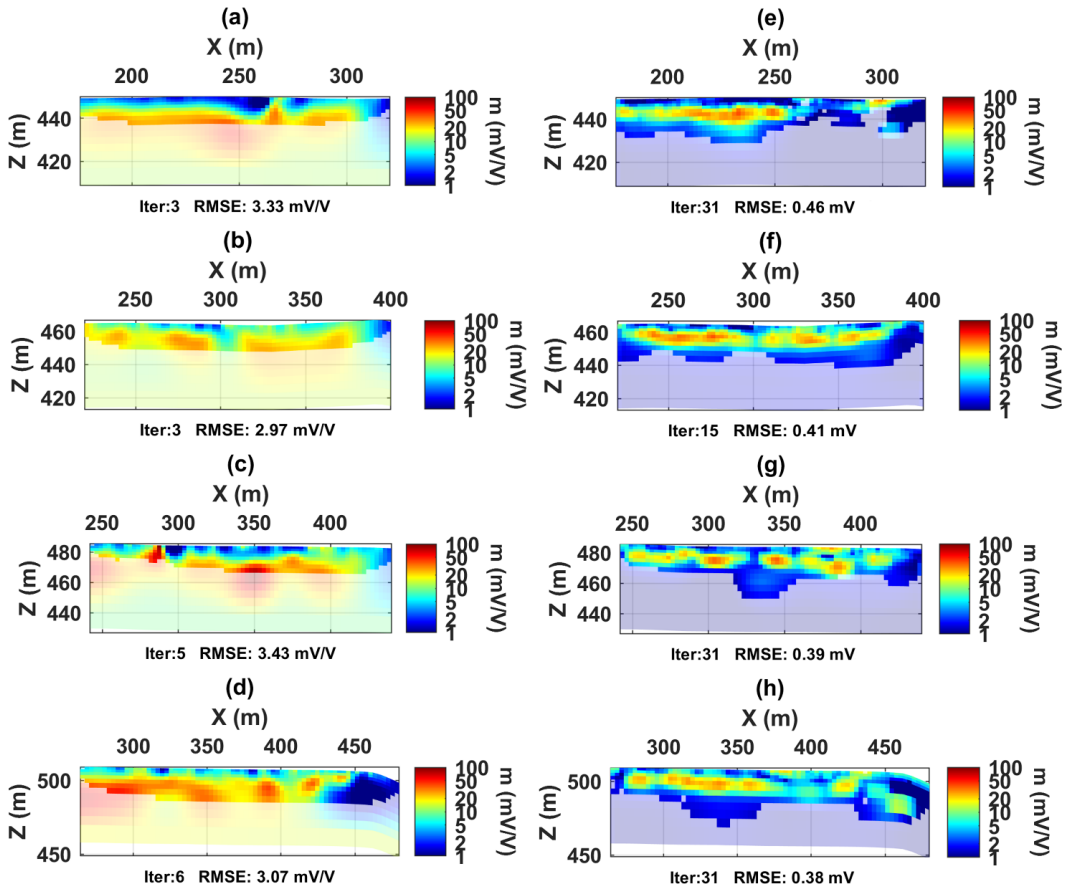


Figure 3.8. Comparison between a-posteriori integral chargeability obtained from (a–d) a conventional two-step inversion and (e–h) the full-decay inversion. Panels correspond to the four profiles (G2, G4, G6 and G9). Shaded regions indicate areas below the adopted coverage cut-off value.

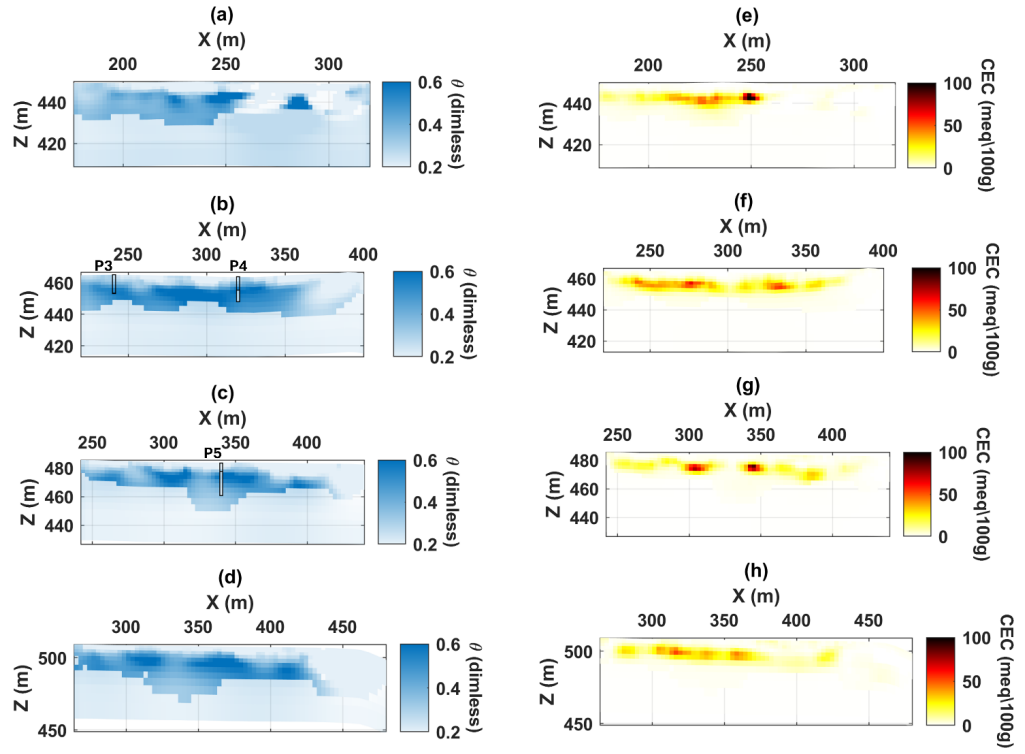


Figure 3.9. Water content (a–d) and CEC (e–h) models retrieved with the relation of Revil et al. (2020) using inverted parameters from full-decay inversion of full-waveform data. Panels correspond to the four profiles (G2, G4, G6 and G9). The piezometric levels logged in wells are superposed with white filled rectangles.

Chapter 4

Global inversion strategies

Paper #2: Global joint inversion of ERT and TDIP data using very fast simulated annealing: uncertainty quantification and application to landfill characterization

Giorgio De Donno¹, Michele Cercato¹, and Davide Melegari^{1,*}

¹ Dipartimento di Ingegneria Civile Edile e Ambientale, “Sapienza” Università di Roma, Rome, Italy

* Corresponding author: davide.melegari@uniroma1.it

The manuscript presented in this chapter is currently under review for publication in the Geophysical Journal International and is reported here in its complete form.

4.1 Abstract

In this study, we present a stochastic global joint inversion framework for two-dimensional electrical resistivity tomography (ERT) and time-domain induced polarization (TDIP) data, aimed at improving model reconstruction and providing statistically consistent uncertainty estimates. Although ERT and TDIP inversion has been extensively investigated, systematic uncertainty analysis in full two-dimensional frameworks remains limited. The proposed approach jointly inverts ERT and IP data for resistivity and chargeability using the very fast simulated annealing (VFSA) on an unreduced 2D parameter mesh, thereby avoiding model simplifications that may bias parameter estimation and uncertainty quantification. The algorithm is implemented using parallel independent runs, base-10 logarithmic reparameterization, and inequality constraints on resistivity values to ensure numerical stability and physical plausibility. Model uncertainty is evaluated through ensemble-based analysis, where posterior covariance is computed from models lying within a misfit tolerance, approximating the high-likelihood region of the parameter space, while correlation is computed from the full ensemble of the accepted models. This formulation enables assessment of parameter variability, identifiability, and resistivity–chargeability coupling without relying on local linearization. The methodology is tested both on a synthetic model and field data acquired over a municipal solid waste landfill, where leachate accumulations represent geophysical targets of interest. Results demonstrate that global exploration of the solution space provides improved characterization of model uncertainty and reveals features that may remain unresolved in deterministic approaches. The proposed framework illustrates how global optimization can be integrated with ensemble-based statistical analysis for multiparameter geophysical imaging, contributing to more robust interpretation of 2D ERT and TDIP data.

Keywords: electrical resistivity tomography (ERT), induced polarization, inverse theory, probabilistic forecasting.

4.2 Introduction

Electrical resistivity tomography (ERT) and time-domain induced polarization (TDIP) are widely used geophysical techniques for imaging subsurface heterogeneity

in natural and anthropogenic environments. ERT has been extensively applied to characterize geological complexity and environmental processes (Samouëlian et al., 2005; Chambers et al., 2006), while TDIP provides complementary information owing to its sensitivity to electrochemical and microbial activity in the subsurface (Flores-Orozco et al., 2020). The joint use of ERT and TDIP has proven particularly valuable in municipal solid waste (MSW) landfills, where strong heterogeneity and evolving hydrochemical conditions control moisture distribution and leachate migration (see De Donno et al., 2024 for a recent review).

The reliability of geoelectrical investigations strongly depends on the inversion framework used to reconstruct subsurface electrical properties. Over the past decades, several forward modelling and inversion approaches have been developed for two- and three-dimensional imaging (e.g. Loke and Barker, 1996; Rücker et al., 2017; Blanche et al., 2020). Nevertheless, most practical applications rely on deterministic local inversion schemes based on smoothness-constrained formulations (Constable et al., 1987; de Groot-Hedlin and Constable, 1990). Although computationally efficient, these approaches are inherently local: they depend on an initial model, may converge toward local minima, and often produce overly smooth representations of the subsurface. In highly heterogeneous systems such as landfills, smoothness constraints may distort anomaly geometry and bias parameter estimation, particularly when sharp resistivity contrasts are present and prior information is limited (e.g. Loke and Barker, 1996; Sen and Stoffa, 2013).

A further limitation of conventional inversion concerns the treatment of model uncertainty. Resolution matrices and depth-of-investigation estimates (e.g. Menke, 1984; Oldenburg and Li, 1999; Günther et al., 2006) provide qualitative insight into model reliability but do not yield statistically consistent confidence intervals or probabilistic descriptions of parameter variability. This shortcoming is particularly critical in landfill environments, where robust uncertainty quantification is essential for environmental monitoring and risk assessment.

Global optimization strategies offer an alternative framework by exploring a broader portion of the parameter space and reducing dependence on initial model assumptions. Such approaches are better suited to address nonlinearities and non-uniqueness inherent in multiparameter inverse problems (Sen and Stoffa, 2013). However, their application to ERT and TDIP inversion remains limited due to the computational burden associated with repeated nonlinear forward simulations (Chunduru et al., 1996). To alleviate this issue, some studies have proposed reduced parameterizations or simplified model representations (Fernández-Martínez, 2015; Aleardi et al., 2022), which decrease computational cost but may constrain spatial flexibility and resolution.

Among global optimization algorithms, Very Fast Simulated Annealing (VFSA) (Ingber, 1989) provides an efficient strategy for exploring high-dimensional nonlinear parameter spaces. Compared to classical Monte Carlo sampling, VFSA achieves faster convergence while maintaining global search capability (Sen and Stoffa, 2013). Despite these advantages, applications of VFSA to full-mesh two-dimensional joint inversion of ERT and TDIP data remain scarce, particularly in contexts where statistical characterization of parameter uncertainty and parameter coupling is required.

In this study, we present a stochastic global inversion framework for the joint

inversion of ERT and TDIP data based on VFSA, applied directly to a full two-dimensional parameter mesh without model reduction. The approach generates ensembles of resistivity and chargeability models, enabling statistical analysis of parameter variability, identifiability, and resistivity–chargeability coupling within a high-likelihood region of the solution space. Building upon previous applications of VFSA to standalone ERT data (Chunduru et al., 1996; Cercato and De Donno, 2023), this work moves beyond single-parameter inversion by extending the approach to a fully coupled multiparameter framework and by introducing a systematic analysis of the posterior model space. The objectives of this study are twofold: (i) to develop and test a stochastic joint inversion strategy for ERT and TDIP data; (ii) to investigate parameter identifiability and coupling, and to evaluate the added value of uncertainty-aware inversion for interpreting complex geophysical targets in landfill environments. The methodology is implemented within an extended version of the VEMI (Versatile algorithm for Electrical data Modelling and Inversion) software, developed in MATLAB (De Donno and Cardarelli, 2017b), which employs finite-element modelling to simulate both resistivity and integral chargeability responses under different acquisition geometries. The proposed approach is preliminary tested through a synthetic example and subsequently applied to the characterization of a MSW landfill in Central Italy, mainly oriented to the detection and monitoring of potential leachate accumulation areas.

4.3 Method

4.3.1 Forward modelling

The resistive response of a subsoil is described by Poisson’s equation under the hypothesis of an external point source:

$$-\nabla \cdot [\sigma_0(\mathbf{r})\nabla V(\mathbf{r})] = I\delta(\mathbf{r}) \quad \forall \mathbf{r} \in D, \quad (4.1)$$

where σ_0 is the DC (direct-current) electrical conductivity (the DC electrical resistivity is $\rho_0 = 1/\sigma_0$), V the electric potential, $\mathbf{r} \equiv (x, y, z)$ the position vector, I the injected current, δ Dirac’s function approaching infinity at $\mathbf{r} \equiv \mathbf{r}_s$, with $\mathbf{r}_s \equiv (x_s, y_s, z_s)$ the position vector of the source and D the 3-D domain. The point-source approximation is valid when the electrode spacing is large relative to both the electrode size and the investigation depth, as is typical in near-surface ERT applications (Cardarelli and De Donno, 2019).

When 2-D conditions apply – i.e. when the resistivity is assumed to be invariant along the y -direction – Eq. (4.1) is Fourier-transformed in the (x, λ, z) space (e.g. Dey and Morrison, 1979):

$$-\nabla \cdot [\sigma_0(x, z)\nabla \phi(x, \lambda, z)] + k^2\sigma_0(x, z)\phi(x, \lambda, z) = I\delta(x)\delta(z) \quad \forall (x, z) \in \Omega, \quad (4.2)$$

being λ the transformed variable, ϕ the transformed potential and Ω the 2-D domain.

The 2D hypothesis holds if the ERT line is approximately normal to the axis of the expected anomalies or 3D effects are not significant. Among the different

numerical approaches employed for the solution of Eq. (4.2) under appropriate boundary conditions, the Gal rkin formulation of the finite-element method has been applied by numerous authors over the past years (e.g. G nther et al., 2006). In this case, the numerical solution of Eq. (4.2) is found as:

$$\mathbf{\Phi} = \mathbf{Y}^{-1}\mathbf{I}, \quad (4.3)$$

which requires the stiffness matrix $\mathbf{Y}$ [$N \times N$], with N the number of nodes, to be invertible. After solving Eq. (4.3), the electrical potential vector $\mathbf{\Phi}$ is back transformed into the spatial domain, yielding the DC potential vector $\mathbf{V}_0$.

Under the assumption of low-frequency polarization, the IP phenomenon is often modeled through the chargeability m defined as:

$$\sigma_0 = \sigma_\infty(1 - m), \quad (4.4)$$

where $0 \leq m < 1$ and σ_∞ is the instantaneous conductivity (Binley and Slater, 2020). This formulation corresponds to the linear approximation of the polarization effect, valid for typical near-surface IP surveys and small chargeability values (Seigel, 1959).

Therefore, for the combined ERT and IP case, we solve Eq. (4.2) twice by using in turn σ_0 and σ_∞ , yielding through Eq. (4.3) two potential vectors $\mathbf{V}_0$ and $\mathbf{V}_\infty$.

Once the potentials are obtained, the predicted apparent resistivity ρ_a^{pre} and chargeability m_a^{pre} can be computed as:

$$\rho_a^{pre} = K \frac{\Delta V_0}{I}, \quad (4.5a)$$

$$m_a^{pre} = \frac{\Delta V_0 - \Delta V_\infty}{\Delta V_\infty}, \quad (4.5b)$$

being K the geometric factor and ΔV the potential difference computed owing to the particular measurement sequence.

4.3.2 Inversion procedure

4.3.2.1 VFSA inversion

In this study, the VFSA algorithm (Ingber, 1989; Sen and Stoffa, 2013) is adopted for the inversion process, owing to its robustness and the relatively small number of parameters required. The algorithm draws random models from a Cauchy-like distribution that is, in turn, a function of temperature. Models that lower energy are always accepted, while models that increase energy are accepted with a finite and temperature-dependent probability. This probability-dependent acceptance criterion, together with an adequate cooling schedule and enough random samples, ensures the inversion process to converge towards the global minimum of the energy function, which is virtually independent of the initial model (Cercato, 2011). In this study, the energy function E is defined, for the model $\mathbf{M} = [\rho_0, m]$ of resistivity (ρ) and chargeability (m) under consideration, as:

$$E(\mathbf{M}) = \frac{\sum_{i=1}^N \left[\log_{10}(d_{a,i}^{obs}) - \log_{10}(d_{a,i}^{pre}(\mathbf{M})) \right]^2}{N}, \quad (4.6)$$

where $\mathbf{d} = [\rho_a^{obs}, m_a^{obs}]$ is the joint observed dataset and N the number of measurements (doubling the number of quadrupoles), with ρ_a^{obs} and m_a^{obs} apparent resistivity and apparent (integral) chargeability, the latter being equal to (Olsson et al., 2019):

$$m_a^{obs} = t_{tot}^{-1} \sum_{g=1}^{N_G} m_g t_g, \quad g = 1, 2, \dots, N_G, \quad (4.7)$$

where m_g is the chargeability value for the g -th gate, being N_G the number of gates, while t_g and $t_{tot} = \sum_{g=1}^{N_G} t_{gate,g}$ the time width for the g -th gate and the total time summed from the first to the last gate, respectively.

Alternative definitions of the energy function, such as the L1-norm, L2-norm and normalized error, were also considered and tested. However, Eq. (4.6) was consistently adopted throughout all the inversions presented in this work, to ensure methodological uniformity. The base-10 logarithm transformation was used, as it is generally considered more appropriate for electric and electromagnetic applications, where resistivity values typically span several orders of magnitude (Kim and Kim, 2011; Sen and Stoffa, 2013).

4.3.2.2 Inversion parameters and *a priori* information

The cooling schedule is calculated according to Ingber (1989):

$$T_k = T_0 e^{[-ck(1/N_P)]}, \quad (4.8)$$

where T_k is the temperature at the iteration $k = 1, 2, \dots, N_S$, with N_S the number of random steps, $c = \gamma \exp\left(-\frac{n}{N_P}\right)$, with N_P the number of parameters (doubling the number of pixels), $\gamma = \log\left(\frac{T_0}{T_f}\right)$ and $n = \log N_S$. Therefore, the number of random steps must be specified in advance to perform global inversion.

For 2D inversion, with typically hundreds of model parameters, is more appropriate to use a constant factor (0.4) instead of $1/N_P$, to achieve a reliable reduction of temperature to the lowest temperature level (Sharma, 2012). In fact, using N_P for a large set of variables, such as for the ERT case, the number of iterations required for convergence increases dramatically (Sharma and Kaikkonen, 1998). Therefore, we decided to adjust this value to 0.4, based on many previous VFSA studies (see Biswas and Sharma, 2015 and references therein).

The initial and final temperatures (T_0 and T_f) are selected to ensure an effective control of the acceptance criterion throughout the inversion process. At high temperatures ($T \approx T_0$), most candidate models are accepted, including those associated with an increase in the objective function, thus promoting a broad exploration of the model space. As the temperature decreases ($T \rightarrow T_f$), the probability of accepting

unfavourable perturbations rapidly diminishes, leading the algorithm towards convergence around the global minimum. Based on preliminary tests on ERT datasets (Cercato and De Donno, 2023), T_0 and T_f were set to 1 and 10^{-9} , respectively. The stochastic search is carried out within predefined bounds on the model parameters ($a_i \leq M_i \leq b_i$), which incorporate prior information and define the admissible region of the parameter space explored during the inversion. Then, the randomly altered model is found as:

$$M_i^{new} = M_i + y_i(b_i - a_i), \quad i = 1, 2, \dots, N_E, \quad (4.9)$$

where $y_i = T_k \text{sign}(u - 0.5) \left[\left(1 + \frac{1}{T_k}\right)^{|2u-1|} - 1 \right]$, N_E is the number of elements, u is a randomly generated number drawn from a discrete uniform distribution between 0 and 1, such that for each decade within the global bounds only 100 resistivity and chargeability values can be picked and both model and constraints are expressed in terms of $\log_{10}$ values. The discrete uniform distribution is adopted to reduce the number of steps required for convergence without affecting the effectiveness of the procedure (Weber, 2000).

To enhance the stability of the inversion and suppress spurious fluctuations, a weighted median filter is applied to the model parameters after each perturbation step. The filtered value of the i -th pixel is computed as:

$$M_i^{newF} = MED [W_U \diamond M_U^{new}, W_L \diamond M_L^{new}, W_C \diamond M_C^{new}, W_R \diamond M_R^{new}, W_D \diamond M_D^{new}], \quad (4.10)$$

where $MED[\cdot]$ denotes the median operator, W represents the assigned weights, and $\diamond$ indicates the duplication operator. Subscripts (U, L, C, R, D) refer to the neighbouring pixels (up, left, centre, right, down). The weighting scheme is adapted according to pixel location (e.g., interior, boundary, or corner cells), with representative configurations including (1, 1, 3, 1, 1) for interior pixels and reduced weights along model boundaries. The use of a median-based filter, as opposed to conventional averaging schemes, allows preserving sharp resistivity contrasts while effectively reducing isolated artefacts (Brownrigg, 1984; Yin et al., 1996). This contributes to a more stable exploration of the model space and improves convergence efficiency (Weber, 2000).

A sufficiently large number of model perturbations is required to adequately sample the *a posteriori* probability distribution (PPD) (Sen and Stoffa, 1996; Bhattacharya and Sen, 2003; Sen and Stoffa, 2013). In this study, the total number of random steps is defined as a function of the number of model parameters N_P , ensuring an adequate exploration of the parameter space. Because a single realization of the VFSA algorithm may not be sufficient to consistently identify the global minimum, multiple independent runs are performed. Specifically, ten inversions are carried out in parallel, each starting from different random initial conditions. The final solution is obtained by averaging the minimum-misfit models from each run, thus improving robustness and reducing the impact of stochastic variability.

4.3.2.3 Uncertainty estimation

Following the approach proposed by Sharma (2012) for stochastic inversion, the model parameters are represented in $\log_{10}$ space, implying that the corresponding physical parameters follow a lognormal distribution. Uncertainty estimation is therefore performed on the $\log_{10}$ -transformed variables.

To estimate the mean model and the associated variability, a subset of statistically acceptable solutions is selected from the ensemble of sampled models. Models are retained if their objective function value lies within a relative tolerance of the global minimum misfit, defined as:

$$E(\mathbf{M}_i) \leq (1 + \delta_E) \min(E), \quad (4.11)$$

with $\delta_E = 0.05$ (5%). This criterion identifies models that provide a data fit statistically indistinguishable from the optimal solution. In cases where this condition yields fewer than 100 models, the tolerance is progressively increased by steps of 2.5% until a minimum ensemble size of 100 models is reached, ensuring stable estimates of posterior statistics. Uncertainty on model parameters is then estimated from the subset of accepted models via the posterior covariance matrix:

$$\mathbf{C} = \frac{1}{N_M} \sum_{j=1}^{N_M} (\mathbf{M}_j - \langle \mathbf{M} \rangle)^T (\mathbf{M}_j - \langle \mathbf{M} \rangle), \quad (4.12)$$

where N_M is the number of accepted models lying within the misfit threshold and $\langle \mathbf{M} \rangle = \frac{1}{N_M} \sum_{j=1}^{N_M} \mathbf{M}_j$ is the mean of the $\log_{10}$ -transformed model, both computed via the above-mentioned criteria. The diagonal elements of $\mathbf{C}$ contain the standard deviations $\sigma_{\log_{10} \rho}$ and $\sigma_{\log_{10} m}$ of the $\log_{10}$ -transformed unknowns. Then, uncertainties were both expressed as absolute and percentage values using the logarithmic error propagation, yielding:

$$\sigma = \log(10) \sigma_{\log_{10}}, \quad (4.13a)$$

$$\sigma(\%) = 100 \log(10) \langle \mathbf{M} \rangle \sigma_{\log_{10}}, \quad (4.13b)$$

where σ is the standard deviation vector including both σ_ρ and σ_m .

The *a posteriori* correlation matrix, which quantifies the interdependence among model parameters, is evaluated using the full ensemble of sampled models, thereby capturing the overall dependency patterns explored during the stochastic search, as:

$$P_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \quad (4.14)$$

where C_{ij} is a generic element (i-th row and j-th column) of the covariance matrix $\mathbf{C}$, defined in Eq. (4.12), but computed on the full ensemble of sampled models.

4.3.2.4 Petrophysical transformation

To derive petrophysical parameters after inversion, we convert the resistivity and chargeability minimum misfit models into water (leachate) content (θ) and cation exchange capacity (CEC) models, using the equations after Revil et al. (2020):

$$\theta = \left[\frac{1}{\sigma_l} \left(\sigma_0 - \frac{M_n}{R} \right) \right]^{1/n}, \quad (4.15)$$

$$CEC = \frac{M_n}{\theta^{n-1} \rho_g \lambda_p}, \quad (4.16)$$

where M_n (in S m^{-1}) is the normalized chargeability (Slater and Lesmes, 2002), defined as the product of the integral chargeability m and the direct current (DC) conductivity σ_0 (approximated as equal to $1/\rho_0$), n denote the cementation exponent which we set to be equal to 2, as the saturation exponent in the second Archie's law (modified after Winsauer et al., 1952), σ_l corresponds to leachate conductivity, ρ_g denotes the grain (waste) density and $R = \lambda_p/B \approx 0.1 \pm 0.02$ after Revil et al. (2017), being λ_p the apparent mobility of the counterions for the polarization (in $\text{m}^{-2} \text{s}^{-1} \text{V}^{-1}$) and B the apparent mobility of the counterions for the surface conduction (in $\text{m}^{-2} \text{s}^{-1} \text{V}^{-1}$). The average leachate conductivity, directly logged in the wells, was equal to $1.3 \pm 0.2 \text{ S/m}$ in 2019 and $1.0 \pm 0.2 \text{ S/m}$ in 2025, while the waste density ρ_g was assumed as constant and equal in a first approximation to 1500 kg m^{-3} (Steiner et al., 2022).

The resulting θ and CEC models allowed for hydro-geophysical interpretation of the subsurface, offering insights into waste saturation patterns and degradation zones with high biological or electrochemical activity.

4.4 Results

In the following sections, we present a synthetic and a real-world example related to the detection of a leachate accumulation zone within an urban waste landfill. For both cases, *a priori* information – such as the true or logged leachate levels and liner location – was not used as a constraint during the inversion. These data were only overlaid on the final models to support the geological and environmental interpretation of the inversion results.

4.4.1 Synthetic example

4.4.1.1 Model geometry and inversion parameters

The synthetic example investigates a typical anthropogenic scenario, representative of geometries and conditions usually found in landfills, where the identification of the waste mass distribution and of leachate accumulation zones is required for environmental and geotechnical (instability) purposes. To this end, we constructed a synthetic model that reproduces the typical geometry of a MSW landfill, including lateral slopes (Fig. 4.1a–d). A conductive and chargeable target ($\rho = 5 \Omega \text{ m}$, $m = 50 \text{ mV/V}$), simulating the leachate-saturated waste, is positioned in the central

portion of the model beneath a covering layer (mixture of dry waste and covering soil), which is more resistive and slightly chargeable ($\rho = 20 \Omega \text{ m}$, $m = 20 \text{ mV/V}$), while the surrounding environment is characterized by a higher resistivity and lower chargeability ($\rho = 50 \Omega \text{ m}$, $m = 5 \text{ mV/V}$). Both the resistivity values and the model geometry are consistent with previous synthetic landfill simulations (see De Donno et al., 2024 for a review). Although our previous studies (e.g. Melegari and De Donno, 2024; De Donno et al., 2025) suggest that IP can decrease with increasing leachate saturation, this relationship was not explicitly reproduced in the present synthetic example. For the sake of simplicity, no saturation-dependent IP variation was imposed, allowing the synthetic model to remain intentionally schematic. The simulated line consists of 16 electrodes spaced 1 m apart, where a dipole-dipole array is used with $a = [1, 5]$ and $n = [1, 6]$ (110 measurements). The geometry of the model is the same as the model proposed by Chunduru et al. (1996) in their pioneering work on VFSA inversion of ERT data. A zero-mean Gaussian noise was added to the synthetic dataset, with standard deviations set to 5% on both apparent resistivity and chargeability datasets.

We carried out 10 independent runs of the VFSA algorithm on a 105-pixels mesh with $2 \cdot 10^6$ random steps for each run with bounds $[a, b]$ equal to $[0.5\text{--}500 \Omega \text{ m}]$ and $[0.5\text{--}500 \text{ mV/V}]$, which extends one order of magnitude below and above the true minimum and maximum values.

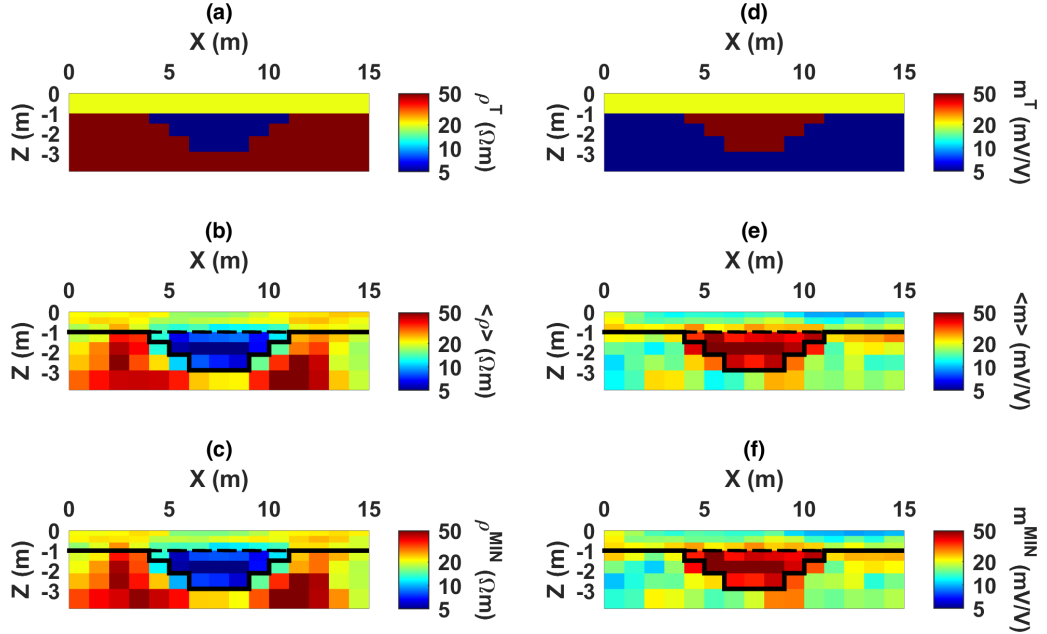


Figure 4.1. Synthetic model: (a) true resistivity model, (b) mean resistivity model, (c) minimum misfit resistivity model, (d) true chargeability model, (e) mean chargeability model, (f) minimum misfit chargeability model. The true position of the bottom liner is marked with a black solid line, as in the case of the true position of the leachate level.

4.4.1.2 Inversion results and uncertainty assessment

The results of the VFSA inversion (Fig. 4.1) yields good reconstructions of the true models, since both mean and minimum-misfit models successfully recover the main conductive–chargeable target, both in terms of magnitude values and geometry. Furthermore, the mean model (Figs. 4.1b,e) does not deviate significantly from the minimum misfit one (Figs. 4.1c,f), confirming the reliability of the inversion. The recovered anomaly is correctly located within the landfill body and remains clearly distinguishable from the more resistive and less chargeable shallow layer, which is well reconstructed, as expected due to the high data coverage and resolution near surface. In contrast, although the resistivity model is able to accurately distinguish between the conductive anomaly and the more resistive background, for chargeability this transition is more gradual, highlighting the common resolution issues associated with the use of integral chargeability in IP in data inversion (Fiandaca et al., 2012).

The uncertainty analysis shown in Fig. 4.2 is expressed as percentage and absolute standard deviations (Figs. 4.2a,b and 4.2c,d, respectively for resistivity and chargeability), correlation matrix (Fig. 4.2e), and misfit evolution with iterations (Fig. 4.2f). The spatial distribution of the standard deviations for resistivity and chargeability reflects the sensitivity pattern imposed by the acquisition geometry and the underlying physical contrasts of the synthetic model. For resistivity (Figs. 4.2a,b), the lowest uncertainties are observed in the shallow portion of the model, where the data coverage and sensitivity are the highest. Uncertainty increases towards the margins and base of the conductive anomaly, where sensitivity decreases and the geometry of the target becomes less constrained. A similar pattern is observed for chargeability (Figs. 4.2c,d), although with slightly higher percentage errors, as expected for the IP response. The core of the conductive–chargeable body is well resolved (low percentage uncertainty), whereas higher uncertainty values occur along the transition zones between the anomalous region and the surrounding background.

The posterior correlation matrix (Fig. 4.2e) displays significant values only for neighbourhood cells where the effect of the filter is active, while exhibiting negligible off-diagonal structure elsewhere, indicating absence of artificial parameter coupling and supporting identifiability of the solution. Conversely, cross-parameter correlations between resistivity and chargeability remain weak throughout the model space, indicating that the inversion does not introduce significant artificial coupling between the two parameters. The evolution of the objective function during the VFSA runs (Fig. 4.2f) shows, after an expected initial exploratory phase characterized by relatively large misfit values, a stable convergence towards a narrow region of the model space corresponding to low misfit values. Although the error function continues to decrease slightly at the end of the exploration, the progressive reduction in misfit dispersion indicates that the algorithm has already reached a stable region of the model space where further improvements are marginal.

4.4.2 Field case study: municipal solid waste (MSW) landfill

4.4.2.1 Site description and data acquisition

The real-world example concerns an ERT/IP survey aimed at detecting potential leachate accumulation within an urban waste landfill situated in Central Italy. This

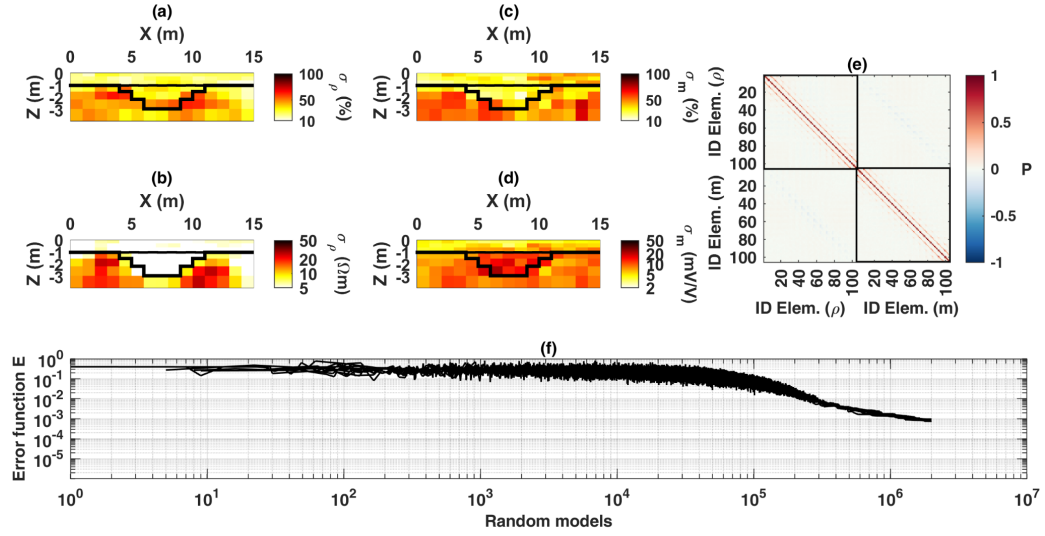


Figure 4.2. Synthetic model: uncertainty estimation in percentage values (a) and absolute values (b) for resistivity, uncertainty estimation in percentage values (c) and absolute values (d) for chargeability, (e) posterior correlation matrix (elements are numbered by rows downward starting from the upper-left corner), (f) error function vs. number of iterations for the 10 independent runs. The true position of the bottom liner is marked with a black solid line, while the true position of the leachate level with a black dashed line.

case study is adapted from Piegari et al. (2023), where the reader can find further details about the geological setting and the layout of the survey lines. This area is of particular interest because leachate accumulation is likely favoured by an impluvium shaped by the pre-existing topographic surface, whose geometry is known from the original design of the landfill. The landfill site was provided with a bottom liner, known to act as an insulating barrier, preventing current flow beyond the landfill boundaries.

Electrical data were acquired in 2019 and 2025 along four profiles (L1–L4 in Fig. 4.3) before and after the execution of a deep drainage network (sub-horizontal drains located at the bottom of the landfill) for lowering the leachate levels. In the former survey (2019) we used the IRIS SyscalPro48 (50% duty-cycle configuration) with a dipole-dipole array ($a = [1, 5]$ and $n = [1, 6]$). After filtering negative decay curves and observations having an experimental standard deviation value higher than 10%, the resulting dataset included 129 measurements. In 2025, we performed a new data acquisition using the ABEM Terrameter LS2 instrument in a 100% duty cycle full-waveform configuration with a multiple-gradient configuration, aiming at monitoring the landfill conditions after the execution of the drainage system as well as improving the quality of the IP data. The 2019 and 2025 pseudosections are not directly paired, because the surveys were acquired with different electrode layouts. In particular, the 2025 acquisition used a longer profile with 81 electrodes, compared with 48 electrodes in 2019. This explains the different number and apparent position of the pseudosection data points. For this reason, a formal time-lapse inversion was not performed, since the differences in electrode layout and profile length would make a direct cell-by-cell comparison unreliable. A paired acquisition geometry will

be adopted in the next monitoring campaign to allow a more rigorous time-lapse analysis.

We selected for global inversion a portion of the L2 line, which corresponds to the right part (starting from $x = 215$ m on the original ERT line), comprising 20 electrodes spaced 10 m apart, and the resulting dataset included 105 measurements. This line was selected since also leachate levels were logged at the time of acquisition in two monitoring wells (P1–P2), providing piezometric control along the profile: P1 was almost dry, P2 registered approximately 8 m (2019) and 7.5 m (2025) of leachate. However, since the leachate level in P2 was very similar in the two campaigns, this information should be considered as a qualitative constraint on the presence and approximate depth of leachate, rather than as a direct validation of the temporal variations observed in the geophysical and petrophysical models. By contrast, the bottom liner superimposed on the following sections is only inferred from the original landfill construction design and is therefore more hypothetical than *a priori*-information such as well data. For global inversion, we assigned the same model bounds as per the synthetic example and a discretized model with 152 pixels. The VFSA algorithm was executed in both cases (2019 and 2025) using the same inversion settings of the synthetic case, ensuring methodological consistency and enabling direct comparison between different cases.

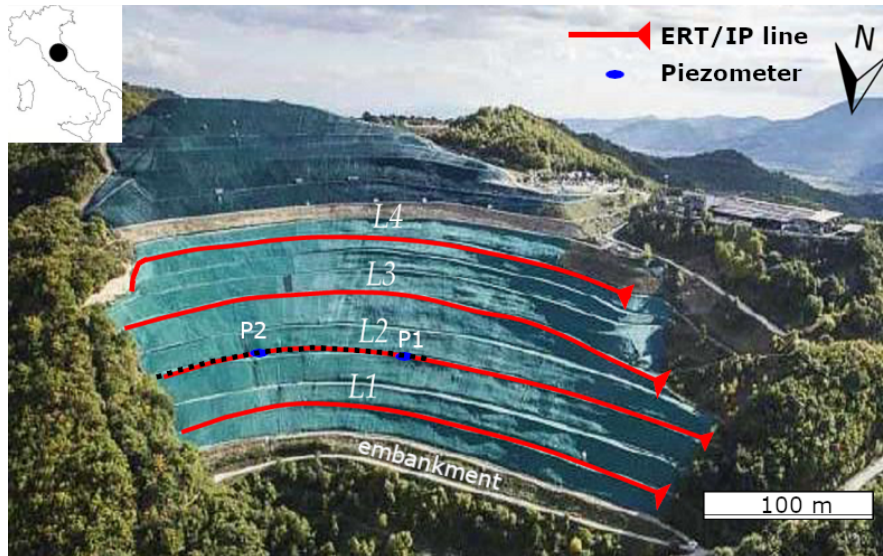


Figure 4.3. Study area.

4.4.2.2 Inversion results and uncertainty assessment for the 2019 dataset

The VFSA inversion results for the 2019 dataset (before drainage) are shown in Fig. 4.4 (data and inverted models) and Fig. 4.5 (uncertainty, correlation, and misfit evolution), with prior information superimposed. As illustrated in Fig. 4.5f, the VFSA runs ensure a rapid convergence toward an almost identical minimum misfit value.

The VFSA inversion (Fig. 4.4) highlights a conductive–chargeable zone within the landfill body, mainly concentrated in its central-to-basal portion, well in agreement

with the expected geometry imposed by the liner and with the leachate levels measured in wells. In particular, the low-resistivity anomaly shows good accordance with the higher leachate level measured at P2 (Figs. 4.4b,c). By contrast, the IP response did not show a clear correlation with the leachate level, since the IP response tends to decrease as saturation increases, as previously introduced. The high resistivity values clearly recall the geometry of the bottom of the landfill (liner, black solid line), particularly on the right side where the data coverage is satisfactory. Additionally, the high chargeability areas are clearly included within the landfill boundaries (Figs. 4.4e,f), thus delineating well the waste mass where strong biochemical activity is expected. Consistently with the assumptions described above, the liner position is only hypothesized, and minor chargeability anomalies outside the landfill may also reflect the lower sensitivity of the IP response. As in the synthetic scenario, the overall agreement between mean and minimum-misfit solutions strengthens the interpretation of the anomaly as a robust feature of the state of the landfill rather than as an inversion artefact.

The uncertainty analysis (Fig. 4.5) clearly highlights a zone of low reliability towards the side and the bottom of the model, as expected due to poor coverage, while intermediate uncertainty values are observed in the middle portion of the model (saturated by leachate), and the most reliable pixels are located in the shallowest layers. In particular, for resistivity (Figs. 4.5a,b) the lowest uncertainties are observed in the shallowest layers of the model, where the sensitivity of the measurements is highest. Uncertainty progressively increases with depth and towards the lateral boundaries of the model domain, reflecting the decreasing resolution of the data. Intermediate uncertainty values are observed within the central portion of the landfill body, where the conductive anomaly associated with leachate accumulation is detected. The chargeability uncertainty distribution (Figs. 4.5c,d) exhibits a comparable spatial pattern. The core of the chargeable anomaly is relatively well constrained, whereas increased variability appears along the margins of the landfill and near the base of the model. These regions correspond to areas where the sensitivity of the IP data decreases and the inversion is more influenced by parameter trade-offs. The correlation matrix (Fig. 4.5e) reveals a similar pattern as per the synthetic example, with weak correlations between resistivity and chargeability parameters, suggesting that the two parameter fields are largely independently constrained by the data. Conversely, significant correlations are visible only for neighbourhood pixels affected by the filter effect.

4.4.2.3 Inversion results and uncertainty assessment for the 2025 dataset

The inverted models for the 2025 dataset (Fig. 4.6) reveal a reduction in the extent and intensity of the conductive–chargeable anomaly relative to the 2019 results. This spatial change is most pronounced in the central and basal portions of the landfill where the 2019 models had shown the strongest response. In the 2025 resistivity models (Figs. 4.6b,c), the conductive areas are restricted to two zones of perched leachate without a lateral continuity, suggesting a lower degree of leachate accumulation after installation of the drainage system. The IP results (Figs. 4.6e,f), show a more spatially coherent pattern, meaning that the main chargeability anomalies are laterally continuous and less affected by isolated high-amplitude artefacts. This is

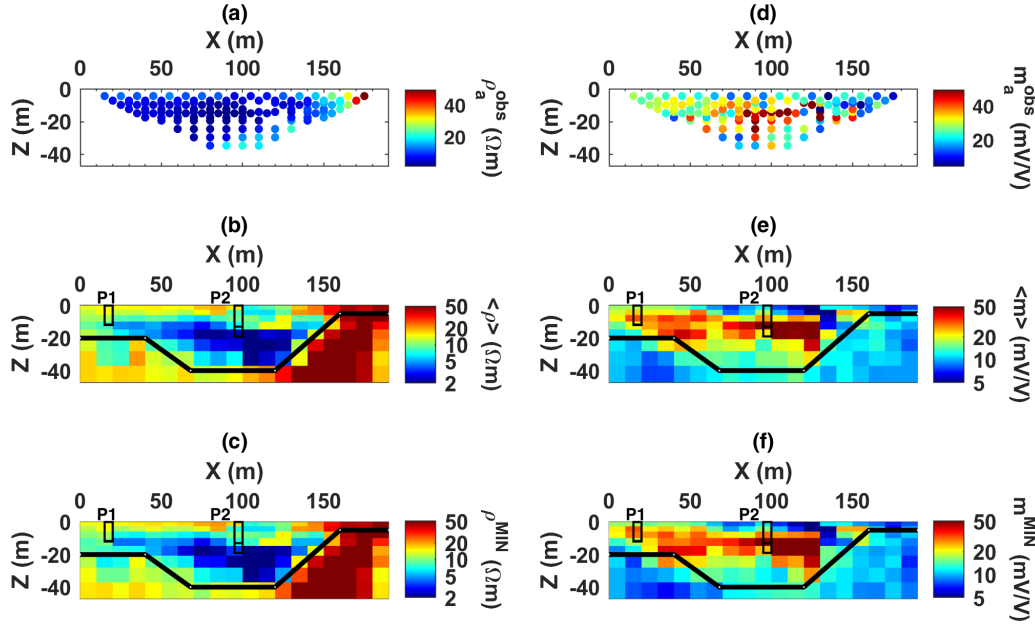


Figure 4.4. 2019 field dataset: (a) apparent resistivity pseudo-section, (b) mean resistivity model, (c) minimum misfit resistivity model, (d) apparent chargeability pseudo-section, (e) mean chargeability model, (f) minimum misfit chargeability model. The estimated liner position is marked with a white solid line. The P1 and P2 wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.

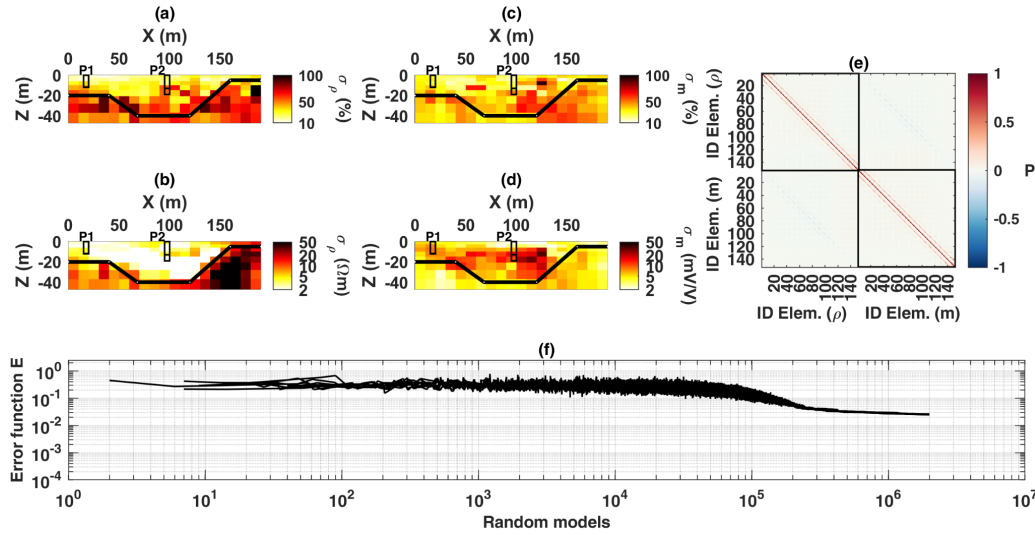


Figure 4.5. 2019 field dataset: uncertainty estimation in percentage values (a) and absolute values (b) for resistivity, uncertainty estimation in percentage values (c) and absolute values (d) for chargeability, (e) posterior correlation matrix (elements are numbered by rows downward starting from the upper-left corner), (f) error function vs. number of iterations for the 10 independent runs. The estimated liner position is marked with a black solid line. The wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.

consistent with the improved quality of the IP dataset, and an overall reduction of chargeability, though also in this case the high chargeability tends to be well-gathered within the landfill boundaries. These changes are likely due to a reduced biochemical response associated with the waste mass over the years and a different type of IP data acquisition (50% vs. 100% duty-cycle). As per the 2019 case, both the minimum-misfit and mean models delineate a localized high-chargeability anomaly between $x = 100$ –130 m, confirming the proper convergence and overall reliability of the inversion results.

Uncertainty analysis (Fig. 4.7) shows low variance in the anomaly core and higher variance in peripheral zones and deeper regions where sensitivity is limited. These patterns mirror for resistivity (Figs. 4.7a,b) those observed in the 2019 dataset, demonstrating stability of the inversion framework and reproducibility of posterior characteristics across surveys. The locally higher percentage uncertainty near the landfill bottom and close to P2 is mainly associated with low parameter values and reduced depth sensitivity. For chargeability (Figs. 4.7c,d), an overall decrease of the standard deviation (both percentage and absolute) can be observed, likely due to an increase of IP data quality owing to the more advanced acquisition system.

The correlation matrix (Fig. 4.7e) remains weakly structured, with only modest off-diagonal terms apart of those induced by the filter, confirming that the algorithm does not introduce artificial parameter coupling so that observed differences between surveys can reflect physical changes. Finally, the convergence of the 10 runs is good also in this case (Fig. 4.7f), with a slight reduction of the minimum misfit compared to the 2019 dataset, owing to the improvement in the quality of IP data.

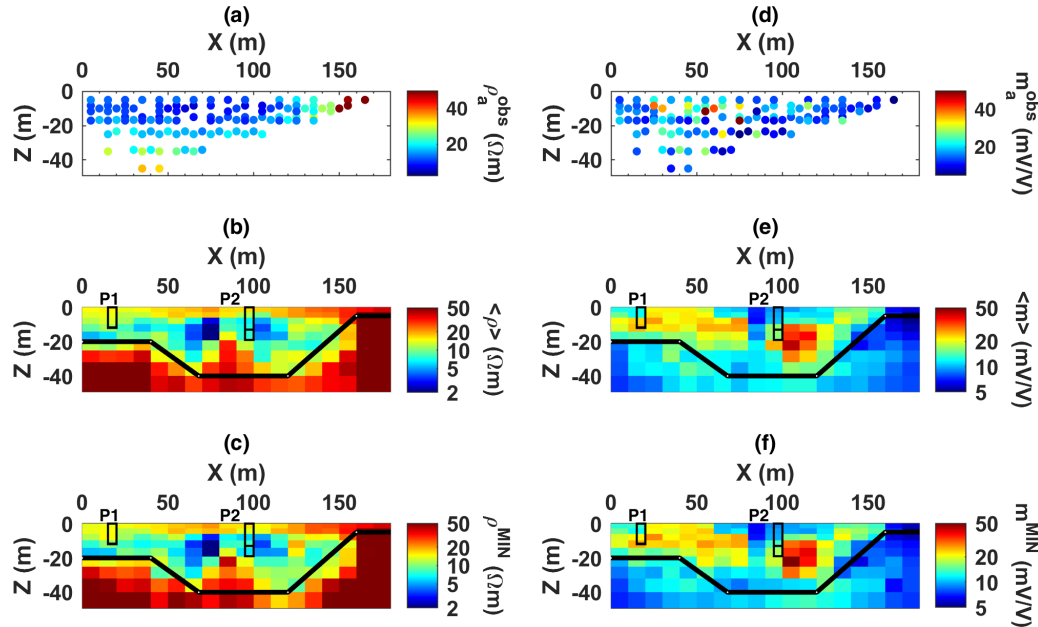


Figure 4.6. 2025 field dataset: (a) apparent resistivity pseudo-section, (b) mean resistivity model, (c) minimum misfit resistivity model, (d) apparent chargeability pseudo-section, (e) mean chargeability model, (f) minimum misfit chargeability model. The estimated liner position is marked with a white solid line. The P1 and P2 wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.

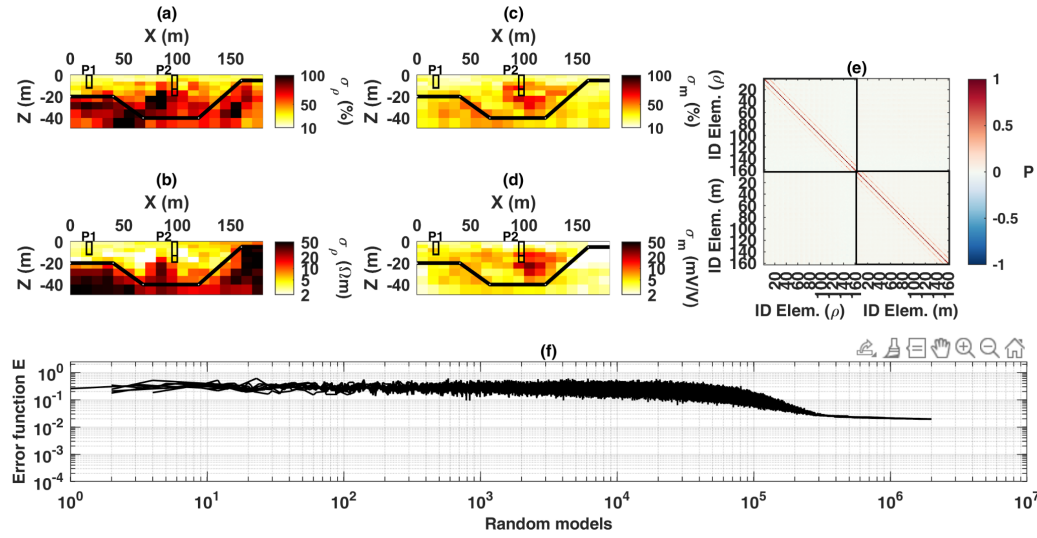


Figure 4.7. 2025 field dataset: uncertainty estimation in percentage values (a) and absolute values (b) for resistivity, uncertainty estimation in percentage values (c) and absolute values (d) for chargeability, (e) posterior correlation matrix (elements are numbered by rows downward starting from the upper-left corner), (f) error function vs. number of iterations for the 10 independent runs. The estimated liner position is marked with a black solid line. The wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.

4.4.2.4 Comparison of petrophysical models

The petrophysical models (Fig. 4.8) derived from the inversion results using Eqs. (4.15) and (4.16), provide a complementary view of the landfill evolution between 2019 and 2025. In 2019, both the estimated water content and the CEC (Fig. 4.8a,b) highlight a spread anomalous area in the central part of the landfill, consistently with the conductive–chargeable body identified in the geoelectrical models. In 2025, the water content model (Fig. 4.8c) shows a clear reduction in the extent and intensity of this anomaly, suggesting that the drainage system effectively decreased leachate accumulation. Meanwhile the CEC distribution (Fig. 4.8d) remains spatially coherent with the 2019 distribution, but appears more localized and separated in two parts, which is likely related both to the weaker biochemical activity in 2025 and to the higher quality of the new IP dataset. In fact, the more spread-out pattern observed in 2019 may partly reflect the lower quality and resolution of the IP dataset, whose uncertainty can propagate into the derived CEC estimate. Overall, the water-content models show strong agreement with the well observations in both 2019 and 2025, consistently matching the areas characterized by higher leachate levels along the profile. In conclusion, the petrophysical transformations are consistent with the inverted resistivity and chargeability models and confirm that the drainage system reduced leachate accumulation, whereas a residual zone characterized by high biochemical activity is still present at $x = 100\text{--}130\text{ m}$ within the waste body.

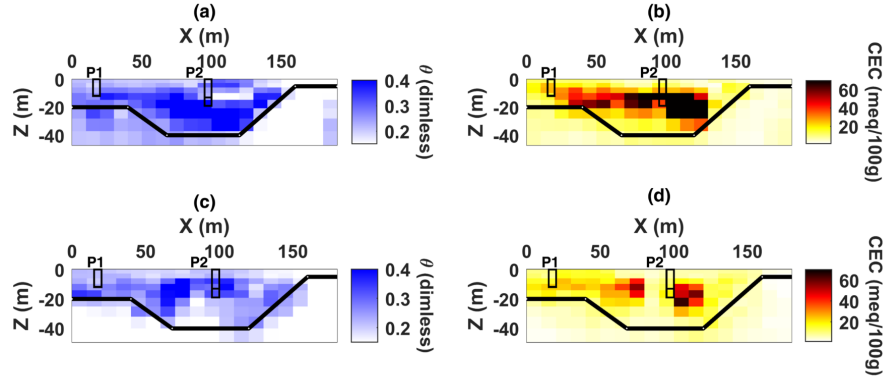


Figure 4.8. Petrophysical models achieved at the landfill site using the relations after Revil et al. (2020): (a) 2019 water content, (b) 2019 CEC, (c) 2025 water content, (d) 2025 CEC. The estimated liner position is marked with a black solid line. The P1 and P2 wells are marked with black rectangles, while the leachate levels logged with a black horizontal line.

4.5 Discussion

4.5.1 Uncertainty assessment and posterior analyses

The VFSA global inversion explores a broad region of the model space before converging toward high-likelihood solutions. This behaviour contrasts with deterministic inversion schemes based on local linearization (Constable et al., 1987; de Groot-Hedlin and Constable, 1990), where uncertainty estimates are derived from the curvature of the objective function around a single minimum. Such linearized diagnostics provide only local information and do not capture alternative models with comparable data fit or possible multimodality of the posterior space (Tarantola, 2005; Mosegaard & Tarantola, 1995).

The uncertainty maps derived from the posterior ensemble (Figs. 4.2, 4.5 and 4.7) are consistent with the sensitivity distribution imposed by the acquisition geometry. In both synthetic and field cases, uncertainties are lowest in shallow regions characterized by high data coverage, and increase with depth and toward model boundaries. These patterns reflect the intrinsic resolution limits of surface electrical methods and the presence of parameter trade-offs in poorly constrained regions (Oldenburg & Li, 1999; Günther et al., 2006). The consistency between synthetic and field results further supports the robustness of the inversion framework.

The posterior correlation matrices (Figs. 4.2e, 4.5e and 4.7e) provide additional insight into parameter coupling. Significant correlations are confined to neighbouring pixels of the same parameter class, producing localized structures attributable to the spatial filtering applied during inversion. In contrast, cross-parameter correlations between resistivity and chargeability remain weak across the model domain, indicating that the inversion does not introduce artificial coupling and that the two parameter fields are independently constrained by the data.

A key aspect is the distinction between the exploratory phase of the algorithm

and the statistical characterization of the posterior solution. Variance computed over the full ensemble reflects algorithmic exploration and should not be interpreted as model uncertainty. In this study, uncertainty descriptors are derived exclusively from models satisfying a misfit threshold, ensuring that the resulting statistics represent admissible solutions consistent with the data (Sharma, 2012). For the landfill dataset, the stability of the posterior basin indicates that the conductive–chargeable anomalies observed in both the 2019 (Fig. 4.4) and 2025 (Fig. 4.6) surveys are robust features rather than inversion artefacts. In several regions, differences between the two surveys exceed the posterior variability, supporting their interpretation as genuine temporal changes following the installation of the drainage system.

4.5.2 Comparison with alternative stochastic inversion approaches

A range of stochastic approaches has been proposed to address the nonlinearity and non-uniqueness of geophysical inverse problems. Among these, Markov chain Monte Carlo (MCMC) methods provide a rigorous Bayesian framework for sampling posterior distributions (Mosegaard & Tarantola, 1995; Tarantola, 2005). While theoretically comprehensive, their application to high-dimensional problems such as full 2D electrical imaging remains computationally demanding, often requiring long chains and carefully tuned proposal strategies. Population-based algorithms such as genetic algorithms (GA) and particle swarm optimization (PSO) have also been widely applied (Sen & Stoffa, 2013). These methods are effective in exploring complex objective-function landscapes but typically require large populations and a substantial number of forward simulations, limiting their scalability for dense model parametrizations.

The VFSA approach adopted here represents a computationally efficient alternative for a multi-parameter model, based on an adaptive annealing schedule that accelerates convergence relative to classical simulated annealing (Ingber, 1989). The consistent misfit evolution observed across independent runs (Figs. 4.2f, 4.5f and 4.7f) indicates stable convergence toward comparable solutions, while the ensemble of accepted models samples a coherent posterior basin. Although VFSA does not aim at exhaustive posterior sampling, it provides a representative ensemble of admissible models at a significantly lower computational cost. In the present application, this balance enables reliable uncertainty estimation and parameter correlation analysis within a full 2D framework, making VFSA a practical compromise between deterministic inversion and fully Bayesian approaches. Further discussion on VFSA parameter selection can be found in Cercato and De Donno (2023).

4.5.3 Implications for environmental geophysics

The landfill case study demonstrates how ensemble-based stochastic inversion improves geophysical interpretation by explicitly quantifying model uncertainty. Unlike deterministic approaches, which provide a single optimal solution (Loke et al., 2003; Oldenburg & Li, 1999), the proposed framework captures the range of admissible models consistent with the data. This allows distinguishing between well-constrained regions and areas affected by limited resolution, supporting more robust and transparent interpretations. The time-lapse comparison between the 2019 and 2025

surveys (Figs. 4.4 and 4.6) highlights the practical value of this approach. The reduction in conductive anomalies is consistent with the installation of the drainage system and indicates a decrease in leachate accumulation. Differences observed in chargeability should instead be interpreted with caution, as they may reflect both changes in acquisition conditions and temporal evolution of the landfill material (Olsson et al., 2015; Kemna et al., 2012).

From a geotechnical perspective, leachate accumulation strongly influences pore pressure and effective stress within the waste mass (Koerner and Soong, 2000). However, leachate flow in municipal solid waste landfills is inherently complex and spatially heterogeneous (Zeiss and Uguccioni, 1994), making it difficult to characterize using only point-scale measurements from monitoring wells. In this context, the spatially distributed information provided by geophysical inversion represents a significant improvement over conventional interpolation-based approach. The models obtained in this study provide a spatially coherent description of leachate distribution that can complement piezometric data and be directly incorporated into geotechnical analyses as heterogeneous saturation inputs. This is particularly relevant for slope landfills, where localized fluid accumulation may significantly affect stability conditions.

Although the present results are derived from a two-dimensional (quasi-2.5D) framework, they highlight the potential of uncertainty-aware geophysical imaging for constraining subsurface fluid distribution. Future developments should therefore focus on fully three-dimensional acquisition and inversion strategies, enabling volumetric estimation of leachate and improved integration between geophysical monitoring and geotechnical modelling. More broadly, the proposed framework is applicable to a wide range of environmental problems characterized by strong heterogeneity and limited prior information, where uncertainty quantification is essential for reliable interpretation.

4.5.4 Petrophysical interpretation

The interpretation of resistivity and chargeability in landfill environments requires careful consideration of their underlying petrophysical controls and of the different physical processes governing the two parameters. While resistivity is primarily sensitive to fluid content and conductivity, induced polarization reflects interfacial electrochemical processes and biogeochemical activity within the waste mass (Slater & Lesmes, 2002; Binley & Kemna, 2005). In particular, integral chargeability tends to decrease with increasing saturation due to enhanced fluid and surface conductivity, which dampens polarization effects (Flores-Orozco et al., 2020). Conversely, elevated chargeability values are typically associated with biogeochemically active zones characterized by high organic content and microbial activity (De Donno et al., 2025). Joint interpretation of resistivity and chargeability helps to reduce this ambiguity by exploiting complementary sensitivities. This is achieved through established petrophysical relationships linking electrical properties to hydrogeological and geochemical parameters, here expressed in terms of integral chargeability (Revil et al., 2020), although alternative formulations based on spectral parameters are also available (Weller et al., 2013).

The petrophysical models (Fig. 4.8) reveal two distinct anomalies associated with

high water content and elevated cation exchange capacity (CEC), respectively. Their spatial separation indicates that zones of high saturation and enhanced biological activity are not coincident, but rather distributed within different portions of the waste body. The agreement between inverted models and piezometric observations further supports the validity of the approach in highly heterogeneous landfill environments.

It should be noted, however, that the present interpretation relies on empirical relationships. Improved calibration of key parameters, such as the cementation exponent and waste density, through laboratory measurements would enhance the quantitative reliability of the derived estimates. More generally, these results highlight the importance of multiparameter inversion frameworks for disentangling overlapping physical processes that cannot be resolved using resistivity data alone.

4.6 Conclusions

Results of the global inversion of ERT and TDIP data on a full two-dimensional mesh demonstrate the capability of the proposed VFSA algorithm to explore the model space efficiently and converge toward stable regions of the objective function. The method preserves sharp resistivity contrasts while providing an ensemble of admissible models that allows statistical assessment of posterior uncertainty.

The VFSA inversion was tested on both synthetic and field examples related to the investigation of municipal solid waste landfills, where prior information is available for validating the geophysical models and where deterministic local inversions may produce biased representations of anomaly geometry. In both cases, the stochastic inversion improves the delineation of the geophysical targets, particularly along the lateral and basal boundaries of the landfill where sensitivity is reduced and data coverage is limited. Improvements are also observed in the shallowest layers, where the ensemble of accepted models consistently reproduces the main structures of the subsurface.

A key advantage of the proposed approach is the possibility of performing a rigorous posterior uncertainty analysis based on the ensemble of models satisfying the misfit criterion. The spatial distribution of uncertainty closely follows the sensitivity pattern of the acquisition geometry, with lower uncertainty in shallow layers and higher values toward the base of the model and near anomaly margins. This information provides a quantitative measure of interpretative reliability and helps distinguish robust model features from structures that are less constrained by the data.

Analysis of the posterior covariance and correlation structure for the synthetic landfill model further indicates that strong correlations occur mainly between neighbouring pixels belonging to the same parameter class, reflecting the spatial filtering effect. In contrast, cross-parameter correlations between resistivity and chargeability remain weak, suggesting that the joint inversion does not introduce significant artificial coupling between the two parameters and that their estimates remain largely independent within the posterior ensemble.

Although the computational effort required for stochastic inversion of field datasets remains significant when performed on standard computing systems and without incorporating prior constraints, ongoing developments in parallel computing

and forward modelling efficiency are expected to substantially reduce these costs. The results presented here indicate that stochastic global inversion combined with ensemble-based uncertainty analysis represents a promising framework for multi-parameter geophysical imaging and for environmental investigations where robust assessment of model reliability is essential.

4.7 Final remarks

The results presented in this chapter also provide an opportunity to clarify the practical meaning and limitations of the VFSA-based uncertainty analysis in landfill applications. As for other simulated-annealing strategies, convergence to the true global minimum is guaranteed only under ideal theoretical conditions, requiring an infinitely slow cooling schedule and an infinite number of iterations. Therefore, the objective of the implementation used in this thesis is not to prove convergence to a unique global minimum, but to explore the model space more broadly than local gradient-based inversion and to identify an ensemble of low-misfit models compatible with the data and the imposed bounds. In this case, models with nearly equivalent misfit are not treated as competing single solutions, but as alternative plausible representations of the subsurface compatible with the data. For this reason, the lowest-misfit model is shown only as a representative realization of the inversion result and is interpreted together with the ensemble mean and the associated uncertainty descriptors. Model equivalence is therefore addressed by analysing the variability of the accepted models and by deriving posterior statistics, rather than by selecting a single deterministic solution as the only valid interpretation.

The maximum number of VFSA steps was used as the main stopping criterion to ensure a consistent comparison among the different runs. Other criteria could also be implemented, such as the stabilization of the objective function or the convergence of selected ensemble statistics. These criteria may reduce the computational cost, but they require careful calibration to avoid stopping the sampling before the relevant part of the model space has been adequately explored. Their implementation represents an interesting future development for improving the efficiency of the VFSA framework.

The model discretization was defined as a compromise between spatial resolution and computational cost. The horizontal cell size was set equal to the electrode spacing, whereas the vertical cell size was gradually increased with depth. This choice provides an adequate resolution in the near-surface and central part of the model, where data sensitivity is higher, while reducing the computation time required by the VFSA inversion. As a result, the deepest parts of the model under the estimated liner position are less resolved and must be interpreted with greater caution, especially where uncertainty is higher.

The VFSA inversion is computationally more demanding than local deterministic inversion because it requires a large number of forward evaluations. In the applications presented here, the number of model parameters was twice the number of model cells, since both resistivity and chargeability were inverted. The number of cells ranged from about 100 to 150 between the synthetic and field cases, corresponding to approximately 200–300 model parameters. Each run was performed using $2 \cdot 10^6$

model perturbation steps. The typical computation time ranged from about 24 h for the synthetic scenario to 48 h for the field cases on a workstation equipped with Intel Xeon(R) Silver 4210R CPU at 2.40GHz with 32 GB RAM and using 10 cores, one for each independent run. A direct comparison with other stochastic inversion algorithms is not provided because computation times depend strongly on the forward solver, parametrization and hardware implementation. VFSA was adopted here because it offers a simple and flexible framework to explore a broad model space with limited a priori assumptions and to obtain an ensemble of plausible models for uncertainty assessment, although future developments should compare its performance with alternative global strategies under controlled conditions. The computational cost also defines the present scalability of the method. Its direct application to very high-resolution 2D datasets would require adaptive parametrization or a reduction of the effective parameter space using prior information. For this reason, the present implementation is a solution for frameworks based on uncertainty-aware interpretation rather than as a routine replacement for fast local inversion in all practical cases. The use of 2D surveys remains an intrinsic limitation in highly heterogeneous landfill bodies, where structures orthogonal to the acquisition tomography may affect the measured response. This limitation was partly assessed in Chapter 5 through the comparison of geometric factors, which provided a practical check of the consistency between the 2D acquisition layout and the expected sensitivity of the measurements.

Finally, the uncertainty descriptors derived from the accepted ensemble have a direct practical use. They allow the interpreter to distinguish anomalies from poorly constrained features and to avoid interpreting areas where the model variability is high. In this thesis, uncertainty information is presented separately from the resistivity and chargeability models, but future developments could integrate these descriptors directly into final interpretative maps, as is the case with the membership value for Chapter 5.

Chapter 5

Machine-learning techniques for integrated interpretation

5.1 Paper #3: A machine learning-based approach for mapping leachate contamination using geoelectrical methods

Ester Piegari^{1,*}, Giorgio De Donno², Davide Melegari², and Valeria Paoletti¹

¹ Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università degli Studi di Napoli Federico II, Naples, Italy

² Dipartimento di Ingegneria Civile Edile e Ambientale, "Sapienza" Università di Roma, Rome, Italy

* Corresponding author

The manuscript presented in this chapter is a selected abstract of the published version on Waste Management: 10.1016/j.wasman.2022.12.015.

5.1.1 Abstract

Leachate is the main source of pollution in landfills and its negative impacts continue for several years even after landfill closure. In recent years, geophysical methods are recognized as effective tools for providing an imaging of the leachate plume. However, they produce subsurface cross-sections in terms of individual physical quantities, leaving room for ambiguities on interpretation of geophysical models and uncertainties in the definition of contaminated zones. In this work, we propose a machine learning-based approach for mapping leachate contamination through an effective integration of geoelectrical tomographic data. We apply the proposed approach for the characterization of an urban landfill. We perform a multivariate analysis on datasets consisting of electrical resistivity, chargeability and normalized chargeability (chargeability-to-resistivity ratio) data extracted from previously inverted model sections. By executing a K-Means cluster analysis, we find that the best partition of the dataset contains ten classes. From such classes and also introducing a distance-based colour code, we get updated cross-sections and provide an easy and less ambiguous identification of the leachate accumulation zones. The latter turn out to be characterized by coordinate values of cluster centroid less than $3\Omega\text{m}$ and larger than 27mV V^{-1} and 11mS m^{-1} . Our findings, also supported by borehole data, show that the combined use of geophysical imaging and unsupervised machine learning is promising and can yield new perspectives for the characterization of leachate distribution and pollution assessment in landfills.

Keywords: Leachate contamination detection, Machine Learning, K-Means clustering Geophysical imaging, Electrical Resistivity Tomography, Induced Polarization Tomography

5.1.2 Introduction

Although there is an increased awareness on the importance of environment protection, urban waste management is one the most important environmental issues. In many countries there is still a little use of recycling or reuse actions, and the majority of municipal solid waste is destined for landfills (WHO, 2015). Waste decomposition

generates leachate that is a highly contaminated liquid consisting of a mixture from organic degradation products, liquid waste and rainwater. Leachate infiltration causes serious environmental issues to groundwater and soils and, therefore, identifying and monitoring its flow pathways has major implications on designing a risk mitigation strategy (Mukherjee et al., 2015; Lavagnolo et al., 2019; Vaccari et al., 2019; Morita et al., 2021; Ergene et al., 2022). To this aim, geophysical methods often represent the only cost-effective, rapid and non-invasive choice for mapping large areas, such as those encountered in urban landfills, down to tens of meters (e.g., Di Maio et al., 2018).

In last decades, many studies have demonstrated that geoelectrical methods can be effective in identifying landfill leachate (e.g., Soupios et al., 2007; Abdulrahman et al., 2016; Bichet et al., 2016; Raji and Adeoye, 2017; Flores-Orozco et al., 2020; Zaini et al., 2022). The geoelectrical devices work through injection in the ground of a direct-current. The measurements of the resulting voltage at different distances from the source, before and after the current switch-off, retrieve information about the resistive and capacitive behaviour of the subsurface (Loke et al., 2013). As leachate is a fluid with high concentrations of ions, it can be successfully imaged by low values of electrical resistivity ρ and high values of chargeability M , sensed respectively by electrical resistivity tomography (ERT) and induced polarization (IP) surveys (Everett, 2013). The contrast between the electrical properties of leachate and those of the surrounding media generally facilitates its identification. However, geophysical inverted models leave ambiguities in defining contaminated zones, particularly in presence of clayey soils, which are often placed at the bottom of the landfill as a low-permeability barrier in combination with a synthetic liner (geomembrane). Therefore, the zones characterized by the highest values of chargeability frequently do not match with the most conductive ones, and the question of how to combine information from different geophysical methods is still open.

Many studies consider the so-called normalized chargeability, M_n (ratio of M and ρ), as it is directly linked to surface polarization (e.g., De Donno and Cardarelli, 2017a; Soupios et al., 2017; Power et al., 2018). High values of M_n are generally related to leachate contamination, even though M_n can be also largely and significantly affected by clay content (Slater and Lesmes, 2002). Additionally, both resistivity and chargeability are related to the saturation levels, and to pick up the threshold of M_n for fully saturated zones (the most dangerous ones) remains a subjective choice. In fact, focusing only on the highest values of M_n may lead to false-positive results in the identification of leachate, as they may not arise from the concomitance of low ρ and high M values. For instance, large values of normalized chargeability may be caused by only large M related to the presence of pockets of clays with high-chargeability or by extremely low resistivity values of leachate-saturated zones. Thus, a residual uncertainty persists in properly identifying leachate accumulation zones.

Over the past decade, different methods for coupled inversions of resistivity data have been proposed to reduce model uncertainties (Audebert et al., 2014; Hellman et al., 2017; Varfinezhad et al., 2022). The objective of this study is to combine ρ , M and M_n inverted data to get one comprehensive cross section integrating information from all geoelectrical data and identifying leachate accumulation zones in an easy and less ambiguous way. To reach this aim, we propose the application of Machine

Learning (ML) algorithms.

ML is a branch of artificial intelligence based on the idea that systems can learn from data, and this learning comes from analysis of data through statistical tools to make predictions or finding patterns in data (Zhang et al., 2022). In recent years, due to the high efficiency in managing and classifying large datasets, clustering algorithms have been applied to a growing number of different fields in geosciences, such as petrophysics (Abdideh and Ameri, 2020), remote sensing and climate science (Lyra et al., 2014; Karpatne et al., 2019; Straus, 2019), geothermal systems (Lindsey et al., 2018; Bernardetti and Bruno, 2019), fault network reconstruction and seismic sequences (Kamer et al., 2020; Cesca, 2020; Piegari et al., 2022).

There are many different types of clustering algorithms, which can be broadly categorised into three types: i) partitioning algorithms, which divide the dataset into a number of groups (clusters) and require the number of clusters as an input data; ii) hierarchical algorithms, which provide a tree-based representation of data points (dendrogram) showing the hierarchical relationship between clusters; iii) density-based algorithms, which groups data points on the basis of their spatial density.

In the frame of leachate pollution assessment, Rana et al. (2018) and Sharma et al. (2020) proposed hierarchical cluster analysis to compute a leachate pollution index, combining information from fifteen physico-chemical parameters. In this study, we use the K-Means algorithm, which is one of the most common partitioning algorithms. In addition to being a fast, robust and simple iterative algorithm, it has the advantage of producing tighter (spherical) clusters than hierarchical and density-based clustering (Shukla and Naganna, 2014; Zhang et al., 2022). K-Means algorithm was used to cluster other geophysical data such as georadar velocity and attenuation (Tronicke et al., 2004) and electrical resistivity and P-wave velocity (Di Giuseppe et al., 2014; Hellman et al., 2017; Di Giuseppe et al., 2018; Bernardetti and Bruno, 2019).

In this study, we present K-Means first application to ERT and IP data. In the following sections, we describe the proposed methodology and show the results of geophysical surveys performed into an urban waste landfill. Then, we report and discuss the results of the K-Means clustering analysis.

5.1.3 Methods

We illustrate our methodological approach in Fig. 5.1.

Traditionally, to characterize leachate contamination by means of geophysical methods, ERT and IP surveys are carried out with the aim of retrieving individual cross sections showing the distribution of ρ , M and M_n after data inversion. In this work, we propose going beyond traditional approaches to get integrated cross sections of geoelectrical data. They do not show the distribution of any specific physical quantity in the subsurface but provide an imaging of the leachate accumulation zones on the basis of triplets of values of ρ , M and M_n . To reach this aim, after inverting the geoelectric data, we combine them in a joint parameter space and perform a K-Means cluster analysis. We group data representing triplets of values of ρ , M and M_n based on their similarity (i.e., Euclidean distance in the joint parameter space) and identify the best partition of the dataset in different classes through

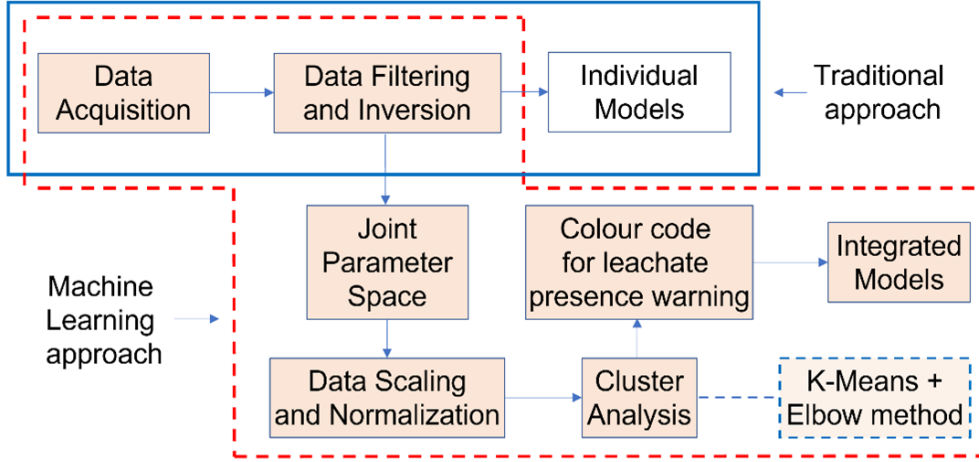


Figure 5.1. Sketch of the traditional and ML based approaches. The red dotted line encloses the proposed procedure.

the elbow method. Finally, we define a colour code to easily identify the warning zones among the different clusters and make an integrated imaging of the electrical properties of the subsoil. In the following subsections, we give details about the inversion procedure and the implemented cluster analysis.

5.1.3.1 Geophysical imaging

5.1.3.1.1 Forward modelling

The resistive response of a 2.5D subsoil (where the conductivity varies only in the x - z plane), is described within a domain D by the Fourier-transformed Poisson's equation under the hypothesis of an external point source located at (x_s, z_s) (Dey and Morrison, 1979):

$$-\nabla \cdot [\sigma(x, z) \nabla \phi(x, z, \lambda)] + \lambda^2 \sigma(x, z) \phi(x, z, \lambda) = I \delta(x_s) \delta(z_s) \quad \forall (x, z) \in D \quad (5.1)$$

where σ is the conductivity ($\rho = 1/\sigma$ is the resistivity), ϕ the transformed electric potential, λ the transformed variable and I the injected current.

Eq. (5.1), subjected to Dirichlet's and Neumann's boundary conditions on surface and lateral and bottom boundaries respectively, is solved numerically. A widespread used technique is the Galérkin formulation of the Finite Element Method (De Donno and Cardarelli, 2017b), which is employed in this work. We used a mesh with rectangular elements with an element size in the x -direction equal to one-half of the electrode spacing and an element size in the z -direction decreasing from one-quarter of the electrode spacing (first row at surface) to one electrode spacing (last row – bottom of the model).

Once the solution of Eq. (5.1) is achieved, potential is back-transformed and the apparent resistivity can be predicted as:

$$\rho_a^{pre} = C \frac{\Delta V}{I}, \quad (5.2)$$

where C is the geometric factor and ΔV the potential difference, both obtained depending on the specific quadrupole sequence.

The capacitive response of a medium can be assessed in the time-domain through the chargeability M , which is proportional to the induced polarization (IP) phenomena occurring in subsoil due to the switch-on or switch-off of an external direct current (DC) source (Seigel, 1959).

In this case, the IP forward solution is given sequentially with the resistivity modelling by calculating the potential V_M resulting from solution of Eq. (5.1) but with the conductivity replaced by $\alpha = \sigma(1 - M)$.

The predicted apparent chargeability M_a is then found as (Oldenburg and Li, 1994):

$$M_a^{pre} = \frac{V_M - V}{V_M}. \quad (5.3)$$

5.1.3.1.2 Data inversion

Time-domain measurements are performed employing a DC electrical source, measuring the apparent resistivity ρ_a and apparent integral chargeability M_a , resulting from the potential decay after current switch-off (Binley and Slater, 2020):

$$\rho_a^{obs} = C \frac{V_p}{I}, \quad (5.4a)$$

$$M_a^{obs} = \frac{\int_{t_i}^{t_f} V_i dt}{V_p \Delta t}, \quad (5.4b)$$

where V_p is the measured voltage during application of the DC current I and V_i the residual voltage after switch-off the electrical current, integrated over a time window Δt defined between times t_i and t_f . Usually, the time window is divided into a few shorter logarithmically spaced gates (often 20), and the integral in Eq. (5.4b) is computed by sum of values achieved for each gate.

Apparent values are inverted for resistivity and chargeability using a Gauss-Newton iterative 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 values in the inverted models. The goodness of fit is evaluated for each line in terms of mean absolute percentage error for resistivity, while for chargeability models is more convenient to use the mean absolute error, expressed in mV/V (De Donno and Cardarelli, 2017b).

The contribution of the surface conduction mechanisms, quantified by the normalized chargeability M_n (Slater and Lesmes, 2002), can be calculated at the end of the inversion process by:

$$M_n = \frac{M}{\rho}. \quad (5.5)$$

5.1.3.2 K-Means clustering

K-Means is a partitioning algorithm that divides the dataset into K predefined non-overlapping groups (clusters) of similar data. The similarity measure is based on a distance-based metric that is typically the Euclidean distance. The algorithm starts assigning a set of K means $\mu_1, \mu_2, \dots, \mu_K$, also called centroids, computes the distance between each point and the K centroids, and assigns each point to the cluster i with the closest centroid μ_i . Actually, it assigns data points to a cluster in order to minimize the sum of the squared distance (SSE) between the data points and the cluster's centroids (Bhattacharya et al., 2021):

$$SSE = \sum_{i=1}^K \sum_{n=1}^N r_{ni} \|d_n - \mu_i\|_2^2, \quad (5.6)$$

where $\|\cdot\|_2$ is the Euclidean L_2 norm, d_n denote the data points, N is the total number of points in the dataset and r_{ni} is a binary variable equal to 1 if d_n is assigned to cluster i and 0 otherwise. Once each point has been assigned to a centroid, the procedure is iterated, i.e., the centroid of each cluster is recomputed as the mean of all data points belonging to the same cluster, and the category of each point is adjusted again until a maximum number of iterations is reached, or the adjustment range is less than a given threshold.

K-Means is a very powerful algorithm that, in addition to its simplicity, has the advantage of having convergence guaranteed, as at each iteration step SSE always decreases (Bhattacharya et al., 2021). However, K-Means finds local minimum of SSE and different initial positions of centroids determine different cluster solutions. To overcome this problem, the algorithm is run many times placing the centroids in different random starting points, recording the variance at each step and selecting the configuration corresponding to the minimum variance. Another drawback of the algorithm is that the clustering depends on the number of K that needs to be specified in advance. The optimal value of K is found by elbow method, if there is not any geologic *a priori* information to constrain the number of clusters (Bhattacharya et al., 2021). This method consists in varying the number of clusters K , then computing the percentage of the explained variance EV for each trial:

$$EV_j = \frac{\sum_{i=1}^j (SSE_i - SSE_{i+1})}{\Delta SSE_{\max}}, \quad j = 1, 2, \dots, K - 1 \quad (5.7)$$

where $\Delta SSE_{\max}$ corresponds to the SSE deviation between 1 and the highest K among those analysed.

The idea is that when increasing K , at some point the addition of another cluster does not significantly improve the modelling of the data, and, thus, the best value of K is chosen as the elbow of the EV curve (Thorndike, 1953). Our cluster analysis was performed by using software packages available in the Statistics and Machine Learning Toolbox of MATLAB.

5.1.4 Site location and geophysical measurements

The proposed approach was applied to an urban landfill, located in central (Fig. 5.2) Italy.

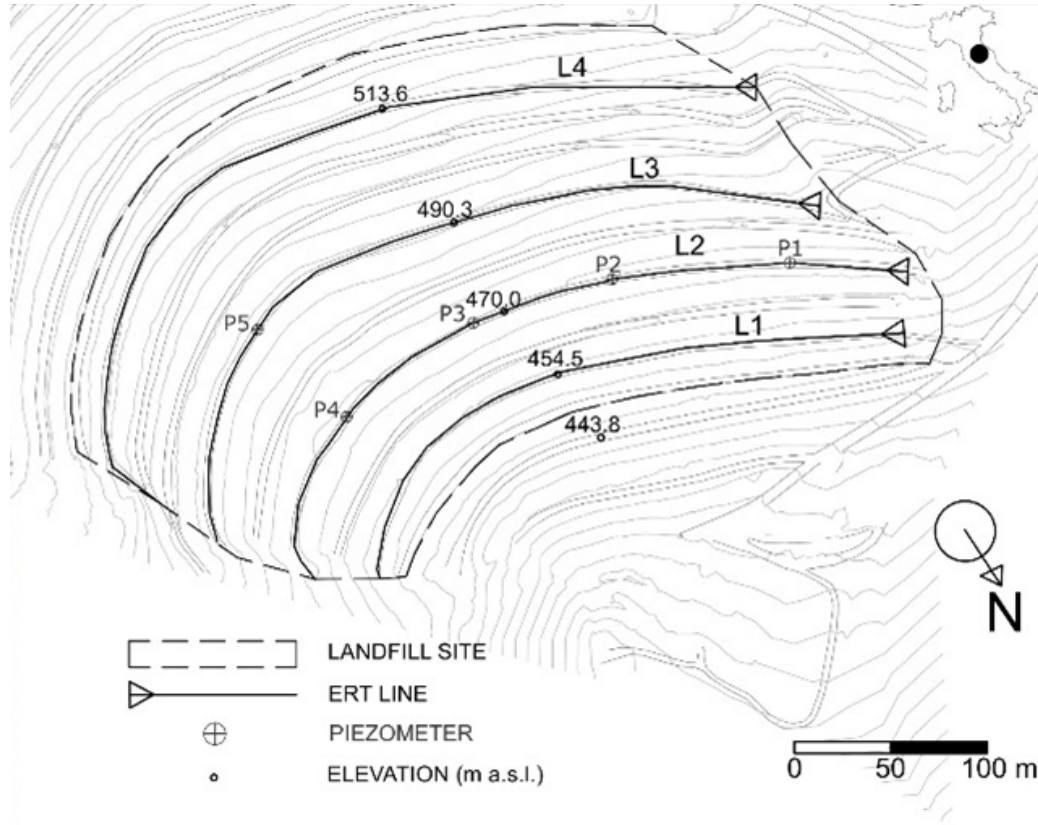


Figure 5.2. Map of the study area for the case study (landfill in central Italy).

The investigated area is a landfill built in the '80s for being used as municipal waste disposal of a medium-size city located in Central Italy. For this landfill, a few more information about the original design is available. The landfill site (Fig. 5.2b), located on a steep slope (slope percent around 40%), was provided with a bottom liner (geo-membrane) overlying the in situ marly-arenaceous flysch. An embankment was built at the bottom (elevation around 445 m a.s.l.) to prevent slope instability phenomena. Geoelectrical investigation was planned for reconstructing the landfill depth and evaluating the leachate accumulation since it could reach a level that may trigger a slope failure. The supposed depth of the buried waste is greater upstream (southern part, elevation: 510-520 m a.s.l.), while reducing downstream (northern zone, elevation: 460-450 m a.s.l.).

The geophysical campaign encompasses four electrical profiles spaced approximately 40-50 m apart (L1 to L4), using the road tracks built for site management (Fig. 5.2b). Five wells are located along L2 and L3 and piezometric levels were logged during the campaign to validate the geophysical models. Time-domain ERT and IP data were acquired by 48 electrodes spaced 5 m apart, using the IRIS Instruments Syscal Pro resistivitymeter. We employed the dipole-dipole array for data acquisition,

as it combines significant depth of investigation and good lateral resolution needed to image the leachate variations. In fact, laying cables outside the landfill for set-up a pole-dipole configuration was unfeasible in such a steep and complex environment. For covering the long lines, we used the roll-along technique by overlapping 36 electrodes for each baseline and data were filtered for outliers, negative DC and/or IP voltage values or decay curves with increasing voltage.

5.1.5 Inversion results

The inverted models are shown in Figs. 5.3a-d (resistivity) and Figs. 5.3e-h (chargeability).

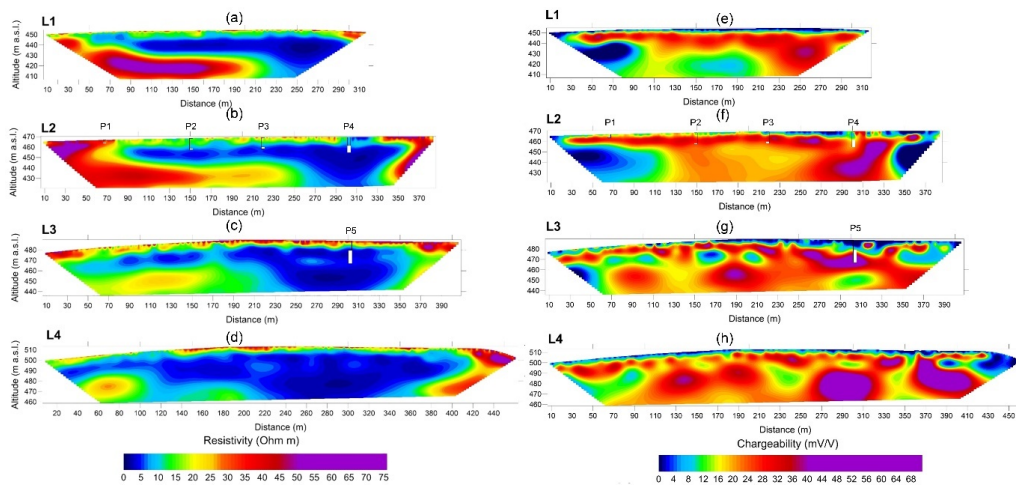


Figure 5.3. Resistivity (a-d) and chargeability (e-h) models for the case study. Absolute errors are 2.2%, 4.2%, 7.4% and 4.5% for resistivity models and 1.1 mV/V, 2.1 mV/V, 3.7 mV/V, and 4.6 mV/V for chargeability models. The piezometric levels in wells (white filled areas) are superposed to the models.

The ERT and IP sections were computed in 2.5D mode by linearization of the curvilinear profiles. The error committed, calculated as the percentage deviation between 2D and 3D geometric factors, is between 0.5% (L1) and 1.5% (L4), due to the low curvature of profiles. We superposed the piezometric levels logged in the available wells to the electrical models (white filled areas in Fig. 5.3), in order to validate the geophysical results.

We reconstruct for the four profiles a three-layer model from top to bottom: i) landfill covering and unsaturated waste, having an average thickness of about 8-10 m, resistivity higher than 15-20 Ωm and chargeability close to zero, with local variation of thickness and resistivity due to the high heterogeneity of the covering soil; ii) saturated waste (leachate), with resistivity lower than 10 Ωm and chargeability higher than 10 mV/V. Strong changes in resistivity and chargeability values are clearly seen throughout sections depending on composition and degree of saturation of the waste mass. The strongest geoelectrical anomalies ($\rho < 3 \Omega\text{m}$ and $M > 35 \text{ mV/V}$) are located in the deepest zones of the landfill between $x = 240 - 350 \text{ m}$ from the beginning of the lines; iii) bottom liner (geomembrane) overlying bedrock (marly-arenaceous flysch), with resistivity values higher than 10 Ωm and chargeability lower

than 10 mV/V. The resistivity values are lower than expected for such a dielectric material (geomembrane), because of the lack of sensitivity at greater depths (De Donno and Cardarelli 2017a). A series of sloped terraces is clearly visible, which ends in the deepest part of the landfill, where we detect the strongest geophysical response. Consequently, the maximum depth of the landfill can be estimated around 60-70 m upstream (L4) and 30-40 m downstream (L1).

Therefore, leachate likely accumulates at the bottom of the landfill and percolates downstream (from L4 to L1) through a preferential pathway, whose lateral extent progressively reduces downstream from about 80 m (L4, between 240 to 320 m) to about 30 m (L1, between 240-270 m). Chargeability models (Fig. 5.3e-h) enhance the significant heterogeneity in leachate accumulation, mainly for L3 and L4 lines (where accumulation zones are larger), because of changes in degree of saturation or in waste composition. The piezometric levels in P4 and P5 wells confirms the geophysical reconstruction, while significant discrepancies can be noticed between chargeability models and leachate levels for P1, P2 and P3. This residual ambiguity leads the way for a more quantitative integration of geophysical data, which encompasses the joint use of both inverted models as an input for the clustering analysis.

5.1.6 Cluster analysis results

The results of the K-Means cluster analysis in the joint parameter space defined by the inverted values of ρ , M and M_n are shown in Figure 5.4. The datasets contain a total number of 10466 datapoints. Since the inverted values of ρ , M and M_n span many orders of magnitude, we consider the $\log_{10}$ -transformed data rescaled in the range interval $[0, 1]$, with the following normalization:

$$x_{new} = \frac{\max(x_{new}) - \min(x_{new})}{\max(x_{old}) - \min(x_{old})} (x_{old} - \min(x_{old})) + \min(x_{new}) \quad (5.8)$$

where x_{old} is the original $\log_{10}$ -transformed value of ρ , M or M_n and x_{new} is the respective normalized value. The normalization allows us both to prevent one feature prevailing over the others and to facilitate the introduction of a colour code warning for leachate presence, as discussed below.

The clustering algorithm was iterated by varying K from 1 to 50 and for each run we considered up to 500 different initial configurations of centroids, choosing the configuration that minimizes the distortion (i.e., the sum of point-to-centroid distances). To retrieve the best number of clusters, K_{best} , we used the elbow method with an explained variance threshold equal to 95% (Figs. 5.4a).

The best partition of the 3D dataset was achieved with 10 different clusters (Fig. 5.4a). Looking at the distribution of data points in the parameter space (Fig. 5.4b), we note that the values of M span over a wide range. The darkest green cluster in Fig. 5.4b that groups data with a shape of straight line is related to the inequality constraints set on the chargeability model during inversion to enforce positiveness of chargeability (minimum value of 0.1 mV/V in this case).

The K-Means cluster analysis allows us to associate triplets of values (ρ , M , M_n) with cluster colour indexes. These latter are used to build integrated cross-section models for each of the investigated profiles. Since ERT and IP data are

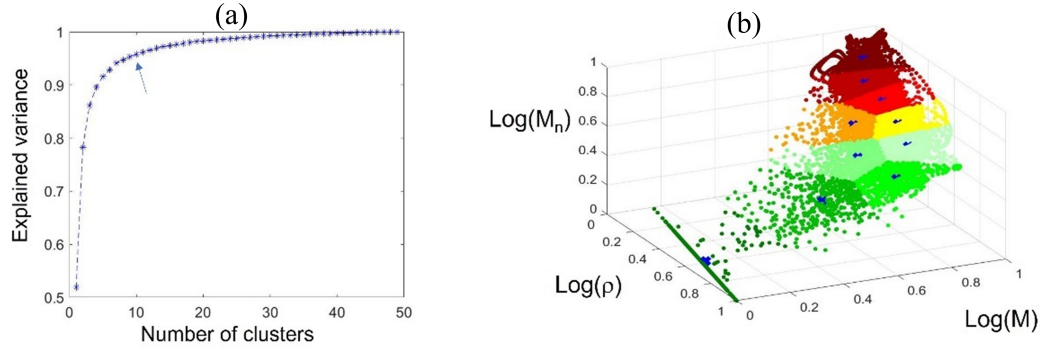


Figure 5.4. Results of the cluster analysis. Panel (a) : explained variance as a function of the number of clusters. The arrows indicate the best choice for K . Panel (b): scatterplot of $\log_{10}$ -transformed geoelectrical data in the normalized space, where clusters are marked by different colours. In each cluster, the location of the centroid is shown by a blue cross.

Table 5.1. Cluster-centroid coordinates reported for the case study. The colour labels are those used in the integrated sections.

Colour	ρ (Ω m)	M (mV V ⁻¹)	M_n (mS m ⁻¹)
Dark green	31.77	0.11	0.003
Green	30.30	1.89	0.06
Green	40.68	9.27	0.23
Green	19.42	4.62	0.24
Green	20.76	26.28	1.27
Yellow	14.02	21.74	1.55
Yellow	9.25	10.08	1.09
Orange	7.77	23.56	3.03
Red	4.59	23.19	5.05
Dark red	2.10	27.32	13.0

acquired along the same profiles, each of the depths of the vertical cross sections are characterized by values of ρ , M and M_n , which are in turn associated to a cluster colour. Thus, we get integrated sections by associating colour indices to depths. It is worth noting that the retrieved final sections differ from the traditional ones as they do not represent the distribution of any specific physical quantity in the subsurface. Instead, combining all information coming from the geophysical surveys, they illuminate a certain number of zones characterized by specific ranges of values of the electrical properties of the investigated subsurface. In the following section, the effectiveness of the proposed ML-based approach in identifying leachate accumulation zones is discussed by comparing the integrated sections with the traditional M_n models.

5.1.7 Discussion

In Figure 5.5 we compare the M_n models, achieved by the ratio of chargeability and resistivity of each pixel in Figs. 5.3, with the sections retrieved by the cluster analysis.

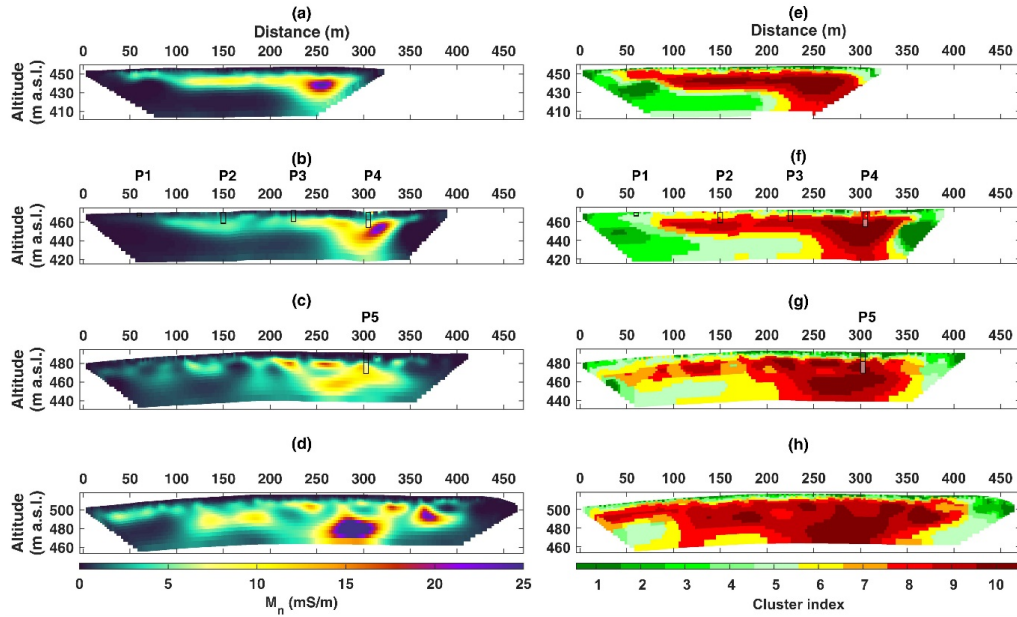


Figure 5.5. Comparison between normalized chargeability models (a-d) and integrated sections after cluster analysis (e-h) for the case study. The piezometric levels are superposed to the models.

Overall, we observe a very good agreement between well data and leachate level predictions from cluster analysis. In many cases the areas characterized by high values of M_n fall within the most hazardous zones identified by the cluster analysis (red clusters). Nevertheless, in many other zones there are significant differences between M_n models and the models obtained by our ML-approach. In fact, there is a lack of lateral continuity of the chargeable zones (i.e. 5.5a), compared to the respective ML images (5.5e). This effect, observed in the M_n sections, increases the uncertainty on the model interpretation based only on inversion results, preventing

a clear detection of the leachate accumulation zones. The effectiveness of K-Means clustering in identifying different groups of data is particularly clear for the L2 and L3 lines (Fig. 5.5), where a quantitative comparison with data from wells is available. The piezometric levels match well with the most hazardous areas (dark red clusters) identified by our clustering analysis, whereas the levels are less correlated with the highest M_n values. The very low level (0.2 m) logged in P1 is not highlighted by both procedures.

One of the main advantages of the proposed procedure is that the number of different zones is directly retrieved from the cluster analysis and the shapes of such well-defined zones are not affected by the choice of the colour scale. In fact, the scale was automatically distributed according to the distance of the centroids from the point (0, 1, 1), which represents the maximum level of contamination, given the electrical properties of leachate (highly conductive and chargeable). We chose to scale colours from red to green as it is used in alert systems to grade the severity of a hazard event. The final output is easy to interpret even for non-experts in the field of geophysics. Typically, leachate is controlled by drilling monitoring wells and our analysis show that the use of geophysical methods combined with ML techniques can provide very accurate information on optimal locations to plan for such wells. The machine learning-based approach can be therefore used to support decision-makers in the waste management sector and reduce the costs of effectiveness of landfill management.

This study has some limitations and there is room for further developments. The main limitations concern the size of the datasets and the choice for the best number of clusters. ML techniques are generally more effective if they use large datasets, and they can provide different results if datasets are differently scaled. We explored different scenarios (Piegari and Paoletti, 2022; De Donno and Piegari, 2022) before developing this procedure. The cluster analyses performed on datasets related to individual profiles provide less accurate results due to the reduced size of the datasets. We found that clustering gives better results if the logarithmic data are scaled in the interval [0, 1] and if the cluster analysis is performed in a 3D space of parameters instead of a 2D space with only ρ and M_n . As regards the choice of the best number of clusters, it depends on the selected threshold value for the explained variance. Although we selected the commonly used elbow method, using the explained variance as a validation index, other methods and/or other validation indices could be also suitable (e.g. Di Giuseppe et al. 2014). Finally, we point out that the proposed identification of the most contaminated zones is only based on the electrical properties of leachate, since they are recognized as the most diagnostic for this purpose (Soupios et al., 2017). However, the proposed multivariate analysis could be applied also to other geophysical datasets, such as i.e. seismic tomography, which might add information about consolidation and compaction of waste materials.

5.1.8 Concluding remarks

In this study, we demonstrated that unsupervised machine learning may be successfully used to integrate data from resistivity and IP methods. The proposed ML-based approach is a flexible tool that can be easily adapted to other case studies also for different type of geophysical data.

For the investigated landfill, by performing a K-Means cluster analysis we were able to: i) achieve integrated model sections combining information from resistivity, chargeability and normalized chargeability data; ii) characterize the subsurface of the investigated landfill by the identification of regions associated with different leachate accumulation warning levels; iii) locate the most hazardous zones (dark red clusters). The suitability of the proposed method for practical applications has been demonstrated by a good agreement between the location of the dark red clusters and the piezometric data available.

Furthermore, the final reconstructions of the investigated landfill are more accurate in the identification of the leachate accumulation zones with respect to the qualitative integration of the traditional geophysical models. Therefore, our findings offer new perspectives for landfill characterization as well as for landfill management by enabling help to predict the leachate flow pathways and enables a more effective planning allocation of financial resources for the development of monitoring and remediation systems (i.e. the drainage network).

5.2 Paper #4: Fuzzy Mapping of Integrated Multi-Parameter Tomographic Data to Detect and Localize Geophysical Targets

Giorgio De Donno¹, Michele Cercato¹, Davide Melegari¹, Valeria Paoletti², Guido Penta de Peppo¹, and Ester Piegari²

¹ Dipartimento di Ingegneria Civile Edile e Ambientale, “Sapienza” Università di Roma, Rome, Italy

² Dipartimento di Scienze della Terra, dell’Ambiente e delle Risorse, Università degli Studi di Napoli Federico II, Naples, Italy

The manuscript presented in this chapter is a selected abstract of the published version on IEEE Transactions on Geoscience and Remote Sensing: 10.1109/TGRS.2024.3445462.

5.2.1 Abstract

In this work, we present an integrated and quantitative approach to detect and localize geophysical target associated with anthropogenic complex scenarios. We complement electrical tomographic techniques and a machine-learning unsupervised algorithm (Fuzzy C-Means) to get a final clustered section validated by well data, where the reliability is evaluated by the membership function. This is a good estimator of the reliability of the proposed procedure, as it ranges 0-1, with 1 reflecting a high reliability of the clustering analysis. This method is applied to the detection of leachate in a municipal solid waste landfill. We also set up a synthetic simulation by reproducing realistic model with similar layering and ranges of geophysical parameters as per the field case. The results show the effectiveness of the method in providing a unique detection of the cluster associated with the desired targets, as highlighted by the matching with direct information. However, the accuracy of the reconstruction is reduced in areas where the resolution of geophysical methods is lower. We demonstrate that the proposed method can achieve a reliable reconstruction if it focuses on searching for the desired target only, rather than a comprehensive reconstruction of the whole layering at the site. This approach can be a valuable automatic tool for improving the benefit-cost ratio of projects where remediation interventions have to be planned.

Keywords: Fuzzy C-Means, electrical resistivity tomography, induced polarization, landfill, leachate detection, reliability assessment.

5.2.2 Introduction

Geophysical methods can provide dense spatial coverage, being at the same time a rapid and non-invasive tool for imaging and monitoring the shallow subsurface (Everett, 2013). During the last decades, research works in this field have been mainly focused on the detection of targets related to the anthropogenic environment, such as contaminated areas, landfills, structures and infrastructures (De Donno et al., 2024) as well as geological features, such as faults, bedrock, cavities (Whiteley et al., 2019). Although geophysical models have been often compared with direct

information coming from boreholes and wells (Bichet et al., 2016; Power et al., 2018), they have been rarely integrated quantitatively and/or accompanied by the assessment of the reliability associated with detected targets (Di Giuseppe et al., 2014; Bernardetti and Bruno, 2019; Mirzakhani and Hashemi, 2022; Zhang et al., 2023). The standalone application of one geophysical method can leave residual ambiguities in the final interpretation, particularly for complex geological and anthropogenic scenarios (Hellman et al., 2017), as variations in a single parameter (electrical resistivity, seismic wave velocity, etc.) can be due to changes in different factors (i.e. lithology, saturation, interstitial fluid, etc.) and, therefore, integration between different geophysical data is often required. An univocal interpretation of geophysical results together with the knowledge of the reliability of the reconstruction may correctly guide retrofitting or remediation activities in the anthropogenic environment or the execution of boreholes and direct inspections (Cardarelli et al., 2018). To this aim, electrical resistivity tomography (ERT) and seismic refraction tomography (SRT) have been jointly used to image the near surface with high resolution and down to significant depths (Hellman et al., 2017). Besides, the induced polarization (IP) method, sensing the capacitive response of the subsurface, can be applied simultaneously to ERT since it is effective for detecting silty/clayey soils or contamination signatures, which are expected to magnify the IP response (Binley and Slater, 2020).

Nowadays, data integration is generally performed by qualitative comparison of the different geophysical models (Soupios et al., 2007), and only a few have recently explored quantitative approaches encompassing joint inversion (Revil et al., 2020) or correlations with petrophysical parameters (Abbas et al., 2022). The last decade has seen a rapid growth of Machine-Learning (ML) techniques for data fusion, due to their potential in processing and interpreting experimental datasets in the Geosciences (Florio and Lo Re, 2018; Karpatne et al., 2019; Cutaneo et al., 2023; Vitale et al., 2023). ML algorithms can be implemented without solving the governing equations, being also able to integrate data of different types. In the field of ML techniques applied to geosciences, unsupervised clustering algorithms have been mostly used for seismic reflection processing, due to their flexibility and ease of practical implementation thanks to the large diffusion in the scientific community of open-source software, often developed in Python (Bergen et al., 2019). Some promising applications have also been performed using supervised ML algorithms, but often limited to only one technique (Liu et al., 2020; Liu et al., 2023), even though their reliability is strongly dependent on the training dataset available and on the choice of parameters for the site-specific problem under examination (Yu and Ma, 2021). Recently, a machine learning-based approach has been proposed to get an effective imaging of multiple geoelectrical data and quantitatively identify leachate accumulation zones (Piegari et al., 2023) and to interpolate seismic velocity and electrical resistivity models for early-stage soil-rock assessment (Dick et al., 2024). Among the unsupervised routines, over the last decades, the Fuzzy C-Means (FCM) algorithm has been successfully applied in the field of Geosciences for identifying polluted sites (Hanesch et al., 2001), interpreting complex geological information (Benjumea et al., 2023), and for complement joint inversion of multiple geophysical data (Sun and Li, 2016). The FCM is known for its ability to handle overlapped data sets and to provide gradual memberships, making it advantageous in scenarios

where data points have ambiguous cluster assignments (Wu et al., 2017).

In contrast to an approximate detection of the main targets, in many civil and environmental engineering applications, the final subsurface reconstruction yielded by geophysical models should be oriented toward the accurate definition of their geometry, including an evaluation of the reliability related to the overall imaging process (Wei et al., 2024). In this sense, during recent years a rigorous uncertainty assessment has been accomplished using global inversion strategies, even if considering a single geophysical method as it is recognized to be highly time-consuming (Lan et al., 2023).

To go beyond previous approaches, we here propose a multi-parameter and multi-technique quantitative method to improve univocity in the detection of the target as well as to assess the reliability of geophysical models related to an anthropogenic scenario. Our approach encompasses the use of geophysical tomographic data (electrical) and ML clustering techniques (FCM) complemented and validated by well data. The proposed method is applied to a single case study in Italy: a municipal solid waste (MSW) landfill located in Central Italy on a steep slope, where detection of leachate accumulation areas is pivotal to prevent instability phenomena. We also set up a synthetic simulation by reproducing realistic model with similar layering and range of geophysical parameters as per the field case.

5.2.3 Method

In this work, we apply a ML approach encompassing a two-step procedure: firstly, a multi-parameter tomographic data inversion is performed from the experimental observations (ERT and IP) to retrieve the 2-D geophysical models (cross-sections); secondly, inverted models are switched to the joint parameter space then applying the FCM clustering algorithm for achieving a final clustered section. This section encloses integrated geophysical information and the assessment of the reliability of the reconstruction using a membership function. Among the tomographic geophysical methods, we selected ERT and IP as they are by far the most applied and sensitive methods for leachate detection (De Donno et al., 2024).

5.2.3.1 ERT and IP: field measurements and data inversion

Electrical galvanic methods, employing a direct current (DC), can sense the distribution of electrical properties of the subsurface, both in terms of bulk (electrolytic) conduction and IP effects. These arise on grain surfaces in the presence of clay minerals and metallic bodies or are activated by biological activity, such as in contaminated sites or within landfills (Binley and Slater, 2020). The former effect is modelled using the ERT method through the electrical resistivity ρ [Ωm] (or its inverse the electrical conductivity σ [S m^{-1}]), while the electrical chargeability M [dimless or mV V^{-1}] that is the ratio between the magnitude of the surface and the bulk conduction, is defined as the model parameter for IP tomography. The contribution of surface conduction can be isolated by computing the so-called normalized chargeability MN [mS m^{-1}]:

$$MN = \frac{M}{\rho}. \quad (1)$$

In this work, we carried out both ERT and IP measurements with a string of 48 electrodes using a DC electrical source operating with a 12V battery. We measured the stationary primary electric voltage V_p [V] with a suitable delay time after the current switch-on (resistive effect) and the voltage decay $V_d(t)$ [V] as a function of time after the current switch-off (IP effect). Therefore, the observed parameters are apparent resistivity ρ_a and apparent chargeability M_a , having the same units of the respective model parameters and achieved as (Binley and Slater, 2020):

$$\rho_a^{obs} = K \frac{V_p}{I}, \quad (2a)$$

$$M_a^{obs} = \frac{\int_{t_i}^{t_f} V_d dt}{V_p \Delta t}, \quad (2b)$$

where K is the geometric factor (depending on the electrode configuration) [m] and I is the current intensity of the external source [A]. Voltage decay is integrated over a time window Δt defined between times t_i and t_f and normalized by the primary voltage to be consistent with the definition of chargeability (Seigel, 1959). Usually, the time window is divided into a few shorter logarithmically spaced gates, and the integral in Eq. (2b) is computed by the sum of potential values for each gate. In this paper, we set both the on- and off-time of the current pulse to 2 s with 2 stacks (repetitions) and we sampled the voltage decay with 20 logarithmically-spaced gates (the first gate starts at 0.04 s and the last gate ends at 1.84 s).

Recorded datasets are inverted for resistivity and chargeability using a Gauss-Newton iterative formulation (Loke and Barker, 1996), where the chargeability dataset is inverted following the linear approximation proposed by Oldenburg and Li (1994). We solved numerically the transformed Poisson's equation by using the finite element method (FEM), fixing Dirichlet's and Neumann's boundary conditions at the surface, lateral and bottom boundaries of the studied domain (2-D vertical cross-section) respectively (De Donno and Cardarelli, 2017b). Data inversion and synthetic simulations were carried out on a FEM mesh made of rectangular elements using the VEMI (Versatile interface for Electrical data Modelling and Inversion) software (De Donno and Cardarelli, 2017b), capable to handle both ERT and IP data. The goodness of fit was evaluated in terms of percentage root mean square error (RMSE) for resistivity, whereas for chargeability it is more convenient to use the mean absolute error, expressed in mV V^{-1} (De Donno and Cardarelli, 2017b).

5.2.3.2 FCM clustering

The FCM is a soft clustering algorithm that assigns each data point a probability score (membership) of belonging to multiple clusters (Bezdek, 2013). Iteratively, FCM updates the cluster centroids through the minimization of an objective function until a convergence criterion is met. In this work, the data points are the $\log_{10}$ -transformed model parameters resulting in the final iteration from the inversion of

geophysical data, which are ρ , M and MN . We chose the $\log_{10}$ transformation as it is the most suitable when the model parameters span many orders of magnitude (Sen and Stoffa, 2013).

The operation of the algorithm can be summarized in the following steps. Initially, FCM calculates the cluster centers (centroids) based on the input data (inverted models):

$$C_i = \frac{\sum_{j=1}^n w_{ij}^b x_j}{\sum_{j=1}^n w_{ij}^b}, \quad (4)$$

where w_{ij} is the membership and represents the degree to which a $\log_{10}$ -transformed data point x_j belongs to the cluster c_i . The membership is updated by using the following equation:

$$w_{ij} = \frac{(1/\|x_j - C_i\|)^{\frac{2}{b-1}}}{\sum_{r=1}^c (1/\|x_j - C_r\|)^{\frac{2}{b-1}}}, \quad (5)$$

where c is the number of clusters and the parameter b is any real number greater than 1, which controls the degree of fuzziness. A larger b results in fuzzier clusters. As b approaches 1, the membership converges to 0 or 1 aligning with the K-means objective function (Di Giuseppe et al., 2014). In this study, we set the fuzzy parameter b equal to 2, which is widely recognized as an appropriate selection (Paasche et al., 2006).

By adjusting the degree of membership and the cluster centers, the algorithm aims to minimize the following objective function:

$$J = \sum_{j=1}^n \sum_{i=1}^c w_{ij}^b D_{ij}^2, \quad (6)$$

where n is the total number of data points and D_{ij} is the distance between the i -th and j -th data points.

The choice of the number of clusters can be done according to *a-priori* information, where well data are available or if the number of layers/zones is known, as for the synthetic simulation, or to established criteria and indices widely used in the field of data science (Lamirel et al., 2016). Therefore, in this work, we fixed the numbers of clusters to be equal to the number of simulated layers/zones for synthetic example, while for field data we used the Calinski-Harabasz (CH) and Bayesian Informative Criterion (BIC) indices. These indices are robust (the optimum value does not change depending on the maximum number of clusters analyzed) and widespread for many application in the geosciences, as they work well for many different types of spatial data distribution (Kryszczuk and Hurley, 2010). The use of multiple indices can help increasing the robustness and accuracy of cluster number estimation in comparison with using a single index and it is also expected to reduce the ambiguity arising when a single index returns more than one optimal point (Kryszczuk and Hurley, 2010).

The CH index (Calinski and Harabasz, 1974) is defined as the ratio between the mean variance between-cluster and the mean variance within-cluster:

$$CH = \frac{\frac{J_1 - J_c}{c-1}}{\frac{J_c}{n-c}}, \quad (7)$$

where J_1 and J_c are the values of the objective function computed for several clusters equal to 1 and c , respectively. The higher its value, the better the data partitioning.

The BIC index (Schwarz, 1978), is based on the principle of minimum description length, whereby a clustering solution is considered better if the data can be encoded more compactly:

$$BIC = -2 \sum_{i=1}^c L_i + m \bullet \log(n), \quad (8)$$

where L is the log-likelihood function of the i -th cluster, m is the number of parameters in the considered c -cluster solution and n the number of observations. A lower BIC value indicates a better fit.

We achieved the optimal number of clusters by running the FCM algorithm with several clusters c ranging between 2 and 10 and selecting the solution yielding at the same time the maximum and minimum values of CH and BIC indices, respectively. For a better comparison between the two indices, we computed, for each cluster solution c , the normalized values $\overline{CH}$ and $\overline{BIC}$ (ranging only between 0 and 1) by:

$$\overline{CH}_c = \frac{CH_c - \min(\mathbf{CH})}{\max(\mathbf{CH}) - \min(\mathbf{CH})}, \quad (9a)$$

$$\overline{BIC}_c = \frac{BIC_c - \min(\mathbf{BIC})}{\max(\mathbf{BIC}) - \min(\mathbf{BIC})}. \quad (9b)$$

where $\mathbf{CH}$ and $\mathbf{BIC}$ in eqs. 9 are vectors containing the values of both indices obtained by varying the number of clusters between 2 and 10. Although there is no mathematical connection between the two criteria that are independent, we can rate the clustering solution as robust if there is a given c for which $\overline{BIC} = 0$ and $\overline{CH} = 1$.

We included the above mentioned formulations in a new MATLAB code able to read the geophysical inversion results from VEMI, apply recursively the built-in FCM function (*fcm.m*) for the computation of CH and BIC indices and finally return a graphical output representing the clustering results shown in the next section.

5.2.4 Results

5.2.4.1 Site description and field survey

The studied MSW landfill, which opened in the '80s and is still active, is located in Central Italy in a natural basin located on a steep slope. The bedrock geological unit

is the marly-arenaceous flysch widespread in the surrounding areas and the landfill is provided with a bottom liner (geomembrane). Mapping the leachate distribution within the waste body is pivotal for assessing the contribution of the leachate to the triggering of slope instability.

The ERT/IP survey (Fig. 5.6) encompassed 4 profiles (L1-L4) acquired using the IRIS Syscal Pro resistivity-meter operating with 48 electrodes spaced 5 m apart with a dipole-dipole array and the roll-along technique for investigating long lines with good coverage (Piegari et al., 2023). Recorded electrode contact resistances were relatively low ($< 1 \text{ k}\Omega$) thus achieving a good current level ($\sim 1.5 \text{ A}$); this condition is favourable for a good signal-to-noise ratio, but data can also be affected by electrode polarization, capacitive coupling, and transient disturbances. Among the four lines we selected for our analysis the L2 line, because we can validate the final reconstruction using the piezometric levels logged in four wells P1-P4 (Table 5.2).

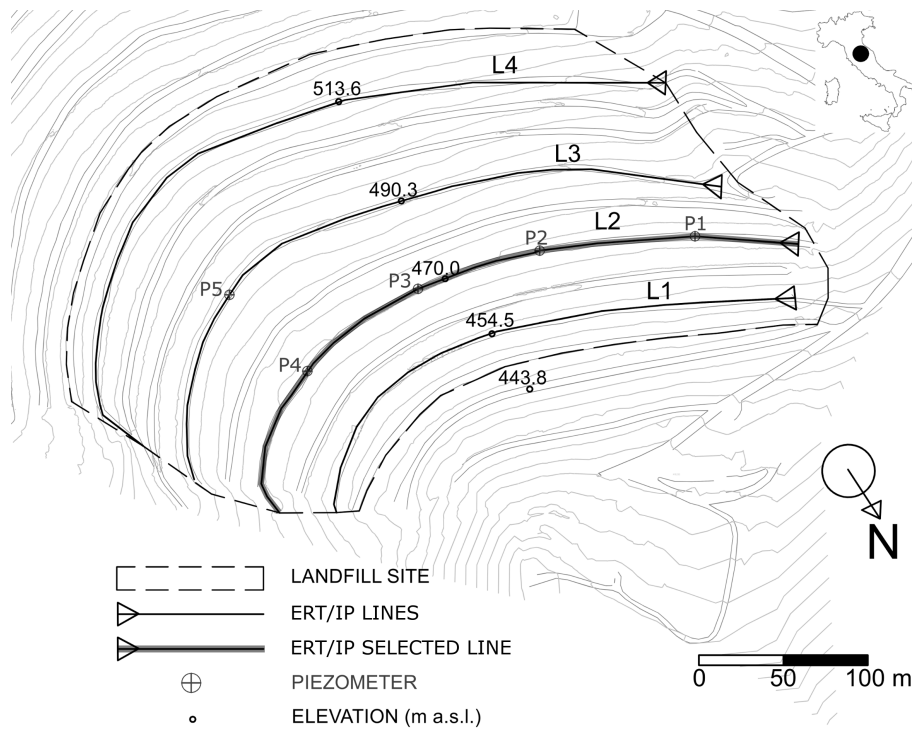


Figure 5.6. Aerial map of the investigated landfill with the ERT/IP lines and the selected line L2 used for the detailed analysis.

Table 5.2. Piezometric levels logged at Site 1 along the selected ERT/IP line L2.

Well	Location along L2 (m)	Bottom of the well (m a.s.l.)	Leachate level (m a.s.l.)
P1	59	462.8	463.7
P2	149	455.8	456.4
P3	227	457.4	458.4
P4	308	452.0	460.1

5.2.4.2 Synthetic data

Firstly, we simulated the typical geometry of a MSW landfill with a five-layer model (Fig. 5.7a), setting the values of the geophysical parameters according to previous synthetic simulations on similar scenarios (Clément et al., 2011; De Carlo et al., 2013; Ibraheem et al., 2021) and to the ranges normally encountered in field surveys (Knight and Endres, 2005; Dahlin et al., 2010; Maurya et al., 2017). The geophysical target is the leachate accumulation area located in the middle-bottom portion of the model (white trapezoid in Fig. 5.7a), having the lowest resistivity ($\rho = 5 \Omega\text{m}$) and the highest chargeability ($M = 50 \text{ mV/V}$) and extending in depth from 14.5 m to 26.5 m (Fig. 5.7a). The bottom layer of the landfill, where the geomembrane, acting as an insulator, (resistivity $\sim 10^7 \Omega\text{m}$) prevents the current flow, is simulated by a very high resistive and not chargeable layer ($\rho = 10^5 \Omega\text{m}$ and $M = 1 \text{ mV/V}$), while the partially saturated shallower layers have intermediate values of ρ and M (Fig. 5.7a).

ERT/IP datasets were simulated using the same electrode spacing, array and parameters as in the L2 field line (line length of 395 m with 80 electrodes spaced 5 m apart), thus yielding the same number of observations. A zero-mean 3% (resistivity) and 0.5 mV/V (chargeability) Gaussian noise was added to the resistivity and chargeability datasets, respectively. Both values are consistent with noise levels normally encountered in the field acquisitions (e.g. Piegari et al., 2023), and the chargeability noise is given as an absolute value since the IP effect is generally low and a percentage error often fails to be representative of the field noise (De Donno and Cardarelli, 2017a).

The results of geophysical inversion are shown in Fig. 5.7b in terms of resistivity, chargeability and normalized chargeability.

The models are generally able to reconstruct the layering of the landfill even though there are some differences among them, mainly due to the lower sensitivity of chargeability to deep targets (bottom of the section) compared to resistivity. This may prevent a univocal location of the desired target.

The application of the FCM algorithm to synthetic data is shown in Fig. 5.8, where we fixed the number of clusters $c = 5$, which is the number of layers.

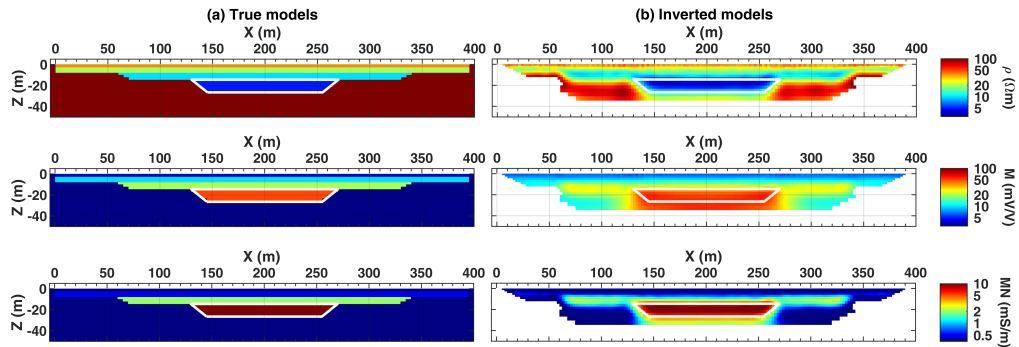


Figure 5.7. Synthetic simulation. The geophysical target (leachate accumulation zone) is marked by the white trapezoid. White pixels indicate no data coverage or unreliable values excluded from the clustering analysis. (a) True models. (b) Inverted models.

In Fig. 5.8a, we report the geophysical parameters after inversion in the 3D

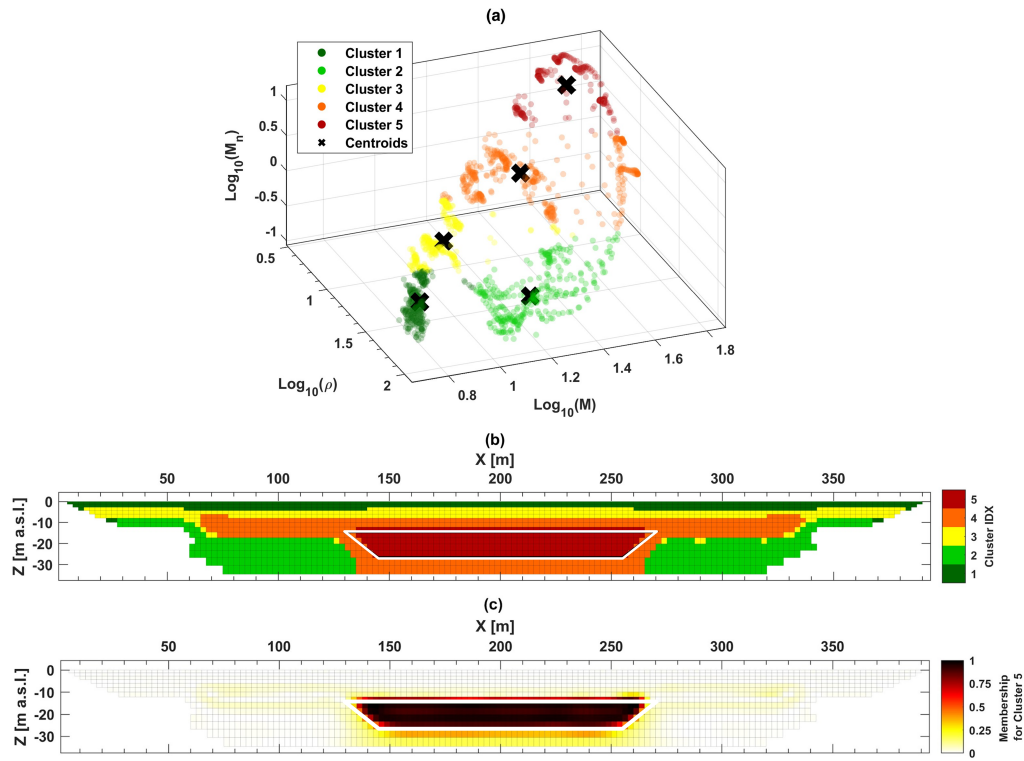


Figure 5.8. FCM clustering results for the synthetic model: (a) 3D $\log_{10}$ parameter space, (b) clustered section, and (c) membership section for the cluster associated with the geophysical target.

$\log_{10}$ space, while in Figs. 5.8b and 5.8c we show, respectively, the clustered section and the membership function for each pixel of the geophysical target (cluster no.5). The clustered section displays a less ambiguous and more accurate identification of the leachate accumulation zone, as confirmed by the good agreement between the true location of target (white trapezoid) and the shape of cluster no. 5 (lowest ρ and highest M and MN), where only the top of the target is slightly shifted toward the surface. Furthermore, the membership section displays high values (> 0.75) within the target while decreasing as a function of depth because of the lower resolution retrieved at the bottom of the model.

5.2.4.3 Field data

Inversion of ERT/IP data for the L2 line highlights a three-layer model from the top to the bottom (Fig. 5.9): i) a shallow (thickness of 3-5 m) resistive and not chargeable layer (covering soil); ii) a conductive and chargeable layer associated with the waste mass; iii) a bottom layer having high resistivity and low chargeability as an effect of the bottom liner, preventing the current flow outside the landfill. The main target is located in the right portion of the model, between $x = 250 - 350$ and $z = 430 - 460$ m above sea level (a.s.l.), where the shape of the bottom of the landfill (old topographic surface) creates an impluvium and leachate can accumulate. However, the conductive and chargeable area associated with the waste mass is also elongated in the left portion of the model ($x < 250$ m), even though it has higher resistivity and lower chargeability values compared to the main anomaly. Therefore, univocal detection of the saturated waste area (which is potentially dangerous since it can trigger instability phenomena for slope landfills) only by a visual comparison of geophysical models is unfeasible.

The application of the clustering algorithm can discern four clusters (Fig. 5.10), following the results of the analysis made through CH and BIC indices (Fig. 5.11).

The main anomaly (cluster no. 4) resembles well (very high membership values in Fig. 5.10c) the leachate level logged in the P4 and P3 wells and, partly, in P2 even if with lower reliability (membership between 0.15 and 0.35 in Fig. 5.10c). No correspondence is found instead between P1 and the clustering analysis, likely due to the lower sensitivity of ERT and IP to image very small leachate volumes, like the ones encountered in that portion of the landfill.

5.2.5 Discussion

During the last decade, there has been an increased demand for a quantitative assessment of the sought geophysical target (both geological and anthropogenic) in terms of shape, position and magnitude, particularly for applications in environmental engineering and when boreholes or direct inspections have to be planned (Balbarini et al., 2018). In this work, the improved information on the subsurface achieved would be reflected in an increased benefit-cost ratio of the whole project, particularly when rehabilitation interventions or remediation activities (i.e. drainage of leachate) have to be executed. The alternatives to the application of ML approaches to merge different datasets in the last few years have almost always been dedicated to the uncertainty assessment by global inversion strategies that are time-consuming and

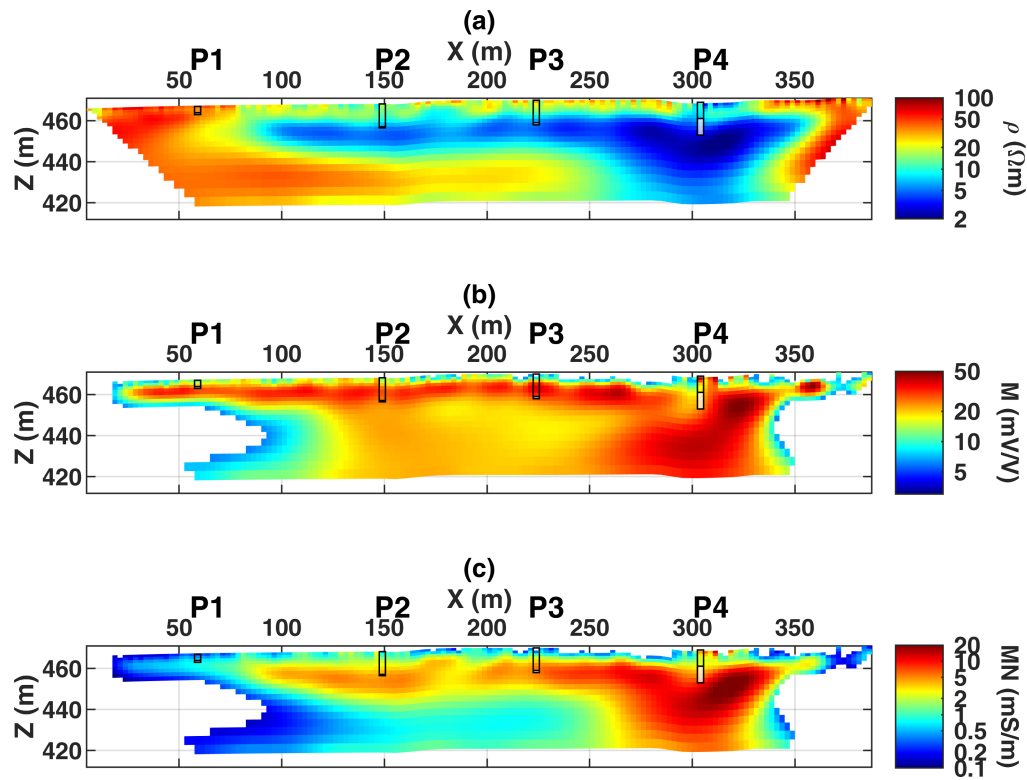


Figure 5.9. Field-data inversion results along line L2: (a) resistivity, (b) chargeability, and (c) normalized chargeability. The leachate levels logged in wells are shown by the white-filled portions of the black rectangles. White pixels indicate no data coverage or unreliable values excluded from clustering.

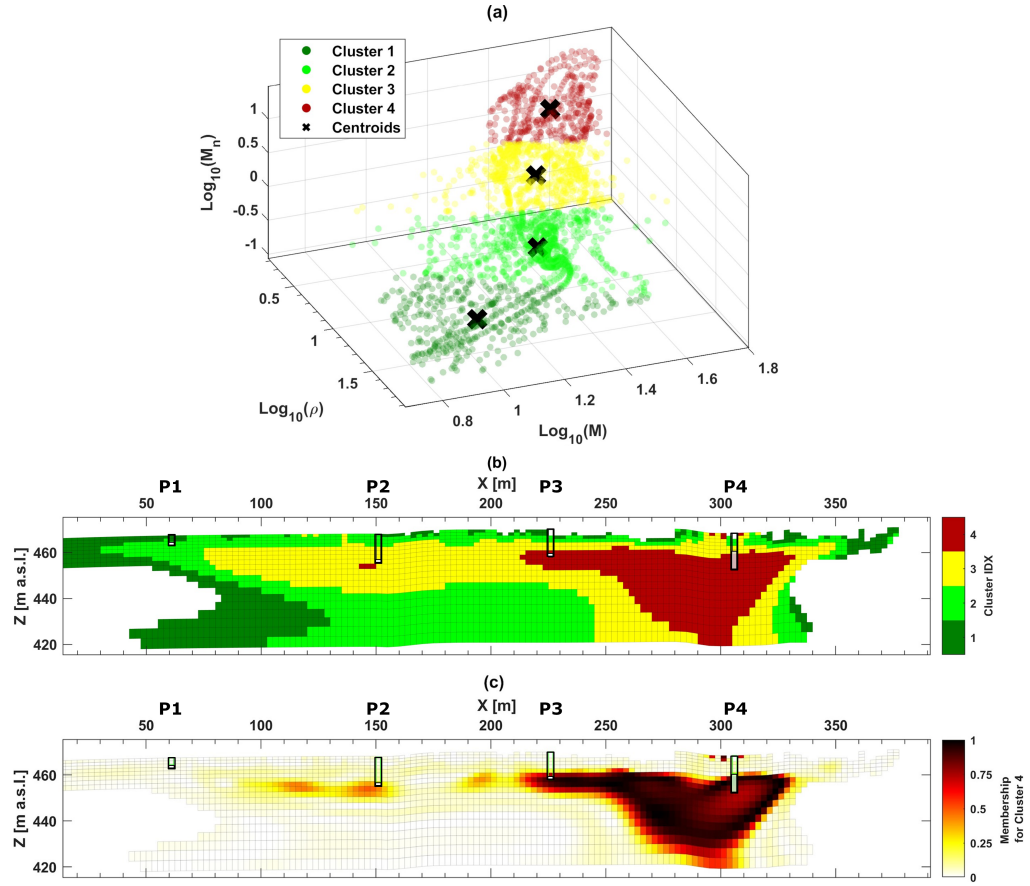


Figure 5.10. FCM clustering results for the field data: (a) 3D $\log_{10}$ parameter space, (b) clustered section, and (c) membership section of the cluster associated with the geophysical target. The well data are superposed to the sections.

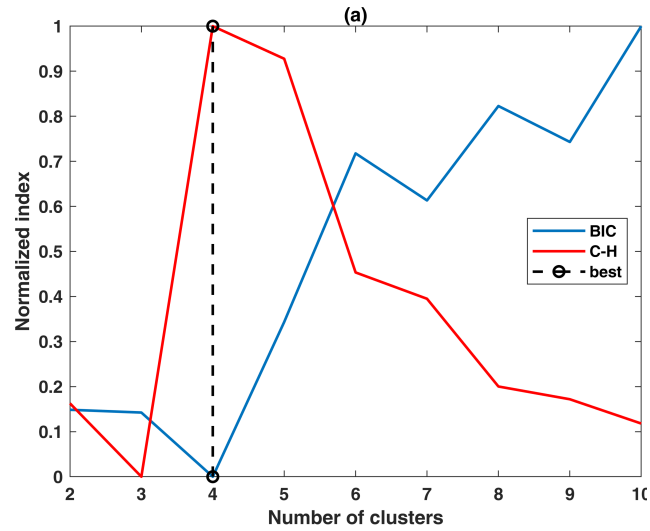


Figure 5.11. Normalized CH and BIC indices used to evaluate the optimal number of clusters for the analyzed examples. The dashed vertical line marks the preferred solution for the field case.

only applied in state-of-the-art to simple situations (Ray and Myer, 2019; Grana et al., 2020; Aleardi et al., 2022). Other options are joint inversion procedures, which require strong assumptions on the structural similarity of the models (Hellman et al., 2017) or the empirical estimation of many petrophysical parameters (Abbas et al., 2022). Additionally, a final interpretation only made through a visual comparison of geophysical models may fail to properly locate the shape and magnitude of the targets, as shown in this work, mainly due to the different sensitivities of the geophysical methods operating in complex geological and anthropogenic scenarios. The correspondence between the shape of the main target differs for resistivity and chargeability models, as the former parameter is mostly affected by changes in saturation and salinity of leachate; while the latter is more sensitive to surface conduction mechanisms that are mainly driven by polarization of the electric double layer occurring in the waste mass (Binley and Slater, 2020). However, the proposed approach can be an effective tool to complement the classical geophysical interpretation as long as it focuses on searching only for one desired target rather than inferring a comprehensive reconstruction of the whole layering at the site. In this sense, the computation of the membership function for the geophysical sought target can be a good estimator of the reliability of the proposed procedure, as its values in field case reflect the expected potential and limits of the geophysical imaging. We detected high membership where the contrast between the geophysical parameters is strong (i.e. main leachate accumulation zone in Fig. 5.10), in agreement with the a priori information for field case and realistic models for simulation. The deviations from the expected target position are reduced in this portion of the models to a few meters or less. Conversely, where the magnitude of the anomaly is weak or the target is deep, the accuracy of our reconstruction is highly reduced (a deviation of some meters from the expected position of the anomaly is also displayed), even though the cluster assignment of the main anomaly is generally correct, as demonstrated for both the real world and synthetic example. Although a rigorous computation of a confidence interval for the reconstruction is not possible without a stochastic inversion of geophysical data, membership values can be practically interpreted as the probability of a model cell belonging to a geologic unit or anthropogenic features (Sun and Li, 2015). To better discuss the effectiveness of membership, we extracted in Fig. 5.12 horizontal profile from the membership sections related to the geophysical target (Figs. 5.10), in comparison with the direct information (wells).

Figure 5.12 shows a horizontal profile of the membership values at the depth of 458 a.s.l., which cuts three wells out of the four wells (P2, P3 and P4). The P2 well is dry at this depth (corresponding membership value close to 0) while leachate was detected in P3 and P4 where high values of membership values ($\sim 80\%$ in P3 and $\sim 85\%$ in P4) are reconstructed. Based on a comparative analysis of all the analyzed model sections, a threshold of 50% (“absolute majority”) on the membership can be suitable for assessing the pixels that “strongly” belong to a cluster, but this topic will be thoroughly investigated in a future work that is beyond the scope of this paper. In any case, the assessment of reliability through the membership values should also take into account the assessment of the number of clusters c . In fact, this choice, in turn, affects the membership values. In our cases the assignment was made to the cluster displaying the highest membership with a minimum membership value for the assignment equal to $1/c = 0.25$ for the field example. In this work,

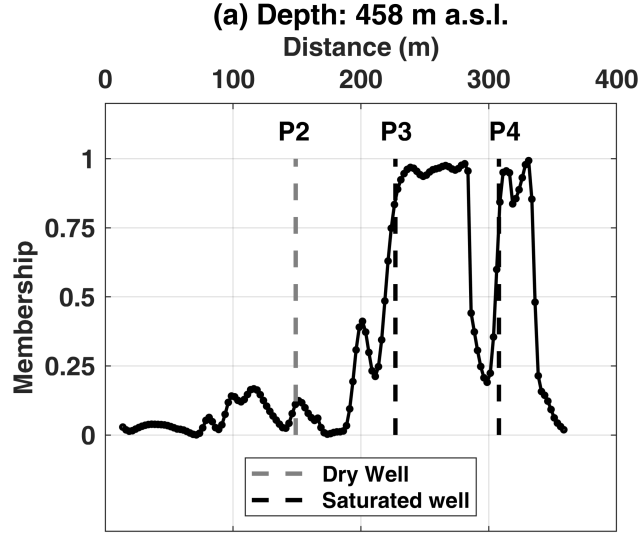


Figure 5.12. Horizontal profile of the membership values extracted at 458 m a.s.l. from the membership section associated with the geophysical target. The profile is compared with the available well information.

we decided to work with two well-established criteria for the computation of the number of clusters (CH and BIC), even though surely other indices proposed in the literature for such purposes can be alternatively used (Schubert, 2023). Although there is no guarantee that both normalized (and independent) indices converge to the same solution, we can state that if this solution exists, the cluster reconstruction will be robust. In the presented examples, CH and BIC seem to work well, since they are not dependent on the maximum number of clusters set in the recursive application of FCM. Furthermore, they tend to privilege simple solutions, which are preferable in engineering applications, while preserving, on the other hand, the heterogeneity of the layering, as demonstrated by the comparison with the direct information.

5.2.6 Conclusion

In this work, we presented a quantitative approach utilizing ML unsupervised algorithms integrating tomographic data to detect a geophysical target through a membership function. It is a good estimator of the reliability of the proposed procedure, as it ranges from 0 to 1, with 1 reflecting a high reliability of the clustering analysis. We applied the proposed procedure to geophysical models achieved from the investigations of electrical (resistivity and chargeability) properties of the subsurface for the detection of geophysical target related to anthropogenic (leachate accumulation areas) scenario. The proposed approach was able to detect and localize the cluster associated with target, as validated by a good agreement with the available direct information (wells). However, the effectiveness of the reconstruction is firmly correlated to the depth of the target and its magnitude compared to the neighborhood layers.

Therefore, the proposed approach can be a valuable automatic tool for many appli-

cations of geophysical techniques in civil and environmental engineering, particularly where remediation interventions have to be planned. In such cases, enhancing the accuracy in specific target detection and assessing the reliability of the reconstruction can help to improve the benefit-cost ratio of the whole project.

5.3 Paper #5: An integrated study of hard and soft cluster analyses for detecting leachate in a MSW landfill site using geoelectrical data

Giorgio De Donno¹, Davide Melegari¹, and Ester Piegari²

¹ Dipartimento di Ingegneria Civile Edile e Ambientale, “Sapienza” Università di Roma, Rome, Italy

² Dipartimento di Scienze della Terra, dell’Ambiente e delle Risorse, Università degli Studi di Napoli Federico II, Naples, Italy

The manuscript presented in this chapter is published in Waste Management and it is reported here in its complete form: 10.1016/j.wasman.2025.01.034.

5.3.1 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, and 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 Ω m, 10–70 mV V⁻¹ and 4.5–37 mS m⁻¹, 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.

Keywords: Leachate, Cluster Analysis, K-means, Fuzzy C-means, Electrical Resistivity Tomography, Induced Polarization Tomography

5.3.2 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., Souplos et al., 2017 and 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 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, 2017a; 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., 2024). 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.

5.3.3 Material and Methods

5.3.3.1 Site description

The investigated site is a MSW landfill built in the '80s at the service of a medium-size city located in Central Italy (Fig. 5.13). 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.

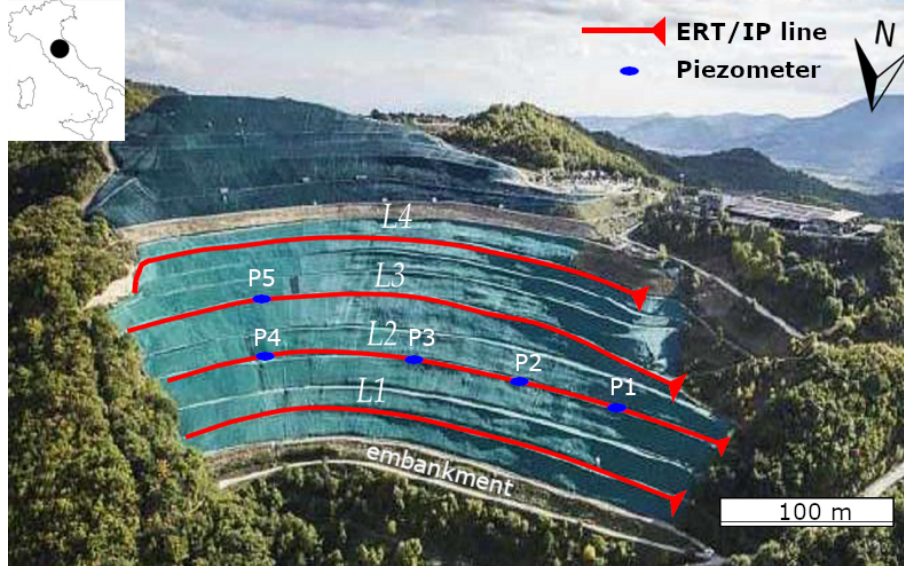


Figure 5.13. View of the studied MSW landfill in Central Italy, showing the locations of the four electrical profiles (red lines, L1–L4). Red arrows indicate the direction of the survey along the lines (from the first to the last electrode). Blue ellipses mark the locations of the five piezometers (P1–P5).

The landfill is equipped with a monitoring network of wells, among which five representative Casagrande-type piezometers were chosen (P1–P5 in Fig. 5.13). 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.).

5.3.3.2 Geophysical data acquisition and processing

We acquired 4 ERT/IP curvilinear profiles (L1–L4) spaced approximately 40 m to 50 m apart (Fig. 5.13) 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 and integral chargeability M , 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 M_N at the end of the inversion process by (e.g., Slater and Lesmes 2002):

$$M_N = \frac{M}{\rho} \quad (5.9)$$

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. 5.13). 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 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:

$$AE_i(\%) = 100 \left| \frac{K_i^{3D} - K_i^{2D}}{K_i^{2D}} \right| \quad i = 1, 2, \dots, M, \quad (5.10)$$

being M the number of measurements (quadrupoles).

The 3D view of the AE for each geophysical measurement (Fig. 5.14) 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 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.

5.3.3.3 Proposed approach integrating hard and soft clustering

Figure 5.15 illustrates the proposed procedure for localizing the leachate accumulation zones integrating hard and soft clustering. Our approach is based on the following

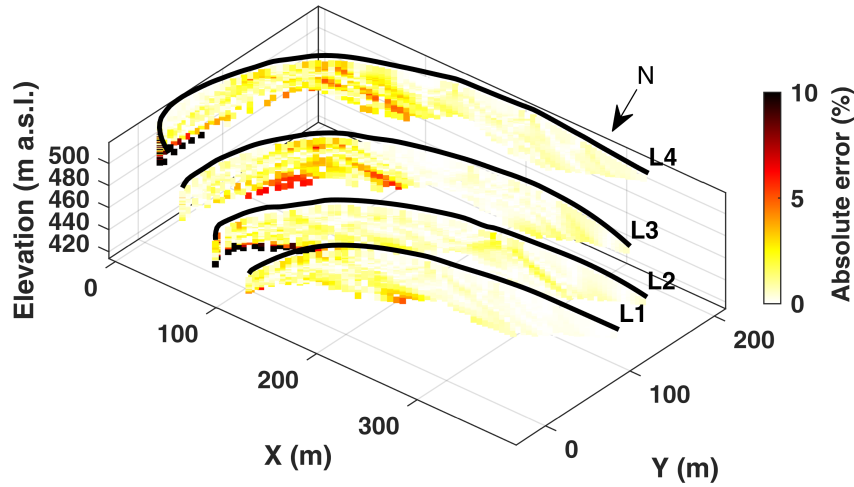


Figure 5.14. Absolute error committed for each measurement due to the linearization of the curvilinear profiles over the studied landfill.

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

5.3.3.3.1 K-means

The *K*-means algorithm is an unsupervised ML partitioning algorithm subdividing a dataset into *K* non-overlapping groups (clusters). The iterative process

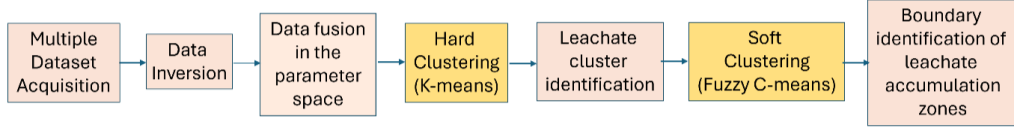


Figure 5.15. Sketch of the proposed procedure to identify leachate accumulation zones through hard and soft clustering starting from geoelectrical tomographic data.

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 subsection (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).

5.3.3.3.2 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 $2/(m-1)$, 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:

$$X_{\rho} = \frac{\log_{10}(\rho) - \min[\log_{10}(\rho)]}{\max[\log_{10}(\rho)] - \min[\log_{10}(\rho)]} \quad (5.11)$$

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.

5.3.4 Results

5.3.4.1 Geophysical models

A 3D view of the inverted geophysical models is shown in Figs. 5.16a, 5.16b and 5.16c 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 ($> 10 \ \Omega \text{ m}$) and low chargeability ($< 10 \text{ mV/V}$), laterally heterogeneous, related to the covering soil and the unsaturated waste; ii) waste mass with different saturation degrees from leachate, having low resistivity ($< 10 \ \Omega \text{ m}$) and high chargeability ($> 10 \text{ mV V}^{-1}$) 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, 2017a). Although the waste mass heterogeneity locally affects the lateral variation of electrical parameters in the middle layer, electrical parameters characteristic of leachate ($\rho \sim 2\text{--}3 \ \Omega \text{ m}$ and $M \sim 30\text{--}40 \text{ mV V}^{-1}$) 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

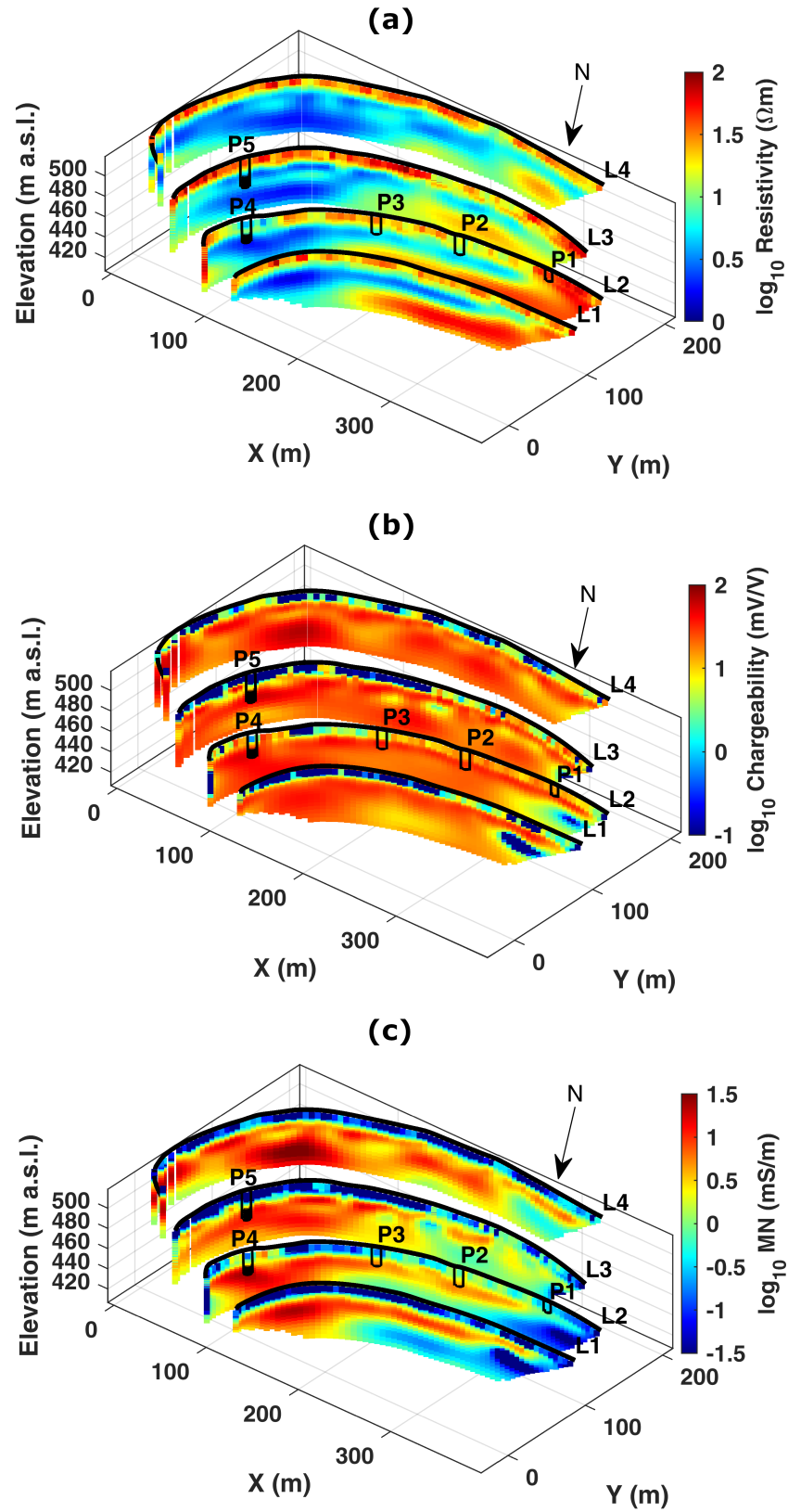


Figure 5.16. 3D view of the resistivity (a), chargeability (b), and normalized chargeability (c) inverted models for the four profiles over the studied MSW landfill. The piezometric levels logged in wells (black filled rectangles) are superposed on the models and their diameters are exaggerated for clarity.

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. 5.16c), 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 M_N values (around 10 mS m^{-1}) 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 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.

5.3.4.2 Cluster analysis

First, we applied the K -means algorithm to get a clustered model of the investigated sections (Fig. 5.17). 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. 5.4 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. 5.17a, while a landfill 3D view is depicted in Fig. 5.17b. 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 M and M_N values, 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. 5.16). 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. 5.18), likely related to leachate accumulation. The data point cloud and the associated memberships for cluster no.10 are shown in Figs. 5.18a and 5.18c, where we can isolate the extreme values (lowest resistivity and highest chargeability) for the four profiles

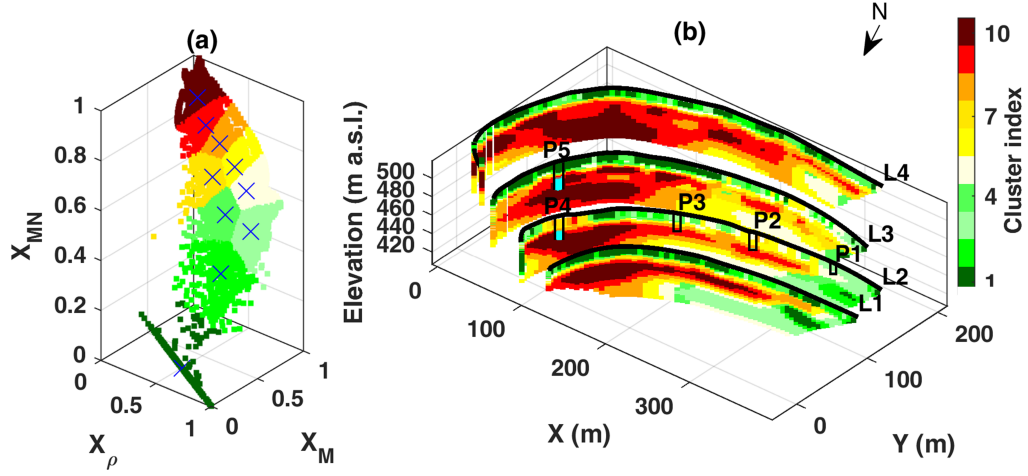


Figure 5.17. (a) K-means clustering in the normalized $\log_{10}$ space, where centroid positions are marked with blue crosses. The dimension of the M_N axis is doubled with respect to the other two axes for clarity. (b) 3D view of the clustered landfill model. The piezometric levels logged in wells (light blue filled rectangles) are superposed on the models.

(filled circles in Fig. 5.18a), which are located in the middle of the accumulation zone (Fig. 5.18c, yellow squares). We were also able to select the boundary points (Fig. 5.18a, filled squares) characterized by low memberships (< 0.5), located close to the cluster no.9, mainly representing the boundary of the accumulation areas (Fig. 5.18c, purple squares).

The 3D views of this dark red cluster and its fuzzy membership, shown in Figs. 5.18c and 5.18d 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 Figs. 5.18c and 5.18d). As shown in Fig. 5.18d, high membership degrees belonging to cluster n.10 are primarily found at the boundaries of the red zones in Fig. 5.18c, 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.

5.3.5 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. 5.18 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. 5.18a that, despite the resistivity values being restricted within a very narrow range ($\rho \sim 1.5\text{--}5 \Omega \text{m}$), both M and MN span a wider range ($M \sim$

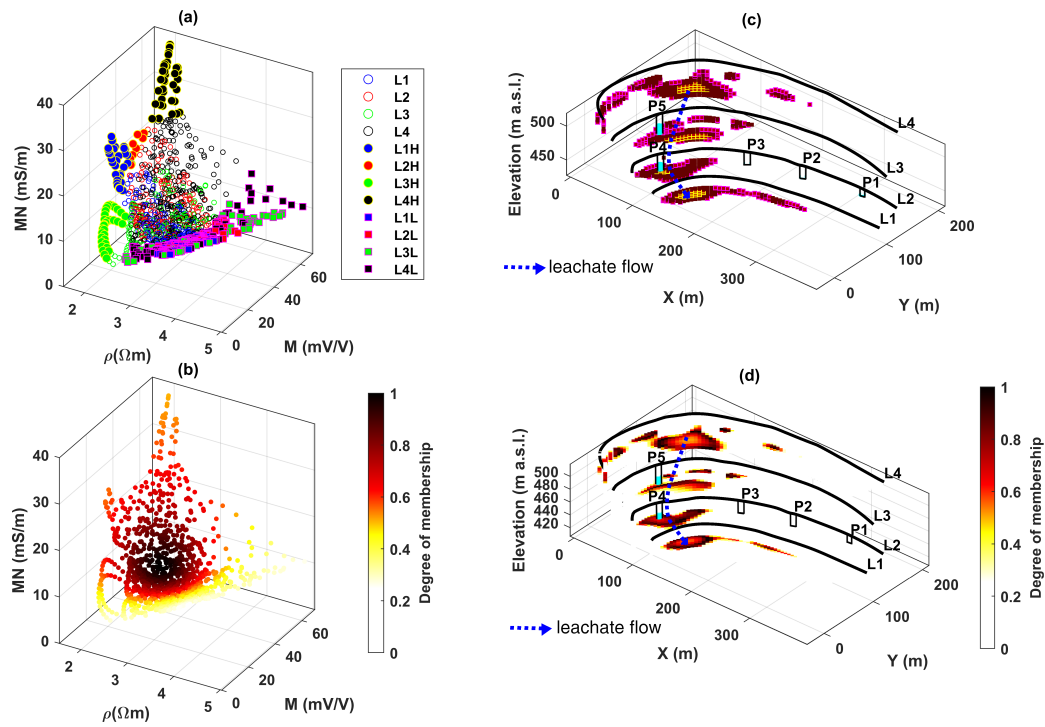


Figure 5.18. Results of the soft-cluster analysis: (a) scatterplot of cluster no. 10 in the unscaled parameter space for the four profiles (L1–L4); the filled circles are the extreme values having the lowest resistivity and highest normalized chargeability, while the filled squares show the boundary values with membership lower than 0.5. (b) Degree of membership related to cluster no. 10. (c) 3D view of cluster no. 10, where the yellow points are the extreme values and the purple points are the low-membership boundary values. (d) 3D view of the membership degree belonging to cluster no. 10. The light-blue filled rectangles indicate the piezometric levels logged in wells. The dotted blue line represents the inferred preferential leachate pathway.

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 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. 5.18), we note how the peaks in the point cloud (filled circles in Fig. 5.18a) clearly display different ranges of the geophysical parameters. The points belonging to L3 (green points) are located at the bottom of the subsection 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, r reaches the lowest values (< 2 Wm) 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. 5.18b 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. 5.18a), 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 Figs. 5.18a and 5.18c).

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

(Figs. 5.19a, b, c), along with the distribution of fuzzy membership for each data point belonging to the dark red cluster (Fig. 5.19d).

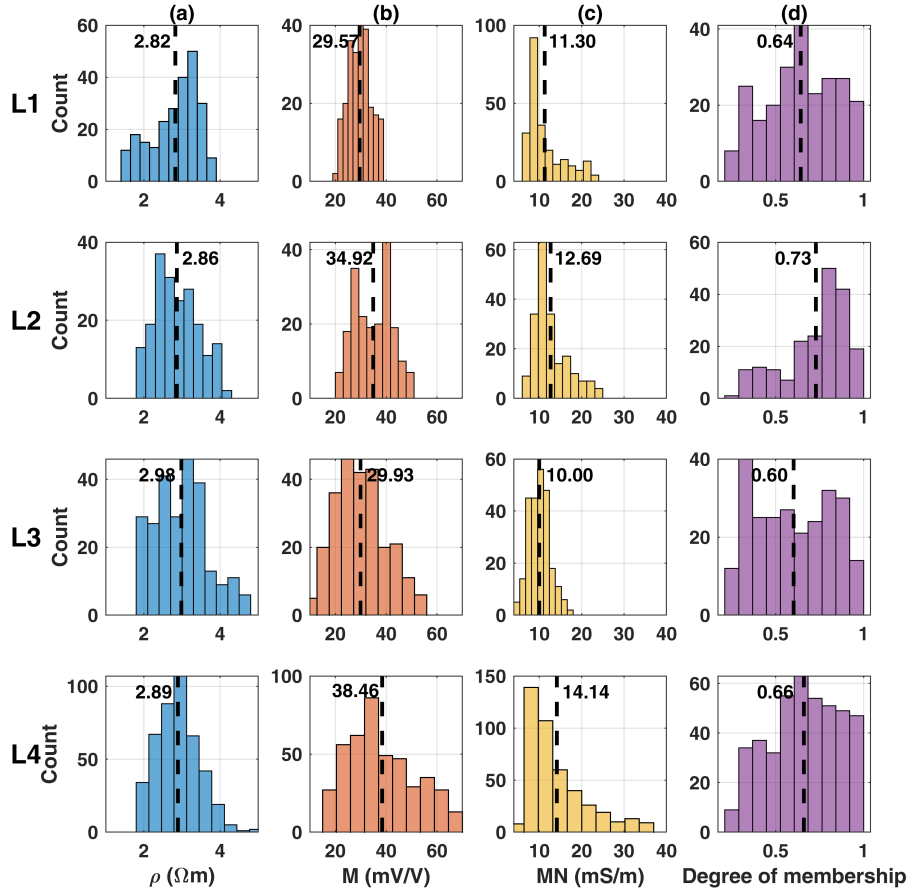


Figure 5.19. Histograms of the three geophysical parameters and fuzzy membership for data in the individual profiles (cluster no. 10): (a) resistivity, (b) chargeability, (c) normalized chargeability, and (d) fuzzy membership. The dashed lines indicate the mean values, which are also superposed on the plots.

The shape of the resistivity distributions (blue histograms, Fig. 5.19a) 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. 5.19b) progressively lose their right-skewed shape and become more focused around the mean value, while the normalized chargeability distributions (yellow histograms, Fig. 5.19c) 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. 5.19b-top), is due to the decrease of resistivity, i.e. increase of surface and fluid conductivity at the bottom of the landfill (Fig. 5.19a-top), counterbalancing the high M_N values (Fig. 5.19c-top). On the other hand, the right-skewed M-distribution for L4 (Fig. 5.19b-bottom) is mainly dominated by the increase of surface polarization clearly visible in the respective MN distribution (Fig. 5.19c-bottom), likely due to a higher biogeochemically activity at the top of the landfill (fresh waste). Finally, the distributions of the membership values (purple histograms in Fig. 5.19) 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 subsection (Fig. 5.18d). 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. 5.20 the comparison between the levels logged in P4 and P5 wells and both the geophysical (ρ and M) (Figs. 5.20a and 5.20c for P4 and P5, respectively) and the clustering (index and membership) parameters (Figs. 5.20b and 5.20d 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. 5.20b, 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 M_N values), located in the middle of the accumulation areas (1.8% of the total pixels, yellow points in Fig. 5.18c), 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. 5.18c) and further drops to 5.10% when considering only membership values higher than 0.75.

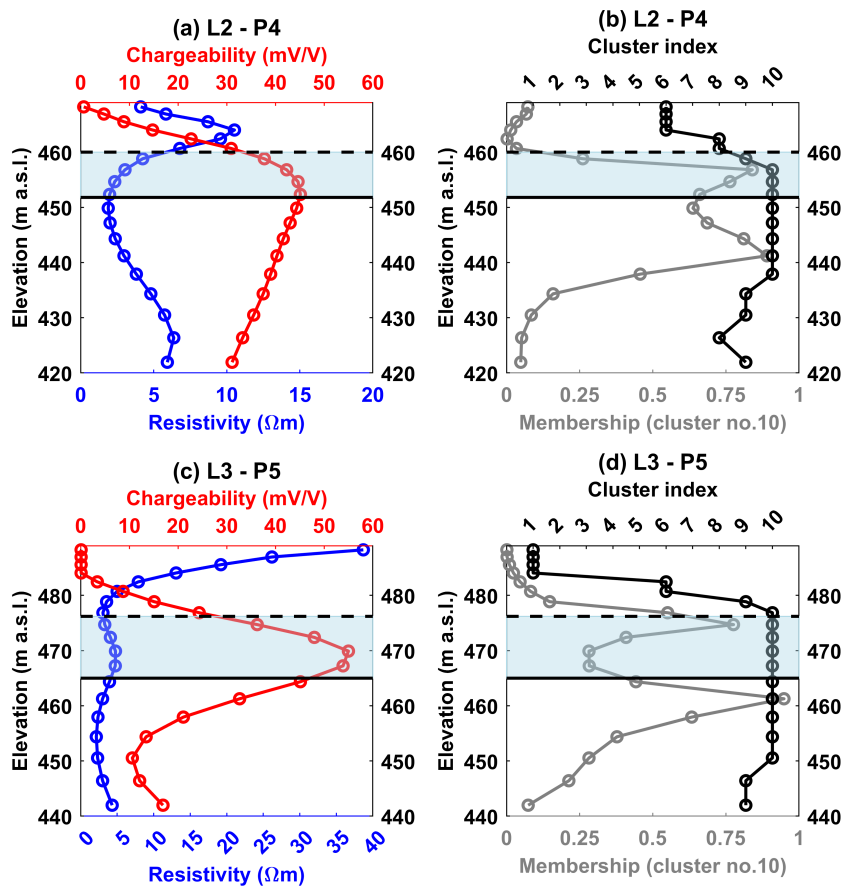


Figure 5.20. Vertical profiles of geophysical parameters ρ and M (a, c), together with clustering index and membership (b, d), extracted at the locations of P4 and P5. The well bottom is marked by a black solid line, the leachate level by a dashed line, and the saturated area is shown by the light-blue filled area.

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 constraints on data (i.e. considering points with a membership lower than 0.5 and with ρ , M and M_N values respectively larger and smaller than those of the cluster centroid), helps defining a range of uncertainty (purple areas in Fig. 5.18c), which is useful for assessing the effectiveness of our procedure.

A full benefit-cost analysis of the presented procedure is beyond the 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. 5.20. 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.

5.3.6 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 M_N , 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. 5.18c.

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.

Chapter 6

Conclusion

This thesis develops three distinct but converging methodological directions for landfill imaging and monitoring using ERT/TDIP data. Each research topic was designed to address a specific limitation that still affects the practical interpretation of geoelectrical data in such complex scenarios. The main scientific motivation was the need to move from qualitative and potentially ambiguous interpretation of geophysical models toward a more reliable, quantitative, and operationally usable tool for mapping and monitoring MSW landfills.

The first research line addressed the problem of TDIP data quality, through the analysis of full-waveform electrical data recordings. This new feature coupled with the use of a 100% duty-cycle can potentially improve not only the cost-effectiveness of the survey but also the data quality, though the extraction of reliable polarization information remains challenging because of harmonics, spikes, drift, coupling effects, and waveform distortions. To overcome this limitation, the thesis developed a dedicated workflow for these advanced TDIP data, including noise filtering, drift correction, quality control, and full-decay-oriented data preparation. This contribution shows that the full-waveform processing phase is a pivotal step, for enabling the full exploitation of TDIP data through a full-decay forward modelling and inversion.

The second topic focused on the use of stochastic inversion for uncertainty assessment, which remains one of the major challenges in geophysical applications. Standard deterministic inversions are efficient and widely used, but they provide only one smooth model among many possible ones, without an explicit characterization of the range of plausible subsurface configurations compatible with the data. This limitation is particularly relevant in landfill settings, where strong heterogeneity, poor depth sensitivity, and complex leachate-related processes increase non-uniqueness. To address this issue, the thesis explored a global inversion strategy based on VFSA, applied to joint ERT and IP data on an unreduced full 2D mesh. The thesis demonstrates the possibility of deriving not only a minimum-misfit solution, but also ensemble-based information on model variability and robustness. This represents an important step toward uncertainty-aware landfill imaging and toward a more transparent use of geophysical models in environmental interpretation.

The third topic addresses model integration, which is often overlooked in practical applications, limiting the effectiveness of multiparameter geophysical investigations. In fact, the results of ERT and TDIP surveys are typically presented as multiple

models, whose comparison is often qualitative and difficult to communicate beyond the geophysical community. To reduce this ambiguity, this thesis presented a set of machine-learning clustering approaches for the integration of multiparameter tomographic outputs. Starting with hard clustering techniques and then extending to target-oriented fuzzy mapping, these methods enable the transformation of separate geoelectrical models into a single integrated representation of the subsurface that can be easily correlated to the environmental targets.

As a final result, these three research directions are closely connected and progressively converging. The first improves the quality of input data, the second strengthens the reliability of the inversion outputs, and the third translates multiparameter models into a quantitative interpretative framework. Although presented as three distinct lines, they collectively define a coherent methodological trajectory for future development. A natural next step is their integration into a single workflow, combining global full-decay inversion with supervised, clustering-based interpretation constrained by well data, aiming to improve both target delineation and the reliability of the final environmental interpretation. Two additional, complementary perspectives concern the direct estimation of petrophysical properties from geoelectrical parameters under landfill-specific conditions, by inverting directly for water content and CEC, and the extension of the inversion framework to time-lapse applications, in order to monitor the temporal evolution of leakage-related processes and improve the dynamic characterization of landfill behavior. Overall, this thesis contributes to applied geophysics for environmental engineering by proposing a set of complementary advances in electrical data processing and inversion, as well as machine-learning-based interpretation. These advances enhance the reliability, interpretability, and operational value of geoelectrical investigations in complex landfill environments, providing a methodological basis for future integrated workflows aimed at more effective leachate detection, delineation, and monitoring.

References

- Abbas, M., Deparis, J., Isch, A., Mallet, C., Jodry, C., Azaroual, M., Abbar, B., & Baltassat, J.-M. (2022). Hydrogeophysical characterization and determination of petrophysical parameters by integrating geophysical and hydrogeological data at the limestone vadose zone of the Beauce aquifer. *Journal of Hydrology*, 615.
- 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 Management*, 109, 231-246.
- Abdideh, M., & Ameri, A. (2020). Cluster Analysis of Petrophysical and Geological Parameters for Separating the Electrofacies of a Gas Carbonate Reservoir Sequence. *Natural Resources Research*, 29(3), 1843-1856.
- 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. *Environmental Earth Sciences*, 75(4), 347.
- Aleardi, M., Vinciguerra, A., Stucchi, E., & Hojat, A. (2022). Probabilistic inversions of electrical resistivity tomography data with a machine learning-based forward operator. *Geophysical Prospecting*, 70(5), 938-957.
- Archie, G. E. (1942). The electrical resistivity log as an aid in determining some reservoir characteristics. *Petroleum Transactions of AIME*, 146, 54-62.
- 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). *Journal of Applied Geophysics*, 111, 320-333.
- Balbarini, N., et al. (2018). Geophysics-based contaminant mass discharge quantification downgradient of a landfill and a former pharmaceutical factory. *Water Resources Research*, 54(8), 5436-5456.
- Benjumea, B., et al. (2023). Geomechanical parameters assessment and geological characterization using fuzzy C means clustering of electrical resistivity and seismic data. *Near Surface Geophysics*, 21(6), 429-443.
- Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, 363(6433), eaau0323.
- Bernardetti, S., & Bruno, P.P.G. (2019). The Hydrothermal System of Solfatara Crater (Campi Flegrei, Italy) Inferred from Machine Learning Algorithms. *Frontiers in Earth Science*, 7, 286.
- Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: the fuzzy c-means clustering algorithm. *Computer & Geosciences*, 2, 191-203.
- Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer

- Science & Business Media.
- Bhattacharya, B.B., & Sen, M.K. (2003). Use of VFSA for resolution, sensitivity and uncertainty analysis in 1D DC resistivity and IP inversion. *Geophysical Prospecting*, 51(5), 393-408.
- Bhattacharya, S. (2021). *A Primer on Machine Learning in Subsurface Geosciences*. Springer.
- Bichet, V., Grisey, E., & Aleya, L. (2016). Spatial characterization of leachate plume using electrical resistivity tomography in a landfill composed of old and new cells (Belfort, France). *Engineering Geology*, 211, 61-73.
- Binley, A., Slater, L. D., Fukes, M., & Cassiani, G. (2005). Relationship between spectral induced polarization and hydraulic properties of saturated and unsaturated sandstone. *Water Resources Research*, 41(12).
- Binley, A., & Slater, L. (2020). *Resistivity and induced polarization: Theory and applications to the near-surface earth*. Cambridge University Press.
- Biswas, A., & Sharma, S.P. (2015). Interpretation of self-potential anomaly over idealized bodies and ambiguity analysis using very fast simulated annealing. *Near Surface Geophysics*, 13(2), 179-195.
- Blanchy, G., Saneiyan, S., Boyd, J., McLachlan, P., & Binley, A. (2020). ResIPy: an intuitive open-source software for complex geoelectrical inversion and modelling. *Computers & Geosciences*, 137, 104423.
- Brownrigg, D.R. (1984). The weighted median filter. *Communications of the ACM*, 27(8), 807-818.
- Bücker, M., Orozco, A. F., & Kemna, A. (2018). Electrochemical polarization around metallic particles-Part 1: The role of diffuse-layer and volume-diffusion relaxation. *Geophysics*, 83(4), E203-E217.
- Bücker, M., Undorf, S., Flores Orozco, A., & Kemna, A. (2019). Electrochemical polarization around metallic particles-Part 2: The role of diffuse surface charge. *Geophysics*, 84(2), E57-E73.
- Byrd, R. H., Gilbert, J. C., & Nocedal, J. (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1), 149-185.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1-27.
- Cardarelli, E., Cercato, M., & De Donno, G. (2018). Surface and borehole geophysics for the rehabilitation of a concrete dam (Penne, Central Italy). *Engineering Geology*, 241, 1-10.
- Cardarelli, E., & De Donno, G. (2019). Advances in electrical resistivity tomography: theory and case studies. In R. Persico, S. Piro, & N. Linford (Eds.), *Innovation in near-surface geophysics* (pp. 23-57). Elsevier.
- Cercato, M. (2011). Global surface wave inversion with model constraints. *Geophysical Prospecting*, 59, 210-226.
- Cercato, M., & De Donno, G. (2018). Focusing on soil-foundation heterogeneity through high resolution electrical and seismic tomography. *Near Surface Geophysics*, 16(1), 67-78.
- Cesca, S. (2020). Seiscloud, a tool for density-based seismicity clustering and visualization. *J Seismol*, 24, 443-457.
- Chambers, J.E. et al. (2006). Electrical resistivity tomography applied to geological, hydrogeological and engineering investigations at a former waste-disposal site. *Geophysics*, 71(6), B231-B239.
- Chunduru, R.K., Sen, M.K., & Stoffa, P.L. (1996). 2-D resistivity inversion using spline

- parameterization and simulated annealing. *Geophysics*, 61(1), 151-161.
- Clément, R., Oxarango, L., & Descloitres, M. (2011). Contribution of 3-D time-lapse ERT to the study of leachate recirculation in a landfill. *Waste Management*, 31(3), 457-467.
- Cole, K. S., & Cole, R. H. (1941). Dispersion and absorption in dielectrics I. Alternating current characteristics. *The Journal of Chemical Physics*, 9(4), 341-351.
- Coleman, T. F., & Li, Y. (1996). An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6(2), 418-445.
- Constable, S.C., Parker, R.L., & Constable, C.G. (1987). Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52(3), 289-300.
- Cutaneo, C., Vitale, A., & Fedi, M. (2023). Unsupervised boundary analysis of potential field data: A machine learning method. *Geophysics*, 88(3), G57-G65.
- Dahlin, T., Leroux, V., & Nissen, J. (2002). Measuring techniques in induced polarization imaging. *Journal of Applied Geophysics*, 50(3), 279-298.
- Dahlin, T., Rosqvist, H., & Leroux, V. (2010). Resistivity-IP mapping for landfill applications. *First Break*, 28(8), 1-10.
- Dahlin, T., & Leroux, V. (2012). Improvement in time-domain induced polarization data quality with multi-electrode systems by separating current and potential cables. *Near Surface Geophysics*, 10(6), 545-565.
- David, A., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *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. *Journal of Environmental Management*, 204, 709-720.
- De Carlo, L., Perri, M. T., Caputo, M. C., Deiana, R., Vurro, M., & Cassiani, G. (2013). Characterization of a dismissed landfill via electrical resistivity tomography and mise-à-la-masse method. *Journal of Applied Geophysics*, 98, 1-10.
- De Donno, G., & Cardarelli, E. (2017a). Tomographic inversion of time-domain resistivity and chargeability data for the investigation of landfills using a priori information. *Waste Management*, 59, 302-315.
- De Donno, G., & Cardarelli, E. (2017b). VEMI: a flexible interface for 3D tomographic inversion of time- and frequency-domain electrical data in EIDORS. *Near Surface Geophysics*, 15(1), 43-58.
- De Donno, G., & Piegari, E. (2022). Clustering analysis of ERT/IP data for leachate mapping in urban waste landfills. *Near Surface Geoscience 2022*, Belgrade, 18-22 September (accepted).
- De Donno, G., & Cercato, M. (2023). Resistivity and full-decay IP inversion for imaging a coastal aquifer prone to saline intrusion: The Pontina Plain case study (Central Italy). *Near Surface Geophysics*, 21(4), 275-287.
- De Donno, G., Melegari, D., Paoletti, V., & Piegari, E. (2024). Electrical and electromagnetic prospecting for the characterization of municipal waste landfills: A review. In A. Anouzla & S. Souab (Eds.), *Technical landfills and waste management: Volume 1: Landfill impacts, characterization and valorisation* (pp. 1-29).
- De Donno, G., Melegari, D., Paoletti, V., & Piegari, E. (2025). An integrated study of hard and soft cluster analyses for detecting leachate in a MSW landfill site using geoelectrical

- data. *Waste Management*, 195, 22-31.
- de Groot-Hedlin, C., & Constable, S. (1990). Occam's inversion in two dimensions. In *SEG Technical Program Expanded Abstracts 1990* (pp. 498-500). Society of Exploration Geophysicists.
- Dey, A., & Morrison, H.F. (1979). Resistivity modelling for arbitrarily shaped two-dimensional structures. *Geophysical Prospecting*, 27(1), 106-136.
- Di Giuseppe, M. G., Troiano, A., Troise, C., & De Natale, G. (2014). K-Means clustering as tool for multivariate geophysical data analysis. An application to shallow fault zone imaging. *Journal of Applied Geophysics*, 101, 108-115.
- Di Giuseppe, M. G., 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). *Journal of Applied Geophysics*, 156, 44-54.
- 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 Management*, 76, 629-642.
- Dick, M. D., Bery, A. A., Okonna, N. N., Ekanem, K. R., Bashir, Y., & Akingboye, A. S. (2024). A novel machine learning approach for interpolating seismic velocity and electrical resistivity models for early-stage soil-rock assessment. *Earth Sci. Inform.*, 1-20.
- Ergene, D., Aksoy, A., & Sanin, F.D. (2022). Comprehensive analysis and modelling of landfill leachate. *Waste Management*, 145, 48-59.
- 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 Management*, 171, 337-349.
- Everett, M. (2013). *Near-Surface Applied Geophysics*. Cambridge: 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. *Engineering Geology*, 241, 76-85.
- Fernández-Martínez, J.L. (2015). Model reduction and uncertainty analysis in inverse problems. *The Leading Edge*, 34(9), 1006-1016.
- Fiandaca, G., Auken, E., Christiansen, A. V., & Gazoty, A. (2012). Time-domain-induced polarization: Full-decay forward modeling and 1D laterally constrained inversion of Cole-Cole parameters. *Geophysics*, 77(3), E213-E225.
- Fiandaca, G., Ramm, J., Binley, A., Gazoty, A., Christiansen, A. V., & Auken, E. (2013). Resolving spectral information from time domain induced polarization data through 2-D inversion. *Geophysical Journal International*, 192(2), 631-646.
- Fiandaca, G., Madsen, L. M., & Maurya, P. K. (2018). Re-parameterisations of the Cole-Cole model for improved spectral inversion of induced polarization data. *Near Surface Geophysics*, 16(4), 385-399.
- Fiandaca, G., Olsson, P. I., Maurya, P. K., Köhl, A., Bording, T., Dahlin, T., & Auken, E. (2022). Heterodox transients in time-domain-induced polarization. *Geophysics*, 87(1), E35-E47.
- 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 Management*, 107, 121-132.
- Florio, G., & Lo Re, D. (2018). Terracing of potential fields by clustering methods. *Geo-*

- physics, 83(4), G47-G58.
- Fraternali, P., Morandini, L., & Gonzalez, S.L.H. (2024). Solid Waste Detection in Remote Sensing Images: A Survey. *Waste Management*, 189, 88-102.
- García-Fiscal, S., & Flores, C. (2018). Sensitivity of galvanic and inductive Induced Polarization methods to the Cole-Cole parameters. *Journal of Applied Geophysics*, 158, 1-10.
- Gaur, V.K., Gautam, K., Vishvakarma, R., Sharma, P., Pandey, U., Srivastava, J.K., Varjani, S., Chang, J., & Wong, W.C. (2024). Integrating advanced techniques and machine learning for landfill leachate treatment: Addressing limitations and environmental concerns. *Environmental Pollution*, 354124134.
- Gazoty, A., Fiandaca, G., Pedersen, J., Auken, E., & Christiansen, A. V. (2012). Mapping of landfills using time-domain spectral induced polarization data: the Eskelund case study. *Near Surface Geophysics*, 10(6), 575-586.
- Grana, D., Azevedo, L., & Liu, M. (2020). A comparison of deep machine learning and Monte Carlo methods for facies classification from seismic data. *Geophysics*, 85(4), WA41-WA52.
- Günther, T., Rücker, C., & Spitzer, K. (2006). Three-dimensional modelling and inversion of DC resistivity data incorporating topography-II. Inversion. *Geophysical Journal International*, 166(2), 506-517.
- 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. *Bioresource Technology*, 319, 124114.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*, 3rd edn. Vol. 1, Elsevier.
- Hanesch, M., Scholger, R., & Dekkers, M. J. (2001). The application of Fuzzy C-means cluster analysis and non-linear mapping to a soil data set for the detection of polluted sites. *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 26(11-12), 885-891.
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE*, 66(1), 51-83.
- Hellman, K., Ronczka, M., Günther, T., Wennermark, M., Rücker, C., & Dahlin, T. (2017). Structurally coupled inversion of ERT and refraction seismic data combined with cluster-based model integration. *Journal of Applied Geophysics*, 143, 169-181.
- 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 Management*, 178, 155-167.
- Ibraheem, M., Tezkan, B., & Bergers, R. (2021). Integrated interpretation of magnetic and ERT data to characterize a landfill in the north-west of Cologne, Germany. *Pure Appl. Geophys.*, 178(6), 2127-2148.
- Ingber, L. (1989). Very fast simulated re-annealing. *Mathematical and Computer Modelling*, 12(8), 967-973.
- 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 Management*, 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. *Environmental Science and Pollution Research*, 30(9), 24779-24789.
- Kamer, Y., Ouillon, G., & Sornette, D. (2020). Fault network reconstruction using agglomerative clustering: applications to southern Californian seismicity. *Nat. Hazards Earth Syst. Sci.* 20, 3611-3625.
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., & Kumar, V. (2019). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on knowledge and data engineering* 31, 8, 1544.
- Kaza, S., Yao, L., Bhada-Tata, P., & Van Woerden, F. (2018). What a waste 2.0: a global snapshot of solid waste management to 2050. World Bank Publications.
- Kemna, A., Binley, A., Cassiani, G., Niederleithinger, E., Revil, A., Slater, L., ... & Zimmermann, E. (2012). An overview of the spectral induced polarization method for near-surface applications. *Near Surface Geophysics*, 10(6), 453-468.
- Kessouri, P., Furman, A., Huisman, J. A., Martin, T., Mellage, A., Ntarlagiannis, D., ... & Placencia-Gomez, E. (2019). Induced polarization applied to biogeophysics: recent advances and future prospects. *Near Surface Geophysics*, 17(6), 595-621.
- Kim, H. J., & Kim, Y. (2011). A unified transformation function for lower and upper bounding constraints on model parameters in electrical and electromagnetic inversion. *Journal of Geophysics and Engineering*, 8(1), 21-26.
- Knight, R. J., & Endres, A. L. (2005). An introduction to rock physics principles for near-surface geophysics. in *Near Surface Geophysics*. Society of Exploration Geophysicists, 31-70.
- Koerner, R. M., & Soong, T. Y. (2000). Leachate in landfills: the stability issues. *Geotextiles and Geomembranes*, 18(5), 293-309.
- Kryszczuk, K., & Hurley, P. (2010). Estimation of the number of clusters using multiple clustering validity indices. In *Multiple classifier systems: Proceedings of the 9th International Workshop* (pp. 114-123). Springer.
- Lamirel, J. C., Dugué, N., & Cuxac, P. (2016, July). New efficient clustering quality indexes. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 3649-3657). IEEE.
- Lan, X., Zou, C., Peng, C., & Wu, C. (2023). Uncertainty Quantification in Intelligent-Based Electrical Resistivity Tomography Image Reconstruction With Monte Carlo Dropout Strategy. *IEEE transactions on geoscience and remote sensing*, 61, 1-16.
- Lavagnolo, M.C. (2019). Landfilling in developing countries. In: Cossu, R., Stegmann, R. (Eds.), *Solid Waste Landfilling: Concepts, Processes, Technologies*, Elsevier, 773-796.
- 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 Management*, 160, 90-100.
- Lindsey, C.R., Neupaneb, G., Spycher, N., Fairley, J.P., Dobson, P., Wood, T., McLing, T., & Conrad, M. (2018). Cluster analysis as a tool for evaluating the exploration potential of Known Geothermal Resource Areas. *Geothermics*, 72, 358-370.
- Liu, H., Shah, S., & Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9), 1635-1647.
- Liu et al. (2020). Deep learning inversion of electrical resistivity data. *IEEE Trans. Geosci. Remote Sens.*, 58(8), 5715-5728.

- Liu et al. (2023). Physics-driven deep learning inversion for direct current resistivity survey data. *IEEE Trans. Geosci. Remote Sens.*, 61, 1-11.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129-137.
- Loke, M. H., & Barker, R. D. (1996). Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method 1. *Geophys. Prospect.*, 44(1), 131-152.
- 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.
- Loke, M. H., Acworth, I., & Dahlin, T. (2003). A comparison of smooth and blocky inversion methods in 2D electrical imaging surveys. *Exploration Geophysics*, 34(3), 182-187.
- Loke, M. H., Chambers, J. E., Rucker, D. F., Kuras, O., & Wilkinson, P. B. (2013). Recent developments in the direct-current geoelectrical imaging method. *Journal of Applied Geophysics*, 95, 135-156.
- Lu W., Chen, J. (2022). Computer vision for solid waste sorting: A critical review of academic research. *Waste Management*, 142, 29-43.
- Lyra, G.B., Oliveira-Júnior, J.F., & Zeri, M. (2014). Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Alagoas state, Northeast of Brazil. *International Journal of Climatology* 34(13), 3546-3558,
- 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.
- Macnae, J. (2025). A physical interpretation of Cole-Cole equations and their ambiguous time constants for induced polarization models. *Geophysical Journal International*, 243(2), ggaf362.
- Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykw, M. A., Kwasi-groch, A., & Majek, K. (2022). Deep learning-based waste detection in natural and urban environments. *Waste Management*, 138, 274-284.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2), 431-441.
- Martinho, E. (2023). Electrical resistivity and induced polarization methods for environmental investigations: an overview. *Water, Air, & Soil Pollution* 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.
- Maurya, P. K., Rønne, V. K., Fiandaca, G., Balbarini, N., Auken, E., Bjerg, P. L., & Christiansen, A. V. (2017). Detailed landfill leachate plume mapping using 2D and 3D electrical resistivity tomography-with correlation to ionic strength measured in screens. *Journal of Applied Geophysics*, 138, 1-8.
- Maurya, P. K., Fiandaca, G., Christiansen, A. V., & Auken, E. (2018). Field-scale comparison of frequency- and time-domain spectral induced polarization. *Geophysical Journal International*, 214(2), 1441-1466.
- Melegari, D., & De Donno, G. (2024). ERT and TDIP survey for mapping of leachate plumes: Application to a MSW landfill in Italy. In *NSG 2024 30th European Meeting of Environmental and Engineering Geophysics*, 2024(1), 1-5. European Association of Geoscientists & Engineers.

- Mendel, J. M., Tan, W., Melek, W., & Ying, H. (2014). Introduction To Type-2 Fuzzy Logic Control: Theory and Applications. 1-376, Wiley-IEEE Press.
- Menke, W. (1984). Geophysical Data Analysis: Discrete Inverse Theory. Academic Press.
- Menke, W. (2015). Review of the generalized least squares method. *Surveys in Geophysics*, 36(1), 1-25.
- Mirzakhani, M., & Hashemi, H. (2022). Semisupervised fuzzy clustering for facies analysis using extended elastic impedance seismic attributes. *Geophysics*, 87(4), N75-N84.
- Morita, A.K.M., Ibelli-Bianco, C., Anache, J.A.A., Coutinho, J.V., Pelinson, N.S., Nobrega, J., Rosalem, L.M.P., Leite, C.M.C., Niviadonski, L.M., Manastella, C., & Wendland, E. (2021). Pollution threat to water and soil quality by dumpsites and non-sanitary landfills in Brazil: A review. *Waste Management*, 131, 163-176.
- Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, 100(B7), 12431-12447.
- Mukherjee, S., Mukhopadhyay, S., & Hashim, M.A. (2015). Contemporary Environmental Issues of Landfill Leachate: Assessment and Remedies. *Critical Reviews in Environmental Science and Technology* 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.
- Nguyen, F., Ghose, R., Isunza Manrique, I., Robert, T., & Dumont, G. (2018). Managing past landfills for future site development: A review of the contribution of geophysical methods. In 4th International Symposium on Enhanced Landfill Mining.
- 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 International*, 38(1), 2164361.
- Oldenburg, D. W., & Li, Y. (1994). Inversion of induced polarization data. *Geophysics*, 59(9), 1327-1341.
- Oldenburg, D.W., & Li, Y. (1999). Estimating depth of investigation in DC resistivity and IP surveys. *Geophysics*, 64(2), 403-416.
- Olsson, P. I., Dahlin, T., Fiandaca, G., & Auken, E. (2015). Measuring time-domain spectral induced polarization in the on-time: decreasing acquisition time and increasing signal-to-noise ratio. *Journal of Applied Geophysics*, 123, 316-321.
- Olsson, P. I., Fiandaca, G., Larsen, J. J., Dahlin, T., & Auken, E. (2016). Doubling the spectrum of time-domain induced polarization by harmonic de-noising, drift correction, spike removal, tapered gating and data uncertainty estimation. *Geophysical Journal International*, Volume 207, Issue 2, November 2016, Pages 774-784.
- Olsson, P. I., Fiandaca, G., Maurya, P. K., Dahlin, T., & Auken, E. (2019). Effect of current pulse duration in recovering quantitative induced polarization models from time-domain full-response and integral chargeability data. *Geophysical Journal International*, 218(3), 1739-1747.
- 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.
- Pelton, W. H., Ward, S. H., Hallof, P. G., Sill, W. R., & Nelson, P. H. (1978). Mineral discrimination and removal of inductive coupling with multifrequency IP. *Geophysics*, 43(3), 588-609.

- Piegari, E., & Paoletti, V. (2021). Analysis of geoelectric data through machine learning algorithms for waste leachate detection. Mediterranean Geosciences Union Annual Meeting, MedGU-21, Istanbul, Turkey, 25–28 November 2021.
- Piegari, E., Herrmann, M., & Marzocchi, W. (2022). 3-D spatial cluster analysis of seismic sequences through density-based algorithms. *Geophysical Journal International* 230, 2073-2088.
- Piegari, E., De Donno, G., Melegari, D., & Paoletti, V. (2023). A machine learning-based approach for mapping leachate contamination using geoelectrical methods. *Waste Management*, 157, 121-129.
- Power, C., Tsourlos, P., Ramasamy, M., Nirvolis, A., & Mkandawire, M. (2018). Combined DC resistivity and induced polarization (DC-IP) for mapping the internal composition of a mine waste rock pile in Nova Scotia, Canada. *Journal of Applied Geophysics*, 150, 40-51.
- 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.
- Rahmani, A., Laner, D., & Melage, A. (2025). A spectral induced polarization (SIP) lab-scale framework for monitoring controlled ammonium desorption from municipal solid waste. *ACS ES&T Engineering*, 5(8), 1953-1963.
- Raji, W.O., & Adeoye, T.O. (2017). Geophysical mapping of contaminant leachate around a reclaimed open dumpsite. *Journal of King Saud University - Science*, 29, 348-359.
- Rana, R., Ganguly, R., & Gupta, A. (2018). Indexing method for assessment of pollution potential of leachate from non-engineered landfill sites and its effect on ground water quality. *Environmental Monitoring and Assessment*, 190, 46, 1-23.
- Ray, A., & Myer, D. (2019). Bayesian geophysical inversion with trans-dimensional Gaussian process machine learning. *Geophys. J. Int.*, 217(3), 1706-1726.
- Revil, A., Coperey, A., Shao, Z., Florsch, N., Fabricius, I. L., Deng, Y., & Gunnink, J. L. (2017). Complex conductivity of soils. *Water Resources Research*, 53(8), 7121-7147.
- Revil, A., Ahmed, A. S., Coperey, A., Ravanel, L., Sharma, R., & Panwar, N. (2020). Induced polarization as a tool to characterize shallow landslides. *Journal of Hydrology*, 589, 125369.
- Rücker, C., Günther, T., & Wagner, F. M. (2017). pyGIMLi: An open-source library for modelling and inversion in geophysics. *Computers & Geosciences*, 109, 106-123.
- Samouelian, A., Cousin, I., Tabbagh, A., Bruand, A., & Richard, G. (2005). Electrical resistivity survey in soil science: a review. *Soil and Tillage Research*, 83(2), 173-193.
- Schubert, E. (2023). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explor. Newsl.*, 25(1), 36-42.
- Schwarz, E. (1978). Estimating the dimension of a model. *Ann. Statist.*, 6(2), 461-464.
- Seigel, H. O. (1959). Mathematical formulation and type curves for induced polarization. *Geophysics*, 24(3), 547-565.
- Sen, M.K., & Stoffa, P.L. (1996). Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophysical Prospecting*, 44, 313-350.
- Sen, M.K., & Stoffa, P.L. (2013). *Global Optimization Methods in Geophysical Inversion*. Cambridge University Press.
- Sharma, S. P., & Kaikkonen, P. (1998). Two-dimensional non-linear inversion of VLF-R data using simulated annealing. *Geophysical Journal International*, 133(3), 649-668.

- Sharma, S. P. (2012). VFSARES—a very fast simulated annealing FORTRAN program for interpretation of 1-D DC resistivity sounding data from various electrode arrays. *Computers & Geosciences*, 42, 177-188.
- Sharma, A., Ganguly, R., & Kumar Gupta, A. (2020). Impact assessment of leachate pollution potential on groundwater: an indexing method. *Journal of Environmental Engineering*, 146(3), 05019007.
- Sharma, K., & Sood, M. (2024). Monitoring, classification and analysis of waste disposal sites using Machine Learning. *Procedia Computer Science*, 235, 1558-1567.
- Shukla, S., & Naganna, S. (2014). A Review on K-Means DATA Clustering approach. *International Journal of Information & Computation Technology*, 4(17), 1847-1860. ISSN 0974-2239.
- Siegel, H. E. (1959). A theory of induced polarization effects (for step-function excitation). In J. R. Wait (Ed.), *Overvoltage research and geophysical applications* (pp. 4-21).
- Slater, L. D., & Lesmes, D. (2002). IP interpretation in environmental investigations. *Geophysics*, 67(1), 77-88.
- Soupios, P., Papadopoulos, N., Papadopoulos, I., Kouli, M., Vallianatos, F., Sarris, A., & Manios, T. (2007). Application of integrated methods in mapping waste disposal areas. *Environ. Geol.* 53(3), 661-675.
- Soupios, P., & Ntarlagiannis, D. (2017). Characterization and monitoring of solid waste disposal sites using geophysical methods: current applications and novel trends. *Modelling trends in solid and hazardous waste management*, 75-103.
- Steiner, M., Katona, T., Fellner, J., & Orozco, A. F. (2022). Quantitative water content estimation in landfills through joint inversion of seismic refraction and electrical resistivity data considering surface conduction. *Waste Management*, 149, 21-32.
- Straus, D.M. (2019). *Clustering Techniques in Climate Analysis*. Climate Science, Oxford University Press.
- Sun, J., & Li, Y. (2015). Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering. *Geophysics*, 80(4), ID1-ID18.
- Sun, J., & Li, Y. (2016). Joint inversion of multiple geophysical data using guided fuzzy c-means clustering. *Geophysics*, 81(3), ID37-ID57.
- 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 Management*, 157, 180-189.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.
- Tarasov, A., & Titov, K. (2013). On the use of the Cole-Cole equations in spectral induced polarization. *Geophysical Journal International*, 195(1), 352-356.
- Thorndike, R.L. (1953). Who belongs in the family? *Psychometrika* 18 (4), 267-276.
- Torres, R.N., & Fraternali, P. (2021). Learning to Identify Illegal Landfills through Scene Classification in Aerial Images. *Remote Sensing*, 13(22), 4520.
- Tronicke, J., Holliger, K., Barrash, W., & Knoll, M. D. (2004). Multivariate analysis of cross-hole georadar velocity and attenuation tomograms for aquifer zonation. *Water Resources Research*, 40(1).
- Urban Plan of Montecorvino Pugliano. (2011). Official Bulletin of the Campania Region, No 1 of 3 January 2011.
- Vaccari, M., Tudor, T., & Vinti, G. (2019). Characteristics of leachate from landfills and

- dumpsites in Asia, Africa and Latin America: an overview. *Waste Manag.* 95, 416-431.
- Varfinezhad, R., Fedi, M., & Milano, M. (2022). The role of model weighting functions in the gravity and DC resistivity inversion. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-15.
- Vitale, G., Gabbriellini, G., & Fedi, M. (2023). Deep learning to estimate the basement depth by gravity data using a feedforward neural network. *Geophysics*, 88(3), G95-G103.
- Wang, J., & Qiao, Z. (2024). A comprehensive review of landfill leachate treatment technologies. *Frontiers in Environmental Science*, 12.
- Weber, Z. (2000). Seismic traveltime tomography: a simulated annealing approach. *Physics of the Earth and Planetary Interiors*, 119(1-2), 149-159.
- Wei, X., Yin, Z., Scheidt, C., Darnell, K., Wang, L., & Caers, J. (2024). Constructing priors for geophysical inversions constrained by surface and borehole geochemistry. *Surveys in Geophysics*, 45(4), 1047-1079.
- Weller, A., Slater, L., & Nordsiek, S. (2013). On the relationship between induced polarization and surface conductivity: Implications for petrophysical interpretation of electrical measurements. *Geophysics*, 78(5), D315-D325.
- Whiteley, J. S., Chambers, J. E., Uhlemann, S., Wilkinson, P. B., & Kendall, J. M. (2019). Geophysical monitoring of moisture-induced landslides: A review. *Reviews of Geophysics*, 57(1), 106-145.
- Whiteley, J. S., et al. (2021). Rapid characterisation of landslide heterogeneity using unsupervised classification of electrical resistivity and seismic refraction surveys. *Engineering Geology*, 290, 106189.
- WHO. (2015). Waste and human health: evidence and needs: WHO meeting report 5-6 November 2015: Bonn, Germany.
- Winsauer, W. O., Shearin Jr, H. M., Masson, P. Y., & Williams, M. (1952). Resistivity of brine-saturated sands in relation to pore geometry. *AAPG Bulletin*, 36(2), 253-277.
- Wu, J., Wu, Z., Cao, J., Liu, H., Chen, G., & Zhang, Y. (2017). Fuzzy consensus clustering with applications on big data. *IEEE Transactions on Fuzzy Systems*, 25(6), 1430-1445.
- Xia, W., Jiang, Y., Chen, X., & Zhao, R. (2022). Application of machine learning algorithms in municipal solid waste management: A mini review. *Waste Management & Research*, 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 Management*, 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.
- Yin, L., Yang, R., Gabbouj, M., & Neuvo, Y. (1996). Weighted median filters: a tutorial. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 43(3), 157-192.
- Yu, S., & Ma, J. (2021). Deep Learning for Geophysics: Current and Future Trends. *Reviews of Geophysics*, 59(3).
- Zaini, M.S.I., Hasan, M., & Zolkepli, M.F. (2022). Urban landfills investigation for leachate assessment using electrical resistivity imaging in Johor, Malaysia. *Environmental Challenges*, 6, 100415.
- Zeiss, C., & Uguccioni, M. (1994). Mechanisms and patterns of leachate flow in municipal

- solid waste landfills. *Journal of Environmental Systems*, 23(3), 247-270.
- Zhang, W., Zhang, Y., Gu, X., Wu, C., & Han, L. (2022). *Application of Soft Computing, Machine Learning, Deep Learning and Optimizations in Geoengineering and Geoscience*. Springer.
- Zhang, P., Chai, J., Cao, J., Qin, Y., Dang, M., Geng, K., & Wei, Y. (2023). Landfill leachate generation mechanism study: a review. *International Journal of Environmental Science and Technology*, 20(8), 9271-9290.
- Zhang, Z., et al. (2023). An integrated machine learning framework with uncertainty quantification for three-dimensional lithological modeling from multi-source geophysical data and drilling data. *Engineering Geology*, 324, 107255.
- 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. *Journal of Hazardous Materials*, 480, 135908.