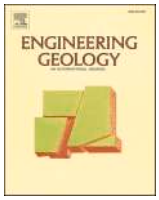
Engineering Geology 337 (2024) 107589

Contents lists available at ScienceDirect

Engineering Geology

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



Geo-constrained clustering of resistivity data revealing the heterogeneous lithological architectures and the distinctive geoelectrical signature of shallow deposits

Paolo Ciampi [a,d,*], Leonardo Maria Giannini [a,d], Giorgio Cassiani [b], Carlo Esposito [a,d], Marco Petrangeli Papini [c,d]

[a] Department of Earth Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[b] Department of Geosciences, University of Padua, Via Gradenigo 6, 35131 Padua, Italy
[c] Department of Chemistry, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[d] CERI Research Center, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy

ARTICLE INFO

Keywords:
Electrical resistivity
Geoelectrical signature
Geology
Geophysics
Heterogeneity
Data-driven clustering

ABSTRACT

For all applications, subsurface models should be consistent with all available geological and geophysical knowledge. Current practices for synergistic interpretation of geological and geophysical approaches often rely on purely qualitative comparisons, resulting sometimes in inconsistent findings. This study introduces a procedure for a statistical and geo-constrained clustering of electrical resistivity data derived from Electrical Resistivity Tomography (ERT) to address this gap, providing a quantitative parameterization for site-specific geoelectrical signatures of litho-stratigraphic architectures. Seventeen boreholes and three ERT surface profiles were employed to link geophysical inversion results to geological criteria. Core samples allowed grain size analyses, while geological-statistical clustering of electrical resistivity, driven by the observation of stratigraphic contacts in drilled boreholes, established a parametric relationship between geology and geophysics. The iterative clustering procedure, utilizing a classification algorithm, geological boundary constraints, and granulometric analyses, discriminated six distinct lithological clusters, capturing the lateral and vertical heterogeneity of shallow deposits. Subsequent spatial grouping of anthropogenic materials delineated lithological structures and facilitated the classification and identification of filling materials, silty sands, clayey sands, and clays and silts, each exhibiting distinct resistivity variations. The geo-driven geophysical clustering revealed complex lithological structures, especially paleo-channels, capturing their unique geoelectric footprints. The iterative clustering of geo-constrained resistivity data emerges as a powerful tool for subsurface exploration, contributing significantly to understanding lithological heterogeneity, quantifying statistically-based geoelectrical parametrization of shallow sediments, and evaluating the resistivity signature of different deposits. By bridging the gap between geology and geophysics, this data-driven approach establishes a benchmark for future applications. For instance, in the context of contaminated sites, it can be applied to identify pollutants versus geological heterogeneities.

1. Introduction

In the realm of subsurface exploration, the synergy of geological and geophysical characterization methods plays a pivotal role in unraveling the unknown structure of the Earth's subsurface (Foged et al., 2014; Høyer et al., 2015; Lelièvre et al., 2009; Looms et al., 2008). Geological surveys provide a foundational understanding of subsurface features and stratigraphy, while geophysical investigations, such as electrical resistivity tomography (ERT) surveys, delve into the hidden complexities of the subsurface providing a complete spatial coverage and high spatial resolution, albeit in the form of physical, rather than geological, properties of different structures (Binley and Slater, 2020; Crook et al., 2008; Loke et al., 2013; Revil et al., 2017). While the direct lithological knowledge of the subsurface obtained through stratigraphic boreholes is

* Corresponding author at: Department of Earth Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy.
E-mail addresses: paolo.ciampi@uniroma1.it (P. Ciampi), leonardomaria.giannini@uniroma1.it (L.M. Giannini), giorgio.cassiani@unipd.it (G. Cassiani), carlo.esposito@uniroma1.it (C. Esposito), marco.petrangelipapini@uniroma1.it (M.P. Papini).

https://doi.org/10.1016/j.enggeo.2024.107589
Received 5 April 2024; Received in revised form 28 May 2024; Accepted 6 June 2024
Available online 7 June 2024

irreplaceable, geophysical investigations, despite their intrinsic uncertainty both in terms of relationship between physical properties and geological characteristics, and in terms of inversion uncertainties, open up the possibility of capturing geostatistical properties in detail (He et al., 2014; Slater, 2007; Stan and Stan-Kłeczek, 2014). Besides, geophysical techniques can also play a crucial role in uncovering subsurface properties, extending beyond mere lithological characteristics (Binley et al., 2015; De Donno and Cercato, 2023; Flores-Orozco et al., 2021; Linde et al., 2015; Maurya et al., 2018; Whiteley et al., 2021). The strong link between geophysical investigations and direct information from drilling is at the heart of exploration for oil and gas or geothermal resources. The approach is much less adopted, and adopted in much rougher manner, in environmental applications, albeit here too it has the same potential. The combination of the observational capability of geological surveys with the high-resolution characterization and data spatialization of geoelectrical methods is a process that incorporates several forms of knowledge. In fact, quantitative interpretation of geophysical data generates pictures of the examined subsoil volume, which must be calibrated against geological features (Ciampi et al., 2022; Gemail, 2015; Samouëlian et al., 2005; Wu and Grana, 2017). This fusion of knowledge, typically deciphered by specialists based on qualitative principles through traditional approaches such as correlating features one by one (Di Maio et al., 2014; Giao et al., 2003; Klose, 2006; Sudha et al., 2009). Current practices involve qualitatively comparing geophysical models to hydrogeological data (Hubbard and Rubin, 2000; Nazaruddin et al., 2017), with limited exploration into the correlation between geological attributes and physical parameters (Abbas et al., 2022; Olierook et al., 2021). As a consequence, interpretations become challenging and can lead to inconsistent, ambiguous, and erroneous findings in reconstructing the spatial heterogeneity of subsurface geological models (Klose, 2006). By integrating stratigraphic observations with widely adopted, easily deployable, and cost-effective geophysical techniques, and utilizing statistically-driven geological and geophysical data models, it becomes feasible to effectively address spatial heterogeneity within these models (Aleardi et al., 2021; Bersezio et al., 2007; Madsen et al., 2023). Notoriously, a statistical approach can combine heterogeneous data sources, managing the distributions of variables associated with and conditioned by the observed data (Giannini et al., 2021; Jenks, 1967). Recent advances in integrated geologic-physical statistical analysis stem from a significant shift in modeling techniques and computational tools (Foged et al., 2014; Michel et al., 2020; Pedregosa et al., 2011; Piegari et al., 2023). In recent years, there has been a notable upswing in the application of machine learning for data fusion, driven by its proficiency in processing experimental datasets (Bressan et al., 2020; Jia et al., 2021; Marzán et al., 2021; Nawaz and Curtis, 2019; Xu and Green, 2023). Clustering algorithms and coupled inversion techniques have also found widespread use (Bouchedda et al., 2012; Hinnell et al., 2010; Whiteley et al., 2021; Yu and Ma, 2021).

Embarking on cutting-edge methodologies, this research aims to advance our comprehension of the relationships between lithological units and the electrical resistivity of formations (Bosh, 1999; Bosh et al., 2001). Although literature reports various ranges of resistivity for soils and Earth materials (Alpaslan, 2021; Chang et al., 2023; Park et al., 2017; Sikandar and Christen, 2012), drawing from the work of Palacky (1988), a gap becomes evident in attributing a typical electrical resistivity to lithological units: in particular, the observable ranges of resistivity are so wide and so widely superimposed as not to allow any general correlation to be drawn. This is due, essentially, to the strong dependence of soil and rock electrical resistivity on factors other than lithology, and particularly water saturation and water ionic content. This highlights a need for a quantitative geophysical lithology parameterization and the identification of site-specific geoelectrical signatures for different deposits. Our approach develops an integrated interpretational methodology for a quantitative description of geological-geophysical structures, bolstered by validation through geological-technical analyses. Several studies in the literature have made attempts to integrate ERT with geotechnical data for subsurface soil characterization (Braga et al., 1999; Cosenza et al., 2006; Stan and Stan-Kłeczek, 2014; Sudha et al., 2009). Recent strides in petrophysical reconstruction from geophysical models, aided by open-source Python codes for data inversion (Rücker et al., 2017), have paved the way for promising endeavors.

While geophysics and geology furnish essential spatial information for modeling heterogeneity, their credibility necessitates validation through direct measurements. To fortify the reliability of both indirect (geophysics) and direct (stratigraphic logs) data, this research implements a statistical and geo-driven clustering procedure for geophysical data. Leveraging stratigraphic contact information from boreholes as a training set, the research employs a geological-statistical clustering of electrical behavior based on grain size analyses. Compared to the approaches commonly used in the literature, which rely purely on qualitative comparison for the joint interpretation of geophysical and geological data, our study showcases a novel procedure for statistical and geo-constrained clustering of electrical resistivity data. This approach synergistically interprets geological and geophysical findings and provides a quantitative parameterization for site-specific geoelectrical signatures of litho-stratigraphic architectures. This strategy, distinct from predictive models and synthetic approximations, signifies an important advancement in the interpretation of the geological-geophysical interplay. This strategy establishes a parametric relationship between geology and geophysics, leading to a simple and objective classification. The approach we delineate for establishing connections between resistivity information and lithologies was implemented in the analysis of shallow sediments at the new high-speed railway station in Bologna, Italy. The sedimentary architecture of the case study has been comprehensively characterized on a broader scale within the context of remediating contaminated groundwater, as outlined in Ciampi et al. (2019). In the following we will concentrate the analysis essentially in the portions of the subsurface better imaged by the acquired ERT data. The geostatistical-based geophysical parametrization of subsurface structures attempts to enhance the discretization resolution of known classification tables and the association of resistivity ranges with different types of deposits reported in the literature. In contrast to conventional methods that heavily depend on interpreting geophysical inverted models, the innovative approach presented effortlessly adjusts to produce comprehensive model sections. It seamlessly integrates varied information sources, establishing a harmonious connection between the lithological attributes of sediments and their resistivity distribution. This method delivers a geological-statistical and data-driven interpretation of electrical behavior, grounded in a profound understanding of geological parameters. Ultimately, the proposed method contributes significantly to formulating a parametric relationship between geology and geophysics.

2. Materials and methods

In the context of the proposed case study, 17 boreholes were drilled and 3 parallel ERT profiles gathered in the same area. These investigations covered a total area of approximately 900 m^2. The geological surveys reach variable depths, ranging from approximately 9 m to 17 m. Twelve core samples (S1-S12) were collected at different depths to perform grain size analyses (ASTM International, 2011), determine Atterberg limits (ASTM International, 2010), and provide soil classification according to the unified soil classification system (USCS) (Table 1).

The 3 ERT lines consist of 72 electrodes each, with an electrode spacing of 1 m. A dipole-dipole skip-4 acquisition scheme was employed (skipping 4 electrodes in each dipole), resulting in dipole lengths of 5 m for both current injection and voltage measurement. Electrodes are arranged in a linear pattern on the ground surface. The dipole-dipole skip-4 acquisition scheme refers to a specific arrangement and spacing of these electrodes for measuring subsurface resistivity. In this scheme, two

Table 1
Core samples collected at various depth intervals from the boreholes drilled in the case study area.

Borehole	Sampling depth range (m)	Sample ID
	6–7	S1
B1	7–8	S2
	8–9.5	S3
	5.5–6.5	S4
B2	7–8	S5
	8.5–9.5	S6
	5.5–6.5	S7
B4	7.5–8.5	S8
	9–10	S9
	5–6.5	S10
B5	7–8	S11
	9–10	S12

pairs of electrodes are used: one pair for injecting current into the ground and another pair for measuring the resulting voltage difference. "Skipping 4 electrodes" means that 4 electrodes are left out between the electrodes in each dipole configuration. With a skip of 4 electrodes, the dipole length is 5 times the spacing between adjacent electrodes. For instance, if each electrode is spaced 1 m apart, the dipole length would be 5 m. A comprehensive reciprocal acquisition was performed to assess measurement errors, following best practices for high-quality data acquisition (e.g. Cassiani et al., 2006, 2014). A complete acquisition of all reciprocity pairs (exchanging the potential with the current electrodes) is crucial for error estimation in data acquisition, allowing for the identification and removal of anomalous elements before data inversion (e.g. Binley and Kemna, 2005; Deiana et al., 2007). ERT inversion was carried out using the ERT inversion programs provided by Lancaster University (ResIPy - Blanchy et al., 2020). An error level equal to 5% was used, as the raw data were filtered out using the same level of reciprocal error, corresponding to about 65% of the used quadrupoles to be saved for the inversion process. Note that the presence of a shallow clay layer in the near subsurface makes deep penetration of electrical current problematic, thus requiring in general larger quadrupoles to enhance the signal-to-noise ratio. To test the suitability of different acquisition schemes, three different approaches were tested along the same line (referred to as Line 1 in the following description), namely: a dipole-dipole skip-2, a dipole-dipole skip-4, and a Wenner-Schlumberger schemes. The results (not reported here for brevity) are very similar – being filtered at the same 5% reciprocal error level, with the Wenner-Schluberger results being a slightly smoother version of the dipole-dipole acquisitions, being the latter essentially indistinguishable from each other. To warrant a suitable resilience towards signal/noise ratio out choice was to adopt the larger dipole-dipole scheme (skip-4) that evidently is not detrimental, in this case, of the resulting spatial resolution.

The information derived from borehole samples, concerning depth of stratigraphic contacts, particle-size characteristics, liquid and plastic limit, plasticity index, and resistivity parameters has been stored in a multi-source geodatabase, a model for the simultaneous storage and management of spatial data (Ciampi et al., 2021). The interpolation of observed geological boundaries during drilling was employed to craft a 3D comprehensive visualization of geological borehole observations and stratigraphic schemes. The interpolation of inverted resistivity data was performed to generate a 2D canvas depicting the spatial distribution of the parameters obtained from ERT investigations. The inverse distance weighted (IDW) method was chosen for interpolating geological-physical data. IDW interpolation is a weighted averaging technique in which the weights of known data points are inversely proportional to their distances from an interpolated unknown point. The distance is elevated to an exponent. The number of adjacent known points and the exponent are user-specified parameters (Wang et al., 2017). In implementing the gridding method, we set 4 neighboring data points and a fixed weighting exponent of 2. Notably, high-fidelity served as an additional filtering option in this process, ensuring that the voxel values in the 3D solid model and the pixel values in the 2D sections align with the observations of stratigraphic contacts in the drilled boreholes and the inverted resistivity in two-dimensional space. Essentially, the interpolation error at points where measurements/data are available is zero. This approach enhances the accuracy and reliability of the overall model (Ciampi et al., 2022). Our work bypasses the analysis of the influence of different gridding methods on geological modeling, a topic extensively covered in the literature (Wang et al., 2017), and instead focuses on the search for a quantitative correlation method for electrical behavior with geological parameters.

Subsequently, attention was focused on the opportunity to link and correlate geophysical data with geological data to understand the resistivity behavior of different lithologies and identify the geoelectrical signature of various deposits. To achieve this, the resistivity values obtained from 1D profiles extracted from the 2D resistivity inverted sections were associated and overlaid with observations of stratigraphic contacts deduced from boreholes B1, B2, B3, B4, and B5, located within a distance of 2 m from the ERT line EF (Fig. 1). This coupled geological-physical attributes approach aims at discriminating geologically data-driven structures based on electrical properties. The contouring of the geoelectrical section (Binley and Slater, 2020), should be harmonized with geological constraints, specifically the stratigraphic discontinuity identified by the geologist during drilling. It must be noted that the resistivity images derived from ERT are naturally a smooth version of reality, as a consequence of both the diffusive nature of the underlying current flow equation, and the regularization towards smooth solutions utilized in the inversion process (Binley and Kemna, 2005). Therefore, the identification of possible sharp geological contacts in the ERT section shall look for plausible resistivity value contours that mimic such interfaces and closely match these geological features.

An iterative clustering process of resistivity data collected along vertical sections of the 2D inverted ERT images, starting with two clusters up to six in this case, halts upon minimizing the differences between the cluster separation line and the stratigraphic discontinuity. The adopted iterative classification method uses a k-means algorithm to cluster stratigraphic units based on resistivity distribution (Han et al., 2012; Jenks, 1967; MacQueen, 1967).

The k-means algorithm is a well-known unsupervised technique for partitioning-based clustering of a n-dimensional vector. k-means clustering may not be the most efficient algorithm for clustering 1D data, however, we decide to implement and adapt k-means for 1D resistivity data clustering considering future developments of the study by introducing new variables. Since k-means was originally designed for multidimensional clustering (Hartigan and Wong, 1979), using it for clustering a single variable like resistivity is an adaptation. The natural break method, for instance, is a native 1D clustering approach, so theoretically, it would appear more appropriate to use this approach (Jenks, 1967). However, the use of k-means opens up the future opportunity to analyze clustering with multiple variables, such as water content and contaminant concentration data. The k-means algorithm aims to minimize the within-cluster variance, effectively trying to make the clusters as compact as possible. As an unsupervised method, it's not possible to define a priori a number k of clusters, although many scientific works report results applied to different fields of study, where the attempt is made to solve this issue (Kumar et al., 2024). Since k-means requires k as an input and doesn't learn it from data, there is no right answer in terms of the number of clusters that we should have in any problem. The geological domain is not an exception, knowledge and expert judgment may help. To try to solve this issue we decide to implement the k-means algorithm in a geo-iter-supervised procedure.

We define the vector space V, representing geophysical measurements:

$$\mathrm{V} = \{\mathrm{v} \in \mathbb{R}^{\mathrm{n}} | \mathrm{v} = (\mathrm{v}_1, \mathrm{v}_2, \ldots, \mathrm{v}_{\mathrm{n}})\}$$

Each vector v includes spatial coordinates and physical measure-

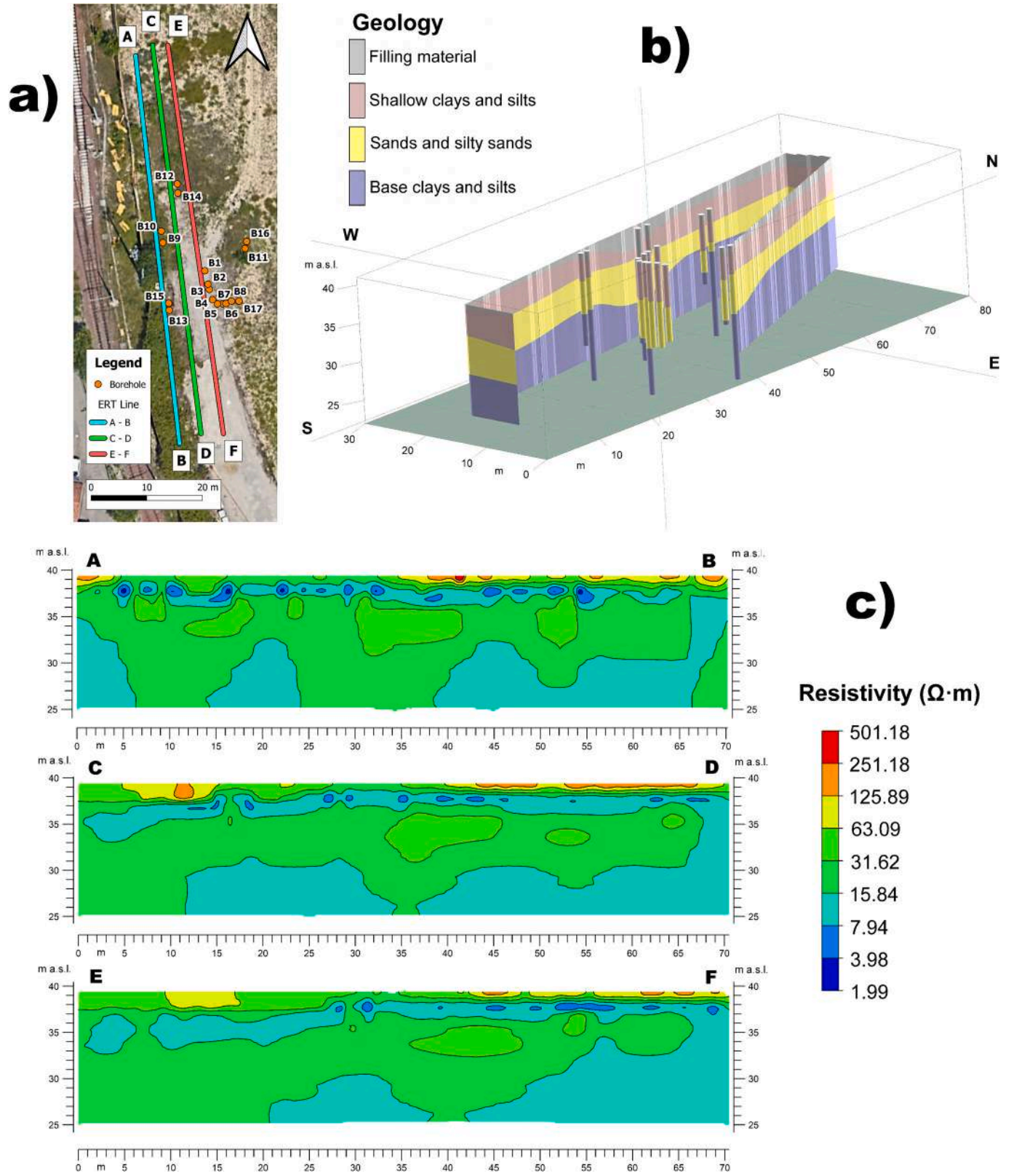


Fig. 1. Localization of ERT profiles and drilled boreholes in the investigation domain (a). 3D geological cross sections and stratigraphic boreholes realized at the study site (b). 2D ERT images acquired along profiles AB, CD, and EF (c).

ments such as electrical resistivity (ρ), forming a n – dimensional vector space where n encompasses both spatial and measurement dimensions (electrode spacing). The geological frontier, F is defined as a subset of V that indicates significant geological changes:

$$F \subset V$$

The method adapts the k-means algorithm to focus on minimization of the Euclidean distance between geographically situated measurement

points and the defined geological frontier, labeling these points with relevant physical measurements, such as electrical resistivity. For a labeled point v, the computation of distance to the geological frontier F relies solely on its geographic location, articulated by the formula:

$$d(v, F) = \min_{f \in F} \|x_v - x_f\|_2$$

Here, x_v and x_f denote the geographic coordinates of a point v and a point f on the frontier F, respectively.

In the case study presented, we solve a 2D geometrical problem with regularly spaced x,y data points, each associated with a value of inverted resistivity. The iterative resistivity data clustering procedure is delineated as follows:

- data initialization: Initialize the analysis by organizing the ERT data into a dataset $\{(x_i, y_i, \rho_i)\}_{i=1}^{N}$, where N represents the total number of data points, x_i, y_i, are the spatial coordinates, and ρ_i is the inverted resistivity value at each point. Geological frontiers, delineating significant discontinuities, are defined by a specific subset of spatial point $\{(x_f, y_f)\}_{i=1}^{M}$, where M indicates the count of such frontier points identified through physical samplings.
- initial cluster number selection k: Select an initial count of clusters k, with the premise that the ideal count will be determined dynamically through iterative analysis.
- clustering and labeling: Apply the k-means algorithm to classify and group the ERT data based on resistivity values ρ_i, assigning each point to a specific cluster.
- formalization of the separation geometry within cluster that is, the determination of the separation geometry and distribution of its constituent data points relative to F, encapsulating a novel approach to cluster differentiation in geophysical data analysis. The separation geometry is articulated as follows:
  a) for clusters exhibiting a linear distribution of data points, indicative of linear geological features, the separation geometry is formalized as a line. This line is mathematically derived through a process of linear regression applied to the data point within the cluster, optimizing the fit to minimize the orthogonal distance to F, thus ensuring an accurate representation of the cluster's alignment with geological boundaries.
  b) in instances where clusters manifest a more complex spatial distribution, extending over a broader area, the separation geometry is delineated as a surface. The determination of this surface involves an optimization process that employs methods such as least squares fitting to define a surface that minimizes the cumulative distances between the cluster's data points and the points constituting F. This surface serves as a geometric abstraction of the cluster's area in relation to the geological frontier, offering insights into the areal extent of geological formations.
- cumulative distance calculation: For each iteration, characterized by a specific cluster configuration with k clusters, we calculate the cumulative distance d_k. This distance represents the total Euclidean distance from all data points to their nearest geological frontier, defined as:

$$d_k = \sum_{i=1}^{N} \min_{f \in F} \|x_i - x_f\|_2$$

This metric quantifies the alignment of data points with respect to the geological frontier.

- iterative optimization and cluster adjustment. To minimize d_k, thereby enhance the geological relevance of the clustering outcome, the number of clusters k is iteratively adjusted as follows:
  a) increase k by one, $k \leftarrow k + 1$, to explore a more granular cluster configuration.
  b) reapply the clustering algorithm with the new k, redefining cluster assignments and recalculating d_{k+1}.
  c) if the change in cumulative distance $\Delta d = |d_{k+1} - d_k|$ falls below a predefined threshold, this indicates that further refinement does not significantly improve the geological interpretability of the clustering.

By adjusting the cluster count based on the sum of distances to these frontiers, the methodology seeks to achieve a balance between the granularity of the subsurface segmentation and the fidelity to geological realities, thereby facilitating a more nuanced and geologically informed interpretation of the ERT data.

Observed geological boundaries are employed as geometric parameters to halt the iteration cycle and separate different lithologies during the iterative search for the optimal number of clusters. The discriminant criterion for determining the geo-fitted number of clusters also derives from grain size analyses conducted on samples collected at the borehole points. To conduct the analysis, we developed a standalone procedure comprising procedural Python code (https://www.python.org/doc/). For data clustering through the k-means algorithm, we employed the Scikit-learn library (Pedregosa et al., 2011). The code operates by computing the minimization of the Euclidean distance between the geological hyperplane, inherited from the geological survey (in this case drilled boreholes and particle size analyses), and the hyperplane derived from the clustering process. The geological-physical data classification method was initially applied to geoelectrical section EF and then replicated on ERT sections AB and CD (Fig. 1).

In a subsequent step, clustering of the anthropogenic materials (rubble on top of the natural geological sequence) was performed to address the well-known heterogeneity of these deposits, both in terms of composition and compaction, and consequently, electrical resistivity. Starting from this assumption, we delimited the filling material by searching for a surface- or multiple surfaces-such that the entropy of resistivity values is maximized and the surface area between the surface (free air) and the underlying geological substrate is minimized. Thus, the clustering of filling materials is performed by grouping points with the maximum variation in resistivity closest to the topographic surface.

The above-described classification of geological-geophysical data, strongly geo-constrained by inherited geological information and combining mathematical-statistical rules, aims at identifying different signatures in terms of resistivity within the geological-technical horizons characterizing the sequence of the study area. The statistical-based geophysical parametrization of subsurface structures attempts to enhance the level of discretization detail of known classification tables and the association of resistivity ranges with different types of deposits reported in the literature (Alpaslan, 2021; Chang et al., 2023; Palacky, 1988; Park et al., 2017; Sikandar and Christen, 2012). The technique presented in this study aims at connecting the lithological characteristics of sediments with their resistivity distribution and deliver a data-driven interpretation of electrical behavior based on knowledge of the geological parameters, contributing to the formulation of a parametric relationship between geology and geophysics.

3. Results

3.1. Hydrogeophysical site model

The geological and geophysical models were introduced earlier (Ciampi et al., 2019) and have been further developed for the purposes of the present study. 3D modeling of borehole log stratigraphy and 2D ERT images show the geometry of subsurface units and their geophysical signatures with a high-resolution characterization (Fig. 1). The geological framework reveals a distinctive alternation of fine-grained and coarse-grained sediments. This litho-stratigraphic sequence, extensively

documented in the literature for alluvial deposits in the Po River plain (Meisina et al., 2022), can be clearly recognized and differentiated down to a depth of 17 m as follows:

- filling (anthropogenic debris) materials, with a thickness varying from approximately 0.2 to 3 m in the boreholes;
- shallow clays and silts: This level is characterized by the presence of silty clays and clayey silts with thin laminations of very fine sand;
- sands, fine sands with silt, and sandy silts, which constitute a shallow aquifer;
- plastic clays and silty clays, serving as a bottom aquiclude (Fig. 1b).

The groundwater table is encountered at a depth of about 3.7 m. Joint hydrogeophysical findings strengthen the understanding of lithological variations. The spatial continuity of the geoelectrical data associates the sandy deposits of the superficial aquifer with a high-resistivity lenticular body, while the low-resistivity layers correspond to clayey horizons.

The lenticular shape of the sands is captured by ERT imaging but it is not clearly delineated by the stratigraphic correlations of geological borehole drillings, that are placed too far apart to image this type of bodies. In addition, the high resistivity variability characterizing the filling material becomes evident. This feature reflects the environmental context of the site, which is heavily anthropized and marked by the presence of road pavements and metallic objects on the ground surface.

3.2. Iterative clustering of geo-constrained resistivity data reveals distinctive lithological architectures

The iterative data-driven classification procedure, ranging from two clusters to six, statistically, parametrically, and quantitatively correlates electrical behavior with the geological parameter. Fig. 2 illustrates the results obtained by applying the developed classification methodology on resistivity section EF (see Fig.1), subsequently replicated on sections AB and CD. The lithological classification based on geophysical signatures originates from overlaying the resistivity values collected along different vertical section of the ERT image with geological information about stratigraphic contacts as from borehole characterization (Fig. 2a). Categorizing resistivity values into two classes distinguishes filling materials with anomalous electrical properties from a geologically essentially undifferentiated domain (Fig. 2b). The contouring of three clusters potentially outlines sandy lenses positioned in the saturated domain and on the smear zone, between the progressive distances of 38–47 m and 52–54 m (Fig. 2c). From a mathematical perspective, the correct solution could appear to be the one with 4 clusters that minimizes the Euclidean distance of the geological constraint represented by the stratigraphic contact (red star) from the hyperplanes delimiting the resistivity clusters (Fig. 2d). However, the step-by-step iteration suggests an increased resolution of discretization for geophysical-lithological classes as the number of clusters increases up to 5 and 6 (Fig. 2d, e, f).

The particle size distribution of the core samples provide the geological-technical parameterization of subsurface deposits and guide the final choice of the number of clusters, already constrained both geologically and geophysically. According to the USCS classification criteria, deposits sampled at different depths generally correspond to silty clayey sands (SC-SM), silty sands (SM), sandy lean clays (CL), and sandy silty clays (CL-ML), characterized by a percentage content of sands ranging from approximately 24% to 84%. The pronounced granulometric-compositional variability in the subsurface is accentuated by the results obtained from the analysis conducted on sample S3, corresponding to soil classified as fat clay (CH) (Table 2).

Samples S1-S12 are nestled within the lenticular structure of heterogeneous sandy deposits that compose the aquifer, clearly delineated by the iterative grouping process into 6 clusters (Fig. 3a). The spatial match of the shallow deposit granulometric parameterization and the clustering procedure constrained by litho-stratigraphical and electrical attributes, defines a cluster number of 6, consistent with all grain-size media properties and the available geological and geophysical knowledge. The replication of the iterative classification procedure with 6 clusters on the geoelectrical sections AB (Fig. 3c) and CD (Fig. 3b) follows the training on the resistivity section EF, and the data-driven choice of a cluster number.

The entire clustering procedure establishes a relationship between the geological data based on borehole logs and the resistivity data resulting from ERT, depicting subsurface lithological architectures and capturing horizontal heterogeneity. The resulting coherent image depicts the geo-constrained clusters of subsurface lithologic structures, puzzled by statistical physics and geological-technical parameterization of lithotypes. From Fig. 3, sandy-silty lenticular and laminar units, sometimes laterally discontinuous, emerge clearly. Additionally, clays and silts, overlapping and underlying the heterogeneous aquifer body, become evident. The final clustering procedure leads to the delimitation of the base surface and grouping of filling materials, and the distinction of anthropogenic materials from the geological substrate (Fig. 4a, b, c).

By linking and restricting geophysical data to geological criteria, a statistical perspective of subsurface lithological structures is defined, providing a quantitative geophysical lithology parameterization, identifying the geoelectrical signature of different deposits, and associating a range of resistivity with different clusters constrained geologically and geophysically. The basic statistical analysis in the form of box plots quantitatively describes the distinctive distribution of electrical responses of the geological media, which are conditioned by the observed data, presenting different classes corresponding to all sampled lithological units. Fig. 5 illustrates the ranges of variability of resistivity for the differentiated lithological classes, revealing a well-distinguishable geoelectric footprint that characterizes and parameterizes the electrical behavior of shallow horizons. The filling material is distinguished by a high resistivity variability, ranging from 18 Ω·m to 1293 Ω·m, due to disturbances to the geoelectric signal introduced by a heavily anthropized environment. The electrical signature attributed to sandy lithologies oscillates between 16 Ω·m and 78 Ω·m, a significantly narrower range than reported in the literature (Alpaslan, 2021; Chang et al., 2023; Palacky, 1988; Park et al., 2017; Sikandar and Christen, 2012), despite the high granulometric variability evident from laboratory analyses. The typical footprint of in-situ clayey-silty deposits is distributed in a range between 1 Ω·m and 20 Ω·m, falling within the variability ranges illustrated in known bar plots (Fig. 5).

Note from Fig. 5 how the local identification of lithotypes on the basis of direct sampling and geoelectrical information tends to avoid the typical overlap of electrical resistivity values of different lithotypes as they appear in the literature. This overlap is an effect of varying moisture content and pore water salinity at different sites, that largely obliterates the differences due to lithology: this is not the case when such an analysis is conducted at a specific site, as clearly demonstrated here. The constitution of a statistics-based relationship on the geologically-geophysically combined analysis procedure of experimental results lead to better tools for understanding the resistivity behavior of different lithologies at the specific sites of interest.

4. Discussions

In the presented research, we employ a comprehensive geological-physical interpretational methodology to explore the geophysical signature of the lithologies constituting the geological sequence of alluvial sediments. This approach integrates statistical techniques and quantitative analysis to describe subsurface structures, lithology, and electrical properties, honoring the spatial distribution of geological parameters provided as prior information on local geology. The aggregation of geoelectrical and geological-technical information was performed in the central portion of the ERT cross section, aligning with a high-resolution tomographic discretization and the target depths of the geophysical imaging. This is intended to leave out the lateral and deep

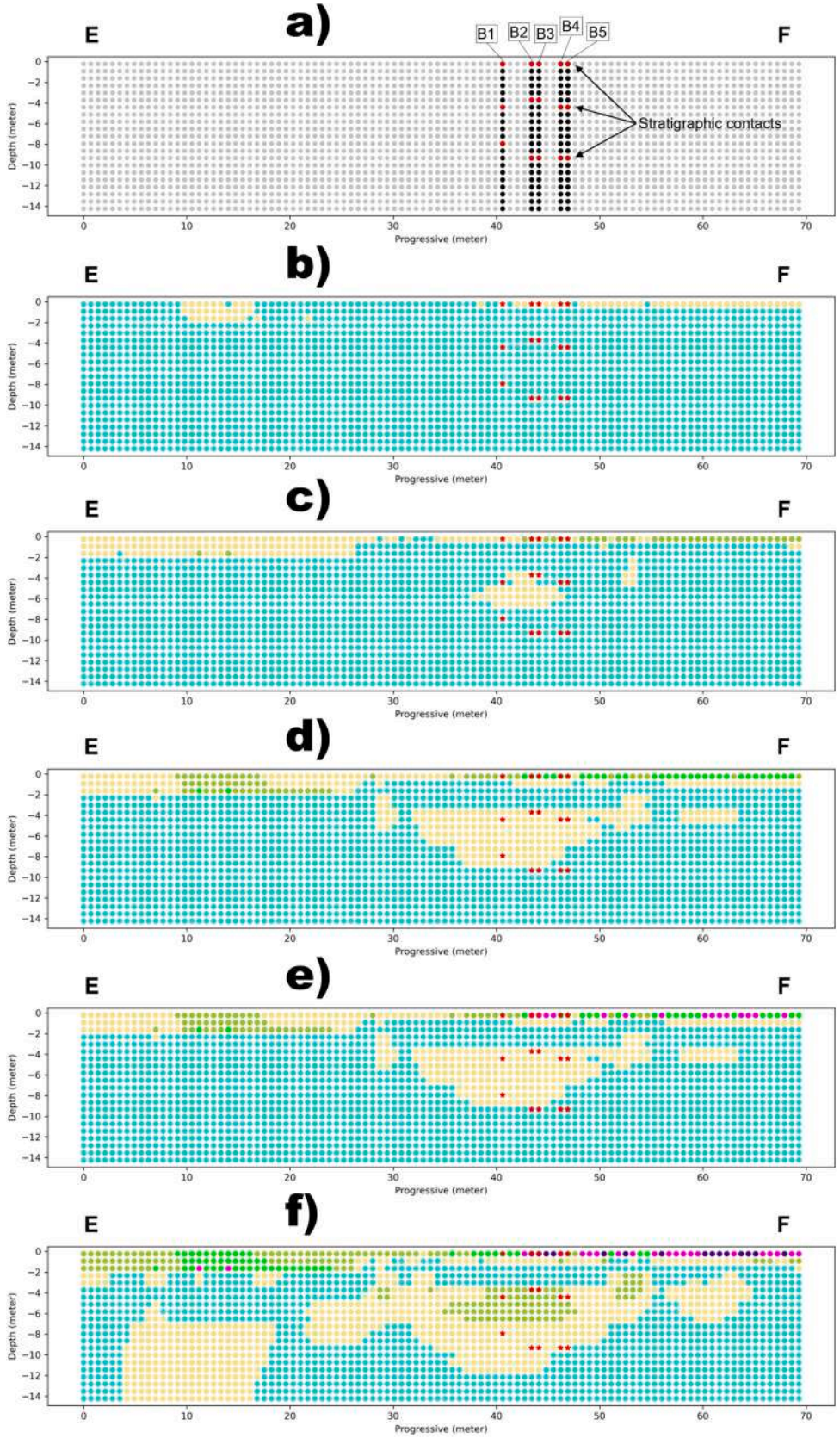


Fig. 2. Placement of drilled boreholes (B1-B5) on section EF and identification of stratigraphic contacts (red stars) (a). Representation of the iterative geo-constrained resistivity data clustering procedure into 2 (b), 3 (c), 4 (d), 5 (e), and 6 (f) classes. Different colors identify various clusters, and the red stars correspond to the stratigraphic contacts (b, c, d, e, f). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Percentage ratio of granulometric fractions, liquid limit, plastic limit, plasticity index, and USCS classification resulting from analyses on the twelve soil samples collected from boreholes.

Sample ID	Gravel (%)	Sand (%)	Silt (%)	Clay (%)	Passing the No. 200 sieve (%)	Retained on No. 200 sieve (%)	Liquid limit (LL) (%)	Plastic limit (PL) (%)	Plasticity index (PI) = LL - PL (%)	Unified soil classification system (USCS) classification
S1	0.7	54.9	18.1	26.3	44.4	55.6	24.1	20	4.1	SC - SM (silty clayey sand)
S2	0	35.6	32.1	32.3	64.4	35.6	28	18.3	9.7	CL (sandy lean clay)
S3	0.2	4.4	37.1	58.3	95.3	4.7	55.8	24.5	31.3	CH (fat clay)
S4	1.6	38.7	24.5	35.2	59.7	40.3	25.4	20.9	4.5	CL - ML (sandy silty clay)
S5	1.1	44.1	28.7	26.2	54.9	45.1	29.3	15.4	11.2	CL (sandy lean clay)
S6	1.6	23.8	41.1	33.5	74.6	25.4	42.4	33.8	8.7	CL (lean clay with sand)
S7	0	84.2	6	9.8	15.8	84.2	39.5	30	9.5	SC - SM (silty clayey sand)
S8	0.1	79.8	10.1	9.9	20	80	/	/	/	SM (silty sand)
S9	8.4	49	23	19.6	42.6	57.4	35.9	28.6	7.3	SC - SM (silty clayey sand)
S10	1	80.7	7.4	10.9	18.3	81.7	28.2	21.8	6.4	SC - SM (silty clayey sand)
S11	0	75.9	12	12	24.1	75.9	/	/	/	SM (silty sand)
S12	6	44.6	23.6	25.8	49.4	50.6	38.8	34.7	4.1	SC - SM (silty clayey sand)

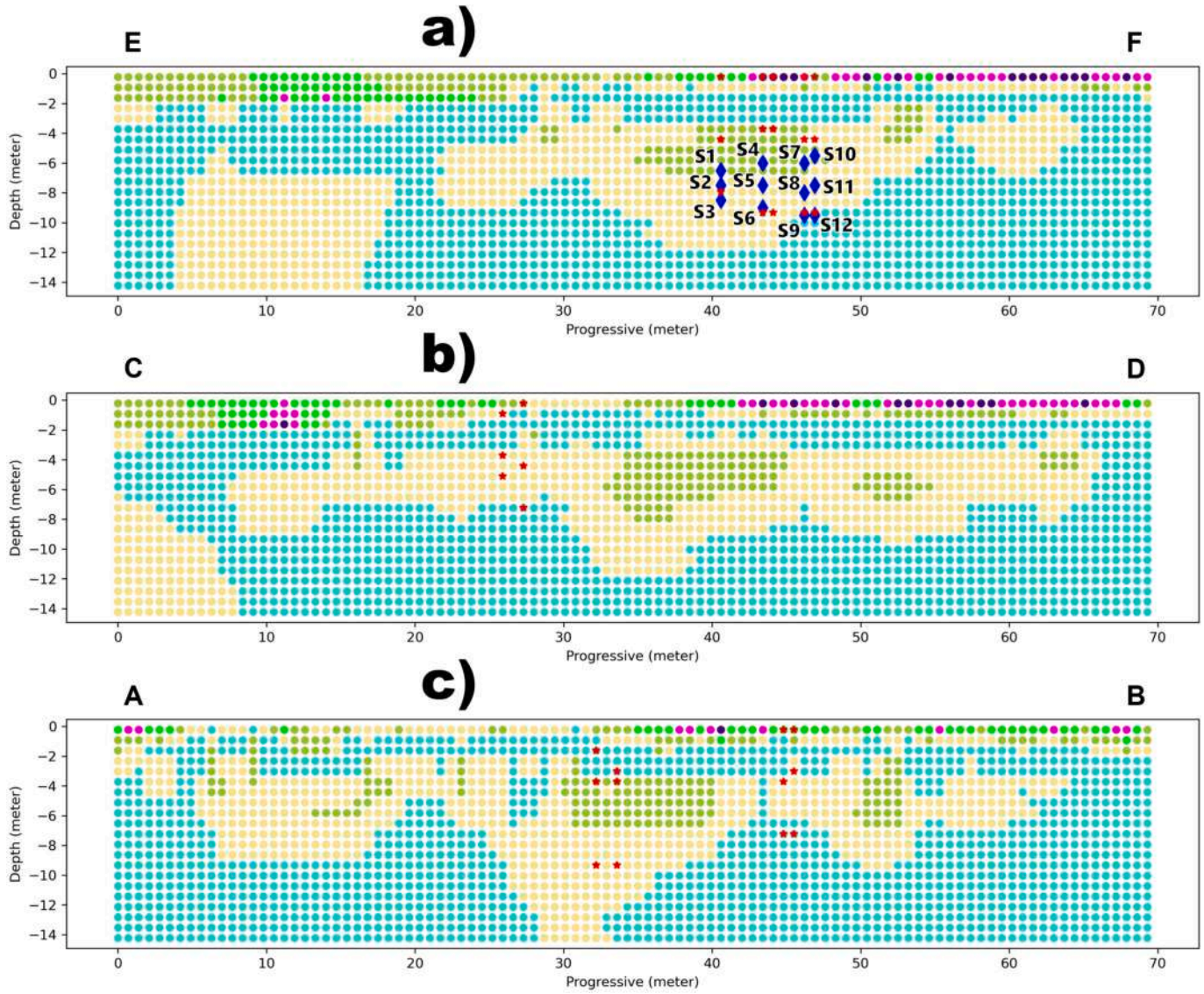


Fig. 3. Placement of samples collected for granulometric analysis (S1-S12) on section EF discretizing 6 clusters of electrical behavior based on knowledge of the geological parameters (a). Corresponding result of the iterative clustering procedure of geo-constrained resistivity data into 6 classes for sections CD (b) and AB (c).

portions of the ERT cross section that can be known to suffer more from inversion errors and poorer resolution (Day-Lewis et al., 2005; Loke et al., 2013; Uhlemann et al., 2018). Through this methodology, we derive an interpretation of the geoelectrical behavior, combining our understanding of geological parameters with the resistivity distribution of different lithotypes. This integrated approach enhances our ability to discern and characterize the intricate relationships within the subsurface geophysical signature and geological architecture. Note that the

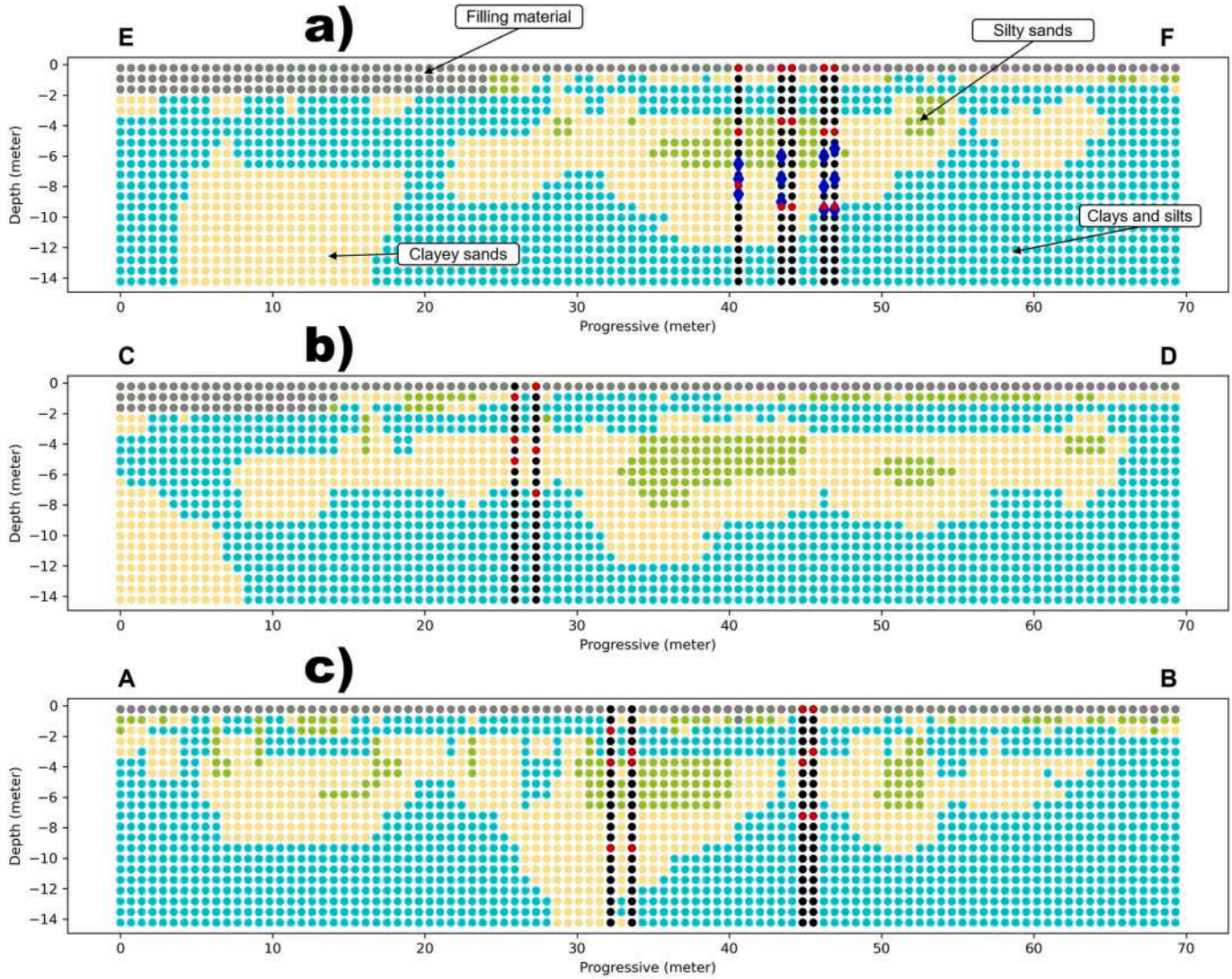


Fig. 4. AB (c), CD (b), and EF (a) 2D profiles depicting the clustering of the statistical-based and geologically-geophysically analysis procedure following the delineation and grouping also of filling materials (debris).

essential contribution of geophysical data, in this context, is to provide a dense and complete spatial coverage, while the local sampling data can only be, at most, interpolated spatially and may easily fail to identify lateral variations in the subsoil structure (compare, e.g., Fig. 1b and c).

The iterative clustering of geo-constrained resistivity data associates an electrical signature with the surface sediments of the study site, characterized by high-definition information. The distinct signatures, expressed in terms of resistivity, delineate geological horizons, aiming to enhance the geological-geophysical discretization detail of classification tables and known intervals in the literature for site-specific studies (Alpaslan, 2021; Chang et al., 2023; Palacky, 1988; Park et al., 2017; Sikandar and Christen, 2012). Despite their technical-scientific significance, these tables are deemed unsuitable for detailed usage due to the broad resistivity ranges associated with each lithological class, and the large overlapping of the same resistivity ranges of very different lithologies.

In the existing literature, a variety of machine learning methods have been proposed to capture the intricate interrelationships between geological models and geophysical evidence (Bressan et al., 2020; Jia et al., 2021; Marzán et al., 2021; Xu and Green, 2023). These methods encompass random forests, support vector machines, artificial neural networks, multilayer perceptrons, decision trees, and are often highly sophisticated. The demand for a substantial dataset for effective learning is apparent, and the selection of the right algorithm can pose challenges. Also, geophysical inversion, employing diverse data-driven and statistical techniques (Aleardi et al., 2021; Bosh, 1999; Bersezio et al., 2007; Foged et al., 2014), such as the Markov chain Monte Carlo (MCMC) approach, qualitative geological constraints, and probability density functions, aims to seamlessly integrate physical reasoning with observational data. First, a coupled hydrogeophysical inversion approach may be susceptible to errors in the hydrological model (Linde et al., 2006; Hinnell et al., 2010) leading sometimes to physically paradoxical results (such as in Linde et al., 2006). In all the aforementioned instances, the definition of linking functions and the selection of methods for learning these functions are predominantly contingent on the type and complexity of the features to be modeled, necessitating an interpretative approach. Furthermore, the form of stratigraphic contacts may not be adequately constrained by the available data (Bosh et al., 2001). In geophysical inversion clustering processes, the choice of the number of clusters emerges as a key element that can be set a priori based on the knowledge of the geological units of the site. In reality, this assumption may not be warranted in regions of complex geology (Singh and Sharma, 2018; Sun and Li, 2015). Also care must be given to accounting for spatially varying factors others than lithology itself, such as water saturation or water salinity. The analytical approach outlined in this study streamlines the establishment of a parametric relationship

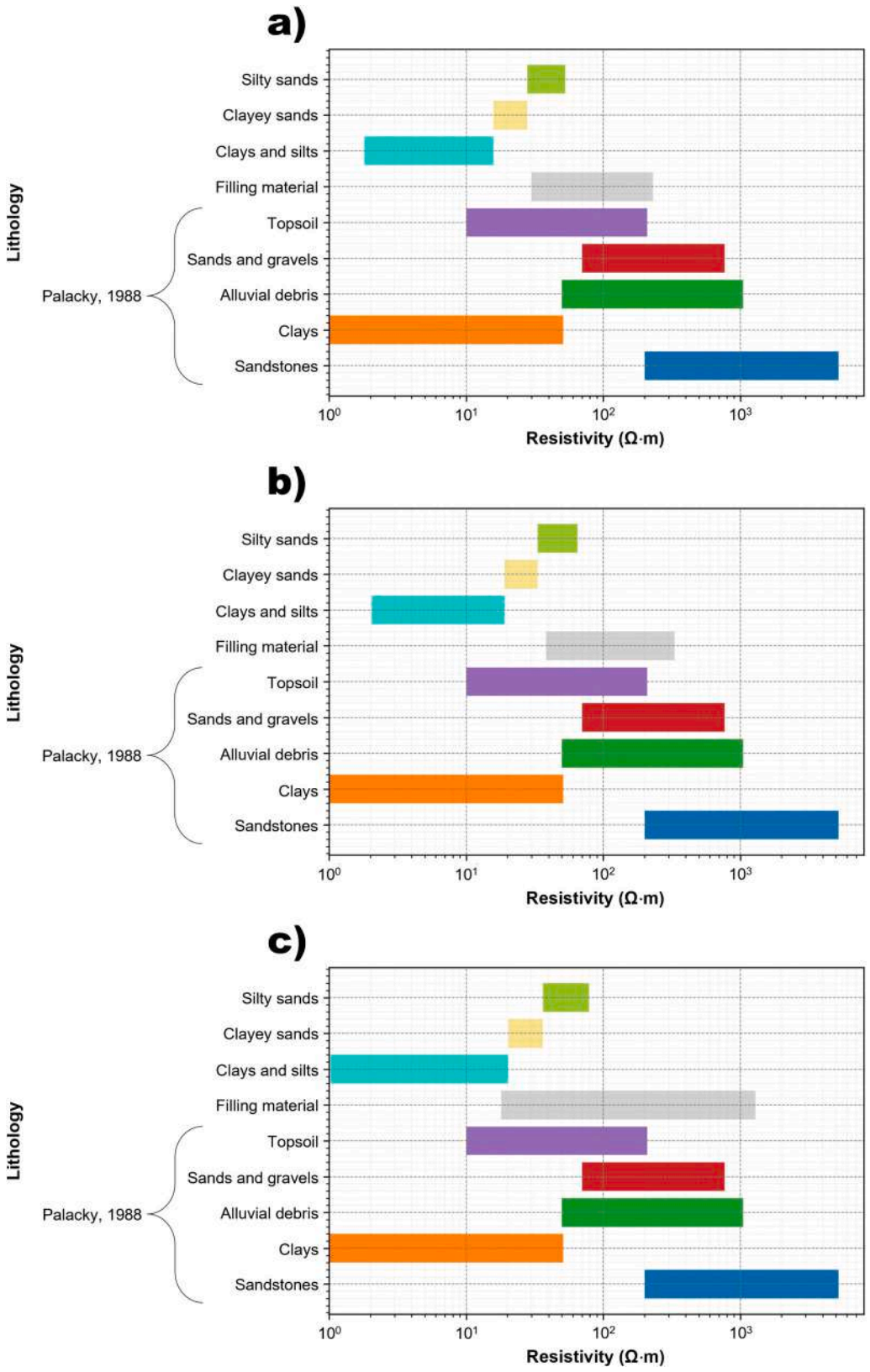


Fig. 5. Bar plots depict resistivity values for clusters differentiated in the EF (a), CD (b), and AB (c) sections through the geologic-geophysical data-driven approach. These plots are overlaid with resistivity variation ranges derived from the literature.

between geology and geophysics across space. Rather than attempting to predict specific parameters, our methodology relies on geoconstrained statistical characteristics of resistivity parameters and allows for the representation of the true and simple model-data relationship. This stands in contrast to synthetic and simulated approximations, offering a more authentic portrayal of lithological patterns with distinct resistivity values and resolving the ambiguity in distinguishing geophysical data from geological characteristics based on observations from drilled boreholes, grain size analysis, and electrical properties. As a result, iterative clustering of geo-constrained resistivity based on real data has the capacity to grasp greater complexity in a problem with less effort, compared to the resources needed to create an accurate predictive model for the same issue. This distinction highlights the efficacy of our method in harnessing the richness and authenticity of real-world data to improve the comprehension and representation of intricate relationships.

The implications of the findings discussed in the text are substantial for the field of engineering geology. Assigning typical, albeit local, electrical resistivity values to lithological units and establishing a link between geophysical imaging and direct geological investigation can significantly contribute to various engineering and geological projects. This advancement has the potential to greatly enhance our understanding of underground geological structures, as well as the geomechanical and hydrological conditions crucial for construction engineering and the mitigation of hydrogeological risks such as landslides (Boyd et al., 2024; Nie et al., 2024).

The results obtained from the statistical clustering procedure, guided by geophysical data and constrained by geological contact information, offer an opportunity for a retrospective analysis of both resistivity data and visual interpretations of field observations regarding stratigraphic contacts. Potential sources of uncertainty, such as inaccurate borehole descriptions, field measurements, and inversion, as well as variations in spatial resolutions and contrasting scales between borehole and geophysical data, may introduce uncertainties (He et al., 2014; Refsgaard et al., 2014). The cluster's hyperplane generally honors the geoconstraint and does not deviate significantly from the stratigraphic boundary observed in the drilled borehole. On the other hand, the lack of alignment between the clustering procedure and granulometric classification, observed in our case for sample S3, can be explained by both the marked horizontal and vertical lithological variability and the different sampling and analysis resolution characterizing ERT imaging and granulometric analyses. High-contrast heterogeneities in the subsurface could be present at a smaller spatial scale than the electrode discretization capacity of ERT imaging (Loke et al., 2013). In this sense, the resolution limitations of the array can lead to information loss where spatial geological variation exceeds the investigative resolution detail. The iterative clustering method paves the way for the geological interpretation of geoelectrical data, which remains constrained by the discriminative capability of ERT. A rearrangement of the array can potentially enhance the quantitative geophysical lithology parameterization capability, making the method increasingly resolute. However, the issues related to the expected signal/noise ratio should also be considered, and these require a larger current dipole size, in contrast with the resolution requirements. In addition, some complex and challenging-to-explain quantitative correlations between ERT data and geological-technical classifications for subsurface soil characterization have been reported in the literature. This highlights the need for further investigations in this direction (Braga et al., 1999; Cosenza et al., 2006; Giao et al., 2003; Sudha et al., 2009). One possible extension of the approach may include accounting explicitly for the spatially variable sensitivity and resolution of any geophysical inverted image. A possible use of a-posteriori sensitivity maps as a weighting factor in the clustering algorithm is envisaged. In the foreseeable future, our endeavor involves expanding and refining our methodology to seamlessly integrate hydrogeological and chemical data. This advancement aims to enhance our capacity for detecting water table levels and directly visualizing pollution, while also discerning geological heterogeneities.

5. Conclusions

The iterative and statistical clustering of geo-constrained resistivity data has emerged as a powerful tool in the quest to unravel the complex subsurface geostructures, shedding light on distinctive lithological architectures and geoelectrical footprints of shallow lithologies. By integrating data from 17 borehole drillings, 3 parallel ERT profiles, and complete laboratory analyses, a comprehensive dataset was assembled. This multi-source picture offers a quantitative parameterization for the site-specific geoelectrical signatures of lithological architectures. The innovative statistical clustering approach, driven by geological boundaries, high resolution ERT data, and informed by particle size analyses, successfully delineated six clusters, capturing the high contrasting heterogeneity of shallow deposits. Notably, the iterative clustering process was strategically halted by aligning with observed stratigraphic discontinuities, showcasing its adaptability to geological features. The derived resistivity clusters not only portrayed distinct lithological structures but also facilitated a quantitative parameterization of the geoelectrical behavior, showcasing unique geoelectric footprints associated with different lithologies. The successive grouping of anthropogenic deposits, coupling spatial clustering with geological-statistical data-driven classification, identifies key geological-physical units. These include filling materials, silty sands, clayey sands, and clays and silts, each differentiated and distinguished by a specific range of resistivity variation for every lithotype. Note that the ranges of resistivity values thus identified at the local site-specific scale do not suffer from the large overlap across different lithologies typical of wide range literature studies. The results, presented through 2D visualizations, offer a comprehensive understanding of subsurface lithological architectures, establishing a link and bridging the gap between geology and geophysics in a simple and data-driven manner. This methodology, distinct from predictive models and synthetic approximations, leverages real-world data richness to authentically represent complex relationships, making it a valuable asset in subsurface exploration and characterization. The study's success in overcoming challenges and uncertainties, such as granulometric variability and spatial resolution disparities, underscores its efficacy in providing an advanced interpretation of the geological-geophysical interplay. Overall, this innovative approach sets a benchmark for the integration of geological and geophysical information, contributing significantly to our understanding of subsurface heterogeneity and pave the way for its application in contaminated contexts for discretization of pollutants in the geological arena.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Paolo Ciampi: Writing – original draft, Visualization, Software, Data curation. **Leonardo Maria Giannini:** Writing – review & editing, Visualization, Software, Methodology, Data curation, Conceptualization. **Giorgio Cassiani:** Writing – review & editing, Investigation, Data curation. **Carlo Esposito:** Writing – review & editing, Validation, Supervision. **Marco Petrangeli Papini:** Supervision, Project administration.

Declaration of competing interest

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

Data availability

The data supporting the findings of this study are available within the article and are available from the corresponding author upon reasonable request.

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. J. Hydrol. 615, 128725 https://doi.org/10.1016/j.jhydrol.2022.128725.

Aleardi, M., Vinciguerra, A., Hojat, A., 2021. A geostatistical Markov chain Monte Carlo inversion algorithm for electrical resistivity tomography. Near Surf. Geophys. 19 (1), 7–26. https://doi.org/10.1002/nsg.12133.

Alpaslan, N., 2021. Determination of borehole locations and saline-water intrusion for groundwater in Central Anatolia Region, Turkey using electrical tomography (ERT) method. Environ. Earth Sci. 80 (24), 810. https://doi.org/10.1007/s12665-021-10117-7.

ASTM International, 2010. Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils, ASTM D4318–10. ASTM International, Montgomery County, PA, USA.

ASTM International, 2011. Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System), ASTM D2487–11. ASTM International, Montgomery County, PA, USA.

Bersezio, R., Giudici, M., Mele, M., 2007. Combining sedimentological and geophysical data for high-resolution 3-D mapping of fluvial architectural elements in the Quaternary Po plain (Italy). Sediment. Geol. 202, 230–248. https://doi.org/10.1016/j.sedgeo.2007.05.002.

Binley, A., Kemna, A., 2005. DC Resistivity and induced polarization methods. In: Rubin, Y., Hubbard, S.S. (Eds.), Hydrogeophysics. Water Science and Technology Library, vol. 50. Springer, Dordrecht, pp. 129–156. https://doi.org/10.1007/1-4020-3102-5_5.

Binley, A., Slater, L., 2020. Resistivity and induced polarization: Theory and applications to the near-surface earth. Cambridge University Press. https://doi.org/10.1017/9781108685955.

Binley, A., Hubbard, S.S., Huisman, J.A., Revil, A., Robinson, D.A., Singha, K., Slater, L. D., 2015. The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales. Water Resour. Res. 51, 3837–3866. https://doi.org/10.1002/2015WR017016.

Blanchy, G., Saneiyan, S., Boyd, J., McLachlan, P., Binley, A., 2020. ResIPy, an intuitive open source software for complex geoelectrical inversion/modeling. Comput. Geosci. 137, 104423 https://doi.org/10.1016/j.cageo.2020.104423.

Bosh, M., 1999. Lithologic tomography: from plural geophysical data to lithology estimation. J. Geophys. Res. 104, 749–766. https://doi.org/10.1029/1998JB900014.

Bosh, M., Guillen, A., Ledru, P., 2001. Lithologic tomography: an application to geophysical data from the Cadomian belt of northern Brittany, France. Tectonophysics 331, 197–227. https://doi.org/10.1016/S0040-1951(00)00243-2.

Bouchedda, A., Chouteau, M., Binley, A., Giroux, B., 2012. 2-D joint structural inversion of cross-hole electrical resistance and ground penetrating radar data. J. Appl. Geophys. 78, 52–67. https://doi.org/10.1016/j.jappgeo.2011.10.009.

Boyd, J.P., Binley, A., Wilkinson, P., Holmes, J., Bruce, E., Chambers, J., 2024. Practical considerations for using petrophysics and geoelectrical methods on clay rich landslides. Eng. Geol. 334, 107506 https://doi.org/10.1016/j.enggeo.2024.107506.

Braga, A., Malagutti, W., Dourado, J., Chang, H., 1999. Correlation of electrical resistivity and induced polarization data with geotechnical survey standard penetration test measurements. J. Environ. Eng. Geophys. 4, 123–130. https://doi.org/10.4133/JEEG4.2.123.

Bressan, T.S., de Souza, M.K., Girelli, T.J., Junior, F.C., 2020. Evaluation of machine learning methods for lithology classification using geophysical data. Comput. Geosci. 139, 104475 https://doi.org/10.1016/j.cageo.2020.104475.

Cassiani, G., Bruno, V., Villa, A., Fusi, N., Binley, A., 2006. A saline tracer test monitored via time-lapse surface electrical resistivity tomography. J. Appl. Geophys. 59, 244–259. https://doi.org/10.1016/j.jappgeo.2005.10.007.

Cassiani, G., Binley, A., Kemna, A., Wehrer, M., Flores-Orozco, A., Deiana, R., Boaga, J., Rossi, M., Dietrich, P., Werban, U., Zschornack, L., Godio, A., Jafargandomi, A., Deidda, G.P., 2014. Noninvasive characterization of the Trecate (Italy) crude-oil contaminated site: links between contamination and geophysical signals. Environ. Sci. Pollut. Res. 21, 8914–8931. https://doi.org/10.1007/s11356-014-2494-7.

Chang, P.Y., Doyoro, Y.G., Lin, D.J., Puntu, J.M., Amania, H.H., Kassie, L.N., 2023. Electrical resistivity imaging data for hydrogeological and geological hazard investigations in Taiwan. Data Br. 49, 109377 https://doi.org/10.1016/j.dib.2023.109377.

Ciampi, P., Esposito, C., Viotti, P., Boaga, J., Cassiani, G., Petrangeli Papini, M., 2019. An integrated approach supporting remediation of an aquifer contaminated with chlorinated solvents by a combination of adsorption and biodegradation. Appl. Sci. 9 (20), 4318. https://doi.org/10.3390/app9204318.

Ciampi, P., Esposito, C., Cassiani, G., Deidda, G.P., Rizzetto, P., Petrangeli Papini, M., 2021. A field-scale remediation of residual light non-aqueous phase liquid (LNAPL): chemical enhancers for pump and treat. Environ. Sci. Pollut. Res. 28, 35286–35296. https://doi.org/10.1007/s11356-021-14558-2.

Ciampi, P., Esposito, C., Cassiani, G., Deidda, G.P., Flores-Orozco, A., Rizzetto, P., Chiappa, A., Bernabei, M., Gardon, A., Petrangeli Papini, M., 2022. Contamination presence and dynamics at a polluted site: Spatial analysis of integrated data and joint conceptual modeling approach. J. Contam. Hydrol. 248, 104026 https://doi.org/10.1016/j.jconhyd.2022.104026.

Cosenza, P., Marmet, E., Rejiba, F., Cui, Y.J., Tabbagh, A., Charlery, Y., 2006. Correlations between geotechnical and electrical data: a case study at Garchy in France. J. Appl. Geophys. 60, 165–178. https://doi.org/10.1016/j.jappgeo.2006.02.003.

Crook, N., Binley, A., Knight, R., Robinson, D.A., Zarnetske, J., Haggerty, R., 2008. Electrical resistivity imaging of the architecture of substream sediments. Water Resour. Res. 44, W00D13. https://doi.org/10.1029/2008WR006968.

Day-Lewis, F.D., Singha, K., Binley, A.M., 2005. Applying petrophysical models to radar travel time and electrical resistivity tomograms: resolution-dependent limitations. J. Geophys. Res. 110, B08206. https://doi.org/10.1029/2004JB003569.

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 Surf. Geophys. 21, 275–287. https://doi.org/10.1002/nsg.12259.

Deiana, R., Cassiani, G., Kemna, A., Villa, A., Bruno, V., Bagliani, A., 2007. An experiment of non-invasive characterization of the vadose zone via water injection and cross-hole time-lapse geophysical monitoring. Near. Surf. Geophys. 5 (3), 183–194. https://doi.org/10.3997/1873-0604.2006030.

Di Maio, R., Fabbrocino, S., Forte, G., Piegari, E., 2014. A three-dimensional hydrogeological–geophysical model of a multi-layered aquifer in the coastal alluvial plain of Sarno River (southern Italy). Hydrogeol. J. 22, 691–703. https://doi.org/10.1007/s10040-013-1087-8.

Flores-Orozco, A., Ciampi, P., Katona, T., Censini, M., Petrangeli Papini, M., Deidda, G.P., Cassiani, G., 2021. Delineation of hydrocarbon contaminants with multi-frequency complex conductivity imaging. Sci. Total Environ. 768, 144997 https://doi.org/10.1016/j.scitotenv.2021.144997.

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

Gemail, K.S., 2015. Application of 2D resistivity profiling for mapping and interpretation of geology in a till aquitard near Luck Lake, Southern Saskatchewan, Canada. Environ. Earth Sci. 73, 923–935. https://doi.org/10.1007/s12665-014-3441-0.

Giannini, L.M., Varone, C., Esposito, C., Scarascia Mugnozza, G., Schilirò, L., 2021. The potential of spatial statistics for the reconstruction of a subsoil model: a case study for the Firenze-Prato-Pistoia Basin, Central Italy. J. Appl. Geophys. 194, 104466 https://doi.org/10.1016/j.jappgeo.2021.104466.

Giao, P.H., Chung, S.G., Kim, D.Y., Tanaka, H., 2003. Electric imaging and laboratory resistivity testing for geotechnical investigation of Pusan clay deposits. J. Appl. Geophys. 52 (4), 157–175. https://doi.org/10.1016/S0926-9851(03)00002-8.

Han, J., Pei, J., Tong, H., 2012. Data mining: Concepts and techniques. Morgan kaufmann. https://doi.org/10.1016/C2009-0-61819-5.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: a k-means clustering algorithm. J. R. Stat. Soc. Ser. C Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830.

He, X., Koch, J., Sonnenborg, T.O., Jorgensen, F., Schamper, C., Refsgaard, J.C., 2014. Transition probability-based stochastic geological modeling using airborne geophysical data and borehole data. Water Resour. Res. 50, 3147–3169. https://doi.org/10.1002/2013WR014593.

Hinnell, A.C., Ferré, T.P.A., Vrugt, J.A., Huisman, J.A., Moysey, S., Rings, J., Kowalsky, M.B., 2010. Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. Water Resour. Res. 46, W00D40. https://doi.org/10.1029/2008WR007060.

Høyer, A.S., Jørgensen, F., Sandersen, P.B.E., Viezzoli, A., Møller, I., 2015. 3D geological modelling of a complex buried-valley network delineated from borehole and AEM data. J. Appl. Geophys. 122, 94–102. https://doi.org/10.1016/j.jappgeo.2015.09.004.

Hubbard, S.S., Rubin, Y., 2000. Hydrogeological parameter estimation using geophysical data: a review of selected techniques. J. Contam. Hydrol. 45 (1–2), 3–34. https://doi.org/10.1016/S0169-7722(00)00117-0.

Jenks, G.F., 1967. The data model concept in statistical mapping. International yearbook of cartography, 7, pp. 186–190.

Jia, R., Lv, Y., Wang, G., Carranza, E., Chen, Y., Wei, C., Zhang, Z., 2021. A stacking methodology of machine learning for 3D geological modeling with geological-geophysical datasets, Laochang Sn camp, Gejiu (China). Comput. Geosci. 151, 104754 https://doi.org/10.1016/j.cageo.2021.104754.

Klose, C.D., 2006. Self-organizing maps for geoscientific data analysis: geological interpretation of multidimensional geophysical data. Comput. Geosci. 10, 265–277. https://doi.org/10.1007/s10596-006-9022-x.

Kumar, A., Kumar, A., Mallipeddi, R., Lee, D.G., 2024. High-density cluster core-based k-means clustering with an unknown number of clusters. Appl. Soft Comput. 111419 https://doi.org/10.1016/j.asoc.2024.111419.

Lelièvre, P.G., Oldenburg, D.W., Williams, N.C., 2009. Integrating geological and geophysical data through advanced constrained inversions. Explor. Geophys. 40 (4), 334–341. https://doi.org/10.1071/EG09012.

Linde, N., Binley, A., Tryggvason, A., Pedersen, L.B., Revil, A., 2006. Improved hydrogeophysical characterization using joint inversion of crosshole electrical resistance and ground penetrating radar traveltime data. Water Resour. Res. 42, W12404. https://doi.org/10.1029/2006WR005131.

Linde, N., Renard, P., Mukerji, T., Caers, J., 2015. Geological realism in hydrogeological and geophysical inverse modeling: a review. Adv. Water Resour. 86 (Part A), 86–101. https://doi.org/10.1016/j.advwatres.2015.09.019.

Loke, M.H., Chambers, J.E., Rucker, D.F., Kuras, O., Wilkinson, P.B., 2013. Recent developments in the direct-current geoelectrical imaging method. J. Appl. Geophys. 95, 135–156. https://doi.org/10.1016/j.jappgeo.2013.02.017.

Looms, M.C., Binley, A., Jensen, K.H., Nielsen, L., Hansen, T.M., 2008. Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data. Vadose Zone J. 7, 238–248. https://doi.org/10.2136/vzj2007.0087.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, No. 14, pp. 281–297.

Madsen, R.B., Høyer, A.S., Sandersen, P.B., Møller, I., Hansen, T.M., 2023. A method to construct statistical prior models of geology for probabilistic inversion of geophysical data. Eng. Geol. 324, 107252 https://doi.org/10.1016/j.enggeo.2023.107252.

Marzán, I., Martí, D., Lobo, A., Alcalde, J., Ruiz, M., Alvarez-Marrón, J., Carbonell, R., 2021. Joint interpretation of geophysical data: applying machine learning to the modeling of an evaporitic sequence in Villar de Cañas (Spain). Eng. Geol. 288, 106126 https://doi.org/10.1016/j.enggeo.2021.106126.

Maurya, P.K., Balbarini, N., Møller, I., Rønde, V., Christiansen, A.V., Bjerg, P.L., Auken, E., Fiandaca, G., 2018. Subsurface imaging of water electrical conductivity, hydraulic permeability and lithology at contaminated sites by induced polarization. Geophys. J. Int. 213 (2), 770–785. https://doi.org/10.1093/gji/ggy018.

Meisina, C., Bonì, R., Bordoni, M., Lai, C.G., Bozzoni, F., Cosentini, R.M., Castaldini, D., Fontana, D., Lugli, S., Ghinoi, A., Martelli, L., Severi, P., 2022. 3D Engineering geological modeling to investigate a liquefaction site: an example in alluvial holocene sediments in the Po Plain, Italy. Geosci 12, 155. https://doi.org/10.3390/geosciences12040155.

Michel, H., Nguyen, F., Kremer, T., Elen, A., Hermans, T., 2020. 1D geological imaging of the subsurface from geophysical data with Bayesian Evidential Learning. Comput. Geosci. 138, 104456 https://doi.org/10.1016/j.cageo.2020.104456.

Nawaz, M.A., Curtis, A., 2019. Rapid discriminative variational Bayesian inversion of geophysical data for the spatial distribution of geological properties. J. Geophys. Res. Solid Earth 124, 5867–5887. https://doi.org/10.1029/2018JB016652.

Nazaruddin, D.A., Amiruzan, Z.S., Hussin, H., Jafar, M.T.M., 2017. Integrated geological and multi-electrode resistivity surveys for groundwater investigation in Kampung Rahmat village and its vicinity, Jeli district, Kelantan, Malaysia. J. Appl. Geophys. 138, 23–32. https://doi.org/10.1016/j.jappgeo.2017.01.012.

Nie, L., Jia, S., Li, Z.Q., Guo, Q., Wang, T., Du, Y., Li, S., Jing, P., 2024. A high resolution detection approach combining probe drilling and horizontal cross-hole resistivity tomography to interpret water conducting channels ahead of the tunnel: a case study in Yunnan, China. Eng. Geol. 331, 107449 https://doi.org/10.1016/j.

Olierook, H.K.H., Scalzo, R., Kohn, D., Chandra, R., Farahbakhsh, E., Clark, C., Reddy, S. M., Müller, R.D., 2021. Bayesian geological and geophysical data fusion for the construction and uncertainty quantification of 3D geological models. Geosci. Front. 12 (1), 479–493. https://doi.org/10.1016/j.gsf.2020.04.015.

Palacky, G.J., 1988. Resistivity characteristics of geologic targets. In: Electromagnetic Methods in Applied Geophysics, vol. 1. Society of Exploration Geophysicists, Tulsa, OK, pp. 53–129. https://doi.org/10.1190/1.9781560802631.ch3.

Park, C.S., Jeong, J.H., Park, H.W., Kim, K., 2017. Experimental study on electrode method for electrical resistivity survey to detect cavities under road pavements. Sustainability 9 (12), 2320. https://doi.org/10.3390/su9122320.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490.

Piegari, E., De Donno, G., Melegari, D., Paoletti, V., 2023. A machine learning-based approach for mapping leachate contamination using geoelectrical methods. Waste Manag. 157, 121–129. https://doi.org/10.1016/j.wasman.2022.12.015.

Refsgaard, J.C., Auken, E., Bamberg, C.A., Christensen, B.S., Clausen, T., Dalgaard, E., Effersø, F., Ernstsen, V., Gertz, F., Hansen, A.L., He, X., Jacobsen, B.H., Jensen, K.H., Jørgensen, F., Jørgensen, L.F., Koch, J., Nilsson, B., Petersen, C., De Schepper, G., Schamper, C., Sørensen, K.I., Therrien, R., Thirup, C., Viezzoli, A., 2014. Nitrate reduction in geologically heterogeneous catchments–a framework for assessing the scale of predictive capability of hydrological models. Sci. Total Environ. 468-469, 1278–1288. https://doi.org/10.1016/j.scitotenv.2013.07.042.

Revil, A., Coperey, A., Shao, Z., Florsch, N., Fabricius, I.L., Deng, Y., Delsman, J.R., Pauw, P.S., Karaoulis, M., de Louw, P.G.B., van Baaren, E.S., Dabekaussen, W., Menkovic, A., Gunnink, J.L., 2017. Complex conductivity of soils. Water Resour. Res. 53 (8), 7121–7147. https://doi.org/10.1002/2017WR020655.

Rücker, C., Günther, T., Wagner, F.M., 2017. pyGIMLi: an open-source library for modelling and inversion in geophysics. Comput. Geosci. 109, 106–123. https://doi.org/10.1016/j.cageo.2017.07.011.

Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A., Richard, G., 2005. Electrical resistivity survey in soil science: a review. Soil Tillage Res. 83 (2), 173–193. https://doi.org/10.1016/j.still.2004.10.004.

Sikandar, P., Christen, E.W., 2012. Geoelectrical sounding for the estimation of hydraulic conductivity of alluvial aquifers. Water Resour. Manag. 26 (5), 1201–1215. https://doi.org/10.1007/s11269-011-9954-3.

Singh, A., Sharma, S.P., 2018. Identification of different geologic units using fuzzy constrained resistivity tomography. J. Appl. Geophys. 148, 127–138. https://doi.org/10.1016/j.jappgeo.2017.11.014.

Slater, L., 2007. Near surface electrical characterization of hydraulic conductivity: from petrophysical properties to aquifer geometries—a review. Surv. Geophys. 28, 169–197. https://doi.org/10.1007/s10712-007-9022-y.

Stan, D., Stan-Kleczek, I., 2014. Application of electrical resistivity tomography to map lithological differences and subsurface structures (Eastern Sudetes, Czech Republic). Geomorphology 221, 113–123. https://doi.org/10.1016/j.geomorph.2014.05.027.

Sudha, K., Israil, M., Mittal, S., Rai, J., 2009. Soil characterization using electrical resistivity tomography and geotechnical investigations. J. Appl. Geophys. 67 (1), 74–79. https://doi.org/10.1016/j.jappgeo.2008.09.012.

Sun, J., Li, Y., 2015. Multidomain petrophysically constrained inversion and geology differentiation using guided fuzzy c-means clustering. Geophys 80 (4). https://doi.org/10.1190/geo2014-0049.1.

Uhlemann, S., Wilkinson, P.B., Maurer, H., Wagner, F.M., Johnson, T.C., Chambers, J.E., 2018. Optimized survey design for electrical resistivity tomography: combined optimization of measurement configuration and electrode placement. Geophys. J. Int. 214 (1), 108–121. https://doi.org/10.1093/gji/ggy128.

Wang, Y., Akeju, O.V., Zhao, T., 2017. Interpolation of spatially varying but sparsely measured geo-data: a comparative study. Eng. Geol. 231, 200–217. https://doi.org/10.1016/j.enggeo.2017.10.019.

Whiteley, J.S., Watlet, A., Uhlemann, S., Wilkinson, P., Boyd, J.P., Jordan, C., Kendall, J. M., Chambers, J.E., 2021. Rapid characterisation of landslide heterogeneity using unsupervised classification of electrical resistivity and seismic refraction surveys. Eng. Geol. 290, 106189 https://doi.org/10.1016/j.enggeo.2021.106189.

Wu, W., Grana, D., 2017. Integrated petrophysics and rock physics modeling for well log interpretation of elastic, electrical, and petrophysical properties. J. Appl. Geophys. 146, 54–66. https://doi.org/10.1016/j.jappgeo.2017.09.007.

Xu, L., Green, E.C., 2023. Inferring geological structural features from geophysical and geological mapping data using machine learning algorithms. Geophys. Prospect. 71 (9), 1728–1742. https://doi.org/10.1111/1365-2478.13371.

Yu, S., Ma, J., 2021. Deep learning for geophysics: current and future trends. Rev. Geophys. 59, e2021RG000742 https://doi.org/10.1029/2021RG000742.