

# Potential shallow aquifers characterization through an integrated geophysical method: multivariate approach by means of k-means algorithms

## *Caratterizzazione di potenziali acquiferi superficiali attraverso l'integrazione di metodi geofisici: applicazione di un codice di calcolo multivariato basato sul k-means*

S. Bernardinetti, S. Maraio, F. N. A. Bruno, V. Cicala, S. Minucci, M. Giannuzzi, M. Trotta, F. Cureda, S. Febo, M. Vacca, E. Guastaldi, T. Colonna, F. Bonciani, E. Tufarolo, F. Brogna, A. Zirulia, O. Milighetti

**Enrico GUASTALDI** ✉

GeoExplorer Impresa Sociale S.r.l.  
Via Bruxelles 10, 52022 Cavriglia (AR)  
guastaldi@geoexplorer.cgtgroup.org

**Stefano BERNARDINETTI**

-CGT Centro di Geotecnologie, Università di Siena  
Via Vetri Vecchi 34, 52027 San Giovanni Valdarno (AR)  
-Dipartimento di Scienze Chimiche e Geologiche,  
Università degli Studi di Cagliari, Via Trentino 51, Cagliari

**Stefano MARAIO, Valentina CICALA**

**Miriana GIANNUZZI, Marilena TROTTA**  
**Francesco CUREDDA, Simone FEBO, Matteo VACCA**  
-CGT Centro di Geotecnologie, Università di Siena  
Via Vetri Vecchi 34, 52027 San Giovanni Valdarno (AR)

**Fabio Nunzio Antonio BRUNO**

-Khalifa University of Science & Technology; Petroleum Institute,  
P.O. Box 2533, Abu Dhabi, United Arab Emirates

**Serena MINUCCI**

**Andrea ZIRULIA**  
-CGT Centro di Geotecnologie, Università di Siena,  
Via Vetri Vecchi 34, 52027 San Giovanni Valdarno (AR)  
-Dipartimento di Scienze chimiche e geologiche,  
Università degli Studi di Cagliari, Via Trentino 51, Cagliari  
-GeoExplorer Impresa Sociale S.r.l., Via Bruxelles 10, 52022 Cavriglia (AR)

**Tommaso COLONNA**

**Filippo BONCIANI**  
-GeoExplorer Impresa Sociale S.r.l., Via Bruxelles 10, 52022 Cavriglia (AR)

**Emanuele TUFAROLO**

-CGT Centro di Geotecnologie, Università di Siena,  
Via Vetri Vecchi 34, 52027 San Giovanni Valdarno (AR)  
-Dipartimento di Scienze Fisiche, della Terra e dell'Ambiente  
Università di Siena, Via Laterino 8, 53100 Siena

**Fabio BROGNA**

-Eurovix S.p.A., Viale Enrico Mattei, 24060 Entratico (BG)

**Omar MILIGHETTI**

-Nuove Acque S.p.A., Via Montefalco 49/55, 52100 Arezzo.

Ricevuto/Received: 16 June 2017-Accettato/Accepted: 26 June 2017  
Pubblicato online/Published online: 30 June 2017

This is an open access article under the CC BY-NC-ND license:  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

© Associazione Acque Sotterranee 2017

**Parole chiave:** integrazione geofisica, ricerca d'acqua, k-means, acquifero superficiale.

**Keywords:** geophysical integration, groundwater research, k-means, shallow aquifer.

**Riassunto:** La necessità di raggiungere una dettagliata caratterizzazione idrogeologica del sottosuolo ed una sua interpretazione per uno sfruttamento sostenibile della risorsa idrica, spesso richiede l'applicazione di più metodi geofisici complementari. Il metodo utilizzato in questo lavoro ha lo scopo di fornire un modello finale univoco che fornisca informazioni utili inerenti le variazioni delle caratteristiche fisico-meccaniche del sottosuolo evidenziate dai differenti metodi geofisici. L'area oggetto di studio ricade nel comune di Laterina (AR); si tratta di un bacino superficiale colmato da depositi lacustri e alluvionali (Pleistocenici ed Olocenici, periodo Quaternario), costituito da un'alternanza di limi, sabbie con percentuali variabili di ghiaie ed argille e dove il substrato è rappresentato da rocce di tipo arenaceo-pelitico (Unità del Monte Cervarola, Dominio Toscano, età Miocenica). Tale bacino superficiale costituisce l'acquifero freatico superficiale di cui è previsto lo sfruttamento nell'immediato futuro. Per implementare il modello geologico ottenuto dal rilevamento geologico di dettaglio sono state utilizzate una tomografia elettrica di resistività ed una tomografia sismica dei primi arrivi in onde P sulla stessa linea di acquisizione, al fine di ottenere 2 modelli tomografici indipendenti ed integrabili. Per mezzo dell'algoritmo k-means è possibile ottenere un'analisi di cluster del dataset bivariato al fine di individuare relazioni tra i due set di variabili. Questo algoritmo permette di individuare un numero n di cluster all'interno dei quali la varianza tra le osservazioni sia minima e sia massimizzata quella tra i diversi cluster. Il numero ottimale di cluster "K", corrispondente alle facies geofisiche individuate, dipende dalla distribuzione del dataset multivariato coadiuvato dalla stima tramite Silhouettes. Il risultato è una tomografia integrata che mostra un numero finito di facies geofisiche omogenee e che permette inoltre di distinguere ed interpretare l'acquifero poroso in modo oggettivo e quantitativo.

**Abstract:** *The need to obtain a detailed hydrogeological characterization of the subsurface and its interpretation for the groundwater resources management, often requires to apply several and complementary geophysical methods. The goal of the approach in this paper is to provide a unique model of the aquifer by synthesizing and optimizing the information provided by several geophysical methods. This approach greatly reduces the degree of uncertainty and subjectivity of the interpretation by exploiting the different physical and mechanic characteristics of the aquifer. The studied area, into the municipality of Laterina (Arezzo, Italy), is a shallow basin filled by lacustrine and alluvial deposits (Pleistocene and Olocene epochs, Quaternary period), with alternated silt, sand with variable content of gravel and clay where the bottom is represented by arenaceous-pelitic rocks (Mt. Cervarola Unit, Tuscan Domain, Miocene epoch). This shallow basin constitutes the unconfined superficial aquifer to be exploited in the nearby future. To improve the geological model obtained from a detailed geological survey we performed electrical resistivity and P wave refraction tomographies along the same line in order to obtain different, independent and integrable data sets. For the seismic data also the reflected events have been processed, a remarkable contribution to draw the geologic setting. Through the k-means algorithm, we perform a cluster analysis for the bivariate data set to individuate relationships between the two sets of variables. This algorithm allows to individuate clusters with the aim of minimizing the dissimilarity within each cluster and maximizing it among different clusters of the bivariate data set. The optimal number of clusters "K", corresponding to the individuated geophysical facies, depends to the multivariate data set distribution and in this work is estimated with the Silhouettes. The result is an integrated tomography that shows a finite number of homogeneous geophysical facies, which therefore permits to distinguish and interpret the porous aquifer in a quantitative and objective way.*

## Introduzione

Le indagini geofisiche sono ormai diventate uno strumento indispensabile a supporto delle ricerche idrogeologiche. La soggettività dell'interpretazione dei risultati geofisici nello studio e nell'individuazione di potenziali acquiferi spesso viene ridotta integrando differenti metodi geofisici, ad esempio utilizzando i metodi sismici per la capacità di individuare condizioni di saturazione in sedimenti sciolti (Haeni 1986; Al-Garni et al. 2010), unitamente ai metodi di resistività elettrica per valutare la porosità e permeabilità dei mezzi. Infatti, la velocità di propagazione delle onde di compressione e la resistività elettrica sono due grandezze geofisiche che ben si prestano alla valutazione delle caratteristiche litostratigrafiche e della presenza di fluidi nel sottosuolo. L'approccio che si propone in questo lavoro verte nel caratterizzare l'acquifero utilizzando contemporaneamente entrambi i parametri geofisici sopracitati, al fine di ridurre le incertezze e le ambiguità nella interpretazione dei dati, conferendo in un risultato finale univoco le caratteristiche evidenziate da entrambe i parametri.

Esistono diversi approcci di elaborazione basati su data set geofisici multipli, in alcuni casi basati sullo stesso parametro geofisico (Lines et al. 1988), in altri tra diversi parametri uniti da relazioni petrofisiche definite a priori (Lees e Van

Decar 1991; Dell'Aversana et al. 2016). Gallardo e Meju (2004) propongono un interessante approccio di inversione congiunta tra velocità sismica e resistività elettrica basato su similitudini strutturali nei modelli tomografici, risolvendo il problema dell'utilizzo di relazioni a priori tra i differenti parametri. Questi approcci, seppur con ottimi risultati, sono vincolati alla ricerca di punti di unione tra le diverse variabili geofisiche, non sempre facilmente ottenibili.

Il metodo che caratterizza questo lavoro si basa invece sull'elaborazione dei dati geoelettrici di resistività e di velocità sismica delle onde di compressione in maniera separata e prevede l'integrazione post – elaborazione tramite un'analisi statistica multivariata. Attraverso l'utilizzo del k – means il data set bivariato è partizionato individuando dei gruppi di dati ad elevata similarità, che consentono di discretizzare il modello tomografico 2D in zone a comportamento geofisico omogeneo.

## Assetto geologico e idrogeologico

L'area indagata, dal punto di vista geologico regionale è collocata nell'estrema porzione meridionale del Bacino Plio-Pleistocenico del Valdarno Superiore (Fig.1), costituito da un half-graben asimmetrico con faglie dirette principali, ubicate

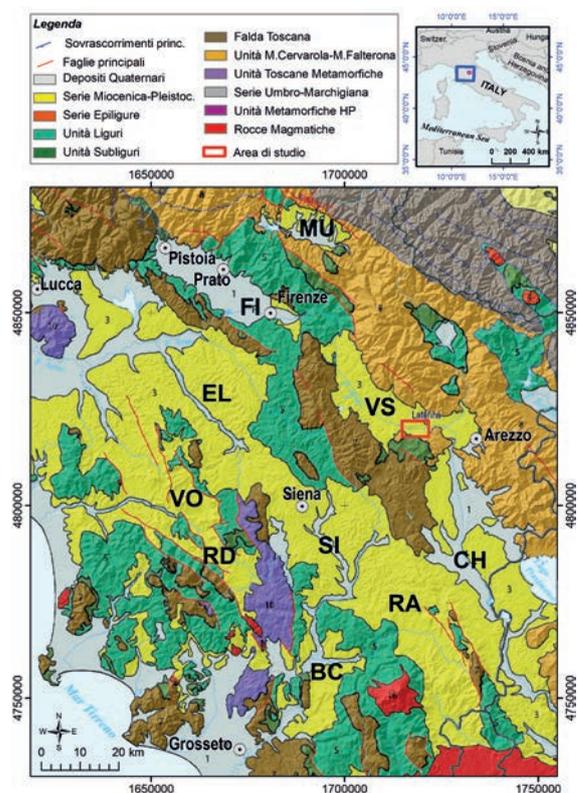


Fig. 1 - Location of the study area in the tectonic plan of Tuscany Apennine area: Neogenic – Quaternary basins of the northern Apennine: BC Baccinello, CH Chiana, EL Elsa, FI Firenze, MU Mugello, RA Radicofani, RD Radicondoli, SI Siena, VO Volterra, VS Valdarno Superiore.

Fig. 1 - Ubicazione dell'area di studio sullo schema tettonico dell'area appenninica toscana: Bacini Neogenici - Quaternari dell'Appennino settentrionale: BC Baccinello, CH Chiana, EL Elsa, FI Firenze, MU Mugello, RA Radicofani, RD Radicondoli, SI Siena, VO Volterra, VS Valdarno Superiore.

alla base dei versanti occidentali del Pratomagno e delimitato ad occidente dalla dorsale dei Monti del Chianti (Bossio et al. 1992; Martini e Sagri, 1993; Boccaletti et al. 1996).

La formazione del bacino valdarnese è legata alla fase distensiva che ha portato all'apertura del Tirreno e segue le deformazioni della fase compressiva, migrate a loro volta verso l'area umbro-romagnola (Boccaletti et al. 1990).

Il modello geologico prevede nell'area di interesse la presenza di 3 unità stratigrafiche (Fig. 2), qui brevemente descritte:

- Depositi alluvionali di età olocenica costituiti da sabbie e ciottolami per uno spessore variabile da 5 m a 20 m (identificati con la sigla "bna" in figura 2). Nello specifico settore d'interesse sono stati individuati due corpi di conoide alluvionale, il maggiore associato al Torrente Bregine e l'altro di minore dimensioni associato ad un piccolo fosso in sinistra idrografica del Torrente Bregine (Fig. 2). I sedimenti che caratterizzano i depositi olocenici delle due conoidi sono costituiti principalmente da sabbie con lenti ghiaiose, tali depositi costituiscono l'acquifero superficiale;
- Limi e argille grigio azzurre villafranchiane con resti vegetali, di spessore variabile da 10 m a 50 m appartenenti al Sub sistema del Torrente Ciuffenna riferibili al Pleistocene medio inferiore (VILa, VILb, VILc, VILh, in figura 2).
- Arenarie oligo-mioceniche in strati decimetrici alternati a peliti e argilliti (FAL4, Fig. 2). Il substrato è costituito dall'Unità del Monte Cervarola-Falterona costituito da

arenarie fini e silti in strati da centimetrici a decimetrici alternati a peliti marnoso siltose. Sono localmente presenti intercalazioni di arenarie a granulometria media e grossolana con spessori medi di circa 1.5 m. I livelli fratturati dell'Unità del Monte Cervarola-Falterona costituiscono un acquifero profondo di notevole interesse.

Nell'area orientale il rilevamento geologico appositamente effettuato ha evidenziato una superficie di faglia ad alto angolo caratterizzata da una direzione circa N/NO-S/SE, con immersione verso il quadrante occidentale, connessa all'attività tettonica distensiva, che ha dato origine al bacino sedimentario intrappenninico Plio-Pleistocenico del Valdarno Superiore. Tale lineamento, è stato evidenziato al margine sud-est del bacino di Laterina, dove alcuni affioramenti documentano le caratteristiche della faglia.

Nell'area di indagine è stata ipotizzata l'esistenza di due acquiferi, sulla base dei valori di soggiacenza misurati nel 2016 (Fig. 2): uno principale ubicato probabilmente ad una profondità di alcune decine di metri all'interno dei depositi fluviali ed un sistema di falde superficiali secondarie ubicate a pochi metri di profondità dal piano campagna in depositi a geometria lentiforme, di granulometria ghiaioso-sabbiosa e la cui ricarica si ritiene sia dovuta principalmente alle piogge o al ruscellamento superficiale.

Lo studio geologico e idrogeologico delle sorgenti presenti nell'area di indagine non hanno permesso una quantificazione degli spessori degli acquiferi, rendendo necessaria pertanto un'integrazione dello studio con tecniche geofisiche ed in particolare geoelettriche e sismiche.

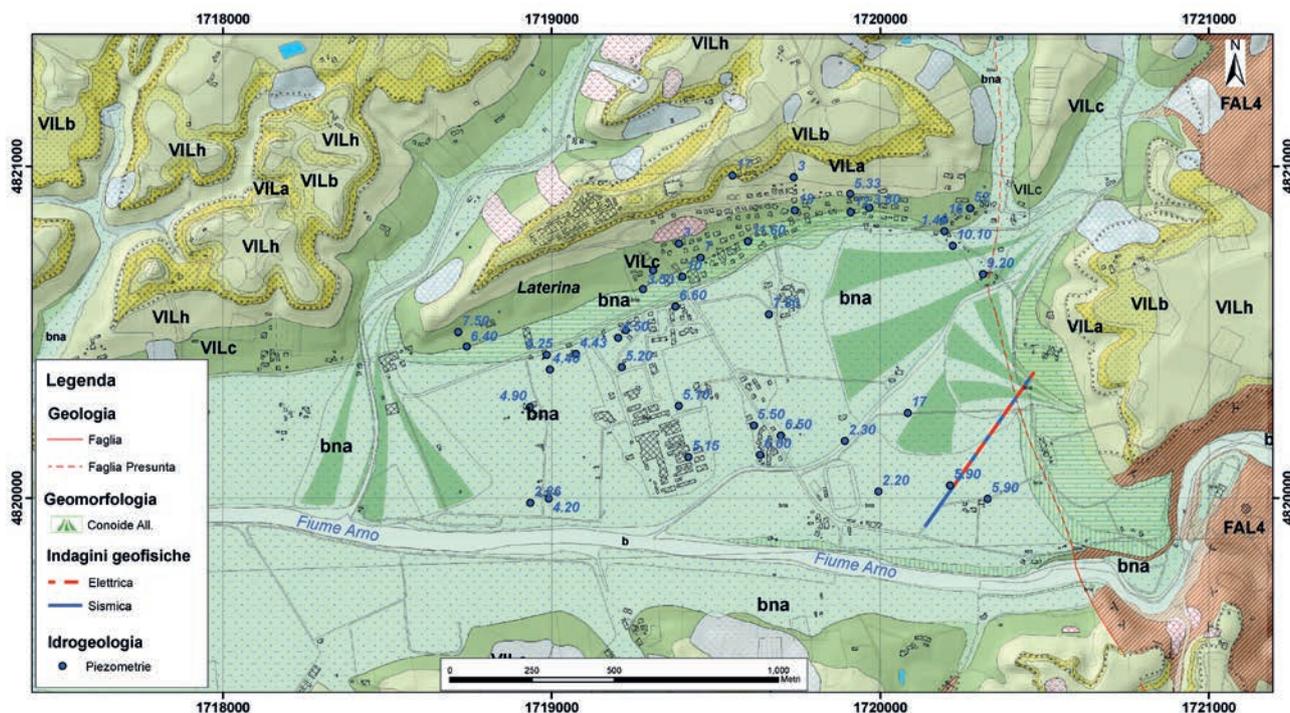


Fig. 2 - Geological map of the investigated area with the location of geophysical surveys and the static level of groundwater measured in available wells. b = actual alluvial deposits; bna = recent deposits; VILa,b,c,h = Villafranchian fluvial-lacustrine deposits; FAL4=Tectonic Unit Cervarola -Falterona.

Fig. 2 - Carta geologica dell'area indagata con ubicazione indagini geofisiche e dei livelli statici delle acque sotterranee misurati nei pozzi accessibili. b = depositi alluvionali attuali; bna = depositi alluvionali recenti; VILa,b,c,h = Depositi Fluvio-lacustri Villafranchiani; FAL4 = Unità tettonica Cervarola-Falterona.

## Acquisizione e analisi dei dati

Il tipo di target, inteso come profondità, estensione e contesto geologico, ci ha condotti verso la scelta di utilizzare come metodi di indagine la tomografia elettrica di resistività e il metodo sismico a rifrazione, due tecniche di indagine già diffusamente applicate allo studio di potenziali acquiferi. Al fine di integrare i risultati, le due tecniche geofisiche sono state applicate sullo stesso profilo e sovrapposte per un tratto di circa 420 m di lunghezza, utilizzando spaziature in fase di acquisizione volte ad ottenere un'elevata risoluzione e comparabilità dei risultati.

### Indagine geoelettrica

I dati di resistività elettrica in corrente continua sono stati acquisiti attraverso l'utilizzo di una base di 96 elettrodi con modalità roll-along (2 spostamenti in avanti da 24 elettrodi ciascuno) e una spaziatura interelettrodica di 3 m. La configurazione quadripolare utilizzata è la Schlumberger reciproco, che consente di realizzare misure di resistività apparente con elevato rapporto segnale/rumore e idoneo all'individuazione di forme geologiche anche complesse.

Tab. 1 - Acquisition parameters for geoelectrical resistivity tomography.

Tab. 1 - Parametri di acquisizione della indagine geoelettrica di resistività.

Indagine geoelettrica	
n° elettrodi attivi	96 + 2 rollalong di 24 elettrodi
Spaziatura interelettrodica (m)	3
Configurazione quadripolare	Schlumberger reciproco
Durata immissione (ms)	500
Vab (V)	400

L'elaborazione del dataset è stata realizzata attraverso una prima fase di rimozione delle misure di resistività apparenti con deviazione standard superiore al 5%, resistività apparenti negative e outlier (Loke 2004). L'inversione tomografica è stata realizzata con il software Res2Dinv, basato sul metodo "smoothness – constrained" ai minimi quadrati (de Groot-Hedlin and Constable 1990; Sasaki 1992). Il modello 2D è suddiviso in elementi rettangolari di dimensioni pari a metà della spaziatura elettrodica utile a modellare in maniera opportuna grandi variazioni di resistività laterali. Il metodo di ottimizzazione del software si basa sulla riduzione dello scarto relativo quadratico medio (RMS) tra pseudo-sezione apparente calcolata e misurata, variando i valori di resistività del modello a blocchi. La tomografia finale è stata ottenuta dopo 4 iterazioni con un valore di RMS pari a 4.1%.

### Indagine sismica

Il dataset sismico è stato acquisito lungo un profilo di 576 metri, utilizzando due stendimenti di 192 canali attivi, equipaggiati con geofoni con frequenza naturale di 40 Hz e equispaziati di 2 m. Le energizzazioni sono state effettuate tramite fucile sismico con interdistanza di 4 m. Tale geometria di acquisizione ha permesso di registrare dati caratterizzati

da un'alta risoluzione spaziale in un ampio intervallo di offset (circa 384 m), entrambi fattori fondamentali per la ricostruzione dettagliata e affidabile del modello di velocità. I parametri di acquisizione sono dettagliati in Tabella 2.

Tab. 2 - Acquisition parameters for seismic refraction data.

Tab. 2 - Parametri di acquisizione dei dati sismici a rifrazione.

Indagine sismica	
Sismografi utilizzati	8
Canali attivi	192
Frequenza geofoni (Hz)	40
Spaziatura geofoni (m)	2
Spaziatura sorgenti (m)	4
Numero array	2
Intervallo campionamento (ms)	1
Tempo di registrazione (s)	1

In fase di elaborazione, la lettura dei primi arrivi è stata effettuata manualmente su circa 27000 tracce; i tempi letti sulle fasi dirette e rifratte sono stati poi invertiti attraverso un algoritmo di ricostruzione di immagine iterativa denominata "SIRT" (Simultaneous Iterative Reconstruction Technique; Gilbert 1972). Il metodo di inversione richiede una stima di un modello iniziale di velocità del sottosuolo, ottenuta attraverso tradizionali tecniche di analisi di rifrazione (Burger 1992). Tale modello di velocità è iterativamente ricostruito sulla base dei tempi letti: il software crea i tempi di arrivo teorici per ciascun modello (raytracing) e l'algoritmo, attraverso l'analisi della differenza tra il tempo misurato e quello letto, modifica le velocità nelle celle del modello. Raytracing e SIRT vengono applicate al modello di velocità finché l'errore RMS (Root-Mean Square) relativo alla differenza tra i tempi di percorso osservati e calcolati viene minimizzato. Il modello di velocità calcolato ha raggiunto, dopo 20 iterazioni, un errore RMS pari a 1.59 ms.

### Analisi integrata

Il metodo di integrazione dei dati proposto, si basa sull'individuazione di zone in cui un dataset multivariato mostra delle similarità tra le diverse variabili, nello specifico le diverse variabili sono rappresentate dalle grandezze geofisiche e i campioni sono i punti di un modello tomografico 2D. Con questo approccio si vanno ad individuare delle porzioni di sottosuolo all'interno delle quali si ha comportamento geofisico omogeneo, che quindi possono essere definite "facies geofisiche". Attraverso l'utilizzo di algoritmi di clustering, si creano dei raggruppamenti all'interno dei quali la dissimilarità tra le osservazioni sia minima mentre sia massima quella tra i diversi cluster.

I dati di partenza sono costituiti da modelli tomografici 2D e da variabili indipendenti una dall'altra (ad es. resistività, polarizzazione, velocità onde P e S, densità, ecc.). Questo processo consente di integrare dati, senza dover fare ipotesi a priori sul legame tra i diversi attributi geofisici utilizzati,

spesso ipotesi sito-dipendenti e fortemente vincolanti. In questo modo si ottiene un processo di integrazione dati del tutto generalizzabile, applicabile in differenti contesti geologici e a partire da variabili in input più idonee alla caratterizzazione del target ricercato.

Come già esposto sopra, l'integrazione proposta è stata effettuata a partire dai modelli tomografici di resistività elettrica e velocità in onde P. Il clustering del data set bivariato è stato effettuato attraverso l'implementazione di un codice di calcolo basato sull'algoritmo k-means (MacQueen 1967).

Il k-means è un algoritmo di clustering partizionale, non gerarchico e non supervisionato, che permette di suddividere un insieme di campioni in K gruppi sulla base delle variabili che li rappresentano. L'obiettivo che l'algoritmo si prepone è di minimizzare la varianza totale intra-cluster e massimizzare quella tra i diversi cluster.

L'algoritmo si inizializza definendo in maniera casuale k centroidi e assegna al j-esimo centroide i campioni più vicini. Si calcola poi il baricentro di ogni gruppo così formato e la varianza interna. Il procedimento di assegnazione dei campioni sulla base delle distanze si reitera, a partire dai nuovi centroidi rappresentati dai baricentri precedentemente calcolati. Il processo si reitera fin quando la varianza interna sia minima e quella tra i cluster sia massima. Il calcolo delle distanze può essere effettuato per mezzo di differenti metodi, ma nel caso in esame è stato utilizzato il quadrato della distanza euclidea. Il numero ottimale dei cluster in cui suddividere il data set non è noto a priori ma è un dato in input necessario ad avviare l'algoritmo. Si procede quindi effettuando l'analisi di clustering per diversi K a partire dal minimo 2 e poi si valuta la bontà dell'analisi di clustering a posteriori. Esistono diversi indici di valutazione della validità del clustering basati su differenti concetti (Halkidi et al. 2001), in questo caso è stato utilizzato il metodo Silhouettes (Rousseeuw 1987) basato sul

calcolo delle distanze intra- e infra-cluster. Questo metodo consente di calcolare un indice che valuta il posizionamento di ogni singolo campione all'interno del cluster a cui è stato assegnato, attraverso il calcolo della media delle distanze intra cluster ed extra cluster. Per l'i-esimo campione appartenente al j-esimo cluster, il valore dell'indice è rappresentato graficamente dalla frontiera dell'area blu (Figura 3). L'indice che ne consegue varia tra -1 e 1: per valori vicini a 1 il cluster è ben formato, per valori prossimi a zero non è chiara la posizione da assegnare all'i-esimo campione e per valori prossimi a -1 il dataset è mal partizionato.

L'applicazione dell'algoritmo k-means ha previsto una fase di omogeneizzazione dei 2 modelli tomografici attraverso la realizzazione di un grid 2D regolare e con variabili resistività elettrica e velocità delle onde P co-registrate. Inoltre, il data set bivariato espresso come logaritmo della resistività e velocità (vedi Figura 6 b); , è stato normalizzato, per rimuovere possibili effetti di dominanza di una delle due variabili geofisiche utilizzate (dovuti a differenti deviazioni standard) (Martinez & Martinez, 2005), attraverso il calcolo della matrice di covarianza e per mezzo della formula di data sphering (Martinez & Martinez 2005):

$$Z = D^{-0.5} Q^T (x_i - \bar{x}), \quad i = 1, \dots, N$$

dove Q sono gli autovettori della matrice di covarianza,  $\Delta$  è la matrice diagonale degli autovalori,  $x_i$  è l'i-esimo campione della variabile geofisica e  $\bar{x}$  è il valor medio della variabile geofisica.

A questa parte di pretrattamento dei dati è seguita l'applicazione dell'algoritmo k-means, per valori di K variabili da 2 a 10, per ognuno dei quali è stato calcolato l'indice di validità Silhouettes. Il clustering ottimale del data set è stato individuato per quel valore di K che ha reso massimo l'indice di validità Silhouettes. In questo modo a partire dalle due tomografie iniziali è stato ottenuto un modello tomografico finale integrato, comprensivo delle capacità di caratterizzazione del metodo resistività elettrica e di quello della velocità sismica in onde P, discretizzato in zone geofisiche a comportamento omogeneo (facies geofisiche) in maniera oggettiva.

## Risultati e discussioni

Il modello di resistività mostrato in Figura 4 evidenzia un range di resistività compreso tra circa 5 ohm-m e 500 ohm-m, con profondità massime tra 50 m e 60 m. Nella tomografia sono evidenti alternanze di materiali resistivi a materiali conduttivi con gradienti di resistività verticali elevati, soprattutto per la parte compresa nei primi 25 m dal piano campagna. Sono altresì evidenti delle variazioni di resistività laterali in corrispondenza della progressiva 50 m e poi un incremento di valori di resistività del corpo superficiale a partire dalla progressiva 170 m. Nella zona più depressa della tomografia è evidente un incremento della resistività,

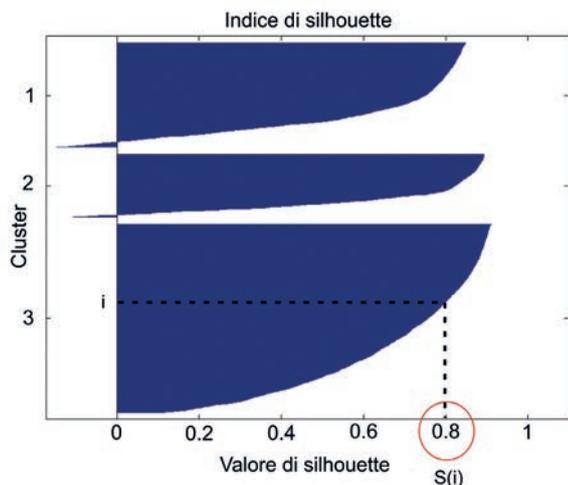


Fig. 3 - An example for the Silhouette index with 3 clusters: i-th sample for the cluster 3 has a Silhouette index value  $S_{(i)}=0.8$ .

Fig. 3 - Esempio dell'indice di Silhouette per un numero di cluster pari a 3: l'i-esimo campione appartenente al cluster 3 ha un valore dell'indice Silhouette pari a  $S_{(i)}=0.8$ .

variazione coerente con quanto evidenziato dalla tomografia sismica.

Come accennato nel paragrafo sull' Acquisizione e analisi dei dati i profili tomografici sono stati sovrapposti per circa 420 m, per cui la Figura 4 mostra soltanto la parte del modello tomografico di velocità interessata dalla sovrapposizione. Il modello mostra valori di velocità in un intervallo tra 300 m/s e 3000 m/s, raggiungendo profondità massime di indagine di circa 50 m. La porzione superficiale del modello (fino a 20 m dal p.c.) è caratterizzata da un forte gradiente verticale di velocità e presenta in corrispondenza della metrica 40 una variazione laterale di velocità che evidenzia un ribassamento del sismostrato con velocità maggiore (approssimativamente 1200 m/s) di circa 10 m. Lo stesso sismostrato mostra poi una tendenza a risalire verso SO. Velocità maggiori di 2700 m/s sono mostrate nel modello tra la metrica 60 e 240 a una quota di circa 140 m s.l.m. Anche in questo caso si nota una brusca variazione laterale di velocità alla metrica 190, che mette in evidenza un ribassamento dei terreni caratterizzati da  $V_p > 2700$  m/s fino alla quota più bassa risolta dal modello di velocità (130 m s.l.m.) in corrispondenza della metrica 240. A seguito del calcolo della matrice varianze-covarianze nel processo di normalizzazione (Tabella 3), l'analisi di integrazione multivariata ha restituito una tomografia suddivisa in 6 zone o "facies geofisiche", all'interno di ognuna delle quali è stata individuata una buona similarità tra le due variabili resistività elettrica e velocità delle onde P.

Il clustering ottimale è stato ottenuto per il numero di partizioni  $K=6$ , che ha reso massimo l'indice di Silhouette.

Tab. 3 - Covariance matrix of the analyzed data set.

Tab. 3 - Matrice di covarianza del data set analizzato.

	Log (Resistività)	Velocità
Log (Resistività)	0.1721	-0.1056
Velocità	-0.1056	0.3864

In Figura 5.a, è rappresentato graficamente l'indice di validità Silhouettes per i sei cluster individuati: forme delle aree (in blu) meno acuminate e con valori più prossimi a 1 sono rappresentative di cluster omogenei e molto ben partizionati (ad es. cluster 1 e 5). Come visibile nel grafico di Figura 5.b, il massimo valore di indice ottenuto nel range analizzato è per  $K=6$ , inoltre l'individuazione di valori minimi per  $K$  prossimi a 2 e a 10 confermano che il range di partizioni sperimentato è opportuno per l'analisi del data set in esame.

In Figura 6.b e Figura 6.c è possibile vedere l'individuazione dei cluster nel data set non trasformato e in quello normalizzato, dove i diversi colori e il rispettivo numero sono descrittivi del cluster in cui ricadono i campioni analizzati. Facendo riferimento al quadro geologico descritto nel paragrafo "Assetto geologico e idrogeologico", partendo dal litotipo stratigraficamente più basso, i campioni del cluster 5 (alta velocità, alta resistività) si ritengono rappresentativi del substrato arenaceo (Figura 6.a, facies geofisica verde); i cluster 4 e 1 sono caratterizzati da medio – elevate velocità ma con valori di resistività bassi, a cui si possono associare formazioni argillose limose costituite dai depositi villafranchiani; i cluster 3 e 6 sono riferibili per range di resistività e velocità

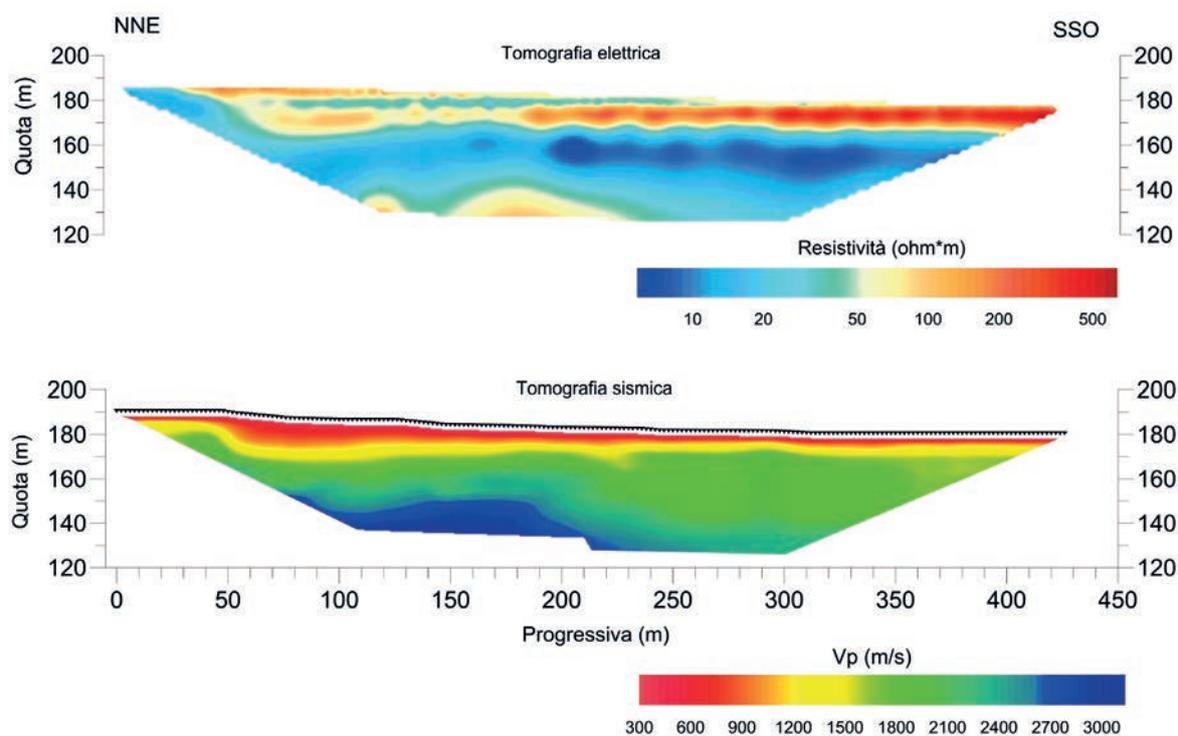


Fig. 4 - Electrical resistivity tomography (above) and P waves seismic refraction tomography (below).

Fig. 4 - Tomografia elettrica di resistività (in alto) e tomografia sismica di velocità delle onde P (in basso).

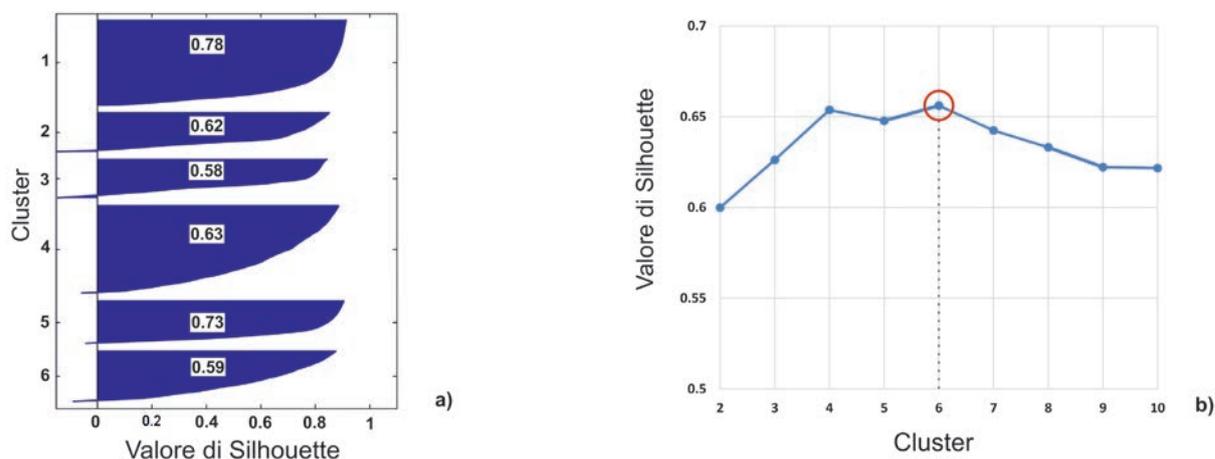


Fig. 5 - Silhouettes as index of clustering evaluation: a) silhouette graph for each cluster with numerical index for  $K=6$ ; the numerical value into blue areas is the average Silhouette index value for each of 6 clusters; b) average values of silhouette index for the  $K$  analyzed, with indication of the maximum value reached for  $K=6$ .

Fig. 5 - Utilizzo dell'indice di valutazione del clustering Silhouettes: a) sono rappresentati i grafici per il clustering ottimale ottenuto con  $K=6$  con valutazione di ogni singola partizione individuata; il valore numerico riportato all'interno delle aree blu esprime il valore medio dell'indice per ognuno dei 6 cluster; b) sono riportati i valori medi dell'indice in funzione dei diversi  $K$  analizzati, si può notare il massimo ottenuto per  $K=6$ .

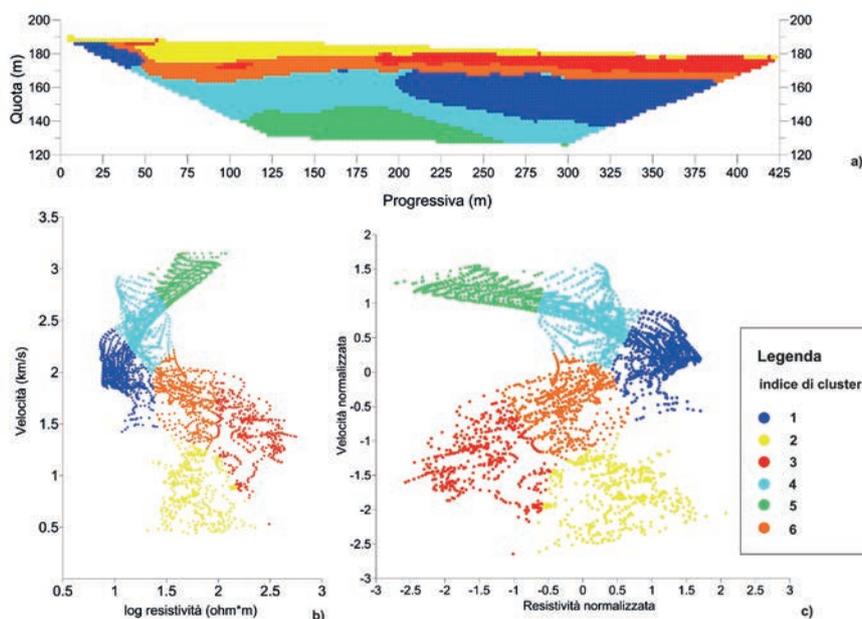


Fig. 6 - Multivariate analysis results with integrated tomography, original data set and normalized data set with data sphering technique: at each of 6 colors and numbers is associated a different geophysical facies. The joint analysis of clusters into the tomography and into the original data set scatter plot allows interpreting the geophysical facies in geological and hydrogeological way.

Fig. 6 - Risultati dell'analisi multivariata con tomografia integrata, data set originale e data set normalizzato attraverso la tecnica di data sphering: con i diversi colori si rappresentano le 6 facies geofisiche individuate attraverso l'analisi integrata. La visione congiunta dei cluster sulla tomografia 2D e nel grafico dei dati non normalizzati consente l'interpretazione delle facies geofisiche in chiave geologica e idrogeologica.

a materiali sabbioso - ghiaiosi saturi. Il cluster 2 è riferibile a depositi superficiali in parte non saturi dati i valori di velocità delle onde di compressione inferiori a 1300 m/s. In particolare il cluster 3 mostra graficamente una distribuzione spaziale non molto omogenea, soprattutto nella zona a minori velocità, con campioni che sembrerebbero più affini con il cluster 2. Questo aspetto è in parte confermato anche dalla Silhouette in Figura 5 che mostra un profilo a 2 diverse pendenze, la prima ai

valori più elevati dell'indice che mostra un'ottima similarità tra le osservazioni raggruppate, mentre il secondo tratto che prosegue verso valori dell'indice Silhouette negativi. Queste considerazioni conducono ad evidenziare con molta probabilità una non buona classificazione per i campioni in oggetto, che date le caratteristiche di velocità e resistività, geologicamente sembrerebbero ascrivibili alla porzione di depositi superficiali non saturi (cluster 2).

I più elevati valori dell'indice silhouette sono stati ottenuti per i cluster 1 e 5, rispettivamente con valori di 0.78 e 0.73 (Figura 5.a), rappresentativi di cluster omogenei e molto ben partizionati. Da un punto di vista geofisico si può osservare una distribuzione dei dati nel piano resistività-velocità poco dispersa e ben definita: tale definizione può essere considerata una risposta geofisica di unità stratigrafiche omogenee nello spazio indagato.

In relazione al modello geologico-idrogeologico concettuale dell'area, tale metodo ha permesso di discriminare gli spessori dell'acquifero poroso superficiale, costituiti da sabbie e ciottolami dei depositi di conoide alluvionale olocenici. Tale acquifero superficiale è chiaramente individuato dai cluster 2, 3 e 6 per spessori massimi di circa 20 m (Figura 6.a), inoltre, i soli cluster 3 e 6 individuano lo spessore saturo dell'acquifero che non supera i 10 m, in accordo con i pochi dati relativi al livello statico della falda freatica misurati nel corso dello studio. Tali distinzioni non sono così ben delineabili né dalla tomografia di resistività (Figura 4.a), né dalla tomografia sismica (Figura 4.b).

## Conclusioni

L'applicazione del codice di calcolo al caso di studio ha permesso di integrare e interpretare in maniera dettagliata e oggettiva le indagini geofisiche eseguite. I risultati dell'analisi confermano quanto evidenziato dal rilevamento geologico e idrogeologico. Lo scopo iniziale di studio del sottosuolo ai fini dell'individuazione di acquiferi saturi, in un contesto geologico ed idrogeologico ben definito, si può ritenere raggiunto grazie alla visualizzazione integrata dei dati resistività elettrica e velocità delle onde di compressione.

Uno dei principali limiti noti dell'algoritmo  $k - \text{means}$  è la capacità di partizionare data set di forma complessa e non sferici, che può condurre ad un clustering non ottimale del dataset. Tuttavia, essendo un algoritmo robusto e di rapida applicazione, in un contesto geologico e geofisico non complesso conduce a risultati validi, univoci, e rappresentativi del sottosuolo. Inoltre, i metodi integrati portano ad un miglioramento delle singole tomografie di resistività e sismica, che possono essere successivamente combinate ad esempio tramite tecniche di imaging multispettrale (Gallardo 2007). Diversamente, l'approccio bivariato numerico qui proposto fornisce come risultato una unica tomografia integrata, in cui i cluster sono univocamente definiti e le litologie ad essi associate sono ben identificabili e paragonabili al modello concettuale dell'area.

Futuri sviluppi saranno rivolti al miglioramento del processo di clustering, attraverso l'utilizzo di algoritmi più efficaci nel partizionamento di data set multivariati complessi e all'utilizzo di ulteriori variabili geofisiche.

## BIBLIOGRAFIA

- Al-Garni, Mansour A., Mohamed G. El-Behiry. (2010) Seismic Refraction Survey to Investigate Shallow Sediments of Wadi Thuwal, North of Jeddah, KSA. *Earth Sciences*. Vol. 22, No. 1, pp: 57-75. DOI:10.4197/Ear.22-1.4
- Boccaletti, M., Bonini, M., Moratti, G., Sani, F. (1996) Tectono-sedimentary Neogene evolution of the post-nappe basins in the Northern Apennines. *Notes et Mém Serv. Géol. Maroc*, n. 387: 117-193.
- Boccaletti, M., Calamita, F., Deiana, G., Gelati, R., Massari, F., Moratti, G., Ricci Lucchi, F. (1990) Migrating foredeep thrust belt system in the Northern Apennines and Southern Alps. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 77: 3-14 [https://doi.org/10.1016/0031-0182\(90\)90095-O](https://doi.org/10.1016/0031-0182(90)90095-O)
- Bossio A., Cerri R., Costantini A., Gandin A., Lazzarotto A., Magi M., Mazzanti R., Mazzei R., Sagri M., Salvatorini G. (1992) I bacini distensivi neogenici e quaternari della Toscana " *Neogene-Quaternary distensive basins of Tuscany*". 76° Riunione Estiva Soc. Geol. It., Firenze, settembre 1992, p. 199-277.
- Burger H. R. (1992). *Exploration Geophysics of the Shallow Subsurface*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 489 pp.
- de Groot-Hedlin, C., Constable, S. (1990). Occam's inversion to generate smooth, two dimensional models from magnetotelluric data. *Geophysics*, 55(2), 1613-1624 <https://doi.org/10.1190/1.1442813>
- Dell'Aversana, P., Bernasconi, G., Chiappa, F. (2016). A global integration platform for optimizing cooperative modeling and simultaneous joint inversion of multi-domain geophysical data. *AIMS Geosciences*, 2(1), 1-31, DOI: 10.3934/geosci.2016.1.1
- Gallardo, L., Meju, M.A., (2004). Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints. *J. Geophys. Res.* 109(B3). DOI: 10.1029/2003JB002716
- Gallardo L., (2007). Multiple crossgradient inversion for geospectral imaging. *Geophysical Research Letters*, 34, L19301, doi: 10.1029/2007GL030409.
- Gilbert P. (1972). Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor Biol*, 36(1):105-117, [https://doi.org/10.1016/0022-5193\(72\)90180-4](https://doi.org/10.1016/0022-5193(72)90180-4)
- Halkidi M., Batistakis Y., Vazirgiannis M. (2001). On Clustering Validation Techniques, *Journal of Intelligent Information Systems*, 17:2/3, 107-145, DOI:10.1023/A:1012801612483
- Haeni, F. P. (1986). Application of seismic refraction methods in groundwater modeling studies in New England: *Geophysics*, 51, no. 02, 236-249, <https://doi.org/10.1190/1.1442083>.
- Lees, J.M., Van Decar, J.C., (1991). Seismic tomography constrained by Bouguer gravity anomalies: applications in western Washington. *Pure Appl. Geophys.* 135, 31-52, DOI:10.1007/BF00877007
- Lines, L., Schultz, A., Treitel, S. (1988). Cooperative inversion of geophysical data. *Geophysics* 53(1), 8-20, <https://doi.org/10.1190/1.1442403>.
- Loke, M. (2004). Res2Dinv v. 3.54 for Windows 98/Me/2000/NT/XP. Rapid 2D Resistivity & IP Inversion using the least-squares method: Software Manual. Res2Dinv ver. 3.53 for Windows 98/Me/2000/NT/XP. Rapid 2D Resistivity & IP Inversion using the least-squares method: Software Manual.
- MacQueen, J.B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of 5th Berkley Symposium on Mathematical Statistics and Probability*, Volume I: Statistics, pp. 281-297.
- Martinez W. L., Martinez A. R., (2005). *Exploratory Data Analysis with MATLAB*. CHAPMAN & HALL/CRC.
- Martini I.P., Sagri M., (1993). Tectono sedimentary characteristics of late Miocene-Quaternary extensional basins of the Northern Apennines, Italy. *Earth Science Reviews*, 34, 197-133, [https://doi.org/10.1016/0012-8252\(93\)90034-5](https://doi.org/10.1016/0012-8252(93)90034-5).
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Sasaki, Y. (1992). Resolution of resistivity tomography inferred from numerical simulations. *Geophysical prospecting*, 40, 453-464, DOI:10.1111/j.1365-2478.1992.tb00536.x.