

DIPARTIMENTO DI SCIENZE BIOLOGICHE, GEOLOGICHE
E AMBIENTALI

Corso di Laurea Magistrale in Geologia e Territorio

Tesi di Laurea Magistrale

Correlazioni tra assetto geologico-strutturale del
distretto cinabrifero del M.te Amiata e sue risorse
geotermiche-epitermali

Candidato:
Lara Capitanio

Relatore:
Prof. Paolo S. Garofalo

Correlatori:
Ing. Fabio Remondino
Dott.ssa Elisa Mariarosaria Farella

Sommario

Abstract.....	2
Introduzione.....	3
Inquadramento geologico	10
2.1 Evoluzione geologica e tettonica del distretto amiatino	10
2.1.1 Storia tettonica.....	10
2.1.2 Assetto strutturale	15
2.1.3 Assetto stratigrafico.....	17
2.2 Risorse minerarie	24
2.3 Risorse geotermiche.....	31
2.3.1 Il sistema geotermico-epitermale del Monte Amiata	31
Dati e metodi	38
3.1 Dataset	38
3.1.1 Struttura.....	38
3.2 Modello 3D.....	44
3.3 Machine learning	50
3.3.1 Regressori.....	51
3.4 Costruzione dell' <i>input file</i> del distretto amiatino.....	55
3.4.1 Predittori.....	56
3.4.2 Incognita (mineralizzazioni).....	67
3.5 Applicazione del Machine Learning.....	69
3.5.1 Apprendimento	69
3.5.2 Test	73
3.5.3 Validazione.....	74
3.6 Elaborazione delle predizioni	75
Presentazione dati.....	77
4.1 Validazione	95
Discussione.....	97
5.1 Valutazione dei regressori	97
5.2 Valutazione dei risultati	98
5.2.1 Analisi di importanza dei predittori.....	99
5.2.2 Valutazione a partire dal dataset di training	99
5.2.3 Casi particolari.....	103
5.3 Variabilità delle correlazioni con la profondità	105
Conclusioni.....	108
Bibliografia.....	110

Abstract

I giacimenti di cinabro (\pm stibina) del M.te Amiata, localizzati in Toscana meridionale, costituiscono un distretto minerario di importanza mondiale, con una produzione storica totale che supera le 117 kt di Hg, prodotte tra il 1850 e il 1982. Nell'area del distretto minerario si trova l'omonimo sistema geotermico, con 6 impianti per la produzione di energia elettrica che producono 121MW equivalenti di energia.

La costruzione della catena a falde appenninica ha fortemente condizionato l'assetto strutturale dell'area amiatina, portando alla costruzione di strutture dovute ad eventi collisionali e post-collisionali che hanno vincolato il flusso di fluidi mineralizzanti all'interno del sistema geotermico-epitermale.

Lo scopo di questa tesi è di individuare e di verificare correlazioni esistenti tra la peculiare distribuzione N-S delle mineralizzazioni a cinabro, le manifestazioni geotermiche e l'assetto strutturale che caratterizza il distretto cinabrifero e il sistema geotermico.

Le correlazioni sono state individuate attraverso l'applicazione di algoritmi Machine Learning (ML) ad un dataset bidimensionale, costruito appositamente con applicazioni GIS per contenere tutti i dati geologici-giacimentologici reperiti in letteratura riguardo al distretto amiatino. È stato costruito anche un modello tridimensionale dell'area di studio basato sulla produzione di quattro solidi che raggruppano le formazioni geologiche presenti nell'area, sulla base delle loro caratteristiche geoidrologiche. Il modello tridimensionale ha permesso di effettuare uno studio di statistica multivariata che permette di capire come variano le correlazioni al variare della profondità.

Attraverso Scikit-learn, sono stati effettuati test sui dati usando differenti modelli di regressione al fine di confrontare diversi approcci statistici basati sulla regressione lineare e sugli alberi decisionali. A partire dai dati noti, sono stati scelti 9 parametri che descrivono le caratteristiche giacimentologiche e strutturali del distretto, denominate predittori, che costituiscono le variabili del calcolo statistico. Sono state testate differenti combinazioni di predittori per capire quale associazione di fattori abbia controllato efficacemente la precipitazione di > 177 kt di Hg.

Sulla base dei risultati ottenuti si può affermare che le tecniche di ML si sono dimostrate utili nell'identificare correlazioni tra i diversi fattori geologico-strutturali che caratterizzano il sistema geotermico del M.te Amiata; la peculiare distribuzione spaziale N-S dei giacimenti del distretto dipende dalla combinazione di un sistema di faglie e di pieghe; i modelli di regressione basati su alberi decisionali (CatBoost e Random Forest) sono complessivamente più performanti e geologicamente significativi rispetto ai modelli lineari.

Questo lavoro suggerisce che il ML rappresenta uno strumento in grado di suggerire nuove e poco sperimentate relazioni tra elementi geologici-giacimentologici di un'area complessa come un sistema geotermico ed è in grado di guidare eventuali fasi successive di studi geologici complessi.

Introduzione

Le mineralizzazioni epitermali a metalli di base e preziosi (Pb, Zn, Cu, Au, Ag) hanno caratteristiche geologiche e giacimentologiche molto simili a quelle dei campi geotermici, e questa similitudine è nota da tempo nella letteratura geologica (White, 1981; Henley & Brown, 1985; Barnes & Seward, 1997). Questo concetto è così radicato nella comunità geologica che è ormai indiscusso il fatto che le mineralizzazioni epitermali rappresentano le espressioni fossili di antichi campi geotermici (Barnes, 2015). Le evidenze su cui si basa questa similitudine non si limitano al fatto che gli ambienti geologici in cui ricorrono le mineralizzazioni epitermali ed i campi geotermici sono gli stessi, ma soprattutto sulle numerose documentazioni di precipitazione di solfuri di metalli di base (sfalerite, calcopirite, galena), sulla precipitazione di Au ed Ag all'interno di pozzi geotermici in produzione e nelle sorgenti calde (Brown, 1986; Brown & Simmons, 2003; Simmons & Brown, 2006; Hardardóttir et al., 2013), sulle caratteristiche alterazioni idrotermali, e su altre evidenze geochimiche.

In aggiunta al connubio epitermale-idrotermale, i sistemi geotermici hanno la caratteristica di ospitare sistematicamente alte concentrazioni di un gruppo di elementi che include Sb, As, Ga, Ge, Tl, B, In e Hg (Barnes, 2015). Alcuni di questi elementi – in particolare Sb, Ga, Ge ed In – hanno notevole rilevanza economica, dal momento che sono considerati critici per l'industria dell'Unione Europea a causa dello scarso approvvigionamento in relazione alla disponibilità geologica (AA.VV., 2020). Nei sistemi epitermali-idrotermali questi elementi ricorrono quasi sempre in proporzioni variabili come elementi minori ed in traccia all'interno delle mineralizzazioni (Brown & Simmons, 2003), in genere come sostituenti isomorfogeni di minerali di metalli di base (e.g., Ga, In e Ge in sfalerite). In alcuni casi però questi elementi formano minerali propri (e.g., antimonite, arsenopirite, realgar, cinabro). La comune presenza di questo gruppo di elementi nell'ambiente epitermale-idrotermale mostra che, in aggiunta alle documentate associazioni geologiche e geochimiche descritte sopra, la distribuzione spaziale degli elementi minori nei campi geotermici e nelle mineralizzazioni fossili è sistematica e prevedibile e rappresenta un utile elemento di conoscenza per la definizione di modelli genetici.

Nell'ambiente epitermale-idrotermale, la determinazione del Hg è considerata in genere un indicatore di sorgenti calde (White, 1981). I minerali mercuriferi documentati nei campi geotermici attivi della Nuova Zelanda (e.g., Broadlands, Rotokawa, Ngawha) e della California (Coastal Range), oltre che nei giacimenti epitermali fossili (ad es., Steamboat Springs, Nevada), sono principalmente il cinabro, il metacinabro ed il Hg metallico (Barnes & Seward, 1997). Studi sulla distribuzione dei minerali a mercurio all'interno di giacimenti epitermali fossili ad Au-Ag sono relativamente scarsi (Steamboat Springs, Nevada: White, 1981). Tuttavia, è stato chiaramente evidenziato che all'interno dei corpi minerari il cinabro ricorre a profondità molto modeste rispetto alla superficie topografica (tra 15 e 25m) ed è disseminato in concentrazioni dell'ordine di 100 µg/g all'interno di calcedonio o nella roccia ospite alterata chimicamente e mineralogicamente. Questo tipo di evidenza è molto importante, perché rappresenta un vincolo per i modelli genetici di giacimenti epitermali.

Con una produzione storica totale eccedente i 14 Mt e generata tra il 1850 ed il 1982 da 14 giacimenti con tenori variabili dallo 0.2 a c. 6 wt% Hg (Tab. 1.1), i giacimenti di cinabro (±stibina) del M.te Amiata rappresentano un distretto minerario di importanza mondiale.

Tabella 1.1. Riassunto delle caratteristiche giacimentologiche dei giacimenti a Hg(\pm Sb) del distretto del M.te Amiata.

Miniera	Periodo di sfruttamento	Tenori wt% Hg	Produzione (t) *	Rocce incassanti	Controlli strutturali	Ore body	Minerali	Ganga
Abbadia S. Salvatore	1847-1982	<1	58944	Calcareniti e calcilutiti (Eocene)	Pieghe, faglie	Caotico e disseminato. Stockwork. Breccie imbutiformi verticalmente continue (camini)	Cinn, metacinn, py, marc	Calcite, argille
Solforate Schwarzenberg	1852-1940							
Solforate	1883-1982	6	29972	Calcareniti e calcilutiti (Eocene)	Faglie in calcarenite con direzione prevalente NW-SE e parallele a stratificazione	Vene, disseminazioni in argille. Replacement bodies in calcareniti	Cinn, py, marc, stibnite	Argille, calcite, gesso, qtz, dawsonite
Morone (Selvena)	1850-1982		6408	Anidriti, dolomite calcari neri (Retico)		Stratabound, disseminato, caotico, breccie imbutiformi verticalmente continue (camini)	Cinn, py, stibnite,	Argille (nere), gesso
Siele	1846-1981	c. 8		Calcareniti e calcilutiti (Eocene)	Quattro direzioni sistematiche registrate in vene stockwork: E-W; N-S; NE-SW; parallele a stratificazione	Stockwork (spessore vene: fino a 20-30 cm)	Cinn, marc	Argille (nere), calcite, bitume
Abetina	1917-1982	<2	4237					
Cornacchino	1872-1921	0.1-0.2	2515	Calcari e radiolariti (Cretaceo inf.)	Masse caotiche allineate c. N-S e Sistema di fratture con direzione NE-SW	Masse caotiche all'interno di calcari e stockwork centimetrici in radiolariti	Cinn, py	Argille (nere, gialle), gesso, calcite, silice
Cerreto Piano	1898-1971	0.3	2170	Arenarie (Pliocene)		Disseminazioni stratabound e concrezioni	Cinn	

Miniera	Periodo di sfruttamento	Tenori wt% Hg	Produzione (t) *	Rocce incassanti	Controlli strutturali	Ore body	Minerali	Ganga
Bàgnore	1920-1976	0.2-0.4	1723	Arenarie (età?)		Stringers e sostituzione stratabound in livelli arenacei	Cinn	
Bagni S. Filippo (Pietrineri)	1902-1979	0.7	517	Dolomie (Retico) Calcari (Lias)		Tenore di Cinn aumenta con profondità	Cinn	Argilla, celestina, gesso
Montebuono	1872-1921	0.3-0.5	172	Arenarie (Sopranumm. oppure Eocene?)		Disseminazioni. Breccie imbutiformi verticalmente continue, principalmente al contatto con gli shales	Cinn	Argille, ossidi di Fe
Cortevecchia	1898-1971		155	Calcari, (Sopranumm., Nummul., Sottonum, Eocene?)	Faglie con direzione prevalente N-S	Caotico, sviluppato a contatto con gli shales. Breccie imbutiformi verticalmente continue	Cinn, py	S nativo
Catabbio	1888-1970	0.01-0.03	34	Marne e argilliti (galestri)				
Monte Labbro	1919-1976	0.2-1.1		Calcari (Lias)	Faglie	Sostituzioni e disseminazioni in Diaspri (età?) e Maiolica (età?)	Cinn	

Note: dati compilati da De Ferrari (1890); De Castro (1914); Savoia (1919); Zucchetti (1964); Arisi Rota et al. (1971); Strappa (1977); Forconi (2011) ed integrati con dati raccolti ad Abbazia S. Salvatore. Tutti i tenori sono espressi in Hg wt% del corpo minerario. Da notare che la composizione delle argille nei depositi è indicata soltanto in pochi casi.

*: tonnellaggio calcolato in base alla produzione di bombole (*flasks*) di Hg prodotto, che è data da Forconi (2011). Le produzioni di Siele e Solforate Schwarzenberg sono aggregate a quelle di Solforate e Cornacchino, rispettivamente.

Questo distretto si trova in Toscana meridionale all'interno dell'omonimo sistema geotermico, che oggi ospita 6 impianti per la produzione di energia elettrica. Gli impianti si trovano nelle località di Piancastagnaio e Bagnore e producono 121 MW equivalenti di energia (Bertani, 2010; Egec Geothermal, 2017), che per importanza strategica sono secondi soltanto al sistema geotermico di Larderello.

Nell'arco della loro storia, i 14 giacimenti del M.te Amiata hanno prodotto c. 117 kt di Hg metallico (Segreto, 1991), segnando profondamente la storia economica e sociale del XX secolo della comunità locale. Questa notevole produzione è stata però rilevante anche per l'economia italiana e globale di quel periodo. Questa rilevanza è indicata da un report interno del Servizio Geologico degli USA redatto subito dopo la II guerra mondiale (Meyer & Mitchell, 1947) che, riguardo la produzione amiatina di cinabro, sosteneva che *“Italy was the leading world-producing country again in 1947... the grade of the ore ranges from 0.6 to 3 wt% and the reserves, as compared to those of the American mines, are enormous”*.

Queste valutazioni suggeriscono che il processo geologico che ha generato i giacimenti di Hg del distretto amiatino rappresenta un evento unico, le cui caratteristiche giacimentologiche sono rilevanti per costruire un modello genetico generale delle mineralizzazioni di ambiente epitermale-geotermico. Le caratteristiche peculiari di questo distretto sono la notevole proporzione di mineralizzazione a Hg e la relativa scarsità degli altri elementi in traccia associati (specialmente As, Ga, Ge, Tl, B, In, Au, Ag). Sulla base dei dati storici (Tab 1.1), soltanto l'Sb è infatti da determinare come presente in concentrazioni economiche-subeconomiche nel distretto (De Ferrari, 1890; De Castro, 1914; Arisi Rota et al., 1971) ed in altre aree limitrofe della Toscana meridionale (Dessau, 1952; Dehm et al., 1983). Quest'associazione Hg±Sb è significativa e vincolante per definire un modello genetico del distretto; tuttavia, non è più possibile studiarla in dettaglio con dati di miniera perché, in seguito alla fine delle attività minerarie, tutti i lavori in sottoterraneo dei 14 giacimenti sono stati resi deliberatamente inaccessibili allo scopo di evitare al massimo l'interazione tra attività antropiche e mineralizzazioni. Di conseguenza, ad oggi tutte le mineralizzazioni e le altre manifestazioni geotermiche del distretto si possono studiare soltanto attraverso gli affioramenti superficiali disponibili, i campioni provenienti da collezioni museali o private, ed i dati storici degli archivi minerari.

Dal punto di vista giacimentologico, i depositi del distretto minerario amiatino mostrano le tipiche caratteristiche attese dai giacimenti epigenetici di questo ambiente. Questi giacimenti occupano un'area vasta del sistema geotermico, che si estende per più di 30 Km in direzione N-S (dall'abitato di Bagni S. Filippo a nord fino a Catabbio a sud), e per c. 15 km in direzione E-W (da M.te Labbro ad ovest fino all'asse Bagni San Filippo-Castell'Azzara ed est – Fig. 1.1). In quest'area sono presenti un numero significativo di mineralizzazioni subeconomiche (Ardigò, 1983), venute di gas e sorgenti calde (Minissale et al., 1997, and ref. therein; Frondini et al., 2009; Nisi et al., 2014; Vaselli et al., 2015; Pierotti et al., 2016; Magi et al., 2019). Più a sud di quest'area sono stati documentati giacimenti minori (Cerreto Piano) ed altre manifestazioni geotermiche (e.g., sorgenti termali di Saturnia), che nell'insieme identificano un'area del sistema geotermico che è ancora più vasta di quella indicata dai giacimenti principali di Fig. 1.1.

Nel distretto, le miniere di Abbadia S. Salvatore, Solforate (Siele) e Morone (Selvena) hanno fornito da sole quasi il 90% della produzione storica, ed hanno la caratteristica di essere allineate lungo una longitudine N-S ben identificabile (Fig. 1.1). Lungo questa direzione si trovano anche gran parte delle altre mineralizzazioni e manifestazioni geotermiche dell'area, che nell'insieme formano un'ampia fascia di manifestazioni orientate lungo una direzione N-S. Questa specifica distribuzione è stata descritta in dettaglio nella letteratura storica (De

Ferrari, 1890; De Castro, 1914) ed in quella più recente (Arisi Rota et al., 1971; Tanelli, 1983), ma non sono mai stati chiariti i processi geologici che l'hanno controllata. Manca quindi un modello genetico generale che spieghi una caratteristica fondamentale dei giacimenti cinabreriferi in relazione agli eventi geologici che hanno controllato il sistema geotermico.

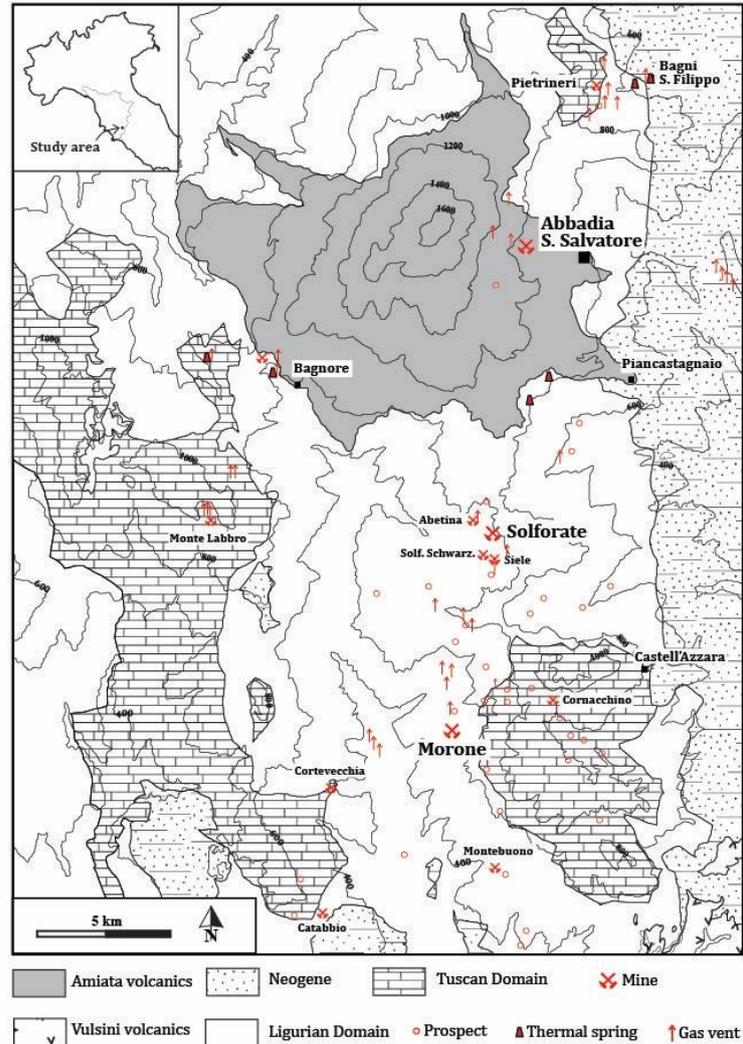


Figura 1.1. Carta geologica del distretto minerario del M.te Amiata e distribuzione delle miniere e *prospect* di cinabro e stibina (modificato da Calamai et al. (1970)). Le tre miniere più grandi sono indicate con simboli più grandi. Le venute di gas e le sorgenti termali sono indicate con simboli distinti.

Esistono soltanto dati indiretti su dimensioni e composizione del plutone che costituisce il motore termico del sistema geotermico (Van Bergen, 1983, 1985; Gianelli et al., 1988; Marinelli et al., 1993). Questi dati sono basati sul sollevamento dei sedimenti pliocenici presenti nell'area, su dati di perforazioni profonde raccolti durante l'esplorazione geotermica, e su dati petrologici relativi a xenoliti metasedimentari tipicamente rinvenuti all'interno delle lave del vulcano. Questi dati suggeriscono che il plutone (possibilmente una sua apofisi) sia presente ad una profondità di 5-7 km al di sotto di M.te Labbro, che le sue dimensioni reali siano di tipo batolitico, che la sua messa in posto sia avvenuta a partire dal Pliocene inferiore (da c. 4 Ma), e che la sua posizione sia asimmetrica rispetto alla distribuzione dei giacimenti e delle manifestazioni geotermiche. Non esiste quindi una corrispondenza diretta tra la localizzazione del plutone in profondità e la distribuzione spaziale N-S delle manifestazioni geotermiche.

Il vulcano M.te Amiata affiora prevalentemente a nord delle manifestazioni geotermiche. È costituito da colate e duomi di composizione trachidacitica/olivina latitica che si sono messe in posto tra 303 e 231 ka (Cadoux & Pinti, 2009; Conticelli et al., 2015). È ritenuto un prodotto di mixing tra un magma acido molto differenziato e parzialmente cristallizzato, uno alcalino alto in K, ed uno mafico sottosaturo, ultrapotassico possibilmente a leucite (Conticelli et al., 2015).

Studi storici delle miniere mostrano alcune caratteristiche giacimentologiche comuni (De Ferrari, 1890; Zucchetti, 1965; Arisi Rota et al., 1971; Branconi, 1976; Ardigò, 1983). Nella miniera di Abbadia San Salvatore la mineralizzazione aveva una forma *stratabound* ed era ospitata nelle calcareniti eoceniche stratificate, alternate con strati di argille sterili del “Sopranummulitico” della Serie Toscana. Questa mineralizzazione consisteva principalmente in disseminazioni, masse irregolari, o reticoli di sottili fratture (*stockwork*) che ricorrevano a basse profondità rispetto al piano locale di campagna (massimo 200 m) ed interessavano l’incassante (che mostrava forte alterazione argillitica) per uno spessore massimo di circa 50 m. Tipologie simili sono state documentate nelle miniere Solforate e Morone ed in quelle minori (e.g., Bagnore, Monte Labbro) all’interno di litologie di diverse composizioni ed età, mostrando che le mineralizzazioni si sono formate senza un controllo litologico specifico.

Nelle miniere il minerale utile era sempre il cinabro, che si trovava in genere associato a marcasite (varietà di FeS_2 bassa T) e stibina (Sb_2S_3). Metacinabro, Hg nativo, realgar ed orpimento erano presenti, anche se non comuni. Il minerale di ganga più comune era la calcite, mentre minori erano le quantità di quarzo, silice amorfa (opale e calcedonio), celestina, fluorite, gesso, e zeolite. La mancanza di datazioni delle mineralizzazioni non permette un’esatta definizione dell’età dei giacimenti, anche se la presenza di cinabro detritico dentro le sabbie plioceniche di Abbadia S. Salvatore, Cerreto Piano e Saturnia (Zucchetti, 1964; Arisi Rota et al., 1971) consente di ipotizzare l’età più antica dell’evento mineralizzante a prima del Pliocene inferiore. Considerato che cinabro ed antimonte precipitano oggi all’interno dei pozzi geotermici del sistema (Morteani et al., 2011), è ipotizzabile che la formazione delle mineralizzazioni a Hg-Sb del sistema epitermale-geotermico sia iniziato nel Pliocene inferiore e continui ancora oggi. Ciò corrisponderebbe ad un’età della circolazione idrotermale di diversi milioni di anni, quindi ben superiore a quella delle vulcaniti.

La letteratura storica sui giacimenti amiatini ha da sempre messo in evidenza lo stretto rapporto tra formazioni di faglie fragili e quella delle mineralizzazioni a cinabro (Tab. 1.1). Lavori recenti (e.g., Brogi et al., 2011) mettono in relazione la formazione dei giacimenti con la genesi di zone di taglio sinistre, in particolare con la formazione di *jog* estensionali e strutture di *pull apart* che si sono formate nel tardo Pleistocene nell’area (giacimento di Morone). La formazione di queste strutture, da sole, non spiega tuttavia la distribuzione nord-sud dell’insieme delle miniere e delle manifestazioni geotermiche del distretto. Quindi, altri fattori geologici in aggiunta alla distribuzione delle strutture fragili devono aver controllato la distribuzione spaziale delle manifestazioni geotermiche. Quali sono quindi i fattori geologici che hanno contribuito in modo decisivo alla distribuzione N-S dei giacimenti di cinabro e delle manifestazioni geotermiche a scala di distretto? In quale modo questi fattori hanno agito in concomitanza ai fattori geochimici che hanno controllato il trasporto e la precipitazione di >117 kt di Hg? È plausibile un confronto tra la storia giacimentologica del distretto e quella dei giacimenti di dimensioni minori presenti in altre zone geotermiche?

Questo lavoro di tesi ha lo scopo di definire possibili correlazioni tra le caratteristiche geologiche-strutturali del sistema geotermico del Monte Amiata e quelle giacimentologiche, allo scopo di vincolare l'insieme dei fattori che hanno determinato la distribuzione spaziale N-S delle sue mineralizzazioni a cinabro (-antimonio) e le manifestazioni geotermiche associate. Lo strumento utilizzato per questa correlazione è il cosiddetto *Machine Learning* (ML), cioè il sottoinsieme dell'Intelligenza Artificiale (AI) che utilizza dati e algoritmi per imitare l'apprendimento umano ed effettuare previsioni mediante l'uso di modelli statistici (Samuel, 1959; Choi et al., 2020).

Oggi, approcci di tipo ML vengono utilizzati in gran parte delle applicazioni scientifiche, dalla robotica ai servizi finanziari, dai trasporti al campo dell'assistenza sanitaria. Tra i diversi campi d'applicazione c'è anche quello delle geoscienze, ed includono in particolare l'esplorazione geotermica, petrolifera e mineraria (Cracknell et al., 2014; Caté et al., 2017; Karpatne et al., 2019; Duplyakin et al., 2020). In ambienti geologici complessi come i sistemi geotermici ed i giacimenti minerali, l'utilizzo del ML permette in linea di principio di individuare relazioni statisticamente significative tra gli elementi geologico-strutturali che caratterizzano l'area di studio - i.e., caratteristiche fisiche e composizionali di rocce e sedimenti, loro assetto strutturale, gradiente geotermico, ecc. - e le variazioni delle proprietà dei fluidi idrotermali che controllano trasporto e deposizione delle *commodities*, i.e., solubilità del cinabro, temperatura del fluido geotermico, ecc.. Queste relazioni potrebbero fornire importanti spunti interpretativi riguardo i principali fattori di controllo delle mineralizzazioni dell'area, e di conseguenza guidare l'esplorazione di potenziali nuove risorse (e.g., mineralizzazioni metallifere, sorgenti calde) ed il monitoraggio del territorio (e.g., definizione di eventuali rischi ambientali). Nel caso specifico del sistema geotermico amiatino, il vantaggio dell'applicazione del ML è il livello di conoscenza geologica elevato delle mineralizzazioni a Hg±Sb e delle manifestazioni geotermiche associate – di cui sono note ubicazione, caratteristiche giacimentologiche, ecc. – accoppiato all'elevato livello di conoscenza geologica dell'area, determinato dal notevole numero di studi geologici e dall'esplorazione geotermica decennale. Questa combinazione suggerisce l'utilizzo di questo sistema geotermico come un caso di studio ideale per l'applicazione del ML.

Sulla base delle premesse elencate sopra, questo lavoro di tesi è stato strutturato in tre parti distinte:

- Costruzione di un database che raccoglie tutti i dati disponibili in letteratura riguardo al distretto cinabrifero amiatino, allo scopo di definire i parametri giacimentologici e i motivi geologici-strutturali dell'area con il più alto livello di dettaglio possibile per la scala di studio;
- Costruzione di un modello tridimensionale dell'area di studio;
- Applicazione di algoritmi di Machine Learning al dataset costruito, al fine di individuare eventuali correlazioni con le principali strutture tettoniche che caratterizzano l'area.

Inquadramento geologico

L'area di studio copre una superficie di circa 287km² e corrisponde alle dimensioni dell'area di studio del lavoro di Cataldi (1967), che comprende al suo interno 15 delle 16 miniere del distretto cinabifero amiatino. L'area è delimitata a nord, est e ovest dai tre principali bacini sedimentari neogenici dell'area, denominati rispettivamente Bacino di Velona, Bacino di Siena-Radicofani e Bacino di Cinigiano-Baccinello. Il confine sud dell'area di studio è costituito dalla valle del Torrente Fiora.

L'area è dominata da una variabilità di strutture e litologie che rendono l'assetto disomogeneo, con la presenza di alti strutturali, rappresentati dalla dorsale di Castell'Azzara-M.te Elmo, la Rupe di Roccalbegna e Cellena, alternati ad aree di basso strutturale, che ospitano le valli dei torrenti Fiora, Paglia, Albegna e Vivo. Sull'area domina il Complesso Vulcanico del Monte Amiata, la cui vetta raggiunge i 1738m s.l.m. ed è costituito da depositi vulcanici pleistocenici che ospitano il più vasto acquifero potabile dell'area che alimenta i tre più importanti torrenti della zona.

2.1 Evoluzione geologica e tettonica del distretto amiatino

2.1.1 Storia tettonica

La zona del Monte Amiata si trova nella parte più meridionale dell'Appennino settentrionale, una catena collisionale di età alpina che si estende dalla Linea Sestri-Voltaggio, che marca il confine con le Alpi Occidentali, alla Linea Olevano-AnTRODoco, che marca il confine con l'Appennino centrale (Marroni et al., 2015).

L'Appennino settentrionale ha avuto origine dalla chiusura dell'Oceano Ligure-Piemontese, uno stretto bacino oceanico di età medio-tardo giurassica situato fra i margini continentali della placca europea, costituiti dal blocco Sardo-Corso, solidale a Europa fino al Miocene, e della microplacca Adria. Questo oceano iniziò a chiudersi nel tardo Cretaceo, come risultato della convergenza tra Europa e Africa (Elter & Pertusati, 1973), ed è stato coinvolto nel prisma di accrezione della subduzione generata per effetto della collisione di crosta oceanica. La successiva collisione continentale, avvenuta tra l'Eocene medio e Miocene superiore, ha portato alla costruzione della catena a falde appenninica. Durante l'Oligocene, la rotazione antioraria del blocco Sardo-Corso ha come conseguenza l'impilamento degli strati rocciosi situati sul margine africano. Tale impilamento è avvenuto grazie allo scollamento e al successivo accavallamento delle rocce sedimentarie presenti ai margini di placca, attraverso livelli meccanicamente deboli (Bosellini, 2005).

La progressiva migrazione del fronte deformativo verso est, in direzione dei domini interni della placca Adria, ha portato all'ispessimento del cuneo orogenico e alla costruzione di una *fold-and-thrust belt*, costituita da una pila di unità tettoniche scollatesi dai domini di Adria e deformate con vergenza E e NE (Barchi et al., 2001). Questa fase compressiva, terminata nel Miocene inferiore, ha portato alla formazione di una pila di nappes orogeniche, costituita da unità tettoniche appartenenti a diversi domini paleogeografici (Bonciani et al., 2005).

Il Dominio Ligure è un dominio oceanico costituito dalla unità tettonica delle Ofioliti e dall'unità tettonica di Santa Fiora, che rispettivamente rappresentano l'oceano Ligure-Piemontese e il dominio di transizione al margine continentale di Adria (Marroni & Pandolfi,

2007). L'unità tettonica delle Ofioliti può essere ricondotta alle Liguridi interne, mentre il Santa Fiora può essere assegnato alle Liguridi esterne orientali (Marroni et al., 2001); manca invece la rappresentazione delle Liguridi esterne occidentali. Il Dominio Subligure è rappresentato dall'unità tettonica Canetolo, costituita da torbiditi carbonatiche depositate nell'Eocene medio al di sopra del margine continentale di Adria. Il Dominio Toscano è rappresentato dalle Falde (Nappe) Toscane, caratterizzate da una successione sedimentaria deposta lungo i margini continentali di Adria dal Triassico al Miocene inferiore.

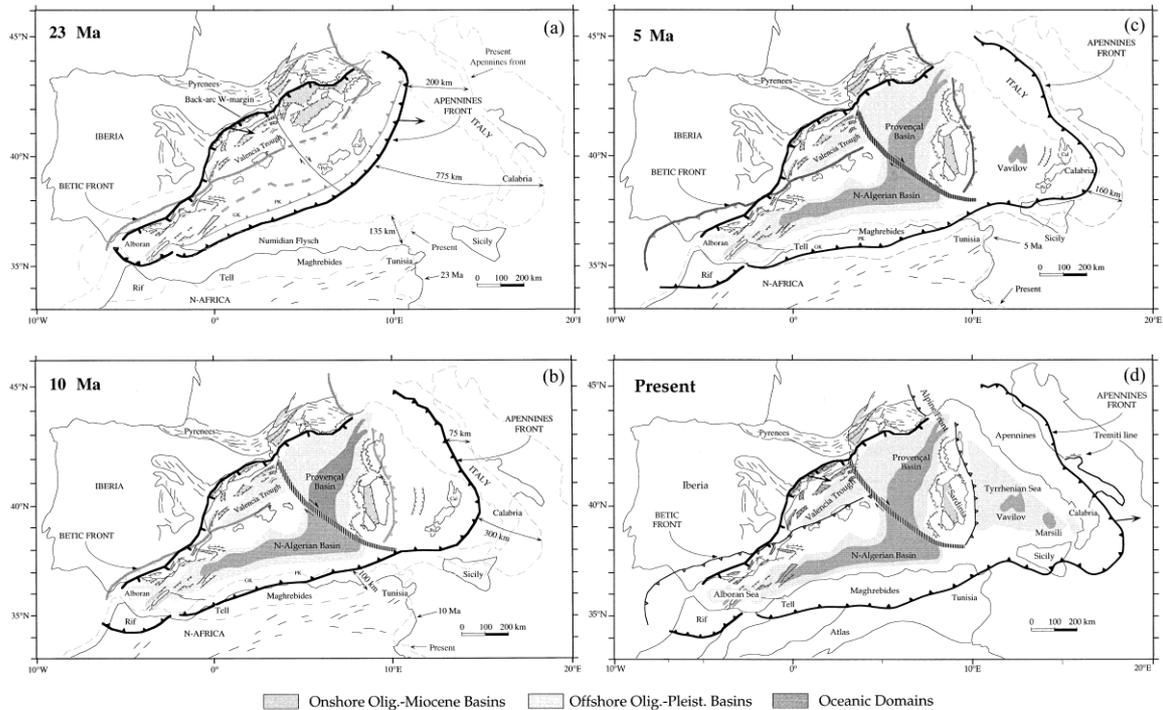


Figura 2.1. Ricostruzione paleogeografica dell'evoluzione tettonica del Mar Mediterraneo occidentale dal Miocene inferiore a oggi (modificato da Gueguen et al. (1998)). (a) Inizio dell'apertura del Bacino Algero-Provenzale nel Miocene inferiore, considerato il bacino di retroarco della subduzione Appenninica. (b) Al Tortoniano, l'apertura di tutti i bacini del Mediterraneo era già quasi completa. (c) Al Pliocene inferiore, la rotazione del Blocco Sardo-Corso è terminata e il Tirreno inizia a sviluppare crosta oceanica. (d) Situazione ad oggi. La subduzione continua nell'area ionica in conseguenza dello *slab-rollback*.

A partire dal Miocene inferiore-medio e fino al Plio-Pleistocene, conseguentemente alla migrazione delle deformazioni verso est e all'apertura del Bacino Algero-Provenzale e del Mar Tirreno, l'area è diventata parte del settore interno dell'Appennino settentrionale, venendo quindi interessata da tettonica estensionale (Fig. 2.1). La tettonica estensionale è legata al collasso dell'inspessita pila tettonica (Brogi, 2008a) e ha generato assottigliamento crostale e sviluppo di bacini sedimentari (Carmignani et al., 1994; Bonciani et al., 2005).

L'evoluzione strutturale e deposizionale del Neogene è stata descritta da Baldi et alii (1994), sulla base di una vasta campagna di rilevamento condotta nell'area delle Colline Metallifere, che ha proposto una suddivisione in tre differenti eventi estensionali. Il primo evento estensionale ha portato alla formazione di strutture con geometria *flat-ramp-flat*, che sono identificabili in faglie normali a basso angolo E-vergenti, che estendono i livelli strutturalmente più alti della catena e hanno portato alla formazione di *megaboudins* (Brogi, 2004b; Bonciani et al., 2005). La forte estensione ha portato a diffuse riduzioni dello spessore delle unità Toscane (Fig.2.2), formando la ben conosciuta "serie ridotta" (Signorini, 1949; Carmignani et al., 1994; Conti & Lazzarotto, 2004). Il secondo evento estensionale, di età compresa tra il Miocene superiore e il Pleistocene, ha interessato la "serie ridotta" e

ha portato alla formazione di faglie listriche e allo sviluppo di strutture *horst* e *graben* a direzione appenninica, con conseguente formazione di bacini sedimentari associati (Martini & Sagri, 1993) riempiti di sedimenti plio-quadernari, delimitati a nord e sud da lineamenti transtensionali antiappenninici. Il terzo evento estensionale, tardo pliocenico, ha portato allo sviluppo di un ulteriore sistema di faglie immergenti verso NE che arrivano ad esaurirsi alla transizione fragile-duttile della crosta profonda (Bonciani et al., 2005), denominata “Orizzonte K” in letteratura (Gianelli et al., 1988).

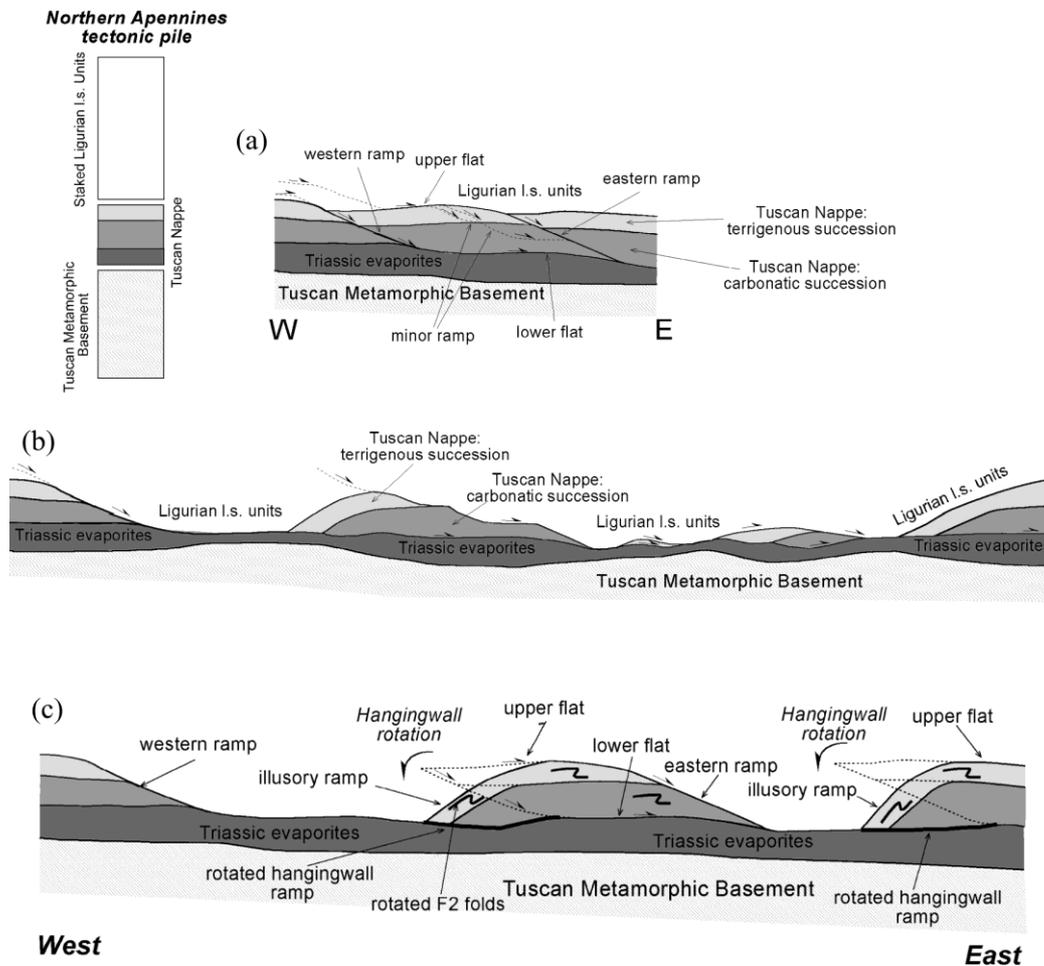


Figura 2.2. Schema di sviluppo della struttura a *megaboudins* prodotta da faglie normali a basso angolo con geometria *flat-ramp-flat*. (a) Stadio iniziale dello sviluppo dei *megaboudins*. (b) Stadio intermedio dello sviluppo di *megaboudins* che pone a contatto le formazioni Liguridi con le evaporiti triassiche del Dominio Toscano (*jog* estensionali). (c) Stadio finale di estensione. I piani di faglia vengono ruotati e ribaltati (modificato da Brogi et al. (2004b)).

Le tre fasi estensionali descritte hanno quindi portato alla formazione di tre diversi orizzonti di scollamento, che sono rappresentati dai livelli meccanicamente più deboli (Bonciani et al., 2005): il primo costituito dai calcari norici della Fm. del Calcere cavernoso, il secondo dalle filladi paleozoiche della Fm. del Verrucano e il terzo che corrisponde all’”Orizzonte K”, a circa 5-6km di profondità.

L'estensione post-collisionale ha generato un importante assottigliamento crostale ed è stata accompagnata da magmatismo che ha portato alla messa in posto di intrusioni magmatiche di derivazione crostale e vulcanismo dal Miocene al Pliocene (Avanzinelli et al., 2009).

Durante la fase post-collisionale si è verificato un importante assottigliamento della crosta continentale fino ad uno spessore di circa 20-22Km e della litosfera fino a circa 30-50km

(Bellani & Gherardi, 2009). Conseguentemente, si è verificata una risalita di materiale astenosferico che ha portato alla messa in posto di un corpo magmatico di morfologia batolitica, il quale ha successivamente alimentato il vulcanismo (Acocella, 2000). La messa in posto di questo corpo magmatico ha comportato un rapido sollevamento crostale (*uplift*) (Dallmeyer & Liotta, 1998) dei sedimenti neogenici. Il sollevamento ha interessato l'area fino al Quaternario, in particolare nella porzione centrale della regione che, nell'area di M.te Labbro, raggiunge i 900 m s.l.m. (Fig. 2.3).

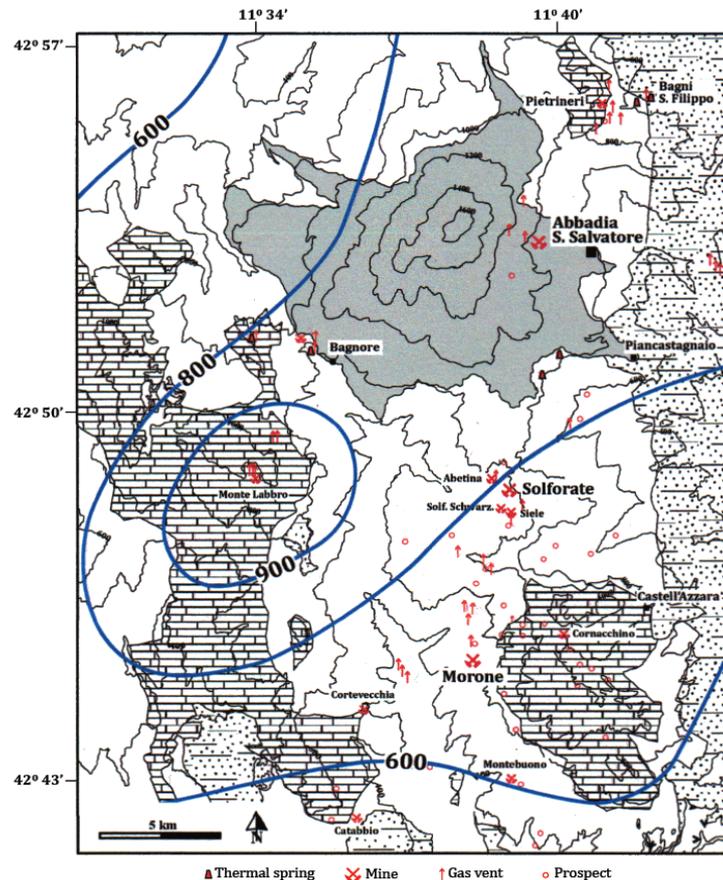


Figura 2.3. Sovrapposizione delle isoipse che marcano il sollevamento post-Pliocene dei sedimenti neogenici (valori in m s.l.m. da Gianelli et al. (1998)) e la carta geologica del distretto minerario del M.te Amiata. Il maggiore tasso di sollevamento, localizzato in corrispondenza di M.te Labbro può essere ricondotto alla presenza di una apofisi del corpo magmatico intruso nella crosta superiore.

Il sollevamento crostale valutato da Acocella et alii (2000) è di circa 2000m, considerando sia il tasso di consolidamento delle argille che la quota iniziale dei sedimenti pliocenici. Si pensa che il corpo magmatico sia di tipo laccolitico, ed intruso nella crosta superiore attraverso due *plug* (Ardigò, 1983; Gianelli et al., 1988; Marinelli et al., 1993). Dalle evidenze degli studi di Gianelli et alii (1988) e Acocella et alii (2000), l'intrusione ha generato anomalie di gravità e di flusso di calore.

Le ultime fasi tettoniche sono state accompagnate da magmatismo, rappresentato da intrusioni granitoidi e vulcanismo effusivo di composizione variabile da riolitica a trachibasaltica (Squarci, 1981). L'attività vulcanica è stata innescata dalla tettonica estensionale plio-quadernaria che ha favorito la messa in posto di corpi magmatici ad una profondità di 6-7km dalla superficie (Gianelli et al., 1988; Cadoux & Pinti, 2009).

L'edificio del Monte Amiata (303-231 ka, Laurenzi et al., 2015) copre un'area di circa 90km² (Barelli et al., 2010) ed ha un volume di circa 14-18km³ (Ferrari et al., 1996). È

costituito dalla sovrapposizione di prodotti di età e composizioni diverse, costituiti da colate e duomi di composizione trachidacitica/olivina latitica.

Durante il Miocene, l'estensione post-orogena ha determinato la segmentazione della pila tettonica preesistente, portando alla formazione, durante il Pliocene, di *transfer-zone* che hanno svincolato e dislocato le strutture preesistenti (Martini & Sagri, 1993; Bonciani et al., 2005; Brogi et al., 2010). L'evidenza di un prevalente controllo strutturale sulla costruzione dell'edificio vulcanico è data dall'interpretazione dell'allineamento dei centri eruttivi in direzione SW-NE (Brogi et al., 2010) (Fig.2.4).

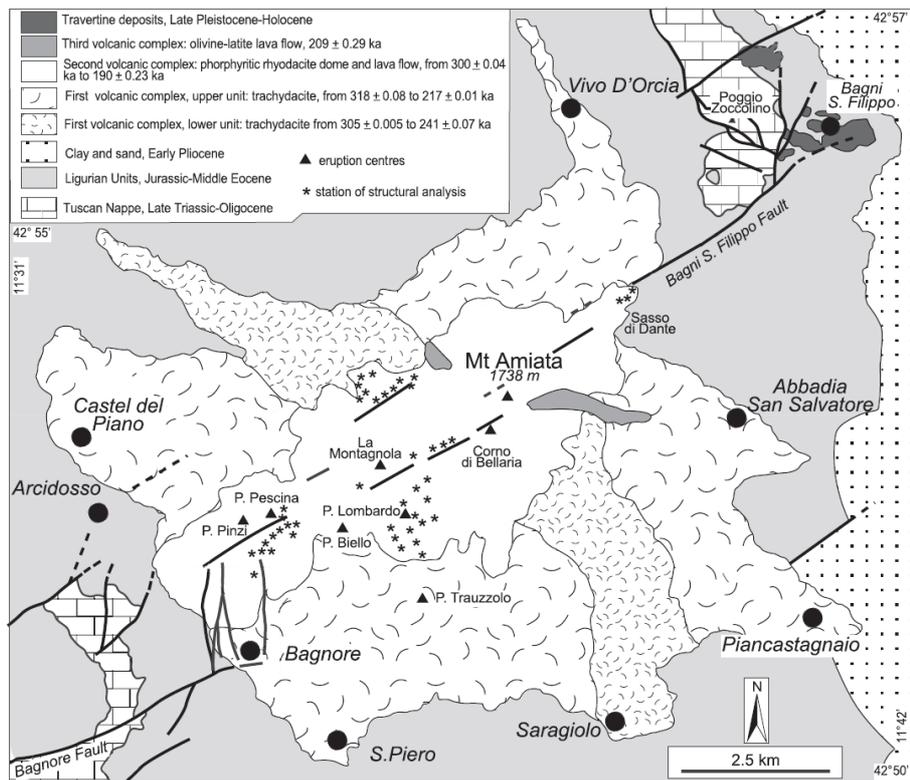


Figura 2.4. Carta geologica schematica del M.te Amiata e delle unità circostanti in cui sono evidenziati i centri eruttivi e i lineamenti strutturali (da Brogi et al. (2010)).

Il lineamento principale che ha comportato la messa in posto dell'edificio vulcanico è espresso dalla presenza di una zona di taglio, identificata nella Faglia di Bagni S. Filippo e nella Faglia di Bagnore, situate rispettivamente ad est e ovest dell'edificio vulcanico. La presenza di strutture vulcano-tettoniche non corrisponde solamente alla zona assiale dell'edificio vulcanico ma comprende un sistema di graben sepolti al di sotto dei depositi vulcanici (Principe et al., 2017). In corrispondenza della terminazione orientale della Faglia di Bagni San Filippo si ha una importante deposizione di depositi idrotermali, costituiti da travertino, che esprimono la manifestazione più superficiale di un deposito *Hot-Spring* (Dessau, 1952; Zucchetti, 1964; Cataldi, 1967).

2.1.2 Assetto strutturale

Le principali strutture tettoniche riconoscibili nell'area di studio sono state analizzate in relazione con la presenza di miniere e manifestazioni geotermiche.

Pieghe

Le principali strutture plicative presenti nell'area del Monte Amiata sono evidenziate da anticlinali e sinclinali che presentano un andamento predominante in direzione N-S (Fig. 2.5).

Lo stile compressivo interessa tutte le falde tettoniche presenti nella successione stratigrafica ma gli effetti sono più evidenti nei terreni del Complesso Alloctono e nelle porzioni più alte della Falda Toscana (Ardigò, 1983).

La formazione delle strutture plicative presenti nell'area può essere attribuita alla principale fase compressiva subita dalla dorsale Appenninica, che ha interessato tutta l'area della Toscana meridionale durante il Miocene. Non è però possibile stabilire con certezza l'età delle strutture compressive.

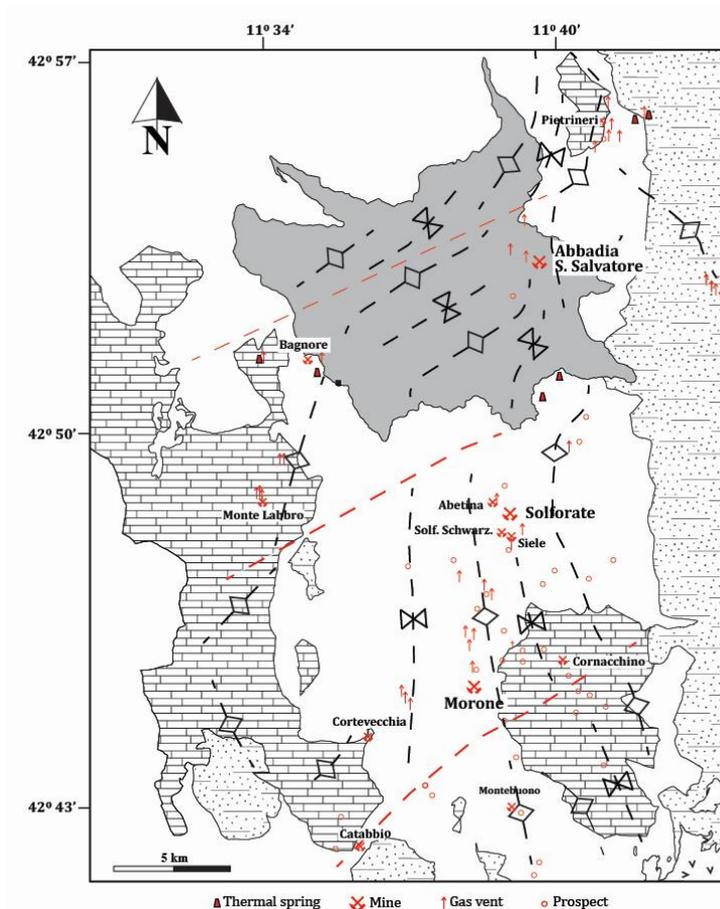


Figura 2.5. Tracce interpretate di antiformali e sinclinali del distretto del M.te Amiata in relazione con depositi, ricerche e venute di gas.

Esistono tre tipologie di meccanismi che controllano le fasi iniziali di crescita di una piega durante l'accorciamento parallelo degli strati: la deformazione longitudinale tangenziale, il flusso flessurale e lo scorrimento flessurale (Fig. 2.6). Il meccanismo che opera è influenzato dalla forza dell'accoppiamento meccanico tra gli strati, dal comportamento meccanico degli strati e dai contrasti di competenza tra gli strati. Ciascuno di questi meccanismi produce un

distintivo *pattern* di distribuzione delle deformazioni intorno agli strati deformati ed è associato a stili di frattura particolari che hanno influenza sulla permeabilità, sulla formazione di vene e localizzazione delle mineralizzazioni durante il piegamento (Cox, 2005).

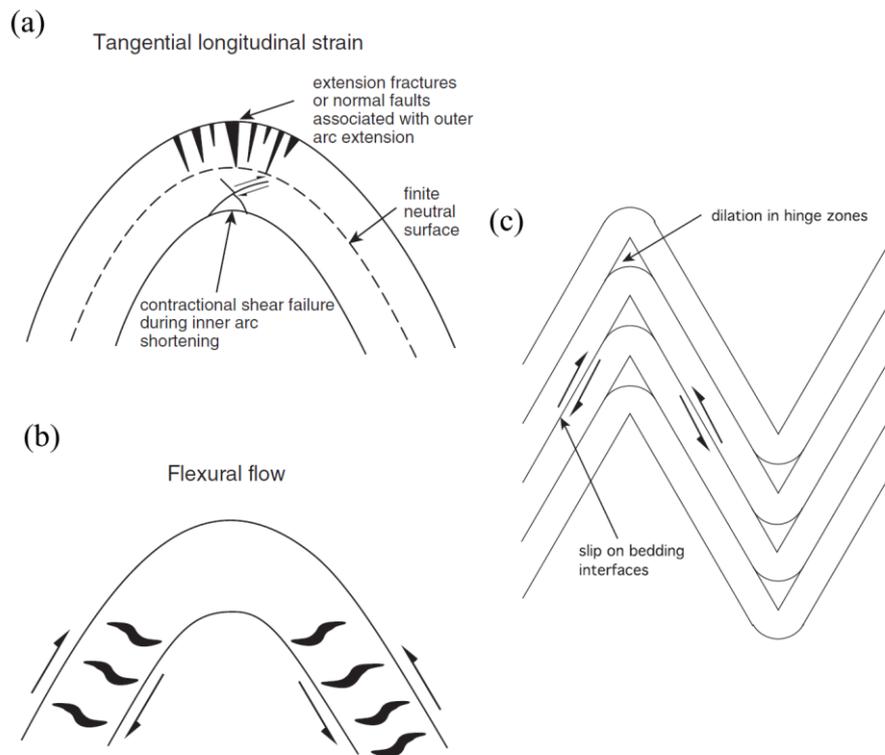


Figura 2.6. Distribuzione di vene e fratture dipendenti dallo stile deformativo che ha influenzato la formazione della deformazione (modificato da Cox (2005)) (a) Distribuzione delle fratture associate alla deformazione longitudinale tangenziale, che genera un aumento di permeabilità nelle aree di intradosso ed estradosso della piega. (b) Sistemi di vene *en echelon* associati al flusso flessurale. (c) Scorrimento flessurale e dilatazione alla cerniera generato in una piega a *chevron* con originale stratificazione sub-parallela.

È necessario tenere conto che l'età delle deformazioni potrebbe non coincidere con l'età del piegamento ma essere antecedente. Sulla base di queste informazioni si può ipotizzare che, a patto che le deformazioni associate al piegamento siano coeve con il piegamento stesso, l'aumento di permeabilità generato da queste strutture permetta un miglioramento del trasporto dei fluidi all'interno del sistema geotermico-epitermale, che in questo modo non sarebbe vincolato solo alle faglie ma anche alla distribuzione delle pieghe e al motivo strutturale.

Faglie

Sulla base degli studi effettuati da Calamai et alii (1970) e da Bonciani et alii (2005), nell'area sono stati riconosciuti almeno quattro sistemi di faglie, distinti sulla base dell'età e dello stile deformativo, che sono:

1. Sistema di faglie inverse (*thrust*) che delimitano le unità della pila tettonica e si sono formate durante le prime fasi collisionali e sincollisionali;
2. Sistema di faglie a basso angolo est-vergenti che comporta la formazione delle strutture a *megaboudins* della "serie ridotta" (evento Langhiano-Serravalliano);
3. Sistema di faglie ad alto angolo che hanno comportato la formazione di strutture tipo *horst-graben* (evento Tortonian-Messiniano);

4. Sistema di faglie ad alto angolo NE-vergenti che rigettano tutte le strutture precedenti (evento Plio-Quaternario).

È presente un sistema di faglie trascorrenti che ha comportato lo svincolo tettonico delle strutture *horst* e *graben*, generato dal movimento relativo dei due lineamenti trascorrenti che delimitano l'Appennino settentrionale, la linea Sestri-Voltaggio a nord e la linea Olevano-Antrodoco a sud. I due lineamenti presentano movimenti relativi in direzione opposta, nel quadro di un movimento generale verso NE.

La comprensione di questi sistemi di faglie permette di capire la complessità della struttura di permeabilità dell'area e del meccanismo di trasporto dei fluidi mineralizzanti.

2.1.3 Assetto stratigrafico

Sulla base di numerosi studi condotti nell'area amiatina e in tutta la Toscana meridionale, si riconosce una pila di unità tettoniche sovrapposte, costruita durante la principale fase compressiva appenninica e dislocata dalla successiva fase estensionale post-collisionale (Bonciani et al., 2005).

L'assetto stratigrafico della regione del M.te Amiata comprende, dal basso verso l'alto:

- Basamento metamorfico;
- Pila di unità tettoniche pre-Neogene, costituite dalle unità tettoniche Toscane, Subliguri e Liguri;
- Depositi sedimentari neogenici;
- Depositi vulcanici pleistocenici del M.te Amiata;
- Depositi di origine idrotermale (travertini).

Complesso Metamorfico Toscano

Costituito da varie formazioni a diverso grado di metamorfismo, comprende il basamento regionale precambriano ed è stato identificato in sondaggi profondi (>1500m) e in xenoliti identificati nelle vulcaniti del Monte Amiata (Van Bergen, 1983; Ferrari et al., 1996; Marroni et al., 2015).

A partire dalle porzioni più profonde si hanno:

- Gneiss (Precambriano)
- Micascisti a plagioclasio e granato (Precambriano)
- Filladi con orizzonti metacarbonatici (Devoniano – Ordoviciano)
- Formazione del Verrucano

Le Filladi con orizzonti metacarbonatici sono caratterizzate da una successione monotona, con raddoppi tettonici, costituita da metarenarie, metapeliti e filladi, con caratteristiche litologico-petrografiche confrontabili con la successione paleozoica che costituisce la Fm. del Farma, affiorante nell'area della Dorsale Monticiano-Roccastrada (Brogi, 2008b).

Unità tettoniche pre-Neogene

Le unità tettoniche di età pre-neogenica comprendono le unità tettoniche del Dominio Toscano e le unità tettoniche del Dominio Subligure e Ligure, definiti rispettivamente Autoctono e Alloctono (Fig. 2.7).

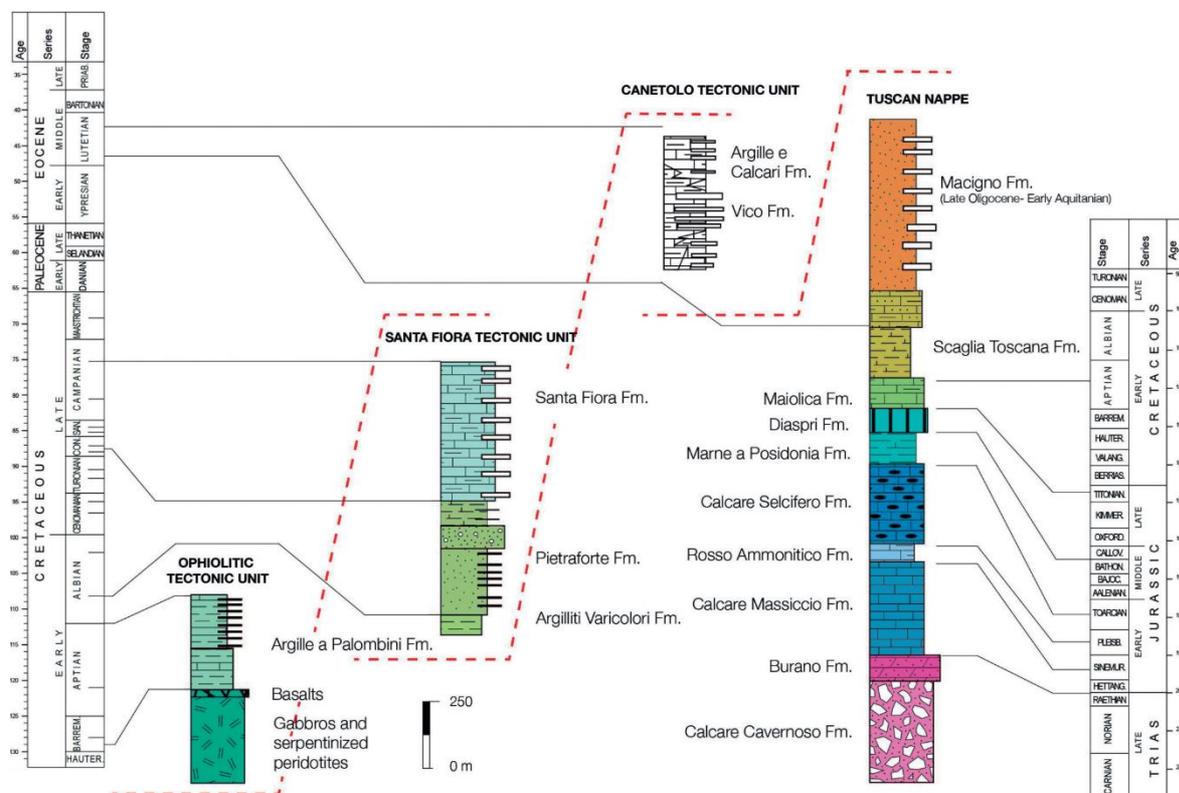


Figura 2.7. Schema cronostratigrafico delle unità tettoniche che costituiscono i Domini Ligure, Subligure e Toscano affioranti nell'area del M.te Amiata (da Marroni et al. (2015))

DOMINIO TOSCANO

Le Nappe Toscane sono la più profonda successione affiorante nella regione amiatina, potente 2500-3000m, costituita da evaporiti (Triassico sup.), carbonati (tardo Triassico – Cretaceo sup.) e torbiditi pelagiche (Cretaceo – Miocene inf.). La successione presenta spessori variabili, dovuti alla sedimentazione all'interno di un bacino sedimentario con substrato irregolare, formatosi durante la fase di *rifting* (Giurassico medio – Triassico) (Brogi, 2004a).

La successione delle Nappe Toscane affiora all'interno dell'area di studio principalmente in due finestre tettoniche: la più grande si estende in direzione NNW-SSE e comprende M.te Aquilaia, M.te Labbro, il centro abitato di Roccalbegna e si estende fino ai rilievi a N di Catabbio; la seconda comprende M.te Civitella, M.te Elmo e il centro abitato di Castellazzara. Altre piccole finestre tettoniche sono situate nella zona di Campiglia d'Orcia, a W di Castel del Piano e a NE di Sorano.

In seguito all'impilamento dovuto alla migrazione verso est delle deformazioni, la successione è scollata dal suo originale substrato e il livello di scollamento è rappresentato dall'unità basale della successione, costituito da anidriti e breccie di età norico-retica, ad oggi quasi completamente trasformate in breccie cataclastiche (Brogi, 2004a).

Stratigraficamente, la successione delle Nappe Toscane è costituita dalle seguenti formazioni, dal basso verso l'alto:

- Calcare Cavernoso
- Anidriti di Burano
- Calcare Massiccio
- Rosso Ammonitico
- Calcare Selcifero
- Marne a Posidonia
- Diaspri
- Maiolica
- Scaglia Toscana
- Macigno

DOMINIO SUBLIGURE

L'unità tettonica Canetolo è una successione sedimentaria di età eocenica, costituita da torbiditi carbonatiche. L'intensa deformazione subita dall'unità tettonica rende il suo spessore difficilmente quantificabile, ma lo spessore apparente è di circa 200-300m.

Come la successione delle Nappe Toscane, anche il Canetolo si trova scollato dal substrato su cui si è originariamente deposto. La stratigrafia dell'unità tettonica Canetolo è costituita da due formazioni torbiditiche coeve, con relazioni eteropiche (Marroni et al., 2015), denominate dal basso verso l'alto:

- Formazione ad argille e calcari
- Formazione di Vico

Secondo Marroni et alii (2015), la relazione eteropica tra le due formazioni è da ricondurre al differente ambiente deposizionale. Dalle evidenze sedimentologiche, probabilmente la Formazione di Vico costituisce un sistema torbiditico formatosi durante la deposizione della Formazione ad argille e calcari.

DOMINIO LIGURE ESTERNO

Le Liguridi esterne sono costituite in quest'area dall'unità tettonica Santa Fiora, una successione torbiditica del Cretaceo superiore (Calamai et al., 1970), scollata dal suo substrato originario.

L'unità tettonica è costituita dalle seguenti formazioni, dal basso verso l'alto:

- Formazione delle Argille Varicolori
- Formazione della Pietraforte
- Formazione di Santa Fiora: relazioni eteropiche con le formazioni sovrastanti

Nella zona della Valle della Senna e sul versante meridionale del M.te Amiata sono inoltre presenti dicchi e *sill* basaltici del Cretaceo, che prendono il nome di selagiti (Marroni et al., 2015).

Lo spessore originario della successione non è facilmente quantificabile, per via dell'intensa deformazione e della distribuzione sparsa degli affioramenti nell'area; tuttavia, si può assumere che la potenza totale non sia inferiore a 1000m.

DOMINIO LIGURE INTERNO

Nell'area amiatina, il Dominio delle Liguridi interne è costituito dai resti della sequenza ofiolitifera e le sue coperture sedimentarie che costituivano la crosta oceanica dell'Oceano Ligure-Piemontese.

Dal basso verso l'altro, la successione comprende:

- Serpentiniti
- Gabbri
- Basalti
- Argille a Palombini

La Formazione delle Argille a Palombini costituisce la copertura sedimentaria delle ofioliti, che si trovano al nucleo delle pieghe isoclinali o in zone di taglio. Lo spessore totale della successione è stimato in 600-700m.

Bacini sedimentari neogenici (Neoautoctono)

I bacini sedimentari formati in seguito alla fase estensionale neogenica sono stati riempiti da sedimenti marini e lacustri, quando l'area del M.te Amiata si trovava in ambiente marino poco profondo e, successivamente, continentale.

Depositi di riempimento, di età Miocene superiore – Pliocene, si trovano all'interno dei tre bacini sedimentari presenti nell'area amiatina. Il Bacino di Cinigiano-Baccinello e il Bacino di Velona, situati rispettivamente a ovest e a nord del M.te Amiata, si trovano al di fuori dell'area di studio, mentre il Bacino di Radicofani comprende tutto il margine est dell'area di studio (Fig. 1.1).

Il Bacino di Radicofani è riempito da una potente successione di depositi, con spessore di circa 2000m, costituita da sedimenti alluvionali-lacustri e marini (Calamai et al., 1970).

Grazie ad una estesa campagna di rilevamento e di interpretazione di sondaggi e sezioni sismiche, la successione sedimentaria è stata descritta come costituita da *mudstone* marini e di ambiente costiero poco profondo, arenarie e conglomerati che presentano relazioni eteropiche (Bonini & Sani, 2002). I depositi affioranti e sepolti del Bacino di Radicofani possono essere suddivisi in termini di unità stratigrafiche a limiti inconformi (UBSU), poiché presentano importanti discordanze angolari.

Depositi vulcanici plio-pleistocenici

All'interno dell'area di studio si trovano due differenti distretti vulcanici (Fig. 1.1). Le vulcaniti situate al margine SE dell'area di studio appartengono al Complesso Bolsena-Latera (Monti Vulsini) e si trovano al contatto con le argille plioceniche. Le vulcaniti sono costituite da basaniti, tefriti leucitiche, tufi e pomice (Peccerillo, 2005).

L'attività vulcanica è iniziata con i Monti Vulsini nel Pliocene Inferiore ed è terminata con il Monte Amiata nel Quaternario (Ferrari et al., 1996). Il complesso vulcanico del Monte Amiata (Fig. 2.8) ha una elevazione massima di 1738 m s.l.m. e si estende su un'area totale di 80 km² (Barelli et al., 2010). È costituito da duomi e colate di composizione trachidacitica, che caratterizzano la prima fase di messa in posto dell'edificio, e da duomi e lave di composizione olivin-latitica, che costituiscono la fase finale di costruzione dell'edificio (Principe et al., 2017).

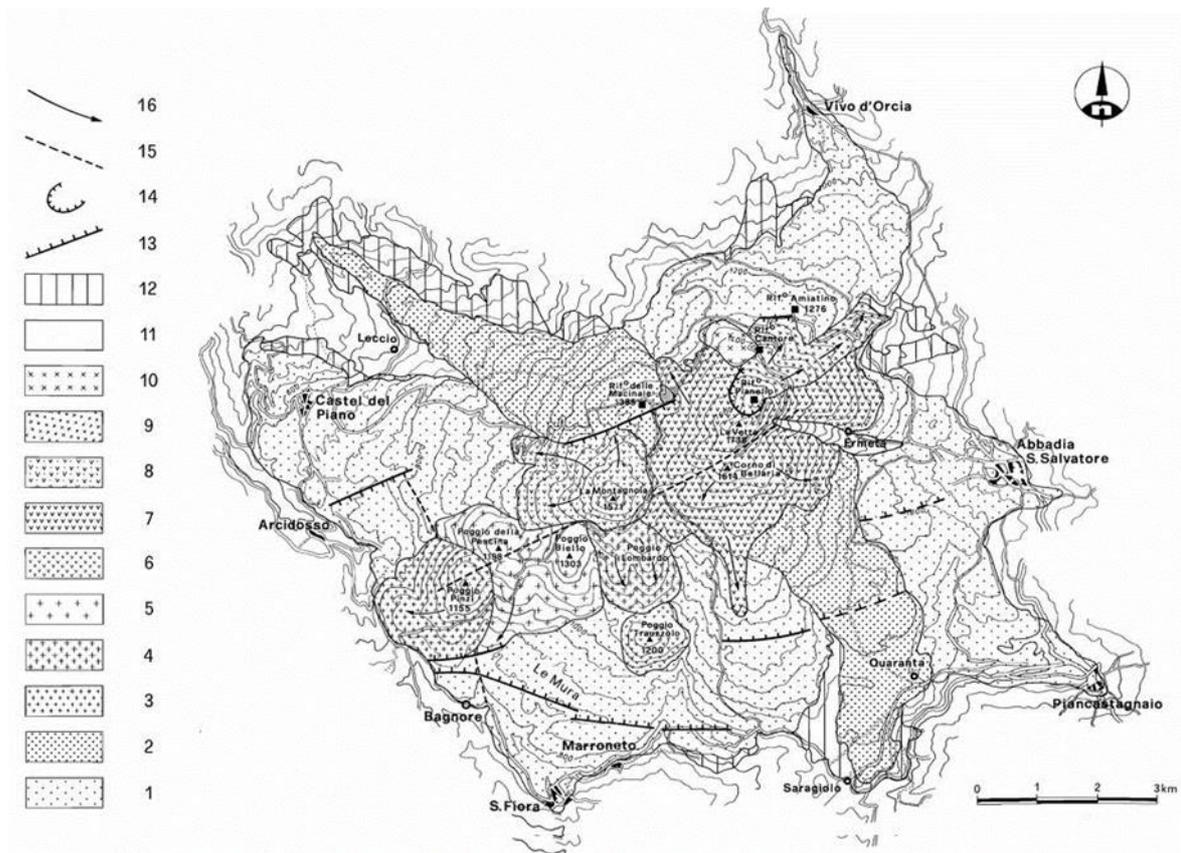


Figura 2.8 Carta geologica alla scala 1:50.000 del Monte Amiata (Ferrari et al. (1996)). 1) Complesso Trachidacitico Basale, unità inferiore; 2) Complesso Trachidacitico Basale, unità superiore; 3) Complesso dei Duomi e delle Lave, unità di Poggio Pinzi; 4) Complesso dei Duomi e delle Lave, unità di Poggio Lombardo; 5) Complesso dei Duomi e delle Lave, unità di Poggio Biello; 6) Complesso dei Duomi e delle Lave, unità di Poggio Trauzzolo; 7) Complesso dei Duomi e delle Lave, unità della Vetta; 8) Complesso dei Duomi e delle Lave, unità della Montagnola; 9) Complesso dei Duomi e delle Lave, unità del Pianello; 10) Complesso dei Duomi e delle Lave, unità di Cantore; 11) Lave Olivin-lattitiche; 12) copertura vulcanica detritica; 13) faglie; 14) aree collassate; 15) faglie presunte; 16) direzioni di flusso.

L'intero complesso vulcanico ospita il più grande e studiato corpo idrico della Toscana Meridionale che prende il nome di Acquifero dell'Amiata. Il serbatoio idrico è caratterizzato da una permeabilità relativamente elevata, dovuta all'intensa fratturazione delle vulcaniti quaternarie (Regione Toscana, 2009; Dini et al., 2010). Il substrato impermeabile dell'acquifero è costituito dalle Unità Liguri.

Il deflusso dell'acquifero è rappresentato da circa 150 sorgenti, situate in corrispondenza dei contatti tra le vulcaniti e le circostanti unità impermeabili. Molte di queste sorgenti sono captate per la distribuzione di acqua nella rete idropotabile (Regione Toscana, 2009).

Nella porzione più a SE dell'area di studio è presente un altro importante acquifero, denominato Corpo Idrico delle Vulcaniti di Pitigliano, costituito da una sequenza ignimbratica prodotta dall'attività magmatica quaternaria del Complesso Bolsena-Latera (Monti Vulsini) (Baldi et al., 1974). Rispetto all'acquifero del Monte Amiata, il corpo idrico di Pitigliano presenta una permeabilità regolata prevalentemente dalla porosità, più che dalla fratturazione (Regione Toscana, 2011).

Travertini

Nelle aree di Bagni San Filippo e Semproniano sono presenti depositi di travertino.

I travertini sono calcari che si formano dove acque calde di falda, ricche di calcio e bicarbonato, sgorgano in superficie (Guo & Riding, 1998). Il degassamento dell'anidride carbonica provoca una rapida precipitazione. La precipitazione rapida e variabile, nonché le marcate variazioni laterali e verticali di facies sono caratteristiche di questi depositi.

Nell'area di Bagni San Filippo il processo di precipitazione del travertino è attivo a tutt'oggi, mentre il deposito nei pressi del centro abitato di Semproniano è di età pleistocenica e non ha evidenze di neoformazione.

Complessi di permeabilità

Uno degli obiettivi di questo lavoro di tesi è quello di costruire un modello tridimensionale dell'area di studio; pertanto, al fine di produrre il modello tridimensionale di un'area di studio così vasta, si è ritenuto opportuno semplificare le formazioni geologiche presenti nella regione amiatina e raggrupparle, seguendo il modello proposto da Cataldi (1967), in quattro complessi sulla base delle loro caratteristiche geoidrologiche.

I complessi, che prendono il nome di complessi di permeabilità (Fig. 2.9), vengono identificati dall'alto verso il basso come:

- **Complesso I: permeabile.** Comprende le vulcaniti del Complesso Vulcanico del Monte Amiata (ignimbriti quarzo-latitiche, reoignimbriti e lave), le vulcaniti del Complesso Vulcanico dei Monti Vulsini (flussi piroclastici appartenenti alla Formazione di Sorano) e i depositi idrotermali che affiorano nell'area di Bagni San Filippo e Semproniano;
- **Complesso II: impermeabile.** Comprende i depositi dell'Alloctono (Ligure e Subligure), i sovrastanti depositi clastici neogenici e i depositi alluvionali quaternari. I depositi clastici del Neogene, costituiti da sabbie e ghiaie permeabili, sono stati inclusi in questo complesso poiché si trovano al contatto con il *flysch* impermeabile.
- **Complesso III: generalmente impermeabile.** Comprende le formazioni geologiche appartenenti alle Nappe Toscane, con un'età compresa tra Cretaceo ed Eocene. Affiora nella parte occidentale dell'area di studio con andamento NNW-SSE, nella porzione a SE e a NE nell'area di Bagni San Filippo;
- **Complesso IV: permeabile.** Include le formazioni geologiche di età compresa tra il Norico-Retico e il Titoniano. Affiora a NE nell'area di Poggio Zoccolino, a SE nella zona di Monte Rotondo e a W in piccoli affioramenti lungo la parte occidentale del Complesso III.

La scelta del modello proposto da Cataldi (1967) è fondata sulla necessità di utilizzare un modello di permeabilità del sistema geotermico-epitermale, basato su dati esplorativi testati e stabiliti, che riduce al minimo l'introduzione di errori prodotti da interpretazioni o stime empiriche.

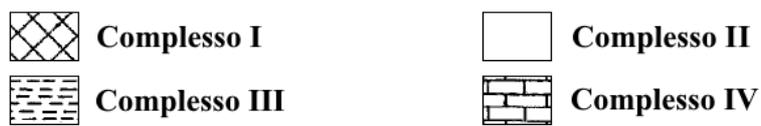
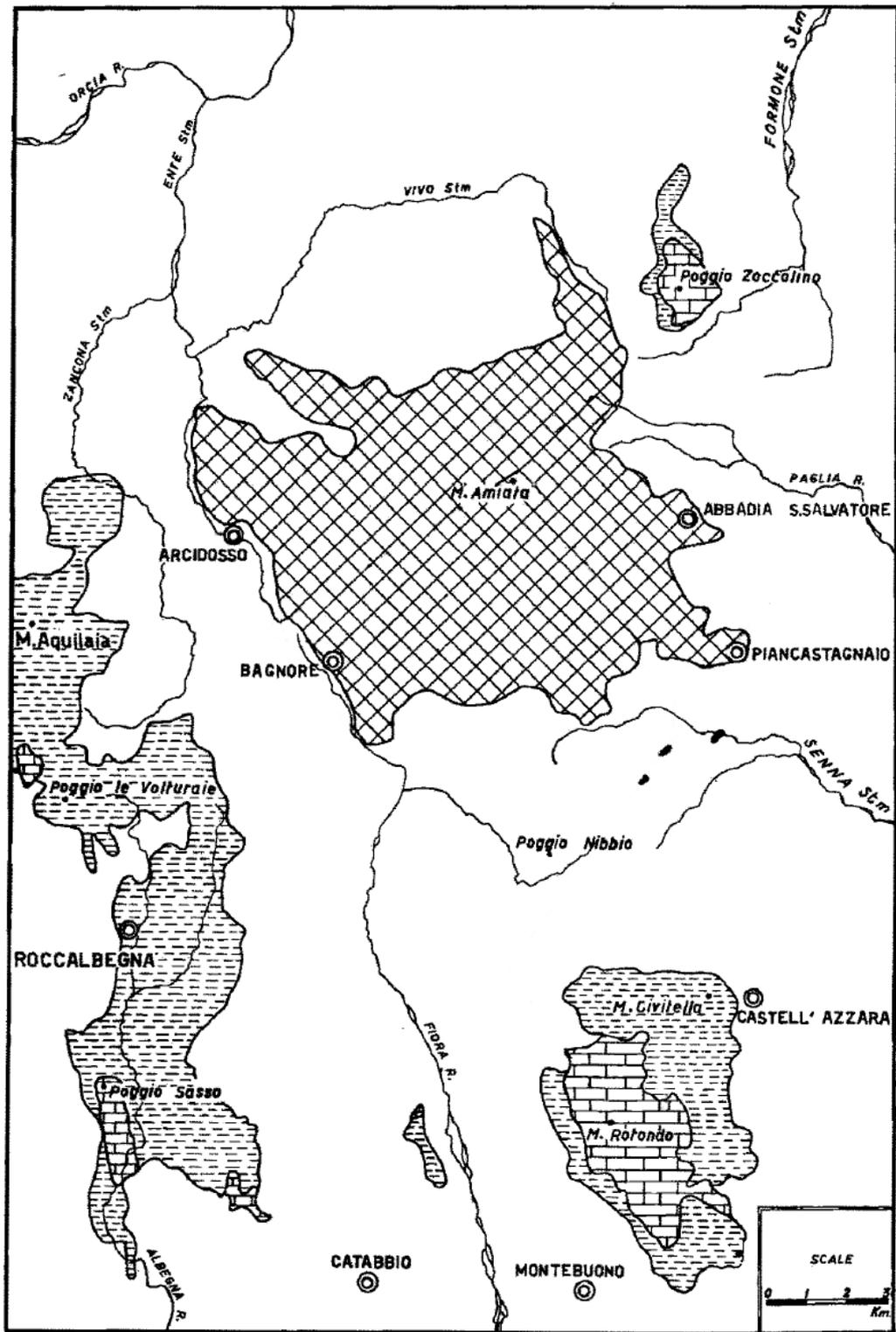


Figura 2.9. Affioramenti dei complessi di permeabilità definiti da Cataldi (1967).

2.2 Risorse minerarie

Per decenni, lo sfruttamento delle risorse minerarie presenti nell'area del Monte Amiata ha permesso la crescita di una fiorente attività economica. I principali minerali coltivati nella zona sono il cinabro e l'antimonite, o stibina, ma si hanno notizie di coltivazione di altri minerali quali siderite, marcasite, realgar, orpimento, dawsonite e zolfo (Forconi, 2011).

Il cinabro (HgS) è il più importante ed abbondante minerale industriale di mercurio. È un solfuro appartenente al sistema trigonale caratterizzato da un legame covalente. I cristalli sono molto rari e generalmente si trova sotto forma di masse microcristalline o terrose di colore scarlatto in vene, incrostazioni o impregnazioni (Mottana et al., 1998).

I principali depositi metalliferi della Toscana meridionale, che rappresentano i maggiori depositi di Hg e Sb, sono associati a due unità geologiche caratteristiche, la prima costituita da metasedimenti silicoclastici paleozoici associati a rocce evaporitiche, la seconda costituita dalle rocce magmatiche della Provincia Magmatica Toscana, nonché ad un controllo strutturale dovuto alle strutture tettoniche estensionali tardo appenniniche. Inoltre, il distretto minerario del Monte Amiata presenta una evidente associazione spaziale e genetica con l'omonimo sistema vulcano-plutonico (Cadoux & Pinti, 2009).

Alcuni autori (Klemm & Neumann, 1984; Brogi et al., 2011) hanno collegato tra loro i depositi di Hg±Sb della Toscana meridionale, implicando un comune processo minerogenetico ed una sorgente di fluido. Tuttavia, Hg e Sb si trovano raramente insieme in quantità economiche (Dessau et al., 1972), sebbene ciascuno sia presente come anomalie geochemiche nei depositi dell'altro metallo. I depositi di Sb contengono spesso Au accessorio (Lattanzi, 1999), che non è invece riportato nei depositi di Hg. Infine, i depositi di Sb mostrano un'associazione privilegiata con il contatto tettonico tra le formazioni triassiche toscane (Calcere cavernoso) e il sovrastante *flysch* ligure (Lattanzi, 1999) mentre i depositi di Hg sono ospitati da una varietà di rocce di età compresa tra Triassico e Quaternario, appartenenti sia alle formazioni Toscane e Liguri che ai sedimenti post-orogenici del Neogene e alle giovani rocce vulcaniche del vulcano del M.te Amiata (Rimondi et al., 2015).

Nell'area amiatina sono presenti 14 miniere, in attività fino al 1982 (Fig. 2.10). Si hanno notizie di coltivazione nell'area a partire dall'epoca etrusca, grazie a ritrovamenti di antichi cunicoli e reperti alla ripresa dei lavori di coltivazione ad inizio Ottocento. Durante la storia estrattiva, nel bacino mercurifero del Monte Amiata sono state estratte 3.101.500 bombole (o *flask*¹), corrispondenti a circa 117 kt di mercurio, per un valore (calcolato al prezzo di mercato del gennaio 2010) pari a 1.5 miliardi di dollari americani (Forconi, 2011).

La mineralizzazione a mercurio si trova sotto forma di vene, impregnazioni diffuse o sostituzioni massicce di notevoli dimensioni nella roccia incassante (Tab. 1.1). Le paragenesi sono generalmente semplici e tipiche dei giacimenti prevalentemente a Hg: il cinabro (HgS), il metacinabro e il Hg nativo sono abbastanza rari. Tra i minerali accessori si trovano solfuri, con pirite e/o marcasite sempre presenti e localmente abbondanti; quantità minori di solfuri di As (realgar e orpimento) sono abbastanza comuni; la antimonite è scarsa (ad esempio, a Morone, dove appartiene a un evento mineralizzante precedente al cinabro); la calcopirite è segnalata alle miniere di Bagnore e Solforate; i solfuri di Ni (millerite e vaesite) sono segnalati nelle miniere di Bagnore e Siele (Dini, 2017).

¹ Una bombola, o *flask*, corrisponde a 34.5kg.

La ganga è tipicamente calcite; il quarzo è relativamente scarso e questo segna un'altra differenza con i depositi di Sb(Au), che sono sempre associati ai diaspri (Rimondi et al., 2015).

Basandosi sulla documentazione storica si possono dividere le tipologie principali di corpi minerali sulla base della morfologia del corpo stesso:

- *Fossoni*, ammassi plastici di minerali argillosi a grana fine contenenti disseminazioni di cristalli di cinabro con altissimi tenori. Sono riconducibili a depositi di sostituzione tipo *mantos* (Principe et al., 2017);
- Caotici, costituiti da *camini* imbutiformi subverticali ad alto tenore, a pianta sub-circolare che raggiungono profondità considerevoli, riempiti da breccie a composizione trachitica e argille. Sono tentativamente interpretabili come *hydrothermal breccia pipes*.
- Caotici di contatto sub-orizzontali con tenore medio-basso, presenti al contatto con le rocce di copertura. Sono denominati *conformable hydrothermal breccia bodies*.
- Disseminazioni a basso tenore, identificabili come vene e *stockwork*.

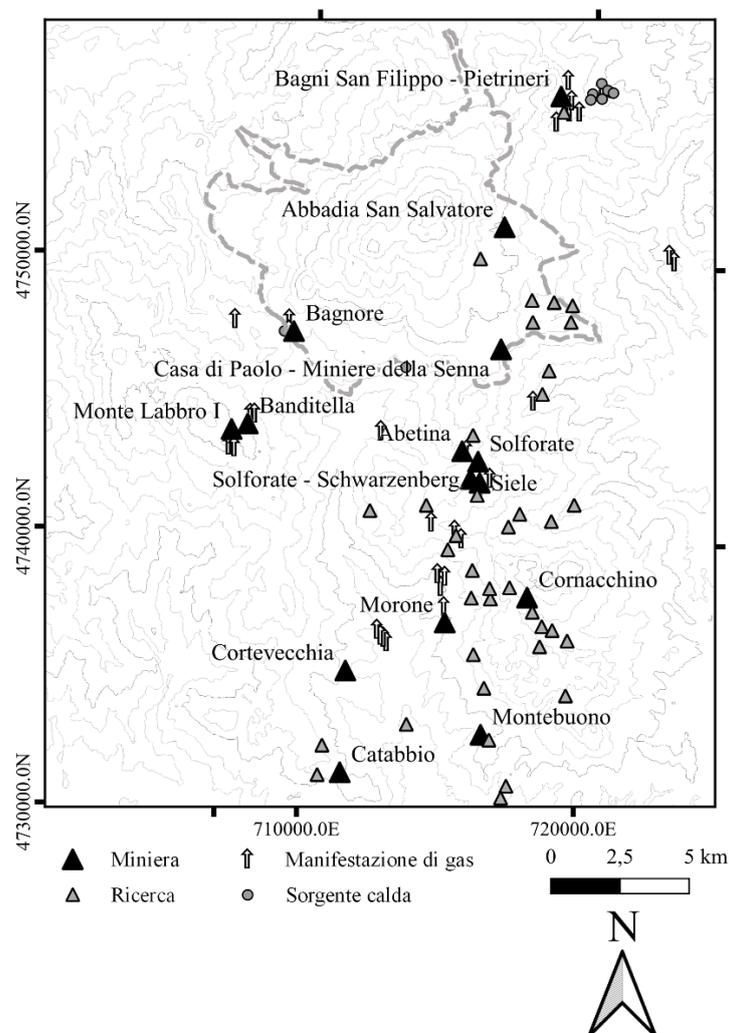


Figura 2.10. Posizione delle miniere del distretto cinabifero amiatino, delle ricerche, delle venute di gas e delle sorgenti clade all'interno dell'area di studio. La miniera di Cerreto Piano di trova a SW dell'area di studio. Il numero totale di miniere posizionate è di 15 in quanto la miniera Casa di Paolo – Miniere della Senna è stata sfruttata per un breve periodo e non sono conservate documentazioni significative, mentre la miniera Banditella costituisce una continuazione della miniera di Monte Labbro.

Evidenze di un controllo strutturale che influenza la distribuzione delle miniere sono presenti sia alla scala del distretto minerario che a quella della singola miniera. In particolare, nella miniera di Abbadia San Salvatore, che costituisce il giacimento più importante del distretto, sono identificabili quattro zone che presentano giaciture differenti (Fig. 2.10), definite in base alla tipologia di corpo minerario.

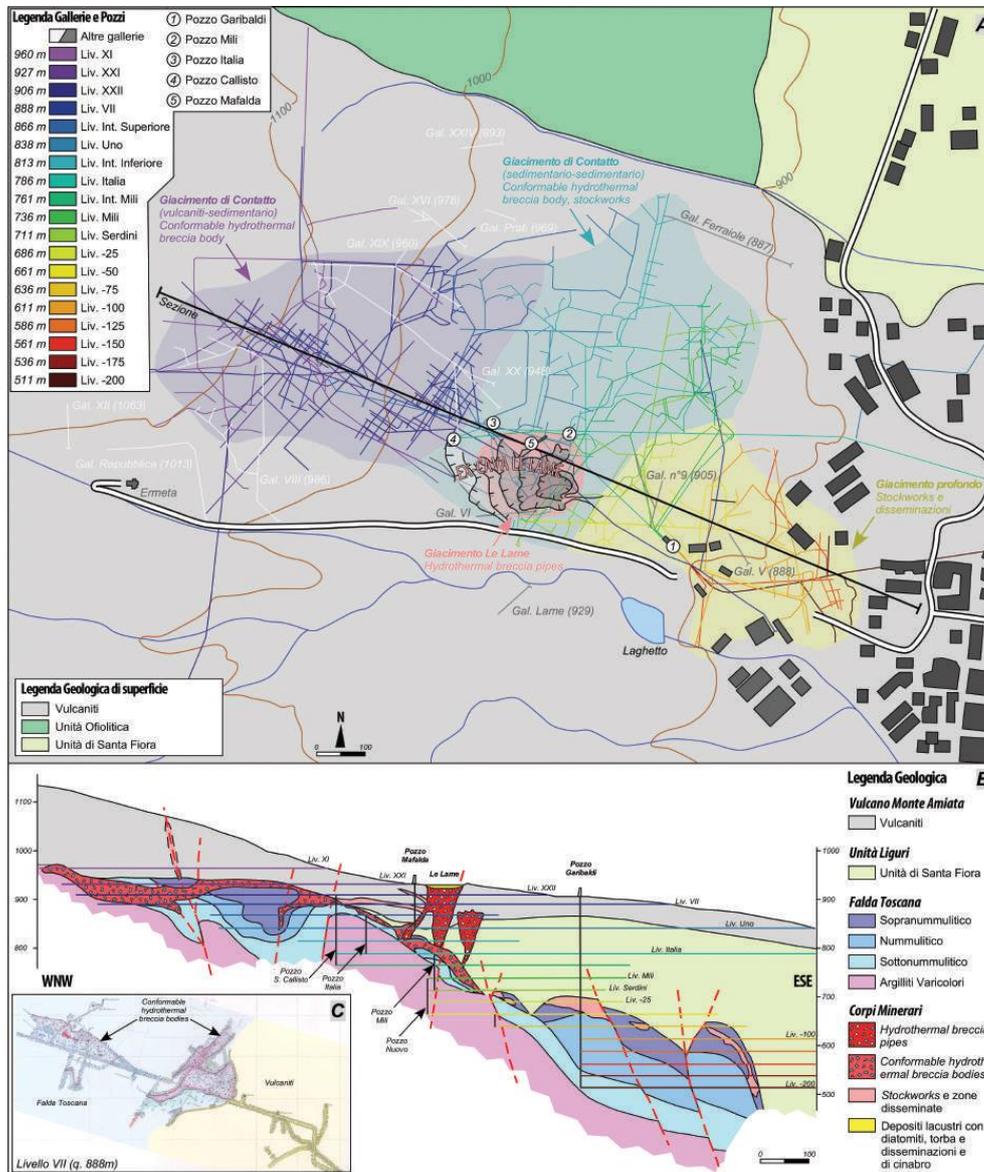


Figura 2.11. (a) Carta geologica superficiale della miniera di Abbadia San Salvatore. Sono identificabili quattro tipologie di corpi minerari, sulla base delle distribuzioni delle mineralizzazioni. (b) Proiezione di miniera e sezione interpretativa dei corpi minerari (da Principe et al. (2017)). (c) Ingrandimento della zona di contatto tra le vulcaniti e le rocce della Falda Toscana.

Dalla sezione interpretativa di Fig. 2.11(b) si nota come la presenza di strutture deformative, quali faglie e pieghe antiformi, influenzi le giaciture dei corpi minerari, in relazione con i contrasti di permeabilità che si verificano al contatto dei diversi Complessi di permeabilità (si veda Fig. 2. 9).

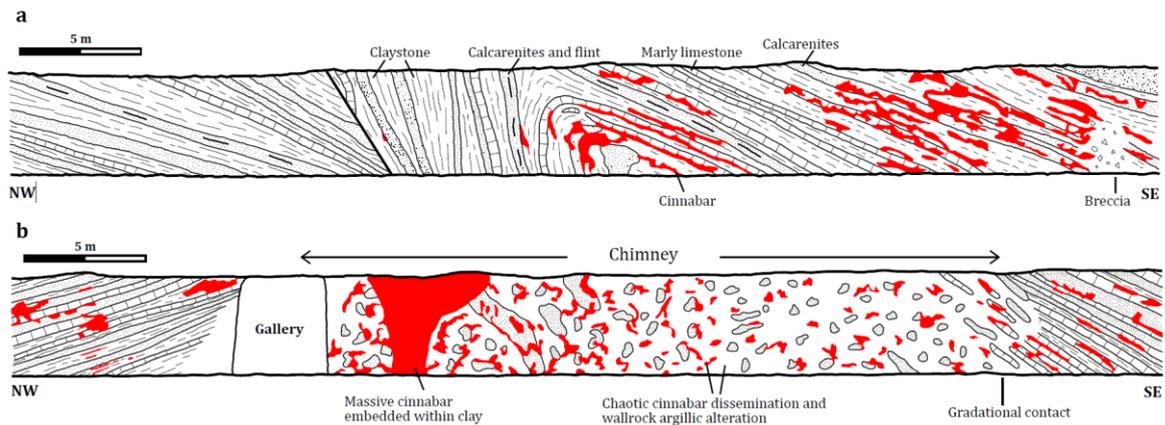


Figura 2.12. Morfologia e controllo strutturale locale dei giacimenti cinabreriferi della miniera di Abbadia San Salvatore (modificato da Arisi Rota et al. (1971)). (a) Breccia e sostituzioni di cinabro in argilliti, marne calcaree e calcareniti in corrispondenza della cerniera di una antiforme NW-vergente. I livelli mineralizzati sono sub paralleli e seguono la stratificazione. (b) Caratteristiche tipiche della mineralizzazione a camino, verticalmente continua. All'interno di una singola galleria, il cinabro si trova sia caotico e massivo nelle stratificazioni delle argille, sia disseminato all'interno delle rocce incassanti intensamente alterate. All'interno del camino, la stratificazione originaria dell'incassante è completamente mascherata dall'alterazione argillitica. I contatti tra camino e incassante sono graduali.

Non sono disponibili studi in letteratura che mettono in relazione la presenza di mineralizzazioni e manifestazioni geotermiche con il motivo strutturale dell'area; tuttavia, è nota da tempo la relazione con le strutture di tipo fragile create durante la fase post-collisionale. Secondo Brogi et alii (2011), la localizzazione dei corpi minerari può essere correlata a faglie estensionali tardo-orogeniche. Dalle evidenze degli studi effettuati all'interno delle miniere amiatine, in particolare quella di Abbadia San Salvatore, si nota come le mineralizzazioni non risultino vincolate solamente all'andamento dei piani di faglia ma anche alle strutture associate al piegamento (Fig. 2.12a). Le giaciture dei corpi minerari sembrano essere correlabili con i modelli di deformazioni associate al piegamento proposti da Cox (2005).

Come descritto in Tab. 1.1, i depositi di Hg amiatini si trovano solitamente in diverse rocce incassanti (calcari, marne, calcareniti e arenarie), aventi in comune una relativa permeabilità; i corpi minerari si trovano tipicamente al contatto con le formazioni sovrastanti, meno permeabili (Fig. 2.12b).

Diversi autori (Klemm & Neumann, 1984; Morteani et al., 2011) concordano sul fatto che la mineralizzazione di Hg al Monte Amiata sia correlata a sistemi idrotermali poco profondi simili all'attuale sistema geotermico della regione. L'unica parziale eccezione può essere rappresentata da alcuni depositi, come quello di Cerreto Piano, presente all'interno di arenarie plioceniche, che si trova circa 14km a SW al di fuori dell'area di studio, per i quali alcuni autori tra cui Zucchetti (1964) hanno ipotizzato che si tratti di depositi secondari alluvionali o residuali.

La genesi di questi depositi può essere ricondotta a processi di mobilizzazione di Hg e Sb e deposizione attivi ancora oggi (Morteani et al., 2011).

Fluidi mineralizzanti

Le strutture prodotte dalla tettonica estensionale post-orogenica sono state sfruttate come condotti per il trasporto di fluidi mineralizzanti (Anderson & Fairley, 2008; Brogi et al., 2011). La mineralizzazione si trova spesso dove una formazione rocciosa relativamente impermeabile bloccava i fluidi in una formazione sottostante più permeabile, provocando il

fenomeno detto *impounding*, con il conseguente ristagno dei fluidi mineralizzanti e impregnazione delle rocce permeabili.

I fluidi mineralizzanti dei sistemi idrotermali sono ritenuti di bassa temperatura e sono considerati di origine meteorica. Tuttavia, dati specifici su questi fluidi sono disponibili solo in studi di calcite non associata a cinabro e formata in vene esterne ai corpi minerari. Lo studio delle inclusioni fluide di alcuni campioni provenienti dalla miniera delle Solforate, effettuato da Brogi e Fabbrini (2009), ha mostrato temperature di omogeneizzazione nell'intervallo 80°-130°C e salinità apparenti di 1,1-1,8% wt. di NaCl equivalente. Questo intervallo di temperature suggerirebbe un'ambiente di deposizione superficiale e di termalità molto bassa del cinabro.

Uno studio effettuato da Gasparrini et alii (2013) su campioni provenienti da vene di calcite della miniera delle Solforate ha evidenziato temperature di omogeneizzazione che variano tra 70° e 120°C, con salinità di circa 5% wt. NaCl equivalente, mostrando inoltre una correlazione di queste vene e la mineralizzazione a cinabro con il fluido mineralizzante. Gasparrini et alii (2013) hanno stimato per i fluidi un $\delta^{18}\text{O}$ compreso tra -5 e -8‰ (SMOW), che indica un contributo meteorico dominante.

In conclusione, i dati sulle inclusioni fluide limitate ai depositi di Hg mostrano temperature di omogeneizzazione e valori di salinità complessivamente inferiori (0,2-6,1 wt% NaCl equivalente; Lattanzi, 1999) a quelli riportati per i depositi di Sb(Au)132-245°C; Lattanzi, 1999).

Origine di Hg

Ad oggi, la comunità scientifica non è concorde su quale sia l'esatta provenienza di Hg in Amiata. Klemm e Neumann (1984) sostennero che la sola separazione di Hg dall'intrusione magmatica in *degassing* non possa costituire una fonte per tutto il Hg presente e proposero un modello basato su una sorgente alternativa di Hg costituita dalle formazioni silicoclastiche del basamento paleozoico toscano. Queste rocce sono scisti neri ricchi di sostanze organiche, che rappresentano i migliori candidati come rocce fonte per grandi depositi di Hg. Varekamp e Buseck (1984) sostengono che l'abbondanza di materia organica possa mantenere un sistema fortemente riducente, promuovendo l'abbondanza di Hg^0_{aq} e la capacità del sistema di trasportare Hg.

Di contro, Lavric e Spangenberg (2003), sulla base degli studi effettuati nel giacimento di Idrija, sostengono che per mobilitare tutto il mercurio contenuto in un giacimento di tali dimensioni sarebbe necessaria una grande cella convettiva, di più di 100km³, per cui concordano con l'ipotesi più realistica di una fonte posizionata in profondità.

Una "connessione con il mantello" è stata invece proposta da Borisenko et alii (2005) sulla base di analisi isotopiche di campioni provenienti da depositi di Hg di tutto il mondo. Tuttavia, per il distretto dell'Amiata il livello di conoscenza non è sufficiente per formulare un'ipotesi definita di roccia sorgente di Hg.

Trasporto di Hg

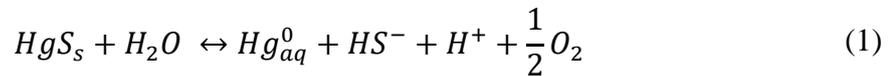
A partire da Dessau et alii (1972), alcuni autori (Varekamp & Buseck, 1984; Morteani et al., 2011) hanno proposto un modello di trasporto di Hg in fase gassosa. La ripartizione preferenziale di Hg in fase gas e di Sb in fase liquida spiegherebbe anche la separazione dei due elementi nei depositi dell'area (Cadoux & Pinti, 2009). Seguendo il modello di questi autori, è possibile che i depositi di Sb(Au) rappresentino porzioni più profonde dei sistemi

idrotermali, di cui i depositi di Hg sono l'espressione meno profonda. Depositati di Sb non scoperti potrebbero quindi trovarsi sotto il Monte Amiata, mentre qualsiasi deposito di Hg sovrastante i depositi di Sb noti sarebbe stato eroso attraverso processi post-deposizionali. Se il trasporto di Hg dalle parti profonde del sistema mineralizzante del Monte Amiata avveniva in fase gassosa, questa fase potrebbe aver depositato cinabro direttamente in prossimità della superficie, oppure potrebbe essersi mescolata con acque meteoriche di bassa profondità, sistemi idrotermali, che poi hanno fatto precipitare il cinabro insieme alle vene di calcite (Rimondi et al., 2015).

Secondo il modello proposto da Varekamp e Buseck (1984), i requisiti generali per il trasporto e la precipitazione efficiente di mercurio sono:

- Roccia sorgente ricca di Hg;
- Raggiungimento di elevata solubilità di Hg alle alte temperature;
- Efficiente diminuzione della solubilità di Hg in ristretto intervallo di condizioni.

Varekamp e Buseck (1984) hanno calcolato l'abbondanza della specie Hg_{aq}^0 nella speciazione di Hg nei fluidi idrotermali, mediante la reazione

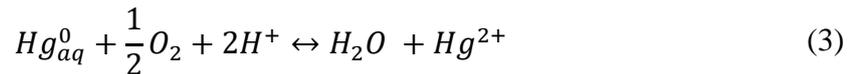


che descrive la reazione di dissoluzione di HgS_s in Hg_{aq}^0 e mostra che all'aumentare di H^+ , HS^- e pO_2 tendono a limitare la solubilità di Hg_{aq}^0 .

L'espressione

$$\log Hg_{aq}^0 = K_r(1) - \log H^+ \cdot HS^- - \frac{1}{2} \log pO_2 \quad (2)$$

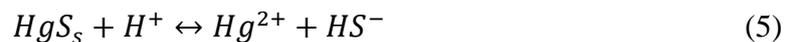
indica la quantità di Hg_{aq}^0 in equilibrio con HgS_s , con $K_r(1)$ la costante della reazione (1). Considerando una soluzione sottosatura in HgS_s , si può riformulare l'equazione (1) nella reazione



con

$$\log \left(\frac{Hg^{2+}}{Hg_{aq}^0} \right) = K_r(3) - 2pH + \frac{1}{2} \log pO_2 \quad (4)$$

e nella reazione



con

$$\log Hg^{2+} = K_r(5) - pH - \log HS^- \quad (6)$$

L'equazione (3) è valida per tutti i sistemi acquosi; per sistemi saturi in HgS_s , la reazione (5) va applicata per prima e i valori risultanti di Hg^{2+} possono essere sostituiti nella reazione (4). Possiamo quindi riassumere che la solubilità di Hg dipende dalle variazioni di temperatura, di pH e di pO_2 e che i principali fattori della precipitazione di HgS_s sono una diminuzione di temperatura, l'aumento di pO_2 o una diminuzione di pH.

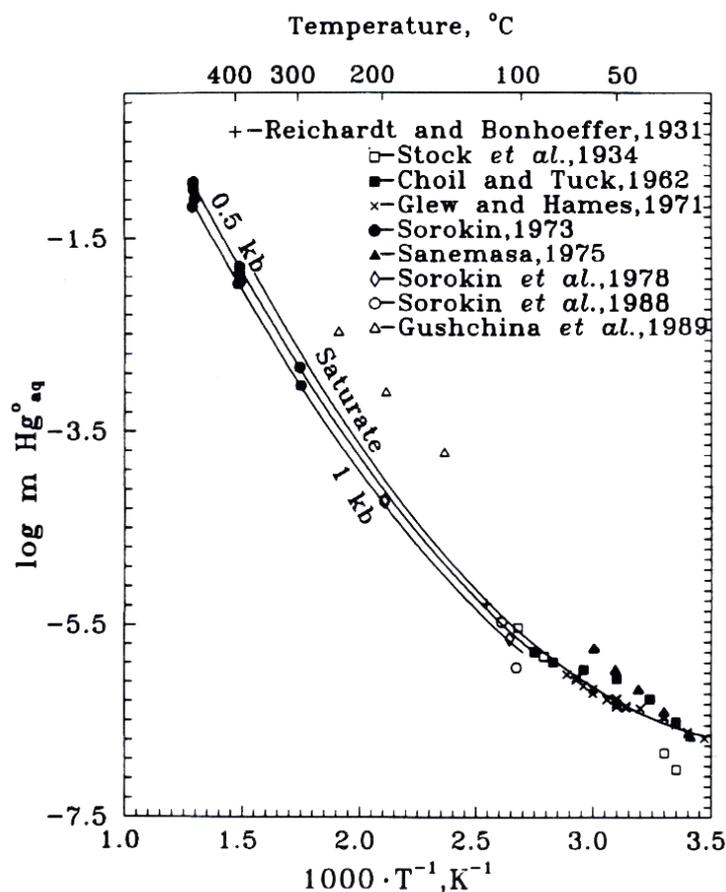


Figura 2.13. Diagramma sperimentale che mostra l'andamento della solubilità di Hg con la temperatura. La solubilità di Hg in acqua è estremamente bassa, ma aumenta con la temperatura. Si noti come la maggior parte degli esperimenti sia stata effettuata per temperature comprese tra 15 e 150°C.

Secondo il modello sperimentale proposto in Fig. 2.13, in accordo con quanto proposto da White (1981) che ha stimato una temperatura massima di 150°C alla quale si ha precipitazione di HgS , è chiaro che a basse temperature la solubilità di Hg_{aq}^0 è molto bassa. Questa bassa solubilità costituisce un vincolo molto importante al processo di trasporto e precipitazione del cinabro in ambiente geotermico, e di conseguenza anche un vincolo al processo genetico. Considerando infatti l'ipotesi di trasporto e deposizione di cinabro in un intervallo di valori di temperatura di 80-100°C (Brogi & Fabbrini, 2009), la massa di fluido idrotermale che trasporterebbe e precipiterebbe le 116775 t di Hg estratto dalle miniere dell'Amiata (Segreto, 1991) sarebbe di $2.6-4.5 \cdot 10^{11}$ t di H_2O , che corrispondono a c. 250-440 km^3 di fluido idrotermale (ad una densità di $1'020 \text{ kg/m}^3$). Questi volumi notevoli, calcolati utilizzando i dati di solubilità di Clever et alii (1985), che considerano l'ipotesi che il fluido sia trasportato dal fluido soltanto come ione Hg_{aq}^0 in soluzione, corrispondono a uno scenario in cui l'efficienza di precipitazione del fluido geotermico è pari al 100%. Di conseguenza, essi costituiscono una stima minima di masse/volumi di fluido mineralizzante.

2.3 Risorse geotermiche

Muffler e Cataldi (1978) hanno definito le risorse geotermiche come “l'energia termica che potrebbe essere ragionevolmente estratta a costi competitivi con altre forme di energia in un determinato momento futuro”.

Un sistema geotermico si viene a creare in particolari zone della Terra che presentano valori di flusso di calore anomali rispetto ai valori crostali. Il flusso di calore riscalda in profondità l'acqua contenuta nelle rocce permeabili (serbatoi o *reservoir*) (Barbier, 1997). Le risorse con il più alto potenziale energetico sono concentrate principalmente sui confini tra le placche, dove sono frequenti manifestazioni superficiali dell'attività geotermica. Tali manifestazioni comprendono venute di gas, sorgenti calde, fumarole, geysers.

Il valore medio di flusso di calore nella crosta continentale è di 59 mW/m^2 , mentre nelle aree geotermiche il valore di flusso di calore può essere fino a 6 volte maggiore, con temperature dei *reservoir* che possono raggiungere 300°C , a seconda della tipologia di sistema geotermico presente.

Per essere definito tale, un sistema geotermico deve essere costituito da quattro componenti fondamentali:

- Sorgente di calore superficiale;
- Serbatoio (*reservoir*) per l'acquifero costituito da rocce permeabili;
- Rocce impermeabili (acquicludo/acquitardo);
- Area di ricarica dell'acquifero.

Quando sono contemporaneamente presenti ed efficaci i quattro componenti, in relazione a un sistema vulcano-plutonico, tale area prende il nome di “sistema geotermico-idrotermale”.

I sistemi ad acqua dominante sono i più diffusi e sono a loro volta divisi in sistemi ad acqua calda, in cui esiste solo la fase liquida e hanno le temperature più basse ($30\text{-}100^\circ\text{C}$), e sistemi a vapore umido, in cui sono presenti miscele di fase liquida e vapore ma con fase liquida volumetricamente predominante e temperature tra 100 e 300°C .

I sistemi a vapore dominante producono vapore saturo o surriscaldato a pressioni superiori a quella atmosferica. Il flusso termico presente in questa tipologia di sistemi geotermici è molto maggiore rispetto a quelli ad acqua dominante. Le temperature rilevate a bocca pozzo variano tra 100 e 200°C .

La produzione di vapore avviene grazie alla presenza della copertura impermeabile, che è fondamentale per mantenere costante il valore di pressione all'interno del serbatoio. Al di sotto della copertura impermeabile prevale l'ebollizione del fluido e la separazione di fase.

2.3.1 Il sistema geotermico-epitermale del Monte Amiata

A partire dal Pliocene, grazie all'intrusione del corpo magmatico, che costituisce il motore termico, nel distretto amiatino si è instaurato un sistema geotermico ad acqua dominante ad alta entalpia, costituito da un sistema convettivo a temperature $>300^\circ\text{C}$.

Il sistema geotermico è costituito da due serbatoi: quello superiore è ospitato nelle formazioni che costituiscono il Complesso di permeabilità IV (Calcere cavernoso, Anidriti di Burano e i carbonati giurassici delle Nappe Toscane), il secondo è ospitato nelle filladi metacarbonatiche che costituiscono il basamento metamorfico toscano (Batini et al., 2003).

Le unità silicoclastiche che costituiscono le Nappe Toscane, Liguri e Subliguri sono la copertura impermeabile del serbatoio geotermico superiore, mentre la Fm. del Verrucano costituisce il livello impermeabile tra il serbatoio profondo e quello superficiale (Fig.2.14).

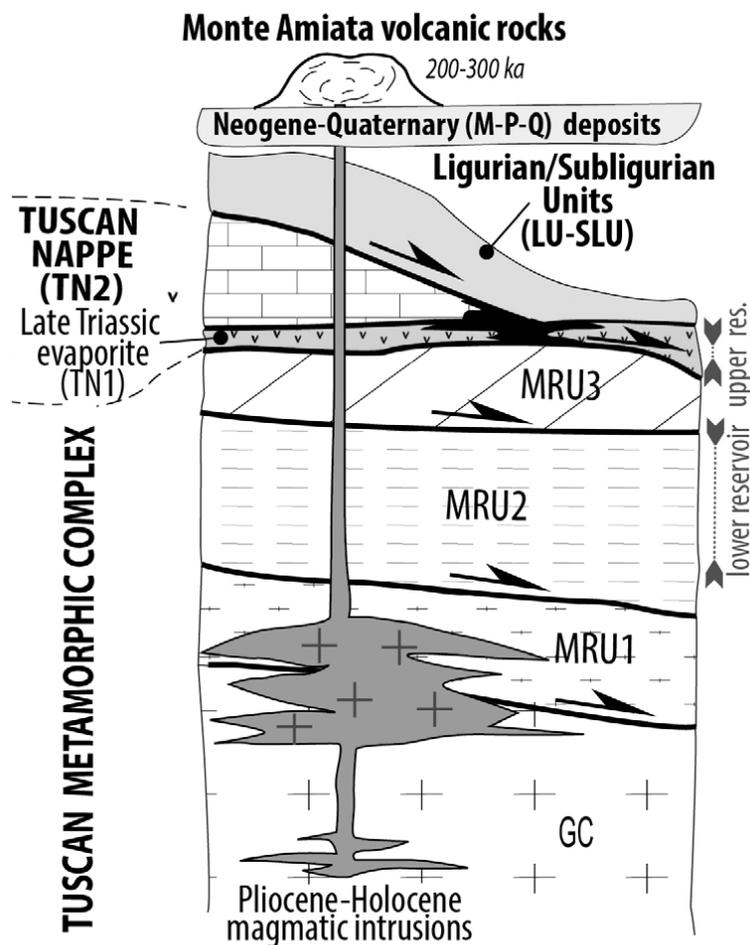


Figura 2.14 Rappresentazione schematica della successione tettono-stratigrafica del sistema geotermico del Monte Amiata (modificato da Sillitoe e Brogi (2020)). La sezione mostra le principali strutture che costituiscono il sistema geotermico. GC: Complesso degli Gneiss appartenenti al basamento metamorfico. MRU(1,2,3): Unità appartenenti alla Unità di Monticiano-Roccastrada (Complesso Metamorfico Toscano) che comprende i micascisti, le filladi del Farma e il Verrucano. TN1: Unità evaporitiche delle Nappe Toscane. TN2: Unità silicoclastiche appartenenti alle Nappe Toscane superiori. LUSLU: Unità appartenenti al Dominio Ligure e Subligure. M-P-Q: depositi neogenici e quaternari (Neoautoctono). Il *reservoir* inferiore è ospitato in MRU2 mentre quello superiore è ospitato in TN1. MRU3 e LU-SLU costituiscono le coperture impermeabili.

L'area di ricarica del serbatoio superiore è espressa dagli affioramenti del Complesso IV. I sistemi di faglie normali e le *transfer-zones* associate costituiscono vie di permeabilità per i fluidi meteorici al di sopra della transizione fragile-duttile e mettono in comunicazione i due serbatoi geotermici (Bellani & Gherardi, 2009; Sillitoe & Brogi, 2021). Questo genera una influenza sulle temperature, innescata dalla circolazione dei fluidi, tra i due serbatoi.

Nell'area è presente un'importante anomalia di flusso di calore ($> 300\text{mW/m}^2$) (Baldi et al., 1995), correlata all'assottigliamento crostale e all'intrusione mafica tardo-pliocenica.

Nell'area di Bagni San Filippo, gli affioramenti del Complesso IV sono associati alla presenza di numerose sorgenti termali, che presentano temperature $>38^\circ\text{C}$ ed elevato contenuto di CO_2 . In quest'area sono presenti anche abbondanti emissioni di gas e, dove

elevate concentrazioni di CO₂ (± H₂S) coincidono con sorgenti termali, si ha l'accumulo di depositi di travertino con deposizione temporalmente continua negli ultimi 65ka (Minissale, 2004; Brogi et al., 2016; Pirajno, 2020).

È nota l'esistenza di una connessione ad alta permeabilità tra le vulcaniti che ospitano l'acquifero freatico e le evaporiti che ospitano il serbatoio geotermico superficiale, data da tre elementi, identificati da Delcroix et alii (2006) in:

- Contatto tra le vulcaniti e la Falda Toscana;
- Condotti vulcanici che tagliano la Falda Toscana;
- Rigetti di faglia dovuti ad eventi di collasso gravitativo dell'edificio vulcanico.

Tali connessioni favorirebbero la circolazione dei fluidi e la possibile ricarica del serbatoio geotermico da parte dell'acquifero freatico.

A partire dagli anni '70, i programmi di perforazione profonda hanno rivelato la presenza di un'ampia fascia permeabile all'interno del basamento metamorfico ad una profondità di 2.5 – 3.5km, 2000m al di sotto del *reservoir* superiore sfruttato fino a quel momento. Le temperature del serbatoio geotermico profondo variano tra 300 e 350°C (Billi et al., 1986; Bertini et al., 1995; Barelli et al., 2010)

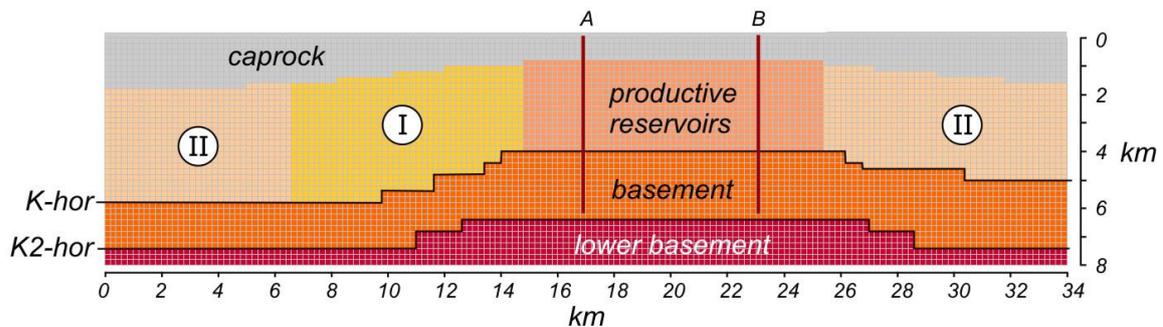


Figura 2.15. Rappresentazione schematica dei serbatoi produttivi. I, II: formazioni geologiche adiacenti all'area produttiva, appartenenti alle stesse unità litologiche, ma con proprietà idrauliche differenti. Il segmento A-B corrisponde alla zona di produzione del sistema geotermico (da Bellani & Gherardi (2009)).

Le indagini sismiche effettuate nell'area a partire dai transetti sismici del progetto CROP, hanno permesso di individuare due riflettori, denominati orizzonte K e orizzonte K2, i cui culmini sono situati a profondità comprese rispettivamente tra 3-4km e 7-9km (Accaino et al., 2005; Bellani & Gherardi, 2009). Come mostrato nella ricostruzione in Fig. 2.15, entrambi gli orizzonti tendono ad approfondirsi nelle porzioni distali rispetto al centro della zona produttiva. L'orizzonte K viene identificato in letteratura come il tetto del basamento metamorfico

In questo contesto, sono presenti giacimenti epitermali la cui formazione è correlata con il sistema geotermico presente. I giacimenti epitermali si sviluppano principalmente in un contesto di margine convergente, in associazione con intrusioni superficiali e vulcanismo effusivo (White, 1981). Si trovano spesso associati a manifestazioni superficiali, come sorgenti calde e depositi di travertino.

Le sorgenti calde, o termali, sono sorgenti in cui la temperatura dell'acqua è maggiore della temperatura media dell'aria locale e hanno valori di 30-40 °C (Des Marais & Walter, 2019). Le sorgenti calde, o termali, si trovano spesso associate a venute di gas, quali fumarole e putizze, in contesti in cui è presente vulcanismo attivo.

Le acque che sgorgano da queste sorgenti possono avere differenti composizioni chimiche ma nella zona in esame sono di tipo solfato-clorurate alcalino-terrose.

Sulla base degli studi geochimici effettuati da Duchi et alii (1987) e Minissale et alii (1997) sulle acque delle sorgenti fredde e termali dell'area amiatina, le composizioni chimiche indicano che per le sorgenti fredde non sembra esserci un contributo da parte di fluidi caldi profondi. Al contrario, per le sorgenti termali sembrano esserci interazioni non chiaramente quantificabili con i fluidi idrotermali del *reservoir* superiore. Tutte le sorgenti termali analizzate da Duchi et alii (1987) presentano una composizione solfato-calcica, che può essere messa in relazione con la dissoluzione della formazione anidritica alla base delle Nappe Toscane oppure con l'ossidazione di solfuri. Nell'area di Bagni San Filippo, in corrispondenza dei depositi idrotermali costituiti da travertino, le acque termali risultano essere l'unico termine di composizione bicarbonato-calcica. Nei campioni più distali rispetto ai depositi di travertino, la composizione risulta solfato-calcica. Si pensa che la precipitazione di CaCO_3 presso la sorgente causi la modificazione chimica della composizione delle acque da bicarbonato-calcica a solfato-calcica e che un processo simile possa verificarsi anche in profondità (Fig. 2.16).

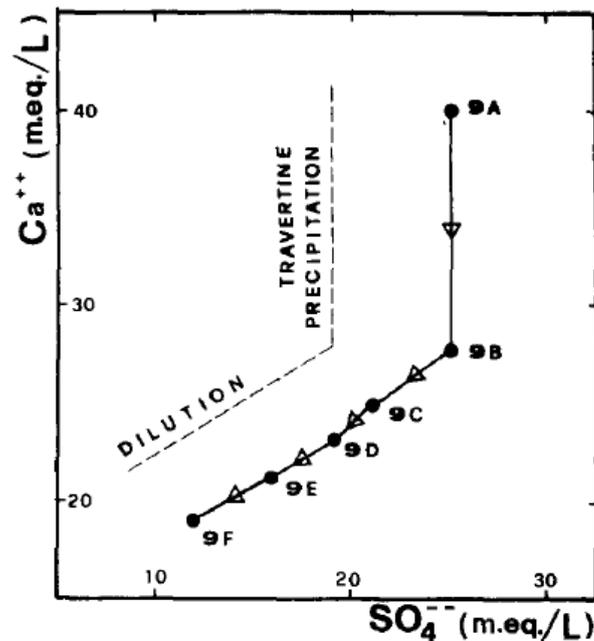


Figura 2.16. Diagramma di variazione della composizione delle acque delle sorgenti di Bagni San Filippo. La direzione della freccia indica la diminuzione di temperatura. (da Duchi et al. (1987)).

La presenza di depositi di travertino è da associarsi a questa tipologia di acque, poiché la sua deposizione è dovuta all'essoluzione della CO_2 dalle acque una volta giunte in superficie (Pirajno, 2020).

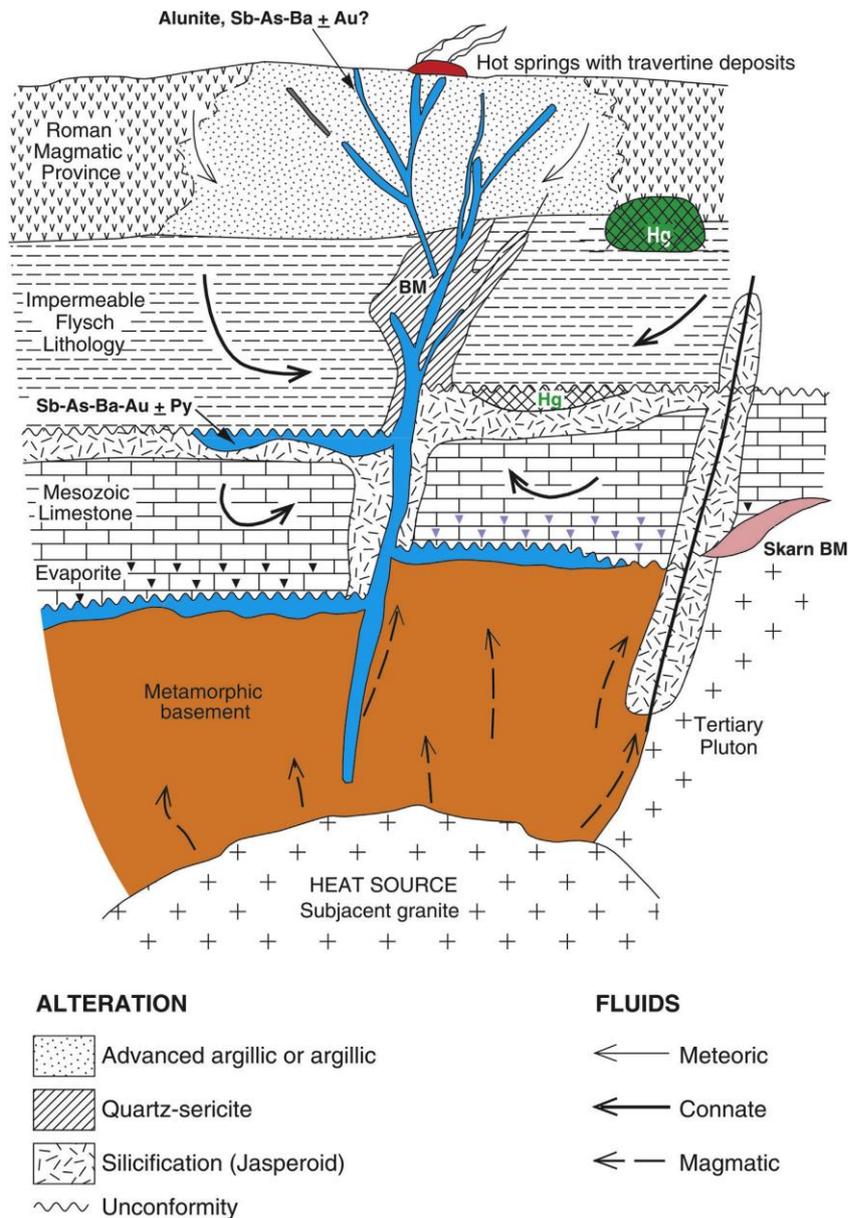


Figura 2.17. Modello concettuale dei giacimenti epitermali toscani che mostra la presenza di un sistema geotermico con depositi epitermali associati (da Pirajno (2020)).

I giacimenti cinabreriferi amiatini presentano strutture associate ad una delle tipologie di sistema minerale epitermale associato a sorgenti calde descritte da Berger (1983).

Secondo il modello concettuale proposto da Pirajno (2020) (Fig. 2.17), i fluidi di origine magmatica risalgono attraverso zone a maggiore permeabilità e si mescolano a fluidi connati all'interno del *reservoir* ospitato nelle formazioni evaporitiche (*reservoir* superiore) e con i fluidi meteorici nelle porzioni più superficiali. Le mineralizzazioni, identificate dalle aree in blu in figura, si trovano ai contatti tra formazioni a differente permeabilità e all'interno delle zone interessate dalle deformazioni fragili. La presenza di tipologie di alterazione idrotermale dominate da silice, quarzo-sericite, o argille nelle litologie incassanti sono diagnostiche per comprendere le paragenesi presenti.

In associazione con i sistemi geotermico-epitermali si possono creare strutture in conseguenza di eruzioni idrotermali (Fig. 2.18).

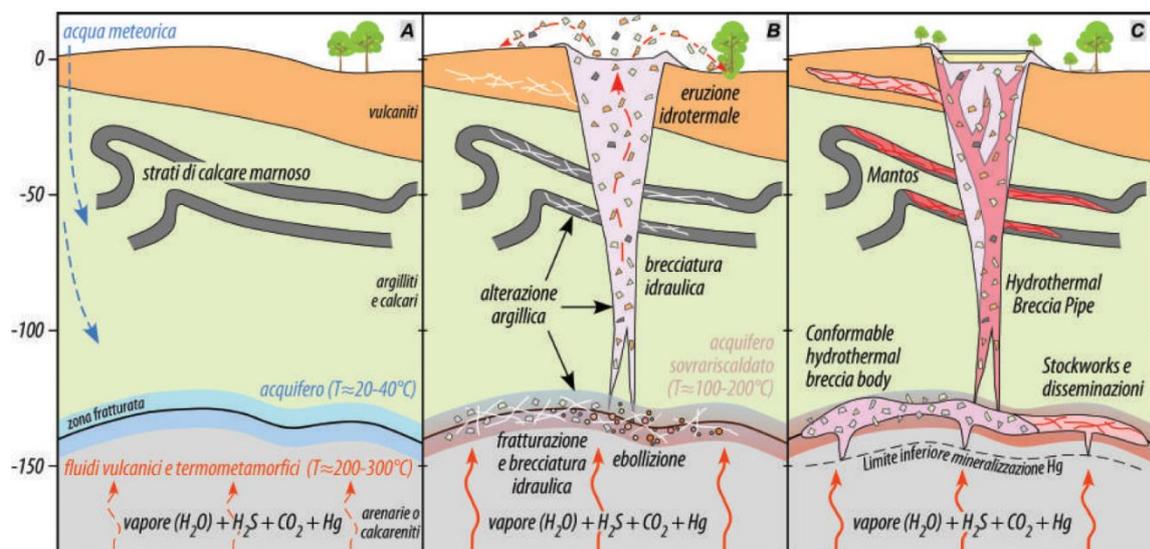


Figura 2.18. Modello concettuale dei depositi associati a eruzione idrotermale in sistemi epitermali per la miniera di Abbadia San Salvatore. Le strutture prodotte in conseguenza del *boiling* dei fluidi assumono la forma di imbuto, con diametro massimo di 30m e profondità di circa 80m, costituiti da breccie e argille. La permeabilità iniziale di queste strutture è molto elevata ma diminuisce rapidamente con l'avanzare dell'alterazione argillitica (da Principe et alii (2017)).

Il processo genetico di un *hydrothermal breccia pipe* può essere riassunto in tre fasi:

- Stato iniziale: le acque meteoriche percolano in profondità dove vengono riscaldate e arricchite in CO_2 , H_2S e Hg grazie al *mixing* con i fluidi magmatici e termometamorfici;
- Fase pre-mineralizzazione: si innesca l'alterazione idrotermale argillitica in seguito all'abbassamento del pH da parte delle acque meteoriche. L'acquifero entra in ebollizione e la copertura impermeabile permette di sviluppare condizioni di sovrappressione. Si genera quindi un processo di eruzione idrotermale e di fratturazione/brecciazione idraulica.
- Fase sin-mineralizzazione: il pH ritorna in condizioni neutre grazie al degassamento di CO_2 e SO_2 ed all'interazione fluido-roccia, permettendo la precipitazione di cinabro, solfuri di metalli di base e carbonati. Vengono mineralizzati tutti i volumi di roccia che hanno subito alterazione o brecciazione.

Mineralizzazioni e sistema geotermico attuale

Le mineralizzazioni a Hg, per quanto riguarda la miniera di Abbadia San Salvatore, sono più giovani delle rocce vulcaniche incassanti. Questo, associato alle sovrapposizioni delle distribuzioni di minerali di Hg con l'attuale sistema geotermico, potrebbe dimostrare un continuum nella deposizione (Morteani et al., 2011). A scala locale, sono numerosi i casi di stretta vicinanza tra la miniera di mercurio e i fenomeni geotermici odierni (sorgenti termali e/o venute di gas), come le emanazioni di gas alla miniera di Solforate o la stretta vicinanza della miniera Pietrineri alle sorgenti termali di Bolloro.

Un'ulteriore evidenza è data dalle emissioni di Hg dalle centrali geotermiche il quale, fino alla fine del secolo scorso, veniva emesso a velocità dell'ordine di 300-400 g/h (Bacci et al., 2000). Attualmente, Hg viene rimosso dai fluidi geotermici utilizzando la tecnologia AMIS (*Abatement of Mercury and Hydrogen Sulfide*) (Lenzi & Caprai, 2016).

Le caratteristiche geologiche-giacimentologiche del sistema geotermico del M.te Amiata sono simili a quelle dei depositi degli Stati Uniti occidentali denominati *Carlin-style* dal nome della località tipo dove ricorrono. I depositi *Carlin-style* sono giacimenti epitermali ad Au, chiamati anche *sediment-hosted gold deposits* oppure *disseminated gold deposits*, che consistono principalmente in disseminazioni di pirite ed arsenopirite arricchite in Au, As, Hg, Sb e Tl all'interno di rocce alterate che formano parti di sequenze sedimentarie di margine continentale di tipo silicoclastico-carbonatico, la cui sorgente dei fluidi non è esattamente definita (Yin et al., 2019). Un'altra somiglianza tra giacimenti *Carlin-style* e giacimenti amiatini è rappresentata dall'ambiente tettonico (ambiente transtensionale a seguito di convergenza del margine) (Dobak et al., 2020).

Ci sono però alcune differenze, poiché nei depositi degli Stati Uniti occidentali la silice è una ganga comune nei depositi tipo *hot spring*, presente sotto forma di sinter siliceo depositato nei pressi della sorgente termale (Rytuba, 1993). Come accennato in precedenza, la silice è invece relativamente rara nei depositi del Monte Amiata. Le sorgenti termali in questa zona depositano tipicamente travertino, con contenuti di Hg fino a 0,45 mg/kg (Duchi, 1976; Guo & Riding, 1998).

Dati e metodi

Per poter applicare uno studio di ML ad un sistema geotermico come quello del Monte Amiata è necessaria un'accurata preparazione dei dati, che prevede la costruzione di un dataset bidimensionale contenente dati provenienti dalla letteratura e la successiva costruzione di un modello tridimensionale che cattura per quanto possibile i suoi quattro componenti fondamentali (i.e., sorgente, serbatoio, acquitardo, area di ricarica) per applicare gli algoritmi di predizione.

3.1 Dataset

Uno degli obiettivi di questo lavoro di tesi è quello di creare un dataset digitale che comprende un numero di dati geologici-giacimentologici di letteratura sufficientemente dettagliati sul distretto cinabrifero del M.te Amiata da permettere una correlazione statistica significativa.

Il materiale utilizzato per costruire il dataset è stato reperito dal Geoportale GEOscopio della Regione Toscana e da altre fonti quali ARPAT, dal Geoportale Nazionale del Ministero dell'Ambiente, ISPRA, dall'Inventario delle risorse geotermiche nazionali - MISE, dalla Banca Dati Nazionale Geotermica (Geothopica – CNR) e dall'OGS Trieste. Infine, si sono anche raccolti i dati provenienti dagli archivi del Museo delle Miniere del Monte Amiata e quelli provenienti dalla campagna di esplorazione geotermica pubblicati da Calamai et alii (1970) e contenenti i primi dati della campagna di prospezione geotermica condotta da Enel nell'area (periodo 1958-1965 considerando la campagna di perforazione). Si noti che questo dataset è stato costruito di proposito su un set limitato di dati di prospezione geotermica allo scopo di testare l'ipotesi che correlazioni significative di dati geologici e giacimentologici sono possibili a partire da pochi dati esplorativi. Quest'approccio rinuncia quindi ad uno studio di dettaglio che sarebbe possibile attraverso l'uso di dataset geotermico esplorativo più esteso e recente, ma trae il suo valore dal fatto che un suo potenziale utilizzo in altri contesti geotermici-epidermali potrebbe guidare le fasi esplorative di risorse geotermiche-epidermali. I dati di questo lavoro, e la loro interpretazione, non vanno quindi intesi come un contributo attuale alla conoscenza del sistema geotermico amiatino, ma come una potenziale guida all'esplorazione geotermica-epidermale nelle sue fasi iniziali.

L'organizzazione del dataset è stata effettuata utilizzando il software *open source* QGIS, che permette di gestire dati, sia vettoriali che immagini *raster*, geografici e geologici.

3.1.1 Struttura

Il dataset è stato costruito cercando di organizzare i dati in maniera logica, suddividendoli in base al tipo di geometria di ogni file vettoriale, o *shapefile*.

Punti

Questa porzione del dataset è stata costruita in modo da contenere tutti i dati puntuali reperiti in letteratura. Il dataset dei dati puntuali comprende sia dati superficiali (sorgenti, venute di gas, localizzazione in carta dei siti minerari e delle centrali geotermiche) che dati di profondità (pozzi geotermici e sondaggi stratigrafici).

Punti superficiali

I punti superficiali sono tutti i punti la cui coordinata z corrisponde alla quota del modello digitale di terreno (DTM). Tali punti sono utili per comprendere l'espressione superficiale del sistema geotermico presente nell'area (Fig. 3.1).

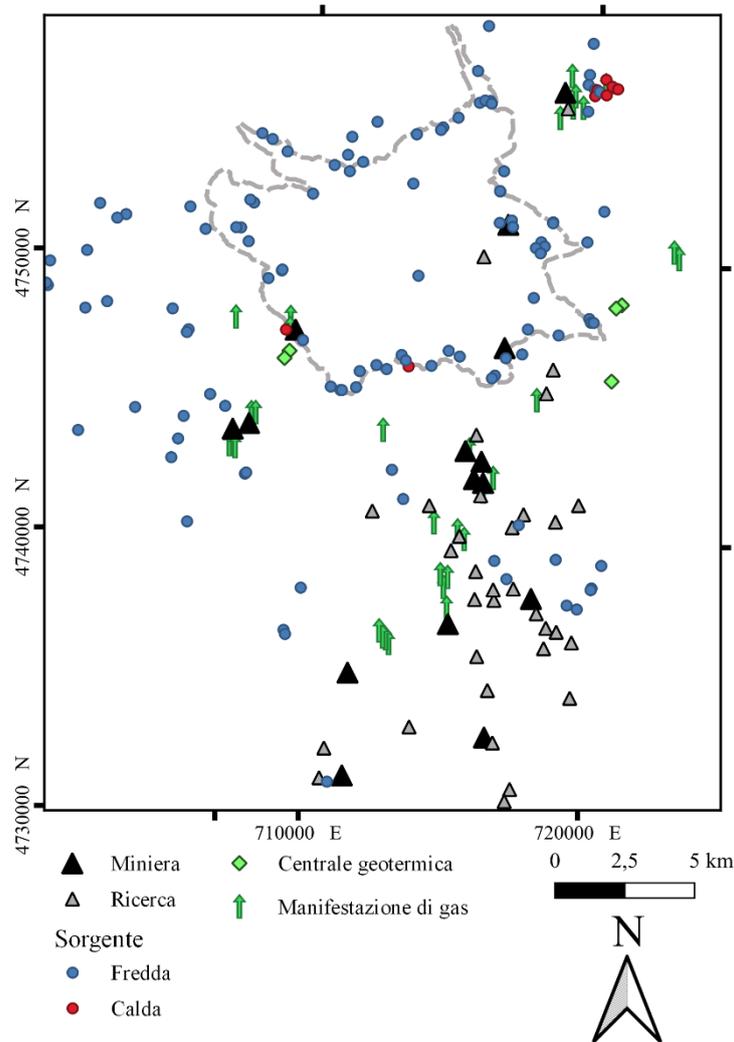


Figura 3.1. Localizzazione dei principali dati puntuali superficiali dell'area di studio.

Sorgenti

La localizzazione delle sorgenti è stata possibile grazie all'utilizzo di due distinti dataset: i dati provenienti dall'Inventario delle risorse geotermiche nazionali (Ministero dello Sviluppo Economico, 1994) e dalla campagna di rilevamento svolta da Magi et alii (2019). Il parametro principale reperito dai dati delle sorgenti è la temperatura, utilizzata in seguito come temperatura superficiale T_0 per effettuare calcoli di solubilità di Hg. Sulla base di questa temperatura, le sorgenti sono state suddivise in due categorie, sorgenti fredde e sorgenti calde o geotermiche, che presentano rispettivamente valori di $T < 38^\circ\text{C}$ e $T \geq 38^\circ\text{C}$.

Venute di gas

Le venute di gas puntuali comprendono manifestazioni di gas e putizze, emanazioni di anidride carbonica accompagnata da idrogeno solforato (Forconi, 2011).

Miniere

Le miniere sono state localizzate in carta a partire dal database BD-RIMI della Regione Toscana, identificate in 15 punti ubicati in corrispondenza degli ingressi principali di ogni sito minerario. Nella carta di Fig. 3.1, sono stati inseriti anche 38 punti denominati “ricerca”, che costituiscono affioramenti di cinabro e stibina o piccoli depositi con bassi tenori che in passato sono stati parzialmente sfruttati, localizzati in carta da Forconi (2011).

Centrali geotermiche

Le centrali geotermiche presenti nell’area si trovano in due zone distinte, a Bagnore, nel Comune di Santa Fiora, e a Piancastagnaio, rispettivamente a SW e a SE del Complesso Vulcanico del Monte Amiata. Le centrali, gestite da Enel Green Power, sono situate in corrispondenza dei massimi valori di flusso di calore registrati nell’area.

Punti in profondità

I dati di profondità (Fig. 3.2) sono stati reperiti dal servizio WMS del Geoportale della Regione Toscana, denominato BDIG – Banca Dati Indagini Geotematiche, che contiene tutte le indagini geologiche effettuate nell’area. L’utilizzo di questa risorsa ha permesso di poter inserire il maggior numero possibile di dati di sottosuolo all’interno del database.

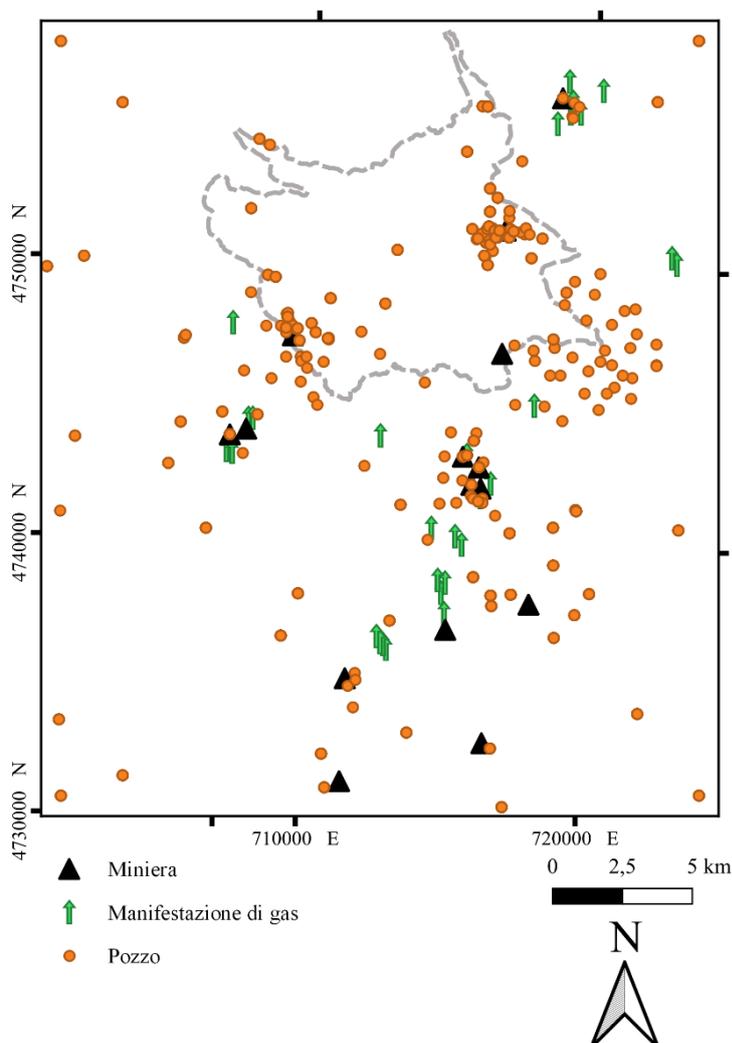


Figura 3.2. Localizzazione dei pozzi analizzati. La maggiore concentrazione di dati si trova in corrispondenza dei siti minerari sfruttati in passato e dei campi geotermici.

I dati di pozzo disponibili provengono da diverse categorie di indagini effettuate nell'area e comprendono pozzi geotermici, pozzi per acqua idropotabile, prove penetrometriche statiche e dinamiche. Dati provenienti da sondaggi a carotaggio continuo sono assenti in letteratura.

Compatibilmente con il livello di dettaglio della caratterizzazione di ogni pozzo, i dati contengono informazioni riguardo alla litologia, alla profondità dei contatti stratigrafici, alla temperatura rilevata in pozzo e alla concentrazione di Hg in soluzione, espresso in ppm.

Tracce

La porzione del dataset denominata “Tracce” è costituita da tutti gli *shapefile* di tipo *polyline* e comprende i contatti stratigrafici e le tracce di faglia – raggruppate sulla base del regime tettonico che le ha generate – reperiti dalla cartografia in scala 1:10'000, gli assi delle pieghe e le tracce delle sezioni stratigrafiche, prodotte in seguito alla campagna di sondaggi documentata nel lavoro di Calamai et alii (1970). All'interno del subset Tracce è stato inserito anche uno *shapefile* denominato “Gallerie” che rappresenta le tracce di gallerie e discenderie di tre miniere presenti nell'area (Bagnore, Morone e Catabbio).

Pieghe

Le tracce degli assi di piega (Fig. 2.5) sono state estratte dal lavoro di Calamai et alii (1970) e digitalizzate.

Faglie

Le tracce di faglia sono state estratte da tre fonti diverse:

- Le faglie cartografabili in superficie, denominate “superficiali”, sono state estratte dalla cartografia geologica 1:10'000 in formato vettoriale (Regione Toscana, 2018) e comprendono faglie normali, faglie inverse e sovrascorrimenti;
- Le faglie che interessano il substrato evaporitico sono state digitalizzate dalla carta del tetto del substrato evaporitico (da Calamai et al. (1970)). In questa carta sono state rappresentate solamente le faglie normali ad alto angolo;
- Le inclinazioni delle faglie normali ad alto angolo che interessano il substrato evaporitico sono state digitalizzate dalle sezioni stratigrafiche utilizzando il software Esri ArcScene, come verrà specificato nei paragrafi seguenti.

Data la carenza di dati di immersione delle faglie, nell'elaborazione del modello 3D le faglie sono state tutte considerate come sub-verticali, non potendo assegnare a ciascuna il valore di inclinazione reale.

Contour

Il subset “Contour” contiene sia gli *shapefile* che sono stati digitalizzati dalle carte reperite in letteratura - digitalizzazione eseguita manualmente dopo l'importazione e la georeferenziazione di ogni carta (o raster) in formato .tiff - i quali comprendono le linee di isogradiante geotermico e le isoipse del tetto della formazione evaporitica, sia i file di isolinee scaricati dal GEOscopio della Regione Toscana e da Geothopica, che comprendono le isolinee del valore di flusso di calore.

È presente anche il file delle isoipse dell'andamento della superficie topografica, estratto dal DTM Orografico in scala 1:10'000 con risoluzione di 10m, derivato dalla Carta Tecnica Regionale in scala 1:10'000 della Regione Toscana. Il DTM è definito “orografico” perché

riporta l'andamento del terreno al meglio della precisione consentita dalla carta da cui è stato derivato.

Raster

La raccolta di immagini raster utilizzate e prodotte in questo lavoro di tesi conta numerosi file, che sono stati suddivisi all'interno del dataset a seconda della provenienza dei dati.

Tra i raster provenienti dalla letteratura, le carte contenute nel lavoro di Calamai et alii (1970) sono state georeferenziate e digitalizzate manualmente al fine di produrre file vettoriali in formato *polyline* che sono stati successivamente inseriti nel subset Contour. Altri raster che non hanno invece richiesto il lavoro di digitalizzazione sono quelli scaricati dal Geoportale GEOscopio, quindi già georeferenziate e fruibili in ambiente GIS.

La georeferenziazione delle immagini è un processo fondamentale senza il quale non si potrebbero utilizzare dati provenienti da archivi cartacei. La georeferenziazione consiste nel posizionamento di alcuni punti di controllo sull'immagine che si vuole importare e sulla cartografia già georeferenzata, come ad esempio le immagini satellitari. Una volta posizionati i punti di controllo, lo strumento di georeferenziazione opera una trasformazione delle coordinate tale per cui tutti i nuovi punti dell'immagine georeferenzata avranno coordinate (x, y) nel sistema di coordinate scelto.

Le immagini raster inserite e digitalizzate all'interno di QGIS comprendono i seguenti prodotti cartografici:

- Carta geologica in scala 1:10'000;
- Carta tettonica in scala 1:10'000;
- Carta del gradiente geotermico, calcolato entro i primi 80m di profondità, con una media di 35m;
- Carta del tetto del substrato evaporitico.

Le immagini raster importate dal Geoportale comprendono:

- DTM 1:10'000 (2017);
- Carte delle temperature a -1000m e -2000m;
- Carta della Anomalia di Bouguer.

La struttura completa del dataset costruito in ambiente QGIS e la metodologia di costruzione della banca dati sono schematicamente descritti nella Figura 3.3.

Attraverso l'interpolazione dei dati vettoriali sono state prodotte carte, utilizzate in seguito per i calcoli statistici, che comprendono:

- Carta di temperatura superficiale sulla base dei dati di temperatura delle sorgenti;
- Carte del gradiente geotermico a -1000m e -2000m;
- Carte del tetto dei Complessi di permeabilità II, III e IV.

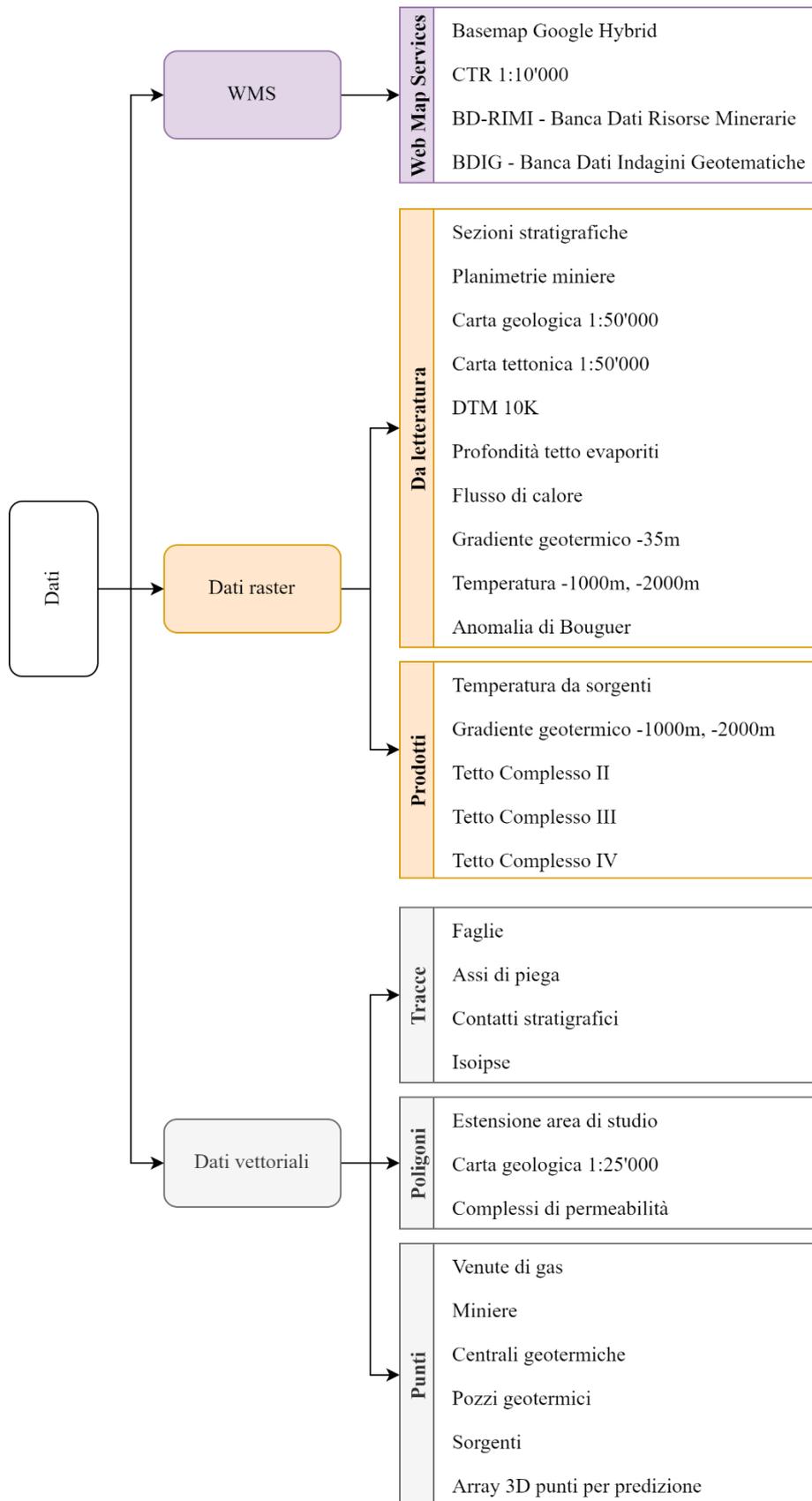


Figura 3.3. Struttura del dataset QGIS. Tutti i dati raster e vettoriali sono stati raggruppati in base alle loro caratteristiche e utilizzati al fine di produrre un database strutturato che rappresenta la banca dati iniziale per le successive applicazioni statistiche.

La produzione di tali raster ha richiesto la necessità di realizzare un set di punti, divisibili nelle seguenti categorie:

- Punti tracciati sulle sezioni stratigrafiche, che determinano la quota della base o del tetto di ogni complesso;
- Punti posizionati alla quota corrispondente al contatto stratigrafico rilevato in pozzo;
- Maglia di punti casuali drappeggiati sul DTM per assegnare le quote degli affioramenti di ogni complesso;
- Maglia di punti casuali drappeggiati sulla superficie raster del tetto del complesso sottostante.

A partire da tali punti, quindi, sono state prodotte due superfici raster di base e tetto per ogni complesso.

Per produrre le superfici raster di base e tetto di ogni complesso sono state studiate differenti tipologie di interpolatori presenti all'interno di ArcScene:

- *Inverse Distance Weighted (IDW)*, tecnica che utilizza una combinazione lineare ponderata di punti di campionamento per determinare i valori delle celle. Questo metodo presuppone che la distanza da una determinata posizione riduca l'influenza della variabile mappata (Fig.3.5a) (Esri, 2022a)²;
- *Kernel interpolation with barriers*, modello di predizione a finestra mobile che unisce i punti ai due lati delle barriere utilizzando la minore distanza tra i punti. Effettua una predizione molto precisa su distanze piccole ma genera errori standard nelle zone in cui la distanza tra i punti è maggiore (Fig.3.5b) (Esri, 2022b)³;
- *Spline with barriers*, strumento che applica il concetto di curvatura minima, metodo lineare che viene applicato ripetutamente su ciascun nodo per creare una approssimazione della superficie minima (Fig.3.5c) (Esri, 2022c)⁴.

Tra tutti gli interpolatori presenti in ArcScene, si è scelto di prendere in esame i tre elencati in quanto permettono di includere nell'interpolazione anche la presenza di faglie, in formato polilinea, considerandole come "barriere". Le superfici raster così generate presentano chiare interruzioni della continuità della superficie.

Tra gli strumenti analizzati, quello che ha migliore corrispondenza con le probabili superfici di contatto tra i complessi di permeabilità è lo strumento *Spline with barriers*. La migliore corrispondenza è data dalla assenza di artefatti nella superficie, che sono stati riscontrati invece utilizzando gli altri due interpolatori.

Il modello tridimensionale è stato quindi prodotto a partire dalle superfici raster che descrivono base e tetto dei complessi di permeabilità. Per ogni complesso, mediante l'utilizzo dello strumento "*Extrude Between*", è stato possibile estrarre un volume tra le due superfici raster di base e tetto in modo da ottenere un solido, costituito da un unico shapefile in formato *multipart*. Sono stati prodotti quattro solidi, uno per ogni complesso di permeabilità (Fig. 3.6-7) che sono poi stati sovrapposti per ricostruire il modello tridimensionale dell'area di studio (Fig. 3.8).

² <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-idw-works.htm>

³ <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-kernel-interpolation-with-barriers-works.htm>

⁴ <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-spline-with-barriers-works.htm>

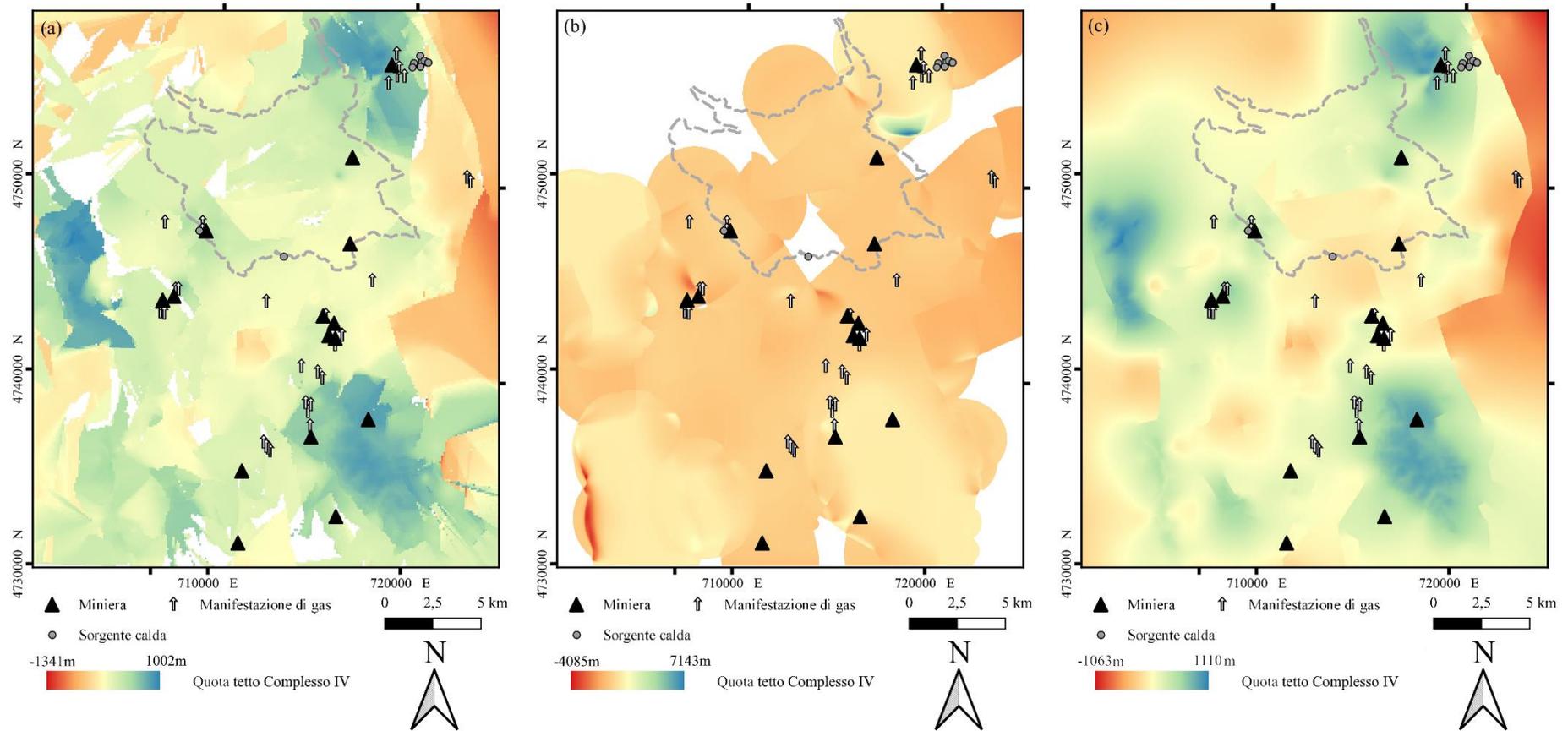


Figura 3.5. Confronto tra gli interpolatori utilizzati per l'interpolazione del tetto del Complesso IV. Le quote sono in m s.l.m. (a) *Inversed Distance Weighted (IDW)*. L'interpolazione tiene conto delle barriere e produce superfici con limiti netti con la presenza di artefatti nell'immagine. Per artefatti si intende la creazione di interruzioni, buchi o picchi di valore di quota, dovuti a una errata interpolazione dei dati. (b) *Kernel interpolation with barriers*. Il risultato dell'interpolazione considera la presenza della barriera solamente in alcune zone. L'andamento delle quote interpolate dipende dal raggio di ricerca. Le quote interpolate non rispecchiano i valori reali dell'area di studio. (c) *Spline with barriers*. Interpolazione che presenta valori più uniformi, interpretando l'elemento barriera in un modo simile all'andamento dei piani di faglia.

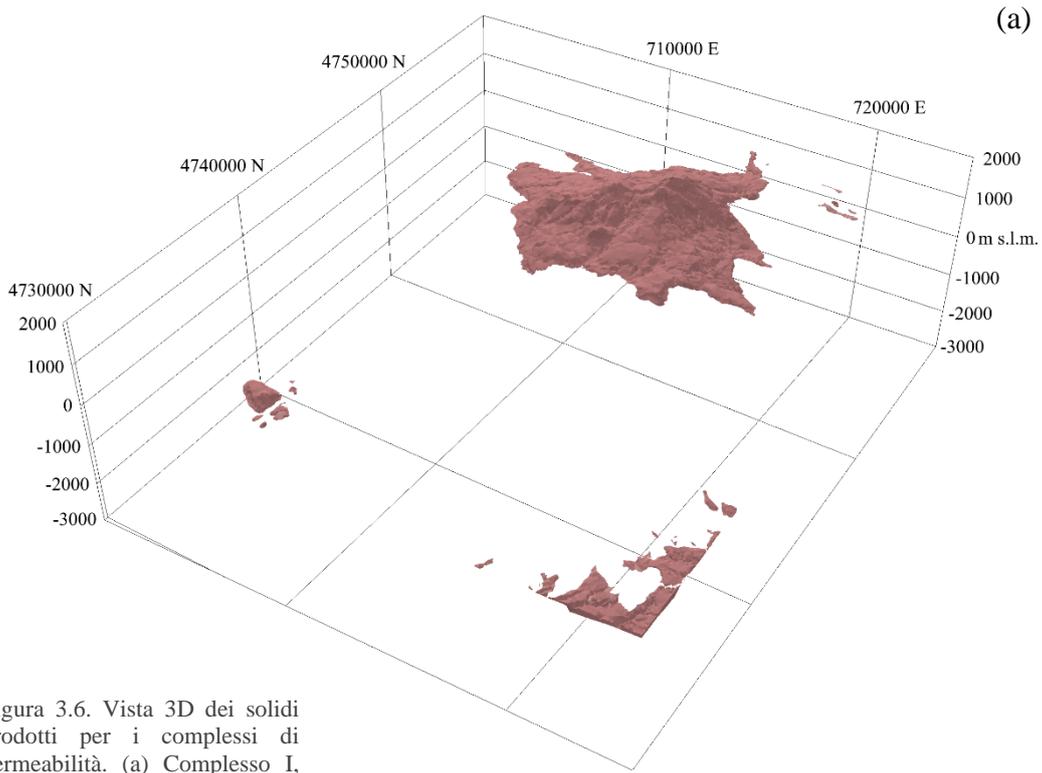
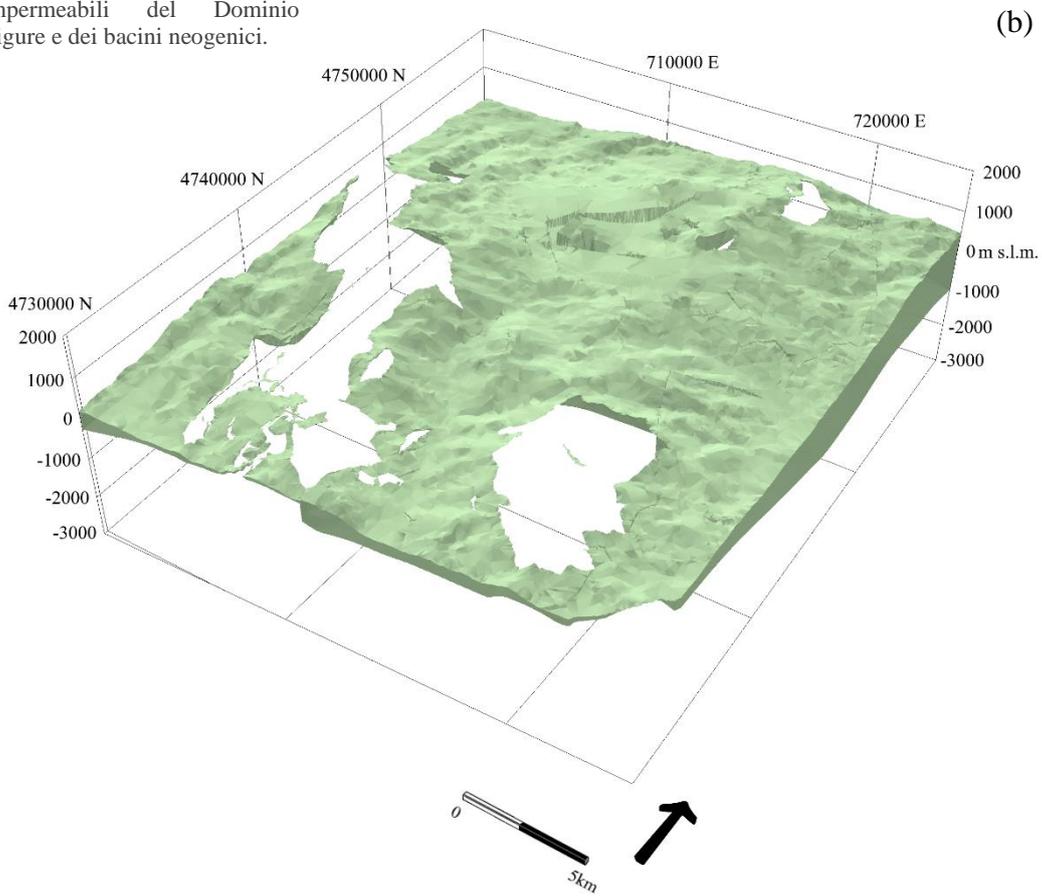


Figura 3.6. Vista 3D dei solidi prodotti per i complessi di permeabilità. (a) Complesso I, comprende le formazioni permeabili costituite da vulcaniti e travertino. (b) Complesso II, comprende le formazioni impermeabili del Dominio Ligure e dei bacini neogenici.



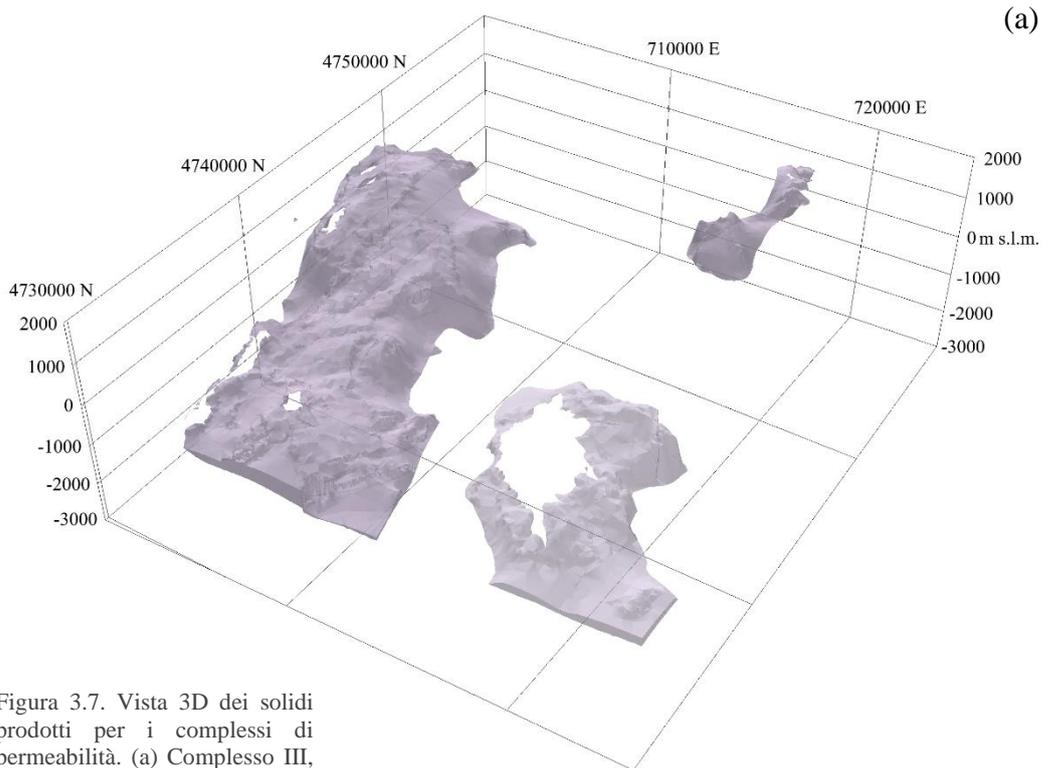
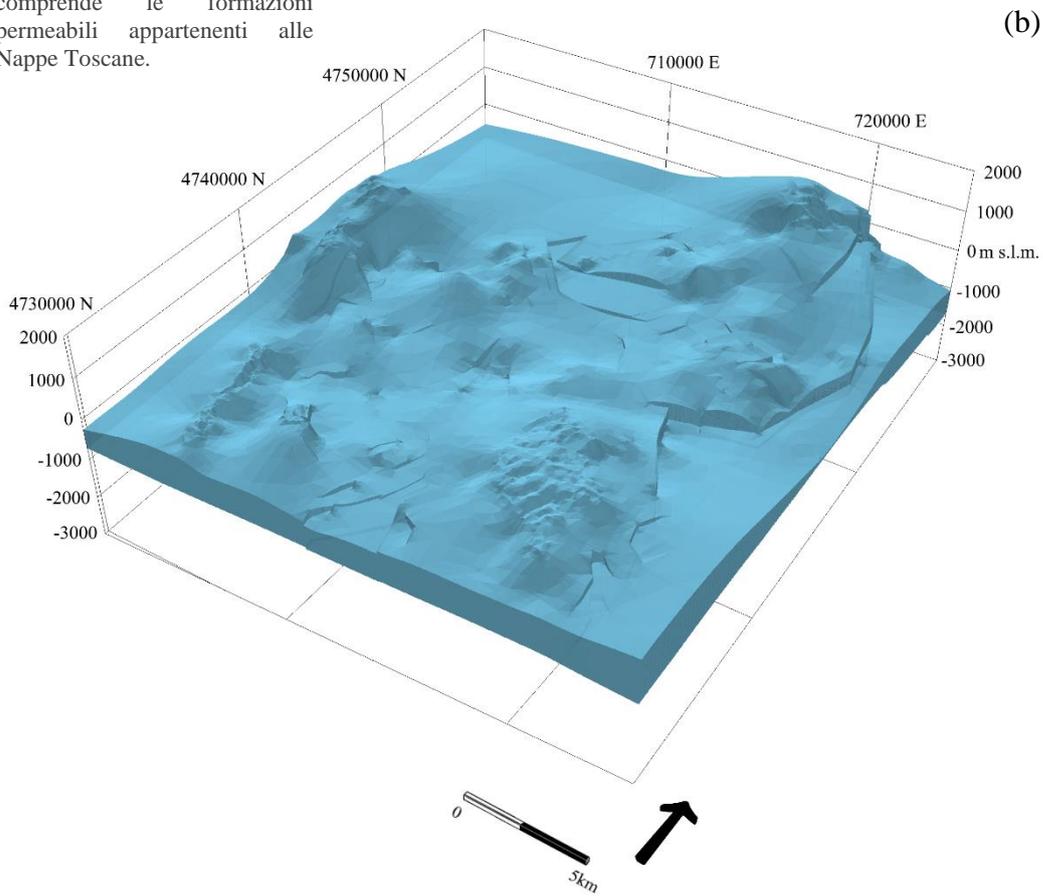
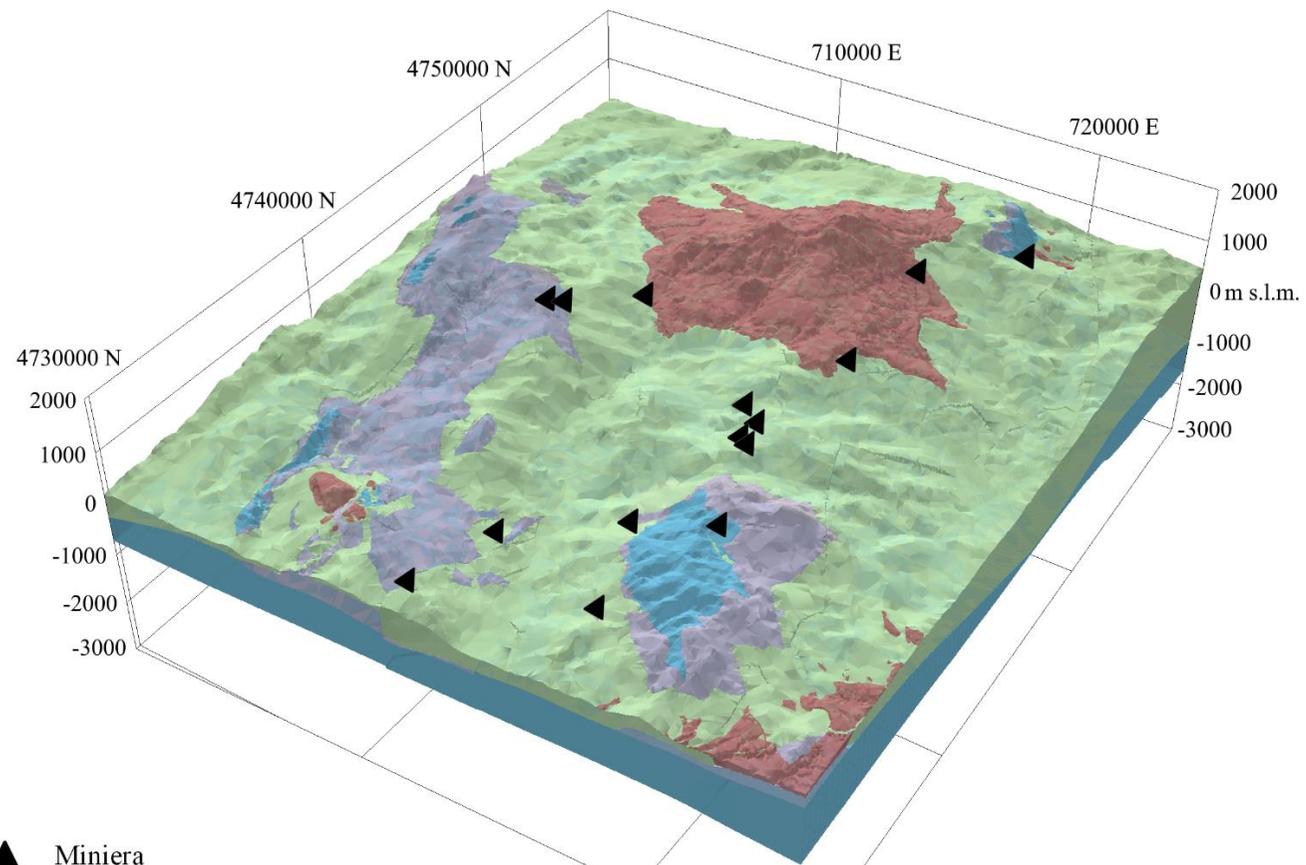


Figura 3.7. Vista 3D dei solidi prodotti per i complessi di permeabilità. (a) Complesso III, comprende le formazioni delle Nappe Toscane superiori, definite generalmente impermeabili. (b) Complesso IV, comprende le formazioni permeabili appartenenti alle Nappe Toscane.





- ▲ Miniera
- Complesso I - permeabile
- Complesso II - impermeabile
- Complesso III - generalmente impermeabile
- Complesso IV - permeabile

Figura 3.8. Vista 3D complessiva dei solidi prodotti per i complessi di permeabilità.

3.3 Machine learning

Il ML è un ramo dell'intelligenza artificiale che identifica modelli nei dati ed effettua previsioni da tali modelli (Caté et al., 2017). La capacità di apprendimento automatico degli algoritmi di ML migliora con l'esperienza (Samuel, 1959).

Oggi il ML viene ampiamente applicato nel campo delle geoscienze, in particolare per ridurre il rischio nell'ambito di esplorazioni geotermiche (Siler et al., 2017; Faulds et al., 2020) e per la prospezione mineraria (Cracknell et al., 2014; Rodriguez-Galiano et al., 2015; Karpatne et al., 2019; Sun et al., 2019).

Gli algoritmi di ML adottano tre principali approcci per l'apprendimento automatico: *Supervised Learning*, *Unsupervised Learning* e *Reinforcement Learning*.

- *Supervised Learning* (SL): questo approccio utilizza una serie di dati di tipo *labelled*, ovvero dati noti ed "etichettati", per addestrare un modello e applicarlo su altri dati per ottenere delle predizioni (Faulds et al., 2020).
- *Unsupervised Learning* (UL): gli algoritmi non richiedono, in questo caso, dati *labelled* in input ma vengono applicati approcci di *clustering* per il raggruppamento dei dati in base alla loro similarità o differenza (Melo & Li, 2021).
- *Reinforcement Learning* (RL): è un approccio di apprendimento basato su una logica di *reward*, ovvero una ricompensa ottenuta in base a quanto la predizione si avvicina alla realtà del dato (Forster, 2021).

In base alla tipologia di dati ottenuti in questa tesi, un approccio di tipo *Supervised* è stato selezionato per la predizione di valori geologici.

Gli algoritmi di *Supervised Learning* effettuano predizioni sia nel caso di regressioni che di classificazioni, restituendo un valore numerico nel primo caso e una classe/*label* nel secondo.

Nell'ambito di questa tesi, per effettuare le predizioni è stato utilizzato Scikit-learn, un modulo Python che integra un'ampia gamma di algoritmi di apprendimento automatico (Pedregosa et al., 2011). Questo pacchetto si concentra sulla possibilità di permettere anche a non specialisti di effettuare predizioni attraverso l'uso del ML. Tale modulo è stato utilizzato in collaborazione con il gruppo 3DOM della Fondazione Bruno Kessler di Trento, che si occupa di sviluppo di metodi e strumenti per l'interpretazione di dati 3D, nonché di metrologia ottica 3D.

Scikit-learn permette di utilizzare algoritmi differenti a seconda del tipo di statistica multivariata che si vuole applicare al proprio modello di dati e si basa sull'elaborazione *batch* di campioni, preferendola all'analisi del singolo campione (Lars et al., 2013).

3.3.1 Regressori

L'analisi di regressione è una tecnica di modellazione predittiva che permette di predire variabili numeriche y , tipicamente chiamate *target*, basata su una o multiple variabili x o predittori (Choi et al., 2020).

Il modello di regressione, quindi, ha l'obiettivo di costruire un'equazione matematica che definisca il target y come funzione del predittore x . Dopo l'apprendimento, o *training*, la funzione può utilizzare nuovi predittori per dedurre il valore y .

I modelli di regressione utilizzati per questo lavoro di tesi sono stati reperiti dalla *library* Scikit-learn, in collaborazione con 3DOM. Diversi regressori, individuati in base alla tipologia dei dati disponibili, sono stati selezionati e testati in questo lavoro di tesi.

Generalmente, in rapporto alla tipologia di dati e al problema da analizzare, è richiesta una personalizzazione dei modelli di intelligenza artificiale (Cracknell et al., 2014), soprattutto in ambito geologico in cui ci si trova a lavorare con matrici di punti tridimensionali (tensori) mentre la maggior parte di tali modelli è stata sviluppata a partire da problemi bidimensionali.

I modelli di regressione testati sono *Ordinary Least Squares Linear regressor* (OLS), *Multilayer Perceptron Regressor* (MPR), *Support Vector Regressor* (SVR), *Random Forest* (RF) e *CatBoost* (CB).

Ordinary Least Squares Linear regressor (OLS)

Questo modello è basato sul metodo dei minimi quadrati, ovvero la minimizzazione della somma dei quadrati delle distanze tra il valore osservato e la variabile. Il modello di regressione lineare permette di analizzare le relazioni tra la variabile indipendente x e la variabile dipendente y . Il modello viene impostato con la seguente equazione:

$$y = a_0 + a_1x + e \quad (7)$$

in cui a_0 è l'intercetta della retta di regressione con l'asse delle ordinate, a_1 è il coefficiente di regressione o pendenza della retta, e è l'errore casuale usato per esprimere l'effetto degli errori casuali sulla variabile dipendente (Rong & Bao-Wen, 2018).

L'analisi di regressione valuta a_0 e a_1 , osservando ogni coppia di valori (x_i, y_i) .

Matematicamente, a partire da coefficienti $a = (a_1, \dots, a_n)$, risolve problemi di tipo:

$$\min_a \|xa - y\|_2^2 \quad (8)$$

Le stime dei coefficienti per i minimi quadrati ordinari sono derivate tenendo conto dell'indipendenza delle caratteristiche. Se le colonne della matrice di progetto hanno una dipendenza lineare, le stime dei minimi quadrati diventano sensibili agli errori casuali, producendo una grande varianza (Hutcheson, 1999).

Multilayer Perceptron regressor (MPR)

Un perceptrone è un algoritmo di ML che prende in input una serie di caratteristiche e i relativi target e tenta di trovare una linea, un piano o un iperpiano che separi le classi rispettivamente in uno spazio bidimensionale, tridimensionale o iperdimensionale (Choi et al., 2020). Queste caratteristiche vengono adattate utilizzando quattro possibili funzioni di attivazione: sigmoide, tangente iperbolica, identità e unità lineare rettificata.

Quando vengono connessi perceptroni multipli, il modello assume una struttura uguale a quella dell'*Artificial Neural Network* (ANN), in cui una rete di “neuroni” contiene uno strato di nodi di input, uno strato di nodi di output (Yu, 2020) e una serie di strati intermedi nascosti. Un nodo è una struttura di dati che tiene traccia dello stato del sistema, del nodo precedente, o genitore, che lo ha generato, dell'azione applicata al nodo genitore per produrre il nuovo nodo e del costo del percorso effettuato per produrre il nodo a partire dallo stato iniziale del sistema (Yu, 2020). Per effettuare predizioni, il modello elabora le informazioni di livello in livello, analizzando e trasformando le informazioni da un nodo a quello successivo (Fig. 3.9). Questo processo viene denominato *feedforwarding* (Murtagh, 1991; Lecun et al., 2015; Choi et al., 2020).

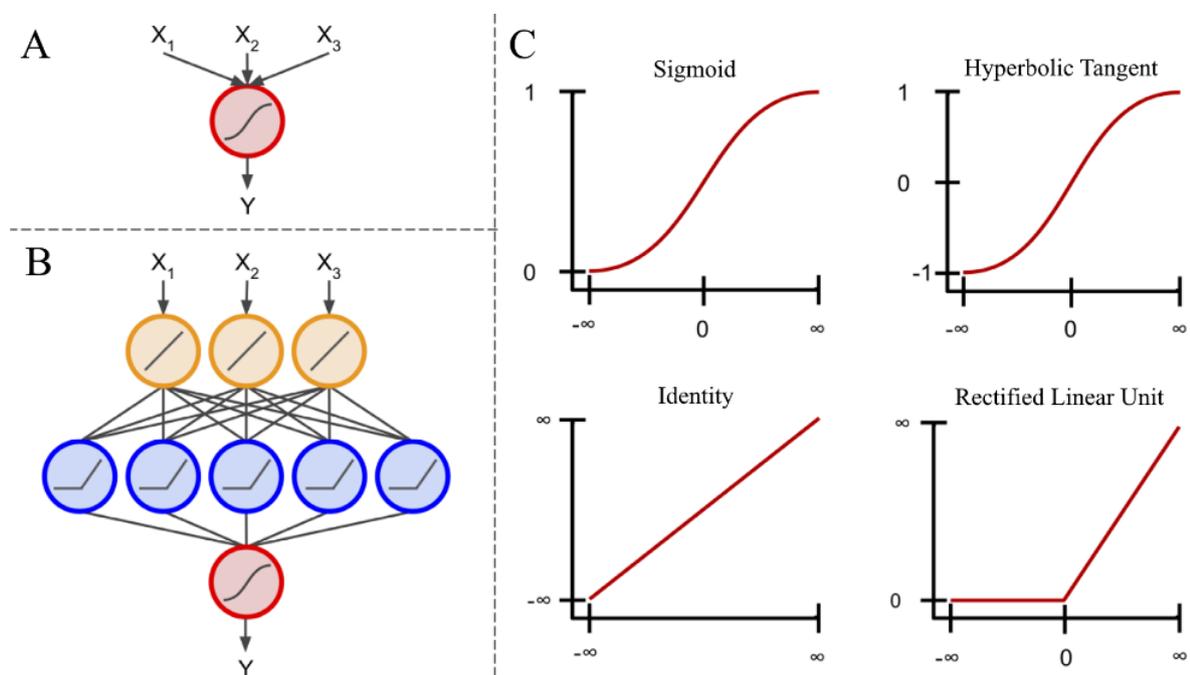


Figura 3.9. Componenti di una rete neurale. (A) Perceptrone, unità di base della rete neurale. In questo caso l'algoritmo utilizza una funzione sigmoide per ridimensionare input multipli in uno solo che va da 0 a 1. (B) Rete neurale che collega più perceptroni, in modo che l'output di un'unità venga utilizzato come input per un'altra. Queste unità non si limitano all'utilizzo della funzione sigmoide. (C) Funzioni di attivazione. Le funzioni sigmoide e tangente iperbolica scalano i dati di input rispettivamente tra 0 e 1 e tra -1 e 1. La funzione identità moltiplica l'input per un valore qualsiasi per produrre un output lineare. La funzione unità lineare rettificata svolge la stessa funzione di quella di identità, assegnando il valore 0 a tutti gli input < 0 (modificato da Choi et al. (2020)).

Ogni strato può contenere un numero qualsiasi di nodi, mentre il numero di nodi di output è vincolato all'obiettivo della predizione. Se si vuole ottenere una regressione, il modello genera funzioni di attivazione che trasformano ogni nodo di input nel nodo di output desiderato (Lecun et al., 2015; Choi et al., 2020).

Support Vector regressor (SVR)

Support Vector (SVR) è un approccio di ML supervisionato che apprende utilizzando una funzione di perdita simmetrica, che penalizza in ugual misura le stime errate (Awad & Khanna, 2015).

Il SVR permette di definire la quantità di errore accettabile nel nostro modello e trova una retta appropriata, o un iperpiano per dimensioni superiori, per adattarsi ai dati (Pedregosa et al., 2011). Confrontato con il metodo OLS, il cui obiettivo è minimizzare l'errore al quadrato, la funzione obiettivo di SVR è quella di minimizzare i coefficienti. La funzione minimizza:

$$\text{MIN } \frac{1}{2} \|w\|^2 \quad (9)$$

dove $|w|$ è il modulo del vettore normale alla superficie approssimata.

Il termine di errore viene gestito da vincoli imposti dall'operatore, in cui è possibile impostare il valore massimo di errore ε , che permettono di definire la regione di influenza della funzione. Tale valore può essere modificato per impostare l'accuratezza desiderata del modello. Per questo motivo, il modello prodotto dipende solamente da un sottoinsieme dei dati di addestramento, poiché la funzione ignora tutti i campioni la cui previsione si trova oltre il margine impostato o molto vicina ad esso (Smola & Schölkopf, 2004).

In un modello multidimensionale, il margine di errore viene definito da un cilindro, denominato ε -tube, che definisce la regione di validità della funzione. La regione esterna al cilindro, denominata ε -insensibile, viene definita dalla funzione di perdita. SVR riformula il problema di ottimizzazione per trovare il cilindro che meglio si avvicina alla funzione a valore continuo, bilanciando la complessità del modello e l'errore di previsione. Quindi, SVR è una funzione multiobiettivo costruita dalla funzione di perdita e dalle proprietà geometriche del cilindro (Awad & Khanna, 2015).

Random Forest regressor (RF)

Random Forest (RF) è un algoritmo di apprendimento supervisionato (*ensemble learning*) basato sull'apprendimento d'insieme (Breiman, 2001), che combina tra loro numerosi alberi decisionali (riducendo varianza e *over-fitting*) per predire il valore di una variabile.

Un albero decisionale (Fig. 3.10) è un metodo di apprendimento supervisionato non parametrico utilizzato per la classificazione e la regressione (Breiman et al., 1984). Un albero ha una struttura costituita da nodi radice, nodi decisionali e nodi terminali. Il nodo radice è il primo nodo dell'albero decisionale e contiene la caratteristica che rappresenta l'intero set di dati. Ogni nodo contiene una singola caratteristica utile per stabilire come avverrà la divisione del dataset in due parti. La divisione crea quindi due nodi, definiti rispettivamente nodo decisionale e nodo terminale. Il nodo decisionale contiene un'altra caratteristica e verrà a sua volta analizzato e diviso, mentre il nodo terminale, denominato "foglia", rappresenta l'output della previsione parziale. Il set di dati viene diviso quindi sulla base della decisione presa al nodo precedente. La costruzione dell'albero decisionale termina quando non sono più disponibili nodi decisionali.

Il processo decisionale si basa sul metodo del partizionamento ricorsivo che prevede la suddivisione ricorsiva dei dati in due partizioni binarie omogenee in base alle decisioni prese a livello di ogni nodo, costruendo in questo modo un albero decisionale.

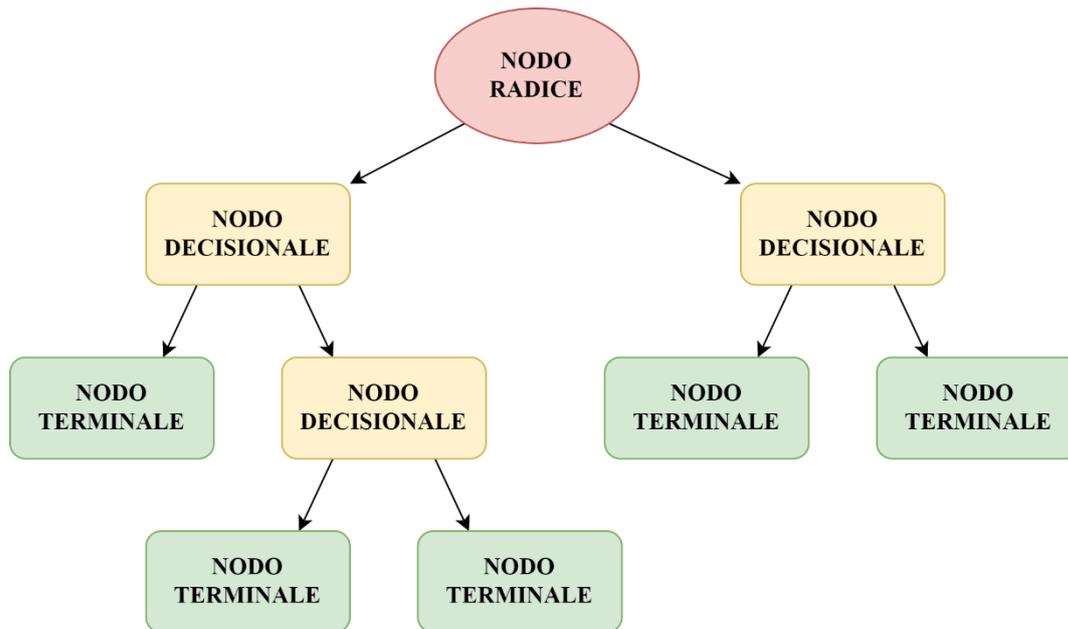


Figura 3.10. Struttura di un albero decisionale. La suddivisione del set di dati inizia dal nodo radice. L'obiettivo è creare un modello che prevede il valore di una variabile target apprendendo semplici regole decisionali dedotte dalle caratteristiche dei dati (modificato da Breiman et al. (1984)).

Il regressore RF utilizza questo approccio per creare una serie di alberi decisionali, facendoli crescere sulla base dei dati di addestramento forniti al modello (Choi et al., 2020). Per evitare la correlazione tra i diversi alberi, RF aumenta la diversità degli alberi facendoli crescere da diversi sottoinsiemi di dati di addestramento creati attraverso una procedura chiamata *bagging*. Il *bagging* è una tecnica utilizzata per la creazione di dati di addestramento che ricampiona in modo casuale il set di dati originale, sostituendo ogni volta un certo numero di valori (Breiman, 2001). In questo modo si ottiene una maggiore stabilità ed accuratezza della predizione, rendendo più robusto il modello di fronte a lievi variazioni nei dati di input.

Ogni albero, quindi, produce una predizione e il valore medio predetto dagli alberi viene utilizzato come predizione finale del modello.

CatBoost regressor (CB)

CatBoost è una implementazione del *gradient boosting*, modello che utilizza alberi decisionali binari come predittori di base (Prokhorenkova et al., 2019), adattandosi consecutivamente a nuovi modelli per fornire una stima più accurata della variabile ricercata.

Il *boosting* avviene costruendo nuovi modelli di apprendimento ad ogni nodo per permettere al modello di “prestare maggiore attenzione” ad eventuali errori commessi nella decisione precedente, costruendo ogni albero decisionale in modo graduale (Dorogush et al., 2018).

La costruzione avviene utilizzando una funzione di perdita predefinita per misurare l'errore in ogni passaggio e correggerlo in quello successivo, producendo alberi adattati fin quando la funzione di perdita viene ridotta al minimo. Poiché l'aumento del gradiente si adatta agli alberi decisionali in sequenza, gli alberi adattati impareranno dagli errori degli alberi precedenti e quindi ridurranno gli errori (Hancock & Khoshgoftaar, 2020).

La funzione principale del *boosting* è, quindi, quella di combinare una sequenza di molti modelli deboli, risultato di ciascun albero decisionale, per creare un modello predittivo forte.

Tabella 3.1. Lista delle funzionalità di ogni modello di regressione.

	Funzione
CB CatBoost regressor	Potenzia le interazioni tra diversi alberi decisionali iterando sugli errori commessi per correggerli al nodo successivo. Produce un modello predittivo a partire da un insieme di modelli predittivi deboli.
OSL Ordinary Least Squares Linear regressor	Minimizza la somma dei quadrati delle distanze tra valore osservato e variabile per tracciare la funzione che interpola al meglio il dataset.
MLP Multilayer Perceptron regressor	Funziona come una rete neurale e combina i diversi percettroni (funzioni sigmoidali di classificazione che separano i dati in classi) per produrre nodi di output
RF Random Forest regressor	Effettua divisioni binarie dei dati per creare alberi decisionali e li fa crescere in base al modello di addestramento. Il risultato è la media dei valori predetti da ogni albero.
SVR Support Vector Regressor	Produce un modello che dipende solo dal subset di addestramento, ignorando i campioni il cui valore predetto è vicino al loro target.

Compilato da Hutcheson (1999); Breiman (2001); Smola and Schölkopf (2004); Pedregosa et al. (2011); Prokhorenkova et al. (2019); Choi et al. (2020); Hancock and Khoshgoftaar (2020).

Riepilogando, ogni modello di regressione, quindi, è stato studiato per adattarsi ai dati di input in maniera differente al fine di produrre predizioni che si avvicinino il più possibile al dato reale. I diversi approcci statistici dei modelli di regressione analizzati sono riassunti in Tab. 3.1.

3.4 Costruzione dell'*input file* del distretto amiatino

Il file di input utilizzato per effettuare le predizioni è stato strutturato a partire da una tabella attributi, esportata da QGIS, che contiene tutto il dataset di punti reperiti in letteratura, insieme a una matrice tridimensionale di punti.

Inizialmente è stata creata una maglia di punti spazati 1km, drappeggiata poi al di sopra del DTM, ovvero posizionando ogni punto alla quota relativa 0. A questa maglia regolare di punti sono stati aggiunti anche i punti superficiali provenienti dalla letteratura, cioè pozzi, sorgenti e venute di gas. Tali punti sono poi stati moltiplicati in profondità, con una distanza relativa di 100m, in modo da formare una matrice di punti tridimensionale, per un totale di 24'398 punti.

Questa scelta è stata dettata dalla volontà di creare un numero ragionevole di punti, che non fossero troppo vicini tra loro, causando problemi di *overfitting*⁵, ma nemmeno troppo distanziati per non rischiare di perdere informazioni.

La matrice di punti è stata tracciata fino alla base del Complesso IV.

⁵ L'*overfitting* di un modello è una condizione in cui un modello statistico inizia a descrivere l'errore casuale nei dati piuttosto che le relazioni tra le variabili (Cohen & Jensen, 1997).

Partire da una struttura tipo “tabella attributi” permette di organizzare facilmente i dati su righe e colonne, in modo tale da attribuire ad ogni punto una serie di caratteristiche.

In seguito, ogni punto è stato corredato da un ID univoco, dal suo set di coordinate EPSG 32632 - WGS 84 / UTM zone 32N e da nove parametri denominati predittori (vedi sotto), che descrivono i parametri geologici e geochimici per quel punto considerato. L’ultima colonna prende il nome di incognita ed è l’oggetto della predizione.

3.4.1 Predittori

Un predittore è una variabile utilizzata per addestrare un modello di apprendimento specifico per prevedere qualcosa (Farella et al., 2021).

Per poter effettuare una predizione è necessario utilizzare differenti parametri che descrivono le caratteristiche geologiche e geofisiche (Faulds et al., 2020) associate ogni punto del dataset di input.

Per assegnare un valore ad ogni predittore della matrice di punti, sono stati utilizzati due strumenti denominati “*Sample raster values*” e “*Near 3D*”, che appartengono rispettivamente all’ambiente QGIS e ArcScene.

Il primo, “*Sample raster values*”, permette di assegnare ad ogni punto del dataset il valore corrispondente di una superficie raster di riferimento, eseguendo un campionamento di tale valore alle coordinate desiderate. È importante menzionare il fatto che questo strumento lavora in due dimensioni, quindi, lungo la verticale, i punti che hanno le stesse coppie di coordinate (x, y) avranno lo stesso valore campionato.

Il secondo, “*Near 3D*”, assegna ad ogni punto un valore di distanza da una *feature class* di riferimento. La distanza viene calcolata nelle tre dimensioni misurando la distanza minima tra ogni punto e la polilinea più vicina.

I predittori sono stati scelti sulla base delle conoscenze geologiche-giacimentologiche dell’area di studio, in particolare in relazione alla presenza di mineralizzazioni a cinabro. Sulla base della tipologia di variabile in oggetto, i predittori sono stati raggruppati in due macrocategorie: predittori giacimentologici e predittori di vicinanza.

Predittori giacimentologici

Le variabili raggruppate nel gruppo dei predittori giacimentologici descrivono caratteristiche geologiche intrinseche ed estrinseche di ogni punto, quali la permeabilità, la solubilità di Hg in funzione della temperatura, la temperatura, il gradiente geotermico, il flusso di calore e l’anomalia di gravità.

Permeabilità

La permeabilità è una proprietà che esprime la capacità del suolo, in condizioni di saturazione, di essere attraversato da un flusso d’acqua, in direzione verticale. La permeabilità dipende prevalentemente dalla porosità primaria del mezzo e dal grado di fratturazione delle litologie interessate, per cui litologie ad elevata porosità o intensamente fratturate, come le vulcaniti, i calcari fratturati e le evaporiti, presentano valori di permeabilità di ordini di grandezza maggiori rispetto a litologie poco porose come i depositi argillosi marini o lacustri.

I dati di permeabilità (κ) dei Complessi II, III e IV sono stati estratti dal lavoro di Bellani e Gherardi (2009), con riferimento al valore di permeabilità orizzontale. Il valore di permeabilità del Complesso I è stato estratto dal rilievo strutturale del M.te Amiata di Delcroix et alii (2006). I valori utilizzati per la costruzione del file di input sono stati inseriti come valore medio. Nell'area di studio, i dati disponibili per la permeabilità sono scarsi ma, sulla base di risultati di perforazione (Barelli et al., 2000), sappiamo che il *reservoir*, denominato come Complesso IV, è quello che presenta i valori maggiori di κ (Bellani & Gherardi, 2009).

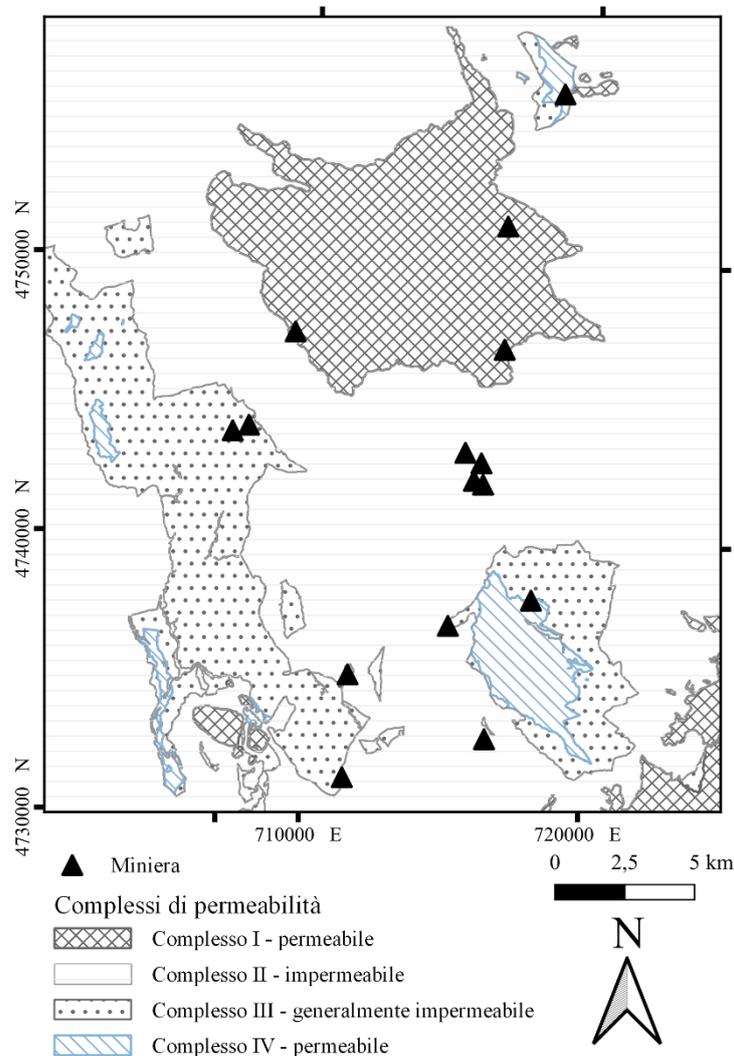


Figura 3.11. Carta degli affioramenti dei complessi di permeabilità. I poligoni sono stati tracciati a partire dalla classificazione dei complessi di Cataldi (1967), sulla base della carta geologica scala 1:25'000.

I valori di permeabilità sono stati assegnati ad ogni complesso come segue:

- Complesso I (C1): $\kappa = 3.16 \times 10^{-4} \text{ m}^2$
- Complesso II (C2): $\kappa = 5.57 \times 10^{-6} \text{ m}^2$
- Complesso III (C3): $\kappa = 2.79 \times 10^{-4} \text{ m}^2$
- Complesso IV (C4): $\kappa = 2.59 \times 10^{-3} \text{ m}^2$

Il valore di permeabilità è stato utilizzato come predittore per identificare le litologie sulla base delle caratteristiche idrauliche e per mettere in relazione il trasporto di fluidi con la presenza della mineralizzazione.

Solubilità di Hg in funzione della temperatura

La variabile che prende in esame la solubilità di Hg in funzione della temperatura è stata utilizzata al fine di comprendere i meccanismi di precipitazione del cinabro, in relazione alle altre variabili prese in esame.

Unico predittore calcolato esternamente a QGIS, il valore di solubilità di Hg in ogni punto della matrice di punti è stato calcolato a partire dai dati sperimentali di solubilità del Hg, espresso come $Hg^{0(aq)}$ riportati nella letteratura più recente (Clever et al. (1985), vedi anche Fig. 2.13). I valori sperimentali di $\log mHg^{0(aq)}$ determinati nell'intervallo 0-350°C (Tab. 2.2) sono stati utilizzati per calcolare una retta di regressione del tipo $\log mHg^{0(aq)} = mT + c$ (Fig. 3.12) e stimare un valore di solubilità del Hg in funzione della temperatura.

Tabella 3.2. Valori di solubilità sperimentali. Dati compilati da Clever et alii (1985) ed elaborati per ottenere il valore di concentrazione di $Hg^{0(aq)}$ nel fluido*. Si noti come a 50°C si ha una concentrazione di 0.12ppm mentre già a 150°C si ha una concentrazione di circa 2ppm, che indica un aumento del valore di solubilità di quasi due ordini di grandezza in 100°C.

T [K]	T [°C]	log m	Molalità [mol/kg]	Concentrazione mg/kg
273.15	0	-6.796275	1.59855E-07	0.032065224
323.15	50	-6.221275	6.00793E-07	0.120513106
373.15	100	-5.646275	2.25801E-06	0.452933328
423.15	150	-5.071275	8.48643E-06	1.702292863
473.15	200	-4.496275	3.18952E-05	6.39785331
523.15	250	-3.921275	0.000119874	24.04552581
573.15	300	-3.346275	0.000450531	90.37208006
623.15	350	-2.771275	0.001693265	339.6520799

* I punti sono stati utilizzati senza tenere conto del dato di pressione.

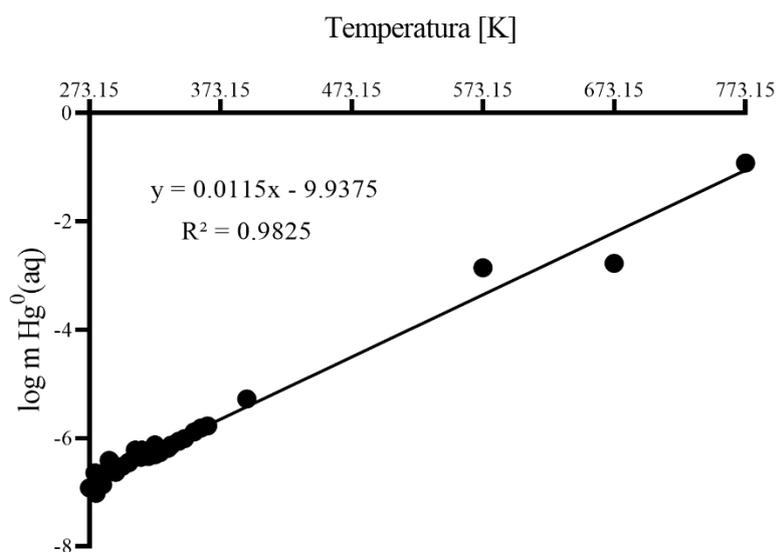


Figura 3.12. Retta di regressione estratta dai dati sperimentali dei valori di solubilità di $Hg^{0(aq)}$ in acqua.

Temperatura

La variabile temperatura è stata inserita nel calcolo di statistica multivariata effettuato dal momento che è uno dei parametri più importanti che controllano la solubilità del Hg.

Il valore di temperatura per ogni punto è stato calcolato a partire da tre parametri: il valore di temperatura alla quota del piano campagna (0m b.g.l.), il dato di temperatura rilevata in pozzo e la temperatura calcolata attraverso il gradiente geotermico misurato.

Il valore di temperatura al piano campagna è stato calcolato attraverso l'interpolazione dei dati provenienti dai dati delle sorgenti amiatine (Magi et al, 2019) che ha permesso di ottenere un dato puntuale di partenza per ottenere una mappa interpolata delle temperature superficiali. Tale mappa è stata prodotta utilizzando un'interpolazione di tipo Kriging.

Il valore di temperatura di ogni punto della griglia tridimensionale è stato calcolato a partire dalle carte delle temperature a -1000m e -2000m, campionando i valori su ciascuna superficie. È stato quindi calcolato il valore di gradiente geotermico per questi intervalli di profondità e successivamente è stato possibile determinare il valore di temperatura di ogni punto.

Le carte delle temperature (Fig. 3.13) mostrano un aumento della temperatura con la profondità localizzato in corrispondenza degli impianti geotermici attualmente in produzione. A 1000m di profondità dal piano locale campagna (Fig. 3.11B) è visibile una grande area in cui la temperatura supera i 210°C nella zona di Piancastagnaio. Tale area rimane visibile anche nella carta a -2000m (Fig. 3.11C), in cui è presente un aumento di temperatura anche in corrispondenza della zona di Bagnore, ma che presenta temperature molto simili a quelle nella zona di Piancastagnaio.

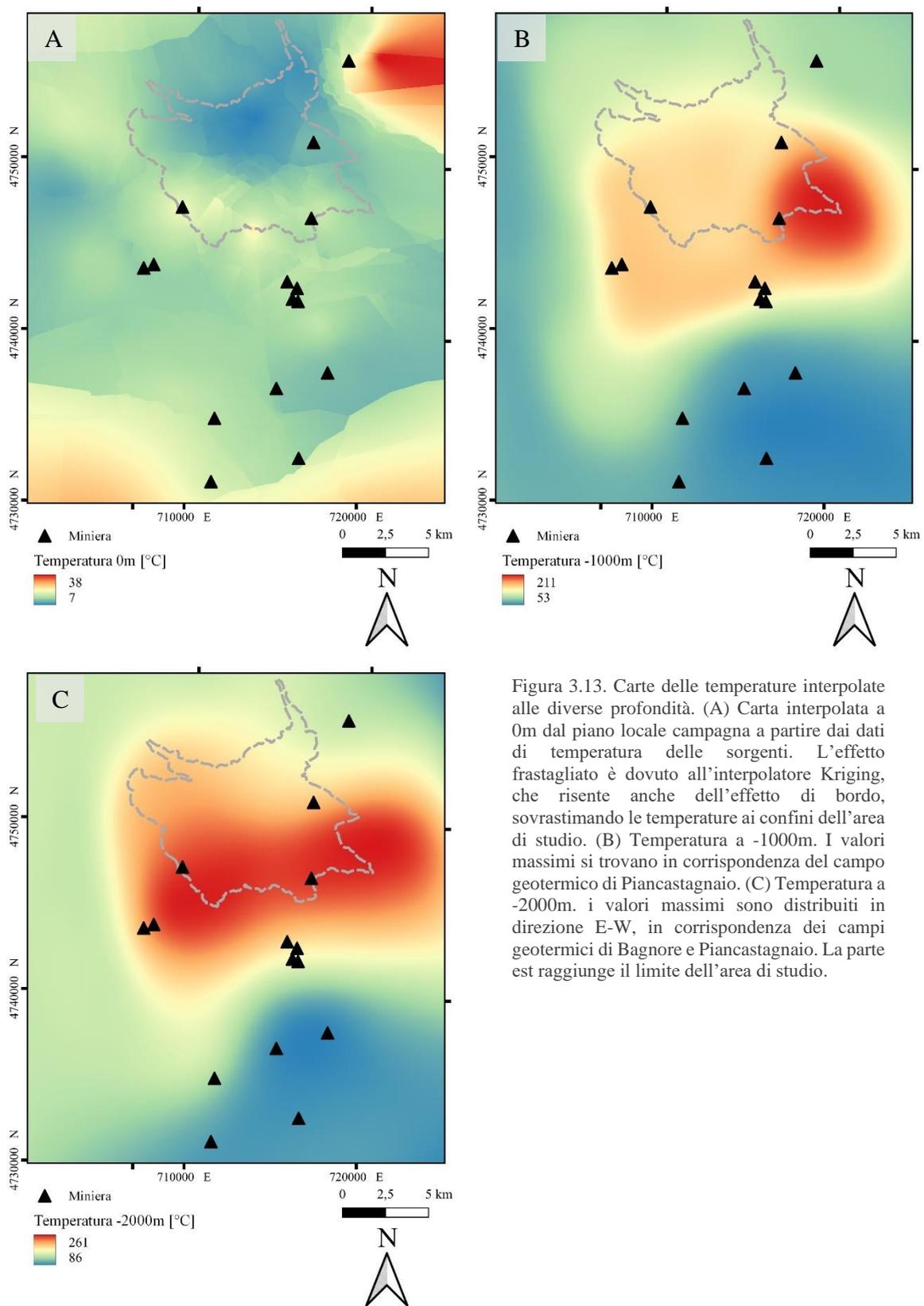


Figura 3.13. Carte delle temperature interpolate alle diverse profondità. (A) Carta interpolata a 0m dal piano locale campagna a partire dai dati di temperatura delle sorgenti. L'effetto frastagliato è dovuto all'interpolatore Kriging, che risente anche dell'effetto di bordo, sovrastimando le temperature ai confini dell'area di studio. (B) Temperatura a -1000m. I valori massimi si trovano in corrispondenza del campo geotermico di Piancastagnaio. (C) Temperatura a -2000m. i valori massimi sono distribuiti in direzione E-W, in corrispondenza dei campi geotermici di Bagnore e Piancastagnaio. La parte est raggiunge il limite dell'area di studio.

Flusso di calore

Il flusso di calore è un flusso di energia per unità di area nell'unità di tempo. La sua unità di misura quindi, secondo il SI, è $[W/m^2]$ (Beardsmore & Cull, 2001). A livello globale, il valore di flusso di calore medio superficiale varia tra 59 e $81 mW/m^2$, con valori maggiori in corrispondenza di dorsali oceaniche e minori in aree continentali. Localmente, si possono verificare importanti anomalie, come ad esempio in corrispondenza dei sistemi geotermici: nell'area del M.te Amiata è documentata un'importante anomalia di flusso di calore che raggiunge valori di $500-600 mW/m^2$. Tale anomalia è riconducibile alla presenza di un plutone granitico in raffreddamento, posto ad una profondità compresa tra 4 e 7 km, che rappresenta la sorgente di calore per il sistema geotermico epitermale del M.te Amiata (e.g., Gianelli et al, 1988).

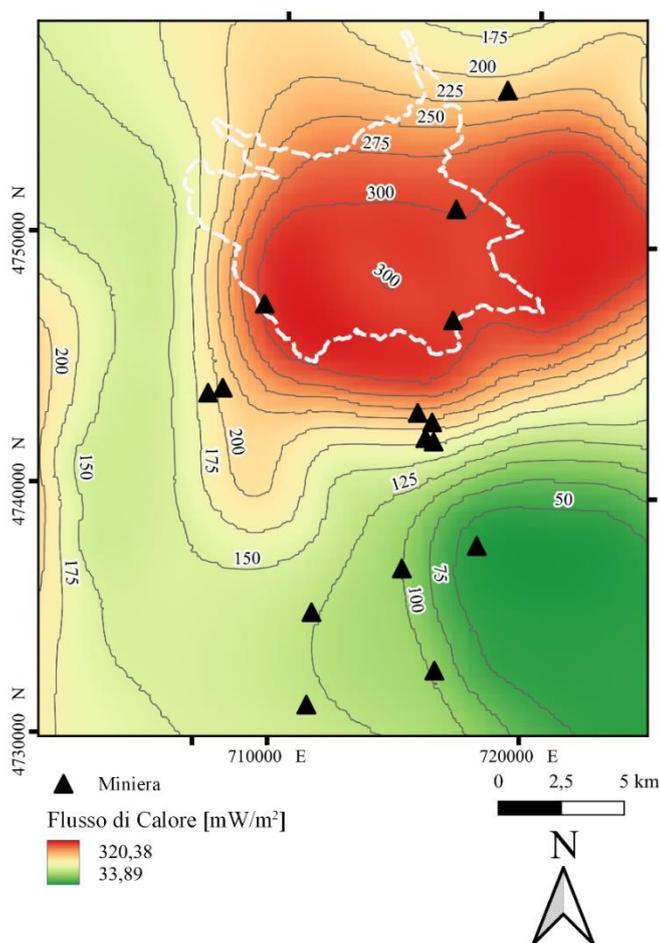


Figura 3.14. Valori di flusso di calore espressi in mW/m^2 per l'area amiatina (da Inventario delle risorse geotermiche nazionali, Ministero dello Sviluppo Economico (1994)).

La carta del flusso di calore (Fig. 3.14) mostra la presenza di una forte anomalia non centrata in corrispondenza del M.te Amiata ma allungata in direzione E-W, con valori massimi ad est del complesso vulcanico. I valori minimi si trovano nella porzione SE dell'area di studio. Al confine ovest dell'area è presente una fascia con valori che superano i $200 mW/m^2$, in corrispondenza delle faglie normali ad alto angolo che hanno portato alla formazione del Bacino di Cinigiano-Baccinello.

Gradiente geotermico

Il gradiente geotermico terrestre è una quantità vettoriale dipendente dalla distribuzione della temperatura nelle tre dimensioni (Beardsmore & Cull, 2001). Considerando una sola dimensione, il gradiente viene definito come la derivata della temperatura rispetto alla profondità e viene calcolato tramite l'equazione:

$$\frac{\partial T}{\partial z} = \frac{(T_2 - T_1)}{\Delta z} \quad (10)$$

Il valore medio di gradiente per la crosta superficiale è di $0.03^\circ\text{C}/\text{m}$, i.e. $30^\circ\text{C}/\text{km}$ (Barbier, 1997). Nell'area amiatina, tali valori di gradiente sono stati misurati da Calamai et alii (1970) solo nelle porzioni più distali rispetto ai centri eruttivi e in corrispondenza dei bacini sedimentari neogenici. I massimi valori di gradiente geotermico, calcolati da entro i primi 80m di profondità, con una media di 35m, si trovano in corrispondenza dei campi geotermici di Bagnore e Piancastagnaio, e hanno valori che superano i 4°C ogni 10m.

Il valore di gradiente geotermico è stato utilizzato per confermare le quote interpolate del tetto dei complessi di permeabilità, tramite il calcolo del gradiente dai dati di pozzo. In particolare, il gradiente calcolato dalle temperature rilevate nel pozzo Senna 2 (Fig. 3.15) ha evidenziato un aumento di pendenza a circa 500m di profondità, che corrisponde alla profondità relativa del tetto del Complesso IV.

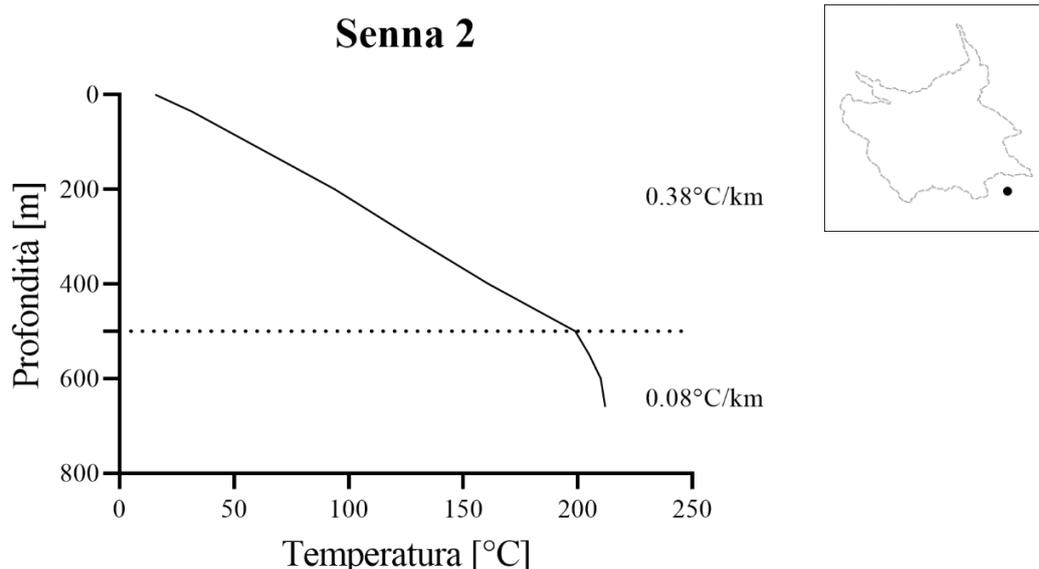


Figura 3.15. Profilo del gradiente geotermico calcolato per il pozzo Senna 2 (da BDIG – Regione Toscana. Sua localizzazione: pannello in alto a destra). La linea tratteggiata a -500m identifica la profondità corrispondente al cambiamento di gradiente: da $0.38^\circ\text{C}/\text{km}$ nel tratto meno profondo a $0.08^\circ\text{C}/\text{km}$ nel tratto più profondo.

Per poter assegnare a ogni punto della griglia il valore di gradiente corrispondente, sono state prodotte due carte di gradiente geotermico medio alle profondità di 1000 e 2000m (Fig. 3.16). Il calcolo è stato eseguito utilizzando i dati di temperatura a 1000 e 2000m di profondità, la carta di temperatura superficiale interpolata dai dati delle sorgenti, i dati delle temperature registrate in pozzo, dove disponibili, e la carta del gradiente geotermico superficiale da entro i primi 35m di profondità.

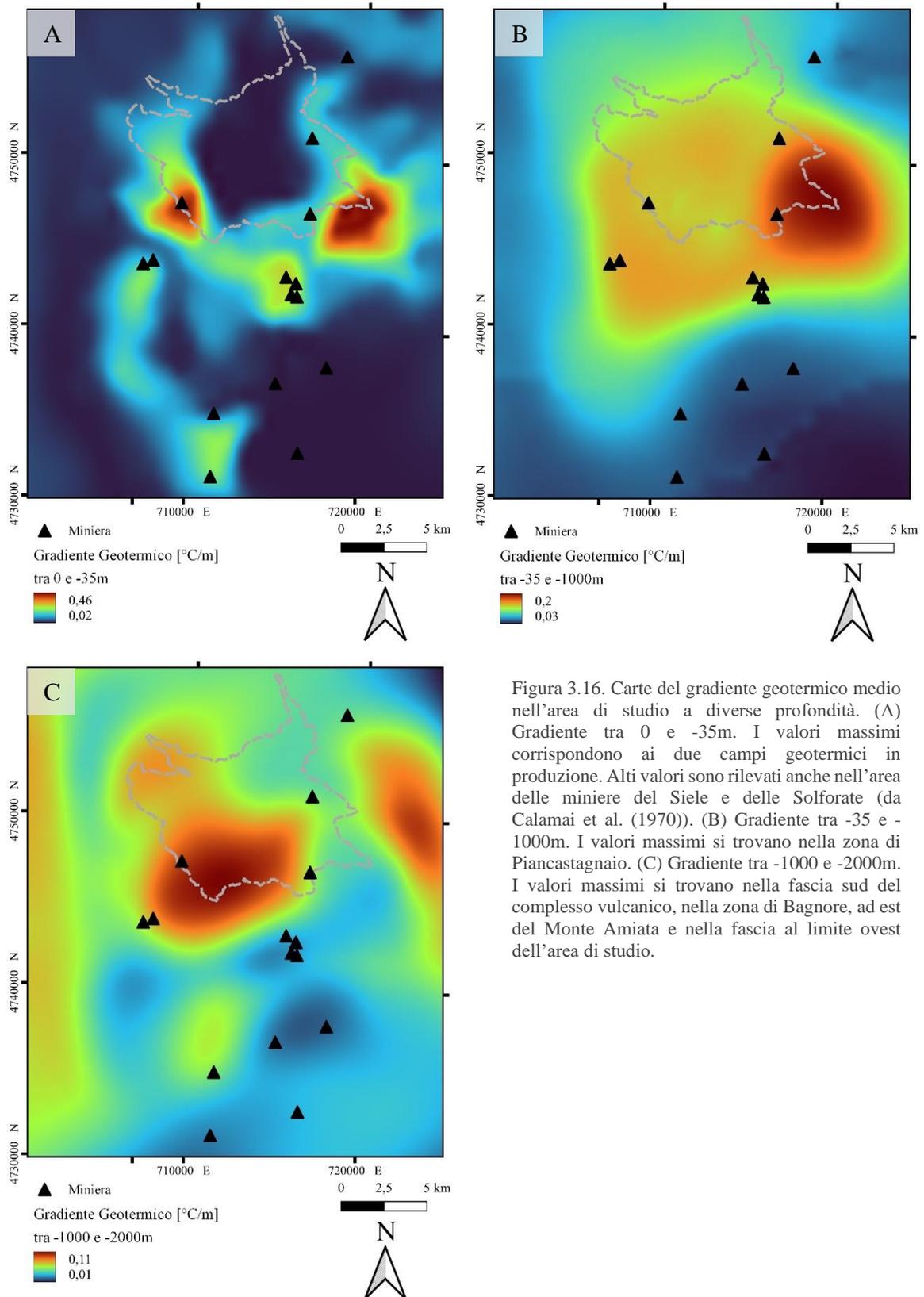


Figura 3.16. Carte del gradiente geotermico medio nell'area di studio a diverse profondità. (A) Gradiente tra 0 e -35m. I valori massimi corrispondono ai due campi geotermici in produzione. Altri valori sono rilevati anche nell'area delle miniere del Siele e delle Solforate (da Calamai et al. (1970)). (B) Gradiente tra -35 e -1000m. I valori massimi si trovano nella zona di Piancastagnaio. (C) Gradiente tra -1000 e -2000m. I valori massimi si trovano nella fascia sud del complesso vulcanico, nella zona di Bagnore, ad est del Monte Amiata e nella fascia al limite ovest dell'area di studio.

Anomalia di Bouguer

LaFehr (1991) ha affrontato il significato delle correzioni nel calcolo delle anomalie Bouguer considerate come la differenza, nel punto di misura, tra il valore osservato e il valore teorico per un dato modello terrestre. Sulla base di tale definizione, l'Anomalia di Bouguer può essere definita come la somma degli effetti di tutte le masse anomale terrestri in un dato punto di misura (Bernabini et al., 1994).

I valori di Anomalia di Bouguer dell'area di studio sono stati reperiti dalla Carta Gravimetrica d'Italia alla scala 1:250'000 in formato WMS, contenente una mappa in formato raster e un livello vettoriale costituito da isolinee con intervallo 2mGal (Ferri et al., 2008). Il valore di Anomalia di Bouguer è stato utilizzato da diversi autori (Gianelli et al., 1988; Bernabini et al., 1994; Orlando et al., 1994; Bellani & Gherardi, 2009) per individuare la posizione del plutone granitico senza ricorrere all'uso di tecniche più costose che richiedano campagne di indagine sul terreno, come ad esempio la tomografia sismica. L'andamento dei valori di Anomalia di Bouguer (Fig. 3.17) è simile a quello delle temperature e del flusso di calore nella parte a sud del M.te Amiata, in accordo con quanto sostenuto da Bellani e Gherardi (2009). Il valore minimo si trova nel margine NE dell'area di studio, in corrispondenza del Bacino di Radicofani.

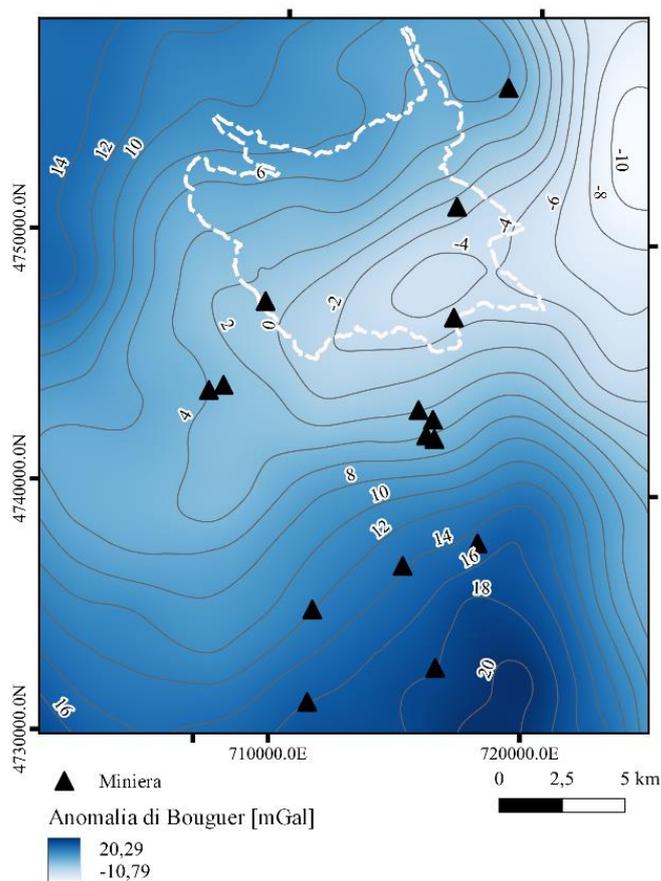


Figura 3.17. Carta dei valori di Anomalia di Bouguer.

In questo lavoro, il valore dell'Anomalia di Bouguer è stato utilizzato per individuare una correlazione tra la posizione dell'intrusione e quella delle miniere.

Predittori di vicinanza

I predittori di vicinanza sono le distanze misurate tra ogni punto della griglia tridimensionale e tre elementi geologici fondamentali dell'area di studio, cioè l'asse di piega più prossimo, la traccia di faglia più prossima ed il tetto del Complesso IV. In dettaglio, si è misurata su un piano orizzontale la distanza tra ogni punto della griglia e l'asse di piega più prossimo, la distanza su un piano orizzontale tra ogni punto della griglia e la traccia di faglia più prossima e la distanza verticale tra ogni punto ed il tetto del Complesso IV. Questi elementi sono stati utilizzati nell'analisi statistica per identificare in modo univoco e quantitativo il dato geologico-strutturale dell'area.

Distanza dall'asse di piega più prossimo

L'andamento delle strutture plicative nell'area del M.te Amiata ha una distribuzione generalmente N-S (Fig. 3.18).

L'importanza di queste strutture è data dal fatto che la presenza di una piega anticlinale permette lo sviluppo di trappole per i fluidi geotermici che, risalendo per contrasto di densità da livelli profondi, vengono a trovarsi in porzioni del *reservoir* sovrastate da un *cap* a bassa permeabilità che ne impedisce il flusso. Vengono così a crearsi, ai nuclei di anticlinali, delle trappole per fluidi che rimangono intrappolati fino ad un nuovo apporto di fluidi provenienti dal basso.

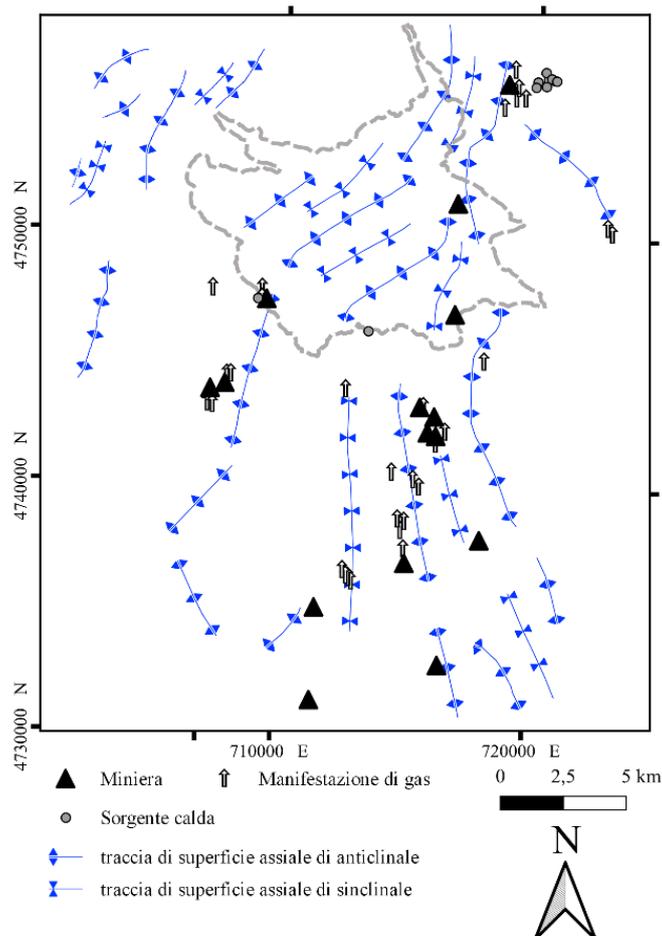


Figura 3.18. Tracce di piano assiale di piega ad andamento prevalente N-S (da Calamai et al. (1970)).

La distanza dall'asse di piega più prossimo è stata calcolata per ogni punto in base alla distanza minima bidimensionale tra la polilinea che descrive la traccia in carta dell'asse di piega e il punto stesso, utilizzando il comando *Near* all'interno di ArcScene.

Distanza dalla faglia più prossima

Le faglie costituiscono vie di trasporto per i fluidi idrotermali mineralizzanti che risalgono dalle zone più profonde e a contatto con i corpi magmatici. Tali vie di trasporto, quindi, sono di fondamentale importanza per i meccanismi che portano alla precipitazione di minerali da fluidi idrotermali (Anderson & Fairley, 2008).

Data la scala dello studio e data la difficoltà in generale a valutare con precisione la struttura di permeabilità di un sistema di faglie eterogeneo e diacrono, non è stato possibile quantificare il valore di permeabilità per ogni sistema presente nell'area (Fig. 3.19). Si è scelto quindi di utilizzare come variabile la distanza di ogni punto dalla faglia più vicina per poter includere nel calcolo statistico il legame tra permeabilità e flusso di fluidi. Minore è la distanza da una faglia, maggiore è stata considerata l'influenza sulla permeabilità. La distanza tra ogni punto e la faglia più prossima è stata calcolata utilizzando lo stesso approccio utilizzato per la distanza con gli assi di piega.

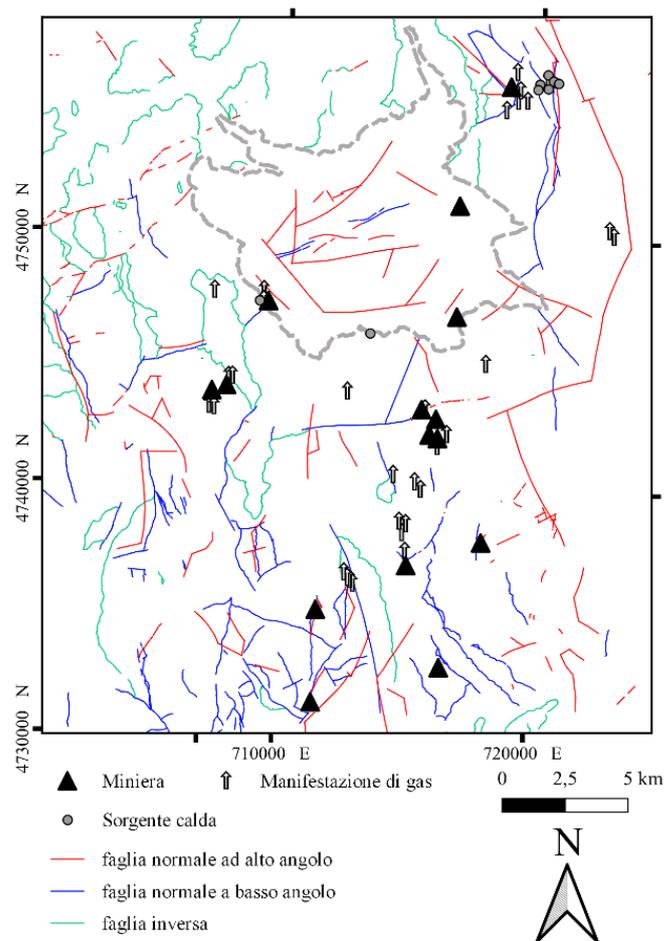


Figura 3.19. Tracce di faglia raggruppate sulla base del regime tettonico che le ha generate. Le faglie inverse sono da ricondurre alla fase compressiva pre-miocenica che ha interessato l'area appenninica; le faglie normali a basso angolo sono il risultato del primo evento estensionale (Langhiano-Serravalliano) che ha portato alla costruzione di strutture *flat-ramp-flat*; le faglie normali ad alto angolo comprendono le faglie generate in relazione al secondo e terzo evento estensionale (Tortoniano-Pliocene sup.) e le faglie vulcanotettoniche.

Distanza verticale dal tetto del Complesso IV

La distanza sulla verticale tra ogni punto del dataset e il tetto del Complesso IV è stata calcolata al fine di testare l'ipotesi di una correlazione tra la profondità del tetto del *reservoir* geotermico superiore, che coincide con il Complesso IV, e le mineralizzazioni a cinabro presenti in prossimità della superficie. La distanza dal tetto del Complesso IV è stata misurata mediante l'utilizzo della funzione *Near 3D* di ArcScene, che permette di calcolare la distanza tra un punto e una superficie raster nelle tre dimensioni.

3.4.2 Incognita (mineralizzazioni)

Per incognita si intende ciò che si vuole predire tramite l'applicazione del ML. Nel nostro caso, l'incognita da ricercare è considerata il valore di probabilità (da 0 a 1) che ogni punto del dataset ospiti le mineralizzazioni a cinabro.

Uno dei principi di funzionamento del ML, in particolare del *Supervised Learning*, è che i dati devono essere di tipo *labelled*, cioè che l'incognita da ricercare deve essere nota in almeno un subset di dati. Per questo motivo, all'interno dell'area di studio sono stati individuati e scelti i principali siti minerari attivi nel passato.

Data la scala di indagine di questo lavoro, e quindi l'impossibilità di effettuare una campagna di rilevamento di dettaglio, si è ritenuto di non considerare tutti gli affioramenti superficiali di cinabro ma di considerare solamente le miniere.

Inizialmente, si è pensato di considerare le miniere come un unico punto, con quota coincidente con il DTM, ma in questo modo i punti per il training sarebbero stati solamente 15, che è un numero troppo piccolo in confronto al numero totale di punti da predire (24398). Per questo motivo si è deciso di considerare, sulla base del lavoro di Forconi (2011), come valori positivi ($y = 1$) anche quelli che prendono il nome di ricerche, ovvero aree mineralizzate a cinabro risultate di scarso interesse economico, ma che in passato sono state oggetto di esplorazione e sondaggi ai fini di ricerca mineraria (Fig. 3.20).

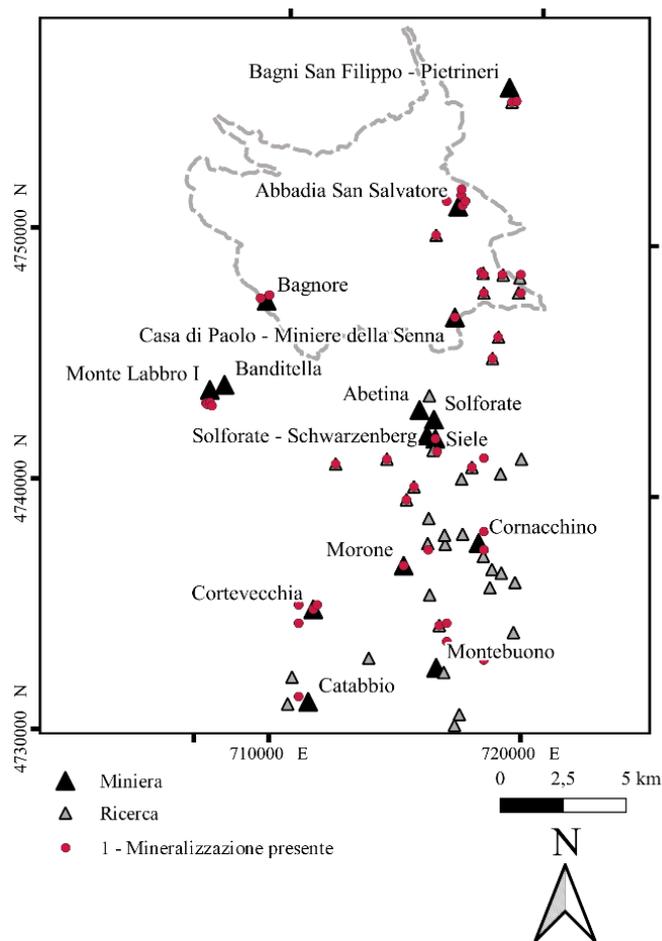


Figura 3.20. Localizzazione delle miniere e delle ricerche del distretto cinabrifero amiatino a cui sono sovrapposti i punti utilizzati come incognite per l'addestramento degli algoritmi di ML. Si noti come non tutti i punti rossi coincidano direttamente con miniere o ricerche poiché, rispettivamente, la localizzazione della miniera marca la posizione relativa all'ingresso principale mentre le mineralizzazioni si possono trovare in un'area poco distante, e non tutte le ricerche hanno identificato risorse di valore economico.

Al fine di aumentare ulteriormente il dataset di training, si è deciso di non identificare le miniere come singoli punti ma di comprendere i punti della griglia che si trovano nelle vicinanze dei siti minerari all'interno di un "volume di buffer". Tale volume è stato calcolato sulla base del diametro massimo di ogni miniera, a partire dalle planimetrie minerarie disponibili in letteratura, alla profondità massima raggiunta dai piani minerari e, dove disponibile, dalle quote corrispondenti alle tracce degli *shapefile* "Gallerie".

Per quanto riguarda le ricerche, il diametro del volume di buffer è stato imposto a 100m. Questa distanza è stata scelta sulla base della minima distanza tra due punti della griglia tridimensionale, che è 100m sulla verticale.

Tutti i punti dell'*array* creato in precedenza che ricadono all'interno di tali volumi sono stati considerati aventi incognita $y = 1$. L'utilizzo di tale approccio ha permesso così di ottenere 185 punti con $y = 1$. Considerato il notevole livello di conoscenza geologica dell'area (v. paragrafo 3.1 Dataset), è molto probabile che i punti utilizzati rappresentino l'effettiva distribuzione delle mineralizzazioni presenti nel territorio amiatino. È quindi da escludere che ci siano aree inesplorate o con un valore di y positivo in altre zone dell'area di studio.

3.5 Applicazione del Machine Learning

L'applicazione del ML avviene seguendo tre fasi, ad ognuna delle quali deve essere associata una certa tipologia di set di dati che prendono il nome di apprendimento, test e validazione dei risultati.

L'approccio utilizzato segue schematicamente il procedimento descritto in Fig. 3.21.

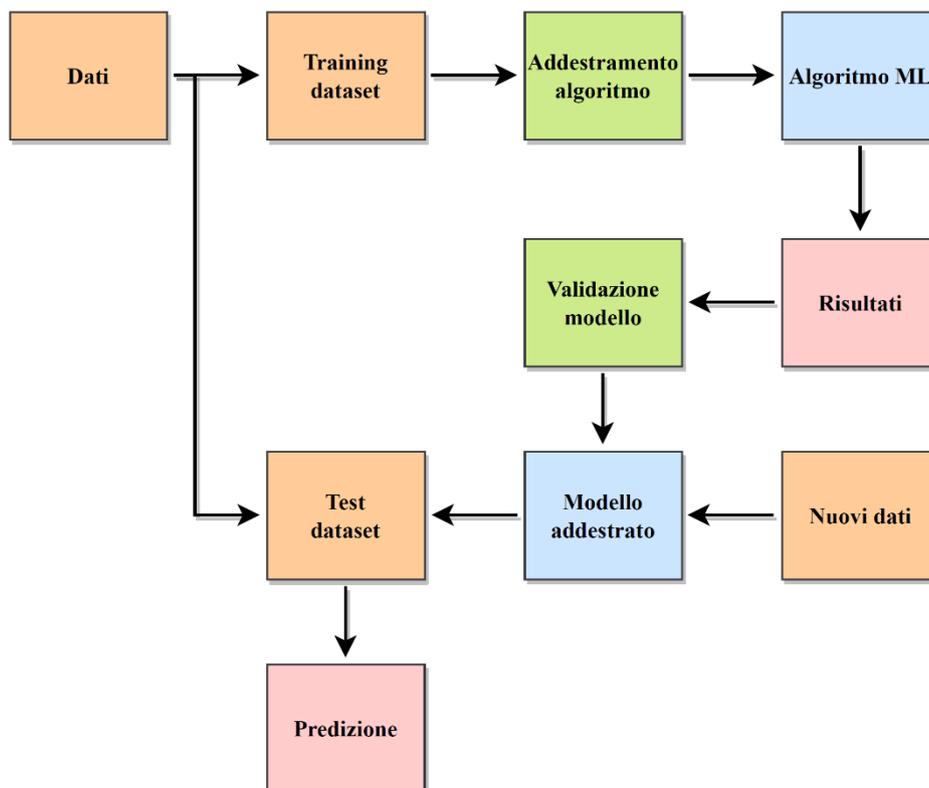


Figura 3.21. Diagramma a blocchi che rappresenta il flusso di lavoro svolto nell'ambito dell'utilizzo degli algoritmi di Machine Learning.

A partire da un dataset iniziale, l'applicazione del ML prevede la divisione del dataset in due subset con funzioni differenti. Il dataset di training viene utilizzato per addestrare l'algoritmo ai dati di input e per effettuare la validazione del modello di regressione, mentre il dataset di test contiene i punti ignoti che devono essere predetti. Successivamente alla fase di addestramento, il modello di regressione è addestrato e può essere applicato al dataset di dati ignoti e a nuovi dati di input per ottenere una predizione.

3.5.1 Apprendimento

Per poter eseguire una predizione, ogni modello di regressione deve essere "addestrato", ovvero deve essere fornito di adeguate istruzioni per poter apprendere tutte le caratteristiche che vengono assegnate ad ogni valore. Tale procedura prende il nome di *training*, o addestramento, e prevede la produzione di un subset dei dati disponibili costituito da dati completi, senza parametri mancanti (Pedregosa et al., 2011).

In questo studio il file di input è stato suddiviso in due parti distinte, una per l'addestramento e una per la predizione, costituite da due differenti insiemi di punti.

Poiché il dataset di addestramento deve contenere un numero simile di punti con incognita $y = 1$ e con $y = 0$, sono stati selezionati 185 punti con $y = 1$ e 244 con $y = 0$ (Fig. 3.22).

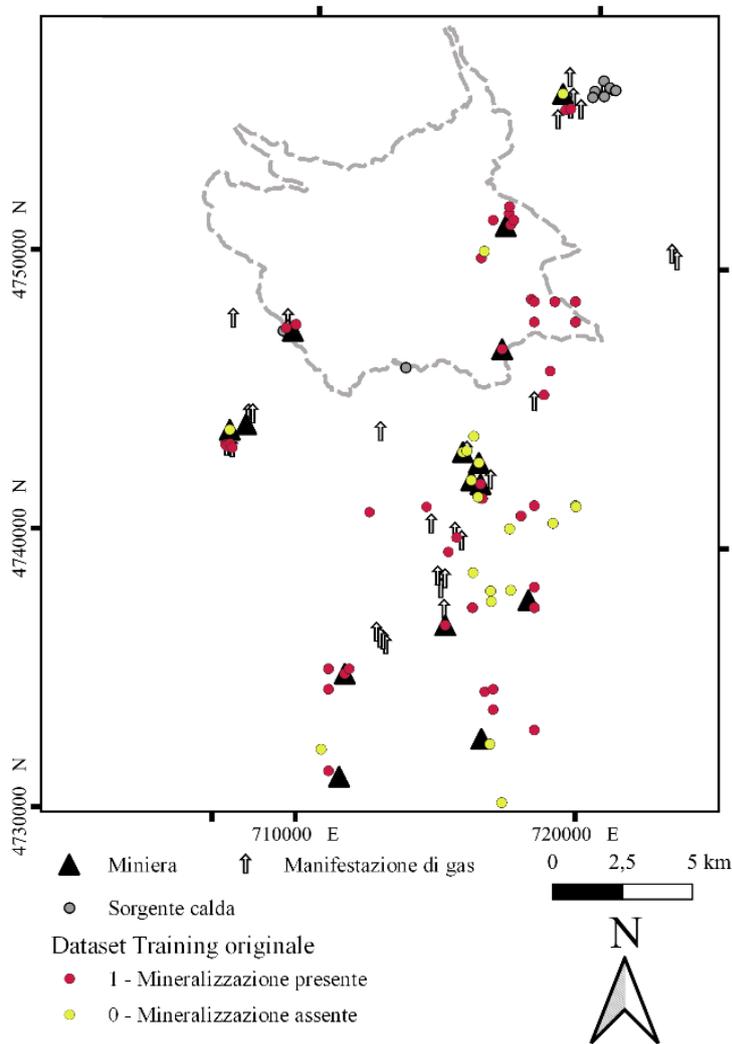


Figura 3.22. Localizzazione in carta dei punti utilizzati contenuti nel dataset di training originale.

È necessario che il subset di addestramento abbia una dimensione, intesa nel nostro caso come numero di punti, comparabile con quella del set di dati di *testing* per evitare di incorrere in errori sistematici che potrebbero influenzare la qualità della predizione (Chawla, 2005).

Vista la scarsa quantità di dati di training, si è deciso di utilizzare un approccio denominato *data augmentation* che ha l'obiettivo di aumentare il numero di valori all'interno del subset di addestramento.

Data augmentation

Il *data augmentation* è una tecnica utilizzata per aumentare la quantità di dati di training ed è basata sull'espansione del set di dati di training attraverso la creazione di copie di dati esistenti o artificiali (Farella et al., 2021). L'espansione del dataset di training si rende necessaria quando i dataset di training e di testing sono sbilanciati, ovvero quando il numero di valori contenuti nel dataset di training è molto inferiore a quello del dataset di testing, producendo quindi predizioni poco sensate (Chawla, 2005).

Generalmente, i nuovi punti aggiunti per l'apprendimento del modello vanno posti al di fuori dell'area di studio, per non creare sovrapposizioni. Nel caso di studio, non potendo aggiungere nuovi punti esternamente all'area a causa della mancanza di dati (alcuni predittori non potevano essere determinati al di fuori dell'area di studio), si è scelto di duplicare i punti esistenti.

I nuovi punti aggiunti al dataset di training sono stati nominati punti aumentati.

Il *data augmentation* è stato effettuato in due modalità:

- A. Copia dei valori del dataset di training iniziale, moltiplicando i punti per 7, per un totale di 3002 punti;
- B. Copia dei 185 dati con $x=1$, moltiplicati per 3, cui sono stati aggiunti 81 punti situati all'interno dei "volumi di buffer", i 244 punti con $y = 0$ e 1064 ulteriori punti con $y = 0$ scelti nell'area di studio in posizioni in cui non sono presenti mineralizzazioni. Il totale dei punti del dataset è di 3110 punti.

Il riepilogo della distribuzione dei valori all'interno dei subset di training aumentati è mostrato in Tab. 3.3.

Tabella 3.3. Numero di punti contenuti in ogni dataset di training.

Dataset	Punti con $y = 1$	Punti con $y = 0$	Totale
Training originale	185	244	429
Training dati aumentati A	1295	1708	3003
Training dati aumentati B	1330	1780	3110

Generalmente, l'approccio di tipo *data augmentation* prevede la creazione di un solo set di dati aumentati, tuttavia si è scelto di produrre due distinti set di dati aumentati per osservare l'influenza di set di punti distribuiti diversamente all'interno dello stesso contesto spaziale.

Training A – Dati aumentati con solo copie del training

L'espansione del set di dati di training moltiplicando i valori presenti nel set originale simula l'approccio Copy-Paste Augmentation descritto da Ghiasi et alii (2021), che viene utilizzato prevalentemente nell'analisi di immagini. Un approccio di questo tipo permette di aumentare il numero di valori all'interno del dataset di training senza avere la necessità di reperire ulteriori informazioni sia all'interno sia all'esterno dell'area di studio.

La disposizione dei punti coincide con quella dei punti di training originali (Fig.3.22).

Training B – Dati aumentati con copie e aggiunte

Il secondo set di dati aumentati è stato costruito moltiplicando i dati con $x = 1$ del set originale e aggiungendo nuovi dati con $x = 0$. Questi ultimi sono stati scelti tra i dati di pozzo reperiti in letteratura, in cui i log riportavano pozzi come sterili.

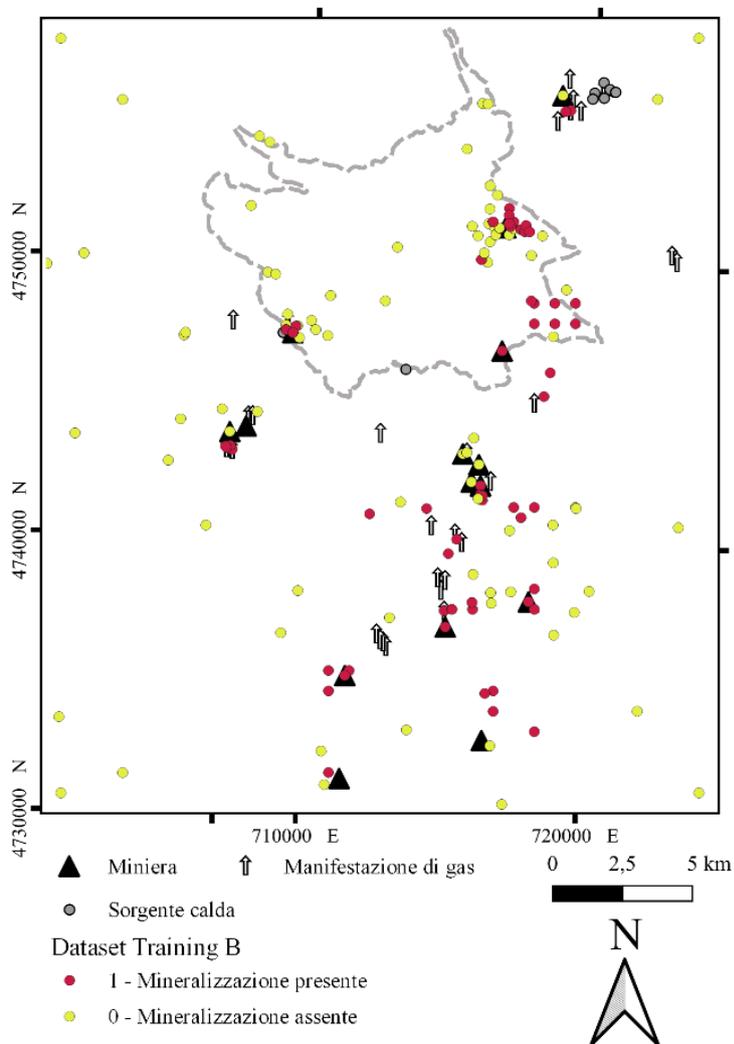


Figura 3.23. Localizzazione in carta dei punti appartenenti al dataset di training aumentato B. Nella zona della ricerca Indovina, i punti sono disposti in maniera regolare perché appartengono alla griglia di punti costruita per effettuare le predizioni.

Come mostrato in Fig. 3.23, i dati aggiunti all'interno di questo subset sono distribuiti più uniformemente all'interno dell'area di studio.

3.5.2 Test

Mediante l'utilizzo di Scikit-learn in ambiente Python, sono stati effettuati differenti "test" sui dati, ognuno dei quali comprende l'utilizzo dei cinque modelli di regressione descritti nel paragrafo 3.3.1.

Ogni test è stato effettuato imponendo all'algoritmo di utilizzare diverse combinazioni di predittori per ogni predizione, la cui selezione è a discrezione dell'operatore.

Le combinazioni di predittori utilizzate per l'esecuzione dei test sono schematizzate in Tab. 3.4.

Tabella 3.4. Combinazioni di predittori utilizzati nei test effettuati.

	Predittori giacimentologici					Predittori di vicinanza			
	Permeabilità	Solubilità	Flusso di calore	Temperatura	Gradiente termico	Anomalia di Bouguer	Distanza pieghe	Distanza faglie	Distanza tetto C4
Test 1	x	x	x	x	x	x			
Test 2							x	x	x
Test 3	x	x	x	x	x	x	x	x	x
Test 4	x	x					x	x	x
Test 5	x	x						x	
Test 6	x	x	x					x	
Test 7	x	x		x				x	
Test 8	x	x		x			x	x	

Sommando i test eseguiti a partire dai tre diversi dataset di training, in totale sono stati eseguiti 24 test, utilizzando per ogni test le seguenti combinazioni di predittori:

- Test 1: considera solo i predittori giacimentologici;
- Test 2: considera solo i predittori di vicinanza;
- Test 3: considera tutti i predittori;
- Test 4: considera permeabilità, solubilità di Hg e i predittori di vicinanza;
- Test 5: considera permeabilità, solubilità di Hg e distanza dalla faglia più prossima;
- Test 6: considera permeabilità, solubilità di Hg, flusso di calore e distanza dalla faglia più prossima;
- Test 7: considera permeabilità, solubilità di Hg, temperatura e distanza dalla faglia più prossima;
- Test 8: considera permeabilità, solubilità di Hg, temperatura, distanza dalla faglia più prossima e distanza dall'asse di piega più prossimo.

Le combinazioni dei primi tre test sono state scelte per prendere confidenza con l'utilizzo di un approccio tramite ML mentre quelle dei test successivi sono state scelte per cercare di discriminare quale insieme di fattori giacimentologici e strutturali può influenzare maggiormente la precipitazione di cinabro.

3.5.3 Validazione

Per valutare il rendimento degli algoritmi di ML e la qualità delle predizioni prodotte si ricorre ad una procedura, denominata *Random Split Training* o *Train-Test split* (Kohavi, 1995; Pedregosa et al., 2011), che prevede una ulteriore suddivisione del dataset di training, utilizzata per la costruzione di un modello di validazione.

Tale procedura consiste nell’inserimento di una colonna, in coda ai predittori, all’interno del set di dati di training, denominata “*split*”, contenente valori 0 e 1. Il valore 1 viene assegnato in maniera casuale, con un rapporto 30/70, ai punti di training (Farella et al., 2021).

I punti le cui celle della colonna *split* contengono un valore uguale a 0 sono rimasti all’interno del dataset di training, mentre i punti ai quali è stato assegnato un valore uguale a 1 sono stati spostati in un dataset di test.

La suddivisione in dataset di training e dataset di test ha permesso rispettivamente di stimare i parametri dei modelli predittivi e valutarne l’accuratezza. Questo procedimento viene in genere utilizzato per valutare la qualità della predizione tramite l’analisi di dati non utilizzati durante la fase di addestramento del modello (Farella et al., 2021). Tramite Scikit-learn utilizzato in ambiente Python, ogni modello di regressione calcola i valori dei parametri di validazione a partire dai dati di training e dai risultati di test.

Poiché è stato utilizzato un approccio di aumento dei dati è necessario specificare che i punti aggiunti non possono rientrare all’interno del dataset di test ma devono essere spostati nel dataset di training, in quanto il loro utilizzo nei calcoli di accuratezza potrebbe invalidare i risultati. Per identificare tali punti, ad ognuno è stato assegnato un valore di *split* uguale a -1.

Le metriche utilizzate per valutare la qualità dei modelli predittivi sono l’errore medio assoluto (MAE), l’errore quadratico medio (RMSE) e il coefficiente di determinazione R^2 (Farella et al., 2021).

Assegnando lo stesso peso a tutti gli errori, il MAE permette di misurare la distanza tra il valore predetto e il valore reale (Willmott & Matsuura, 2005). È descritto dalla formula:

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (11)$$

con

$$|e_i| = |x_i - y_i| \quad (12)$$

dove y_i è il valore predetto e x_i è il valore reale. MAE utilizza la stessa scala dei dati misurati.

L’RMSE è una misura d’errore assoluto che amplifica i valori maggiori al fine di individuare metodi che presentano gli errori più significativi (Chai & Draxler, 2014) ed è descritta dalla formula:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (13)$$

L’ R^2 , definito come coefficiente di determinazione, descrive la correlazione tra la variabile dipendente e la variabile indipendente. Può essere interpretato come la proporzione della

varianza nella variabile dipendente che può essere predetta tramite le variabili indipendenti (Chicco et al., 2021).

Tanto più il valore di R^2 è prossimo all'unità, tanto maggiore sarà la correlazione tra il dato reale e quello predetto. È descritto dalla formula:

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \quad (14)$$

Per poter stabilire quale tra i regressori utilizzati per le operazioni di ML sia il migliore, è necessario fissare un obiettivo di accuratezza. Principalmente si utilizzano MAE e RMSE per valutare l'accuratezza di un modello di regressione poiché quantificano la distanza media tra il valore predetto e il valore reale, permettendo una facile comprensione dei risultati. Per questo motivo, più bassi saranno i valori di MAE e RMSE, maggiore sarà l'accuratezza della predizione. Sulla base dell'obiettivo di accuratezza proposto da Lary et al. (2016), l'accuratezza del modello può essere definita buona per valori di coefficiente di determinazione (R^2) compresi tra 0.80 e 0.97.

Un ulteriore parametro di valutazione delle predizioni, tipico degli approcci di ML, consiste nella valutazione dell'importanza dei singoli predittori sul calcolo statistico. Questa procedura viene effettuata automaticamente da Scikit-learn, che produce una stima percentuale del peso relativo che ogni predittore ha sul calcolo del risultato, e prende il nome di recursive feature elimination (RFE). Lo scopo di questa procedura è quello di identificare ed eliminare i predittori irrilevanti ai fini di calcolo per velocizzare la procedura di predizione complessiva (Farella et al., 2021). In questo studio, la tecnica RFE non è stata utilizzata per eliminare variabili irrilevanti ma piuttosto per individuare le variabili che hanno maggiore importanza sul calcolo, a partire da combinazioni diverse di predittori. Le percentuali di importanza dei singoli predittori sulle predizioni possono essere utilizzate per costruire istogrammi utili al confronto delle varie combinazioni di predittori.

3.6 Elaborazione delle predizioni

I risultati delle predizioni effettuate tramite l'applicazione del ML al file di input descritto nelle sezioni precedenti sono stati elaborati sia all'interno dell'ambiente QGIS, sia tramite l'utilizzo del software Esri ArcGIS Pro, che permette di effettuare elaborazioni geostatistiche tridimensionali più avanzate rispetto a quelle disponibili in ArcScene. La scelta di utilizzare entrambi i software è stata dettata dalla necessità di produrre carte bidimensionali per mostrare i risultati ottenuti combinata alla sperimentazione di differenti applicazioni geostatistiche in combinazione con rappresentazioni tridimensionali di dati geologici.

I risultati delle predizioni sono stati utilizzati per costruire un file di output che è identico a quello di input, ma che ha in aggiunta i valori delle 5 predizioni effettuate dai regressori. In una fase successiva, i file di output di ogni test sono stati caricati in QGIS per la rappresentazione bidimensionale.

Per ogni test e per ogni regressore sono stati prodotti due insiemi di punti, uno contenente i valori predetti al piano campagna ed un altro i valori predetti alla profondità di 100m.

Una volta identificati i punti, si è effettuata l'interpolazione utilizzando un interpolatore Kriging che utilizza un modello di semivariogramma sferico con raggio di ricerca fisso a distanza di 1000m. In questo modo, per ogni test sono state prodotte cinque immagini per entrambe le quote analizzate, che rappresentano il valore risultante dalle differenti operazioni di regressione effettuate attraverso gli algoritmi di ML. Questo procedimento è

stato ripetuto per tutti i file di output, per un totale di 170 interpolazioni prodotte che hanno prodotto altrettante immagini. Tali immagini (vedi capitolo seguente) sono state raggruppate mediante un codice assegnato ad ogni test in base alla combinazione di predittori utilizzati per la predizione.

Una volta prodotte, le carte di probabilità di presenza di mineralizzazioni sono state studiate dal punto di vista grafico e valutate utilizzando i parametri di valutazione della bontà delle predizioni. Di seguito, si è scelto di presentare soltanto le immagini che mostrano la migliore correlazione con i dati di input. I test con migliore correlazione sono poi stati inseriti, in formato vettoriale, in ambiente ArcGIS Pro per essere analizzati effettuando operazioni geostatistiche. Queste analisi sono volte alla produzione di modelli tridimensionali mediante l'interpolazione di dati puntuali considerando non solo le coppie di coordinate (x, y) ma considerando anche i dati sull'asse verticale.

La funzione utilizzata per produrre interpolazioni di questo tipo è Kriging Bayesiano Empirico 3D (EBK3D), cioè un metodo di interpolazione geostatistica che automatizza gli aspetti più complessi della costruzione di un modello Kriging valido, vincolando la scelta dei parametri di ricerca alla tipologia dei dati di input. Il EBK3D utilizza quindi un modello predittivo per interpolare tra loro i punti più vicini e calcolare un semivariogramma per fare previsioni in posizioni in cui non sono presenti dati (Krivoruchko, 2012).

La risoluzione in output di un EBK3D dipende dalla risoluzione dei dati di input ma può essere forzata fino alla risoluzione verticale di 1m, a discapito di lunghi tempi di elaborazione software. Per questo motivo si è scelto di esportare i dati tridimensionali attraverso superfici raster, spaziate 100m, e successivamente elaborate attraverso ArcScene per produrre modelli tridimensionali che mostrano la distribuzione dei valori di probabilità a profondità fino a 1000m.

Presentazione dati

Le immagini prodotte attraverso l'interpolazione Kriging delle predizioni, mostrate e confrontate tra loro, sono state raggruppate sulla base del test effettuato, espresso come combinazione di variabili (predittori).

Per ogni test, sono state prodotte cinque immagini per i due gruppi di quote analizzati, piano campagna e -100m rispetto al piano campagna, che rappresentano il valore risultante dalle differenti operazioni di regressione effettuate attraverso gli algoritmi di ML.

Ogni figura, che rappresenta un test, è composta quindi da cinque immagini che descrivono i risultati dei diversi modelli di regressione utilizzati per lo stesso set di dati di training e le stesse combinazioni di predittori. Le combinazioni di predittori per ogni test sono elencate in Tabella 3.4.

I cinque regressori sono disposti come segue:

- (a) e (f) CatBoost regressor (CB);
- (b) e (g) Ordinary Least Square regressor (OLS);
- (c) e (h) Multilayer Perceptron regressor (MPR);
- (d) e (i) Random Forest regressor (RF);
- (e) e (l) Support Vector Regressor (SVR).

Ad ogni test è stato assegnato un codice identificativo sulla base del dataset di training utilizzato per l'addestramento degli algoritmi:

- T_{0-8} = Training originale
- TA_{1-8} = Training aumentato A
- TB_{1-8} = Training aumentato B

Dei 24 test effettuati solamente 17 hanno prodotto risultati. I test TA_2 , TA_4 , TA_5 e i test TB_1 , TB_2 , TB_3 , TB_4 non hanno prodotto risultati.

Sulla base della profondità relativa, le immagini sono state poste in modo da proporre un confronto non solo tra i diversi modelli di regressione utilizzati a parità di variabili calcolate ma anche per comprendere come cambiano le distribuzioni di probabilità al variare della profondità.

Nelle immagini vengono mostrate le posizioni delle miniere esistenti e delle manifestazioni geotermiche. Non sono state inserite le tracce delle faglie e delle pieghe perché non sarebbero visibili nelle immagini prodotte. Immagini a più alta risoluzione sono riportate in allegato alla tesi.

La valutazione della qualità delle correlazioni tra le predizioni e i dati esistenti si mostra attraverso i parametri di accuratezza di ogni test effettuato.

T01

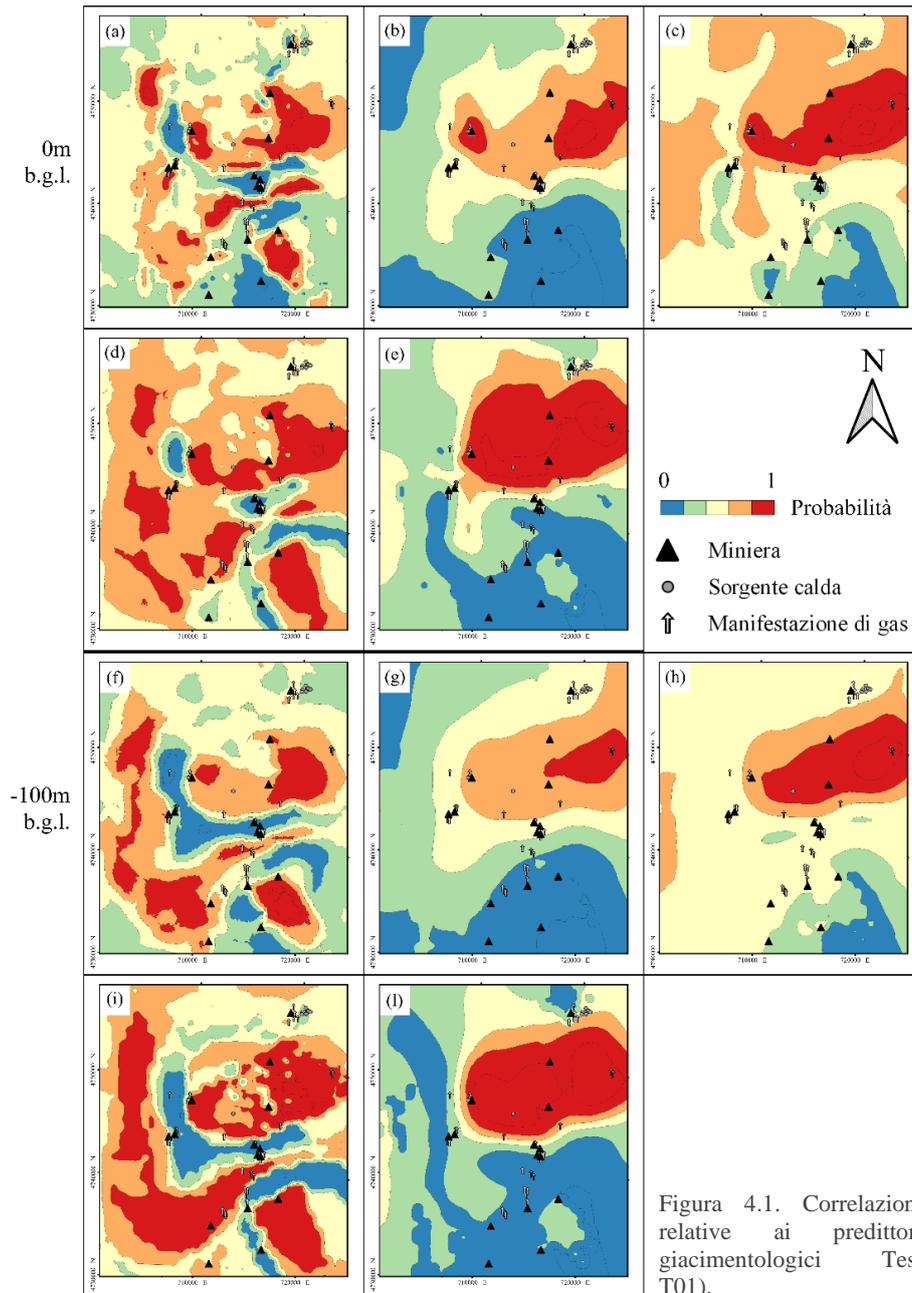
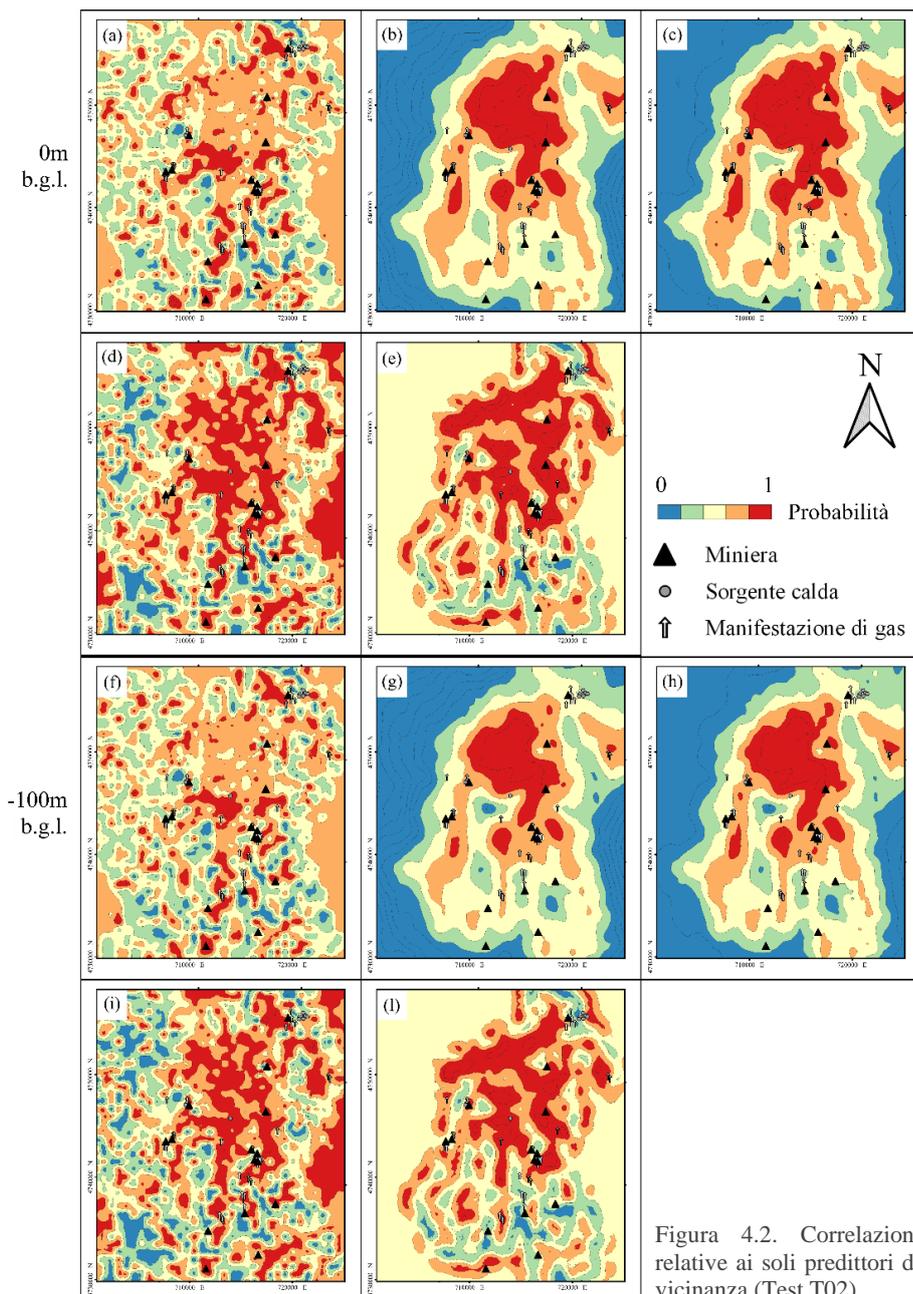


Figura 4.1. Correlazioni relative ai predittori giacimentologici Test T01).

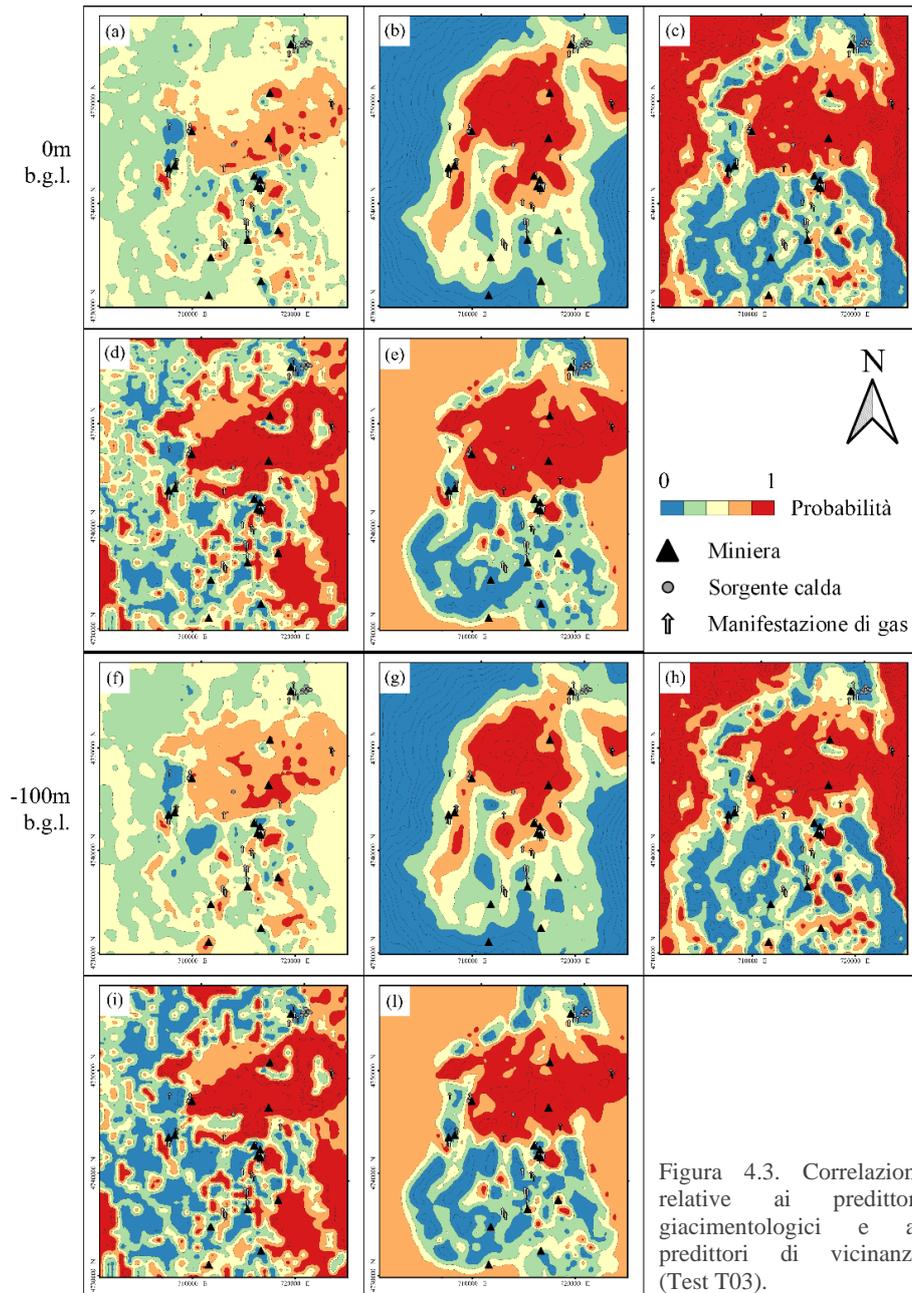
Il test T01 (Fig. 4.1 – che considera solo predittori giacimentologici) ha prodotto pattern di distribuzione di probabilità che non presentano in nessun caso correlazioni con la distribuzione spaziale delle miniere. I regressori Ordinary Least Square (Fig. 4.1b), Multilayer Perceptron (Fig. 4.1c) e Support Vector (Fig. 4.1e) mostrano grandi aree ad alta probabilità che comprendono il M.te Amiata ed il territorio ad est, con un andamento simile a quello del flusso di calore (Fig. 3.14). I regressori CatBoost (Fig. 4.1a) e Random Forest (Fig. 4.1d) mostrano aree più piccole ad alta probabilità e localizzate in aree distali rispetto a quelle degli altri regressori. Alla profondità di 100m nessun regressore riproduce la distribuzione spaziale delle miniere e tutti mostrano distribuzioni di probabilità sensibilmente diverse da quelle calcolate in superficie. Ad esempio, con il regressore Ordinary Least Square (Fig. 4.1g) scompare la zona ad alta probabilità in corrispondenza del campo geotermico di Bagnore mentre con il regressore CatBoost (Fig. 4.1f) e con Random Forest (Fig. 4.1i) si formano alcune aree ad alta probabilità nella fascia ovest dell'area di studio.

T02



I regressori CatBoost (Fig. 4.2a) e Random Forest (Fig. 4.2d) del test T02, che considera solo i predittori di vicinanza, mostrano una distribuzione notevolmente frammentata dei valori di probabilità e che non riproduce la distribuzione spaziale delle miniere. Invece i regressori Ordinary Least Square (Fig. 4.2b), Multilayer Perceptron (Fig. 4.2c) e Support Vector (Fig. 4.2e) mostrano una distribuzione delle probabilità che è continua lungo una direzione N-S ed ha una buona correlazione con le miniere e le manifestazioni geotermiche. In particolare, regressori (b) e (c) mostrano distribuzioni di probabilità molto simili tra loro e generano due vaste aree con alti valori centrate in corrispondenza del Monte Amiata. Il regressore Support Vector (e) mostra un andamento simile ai precedenti ma genera aree ad alta probabilità di minori dimensioni e che si estendono anche nella parte sud dell'area di studio. Alla profondità di 100m (Fig. 4.2f-l), nessun regressore ha calcolato probabilità apprezzabilmente diverse da quelle calcolate per i punti in superficie.

T03



Nel caso del Test 03 (che considera tutti i predittori) i regressori Ordinary Least Square (Fig. 4.3b) e Support Vector (Fig. 4.3e) mostrano una discreta ma discontinua correlazione con la distribuzione delle miniere dell'area, mentre gli altri tre mostrano valori di probabilità elevata in zone irregolari e senza una sistematica correlazione con le miniere. La migliore distribuzione sembra essere quella del regressore CatBoost (Fig. 4.3a), dal momento che le aree discontinue con probabilità $>80\%$ mostrano la maggiore prossimità con le miniere. Alla profondità di 100m (Fig. 4.3f-l) i regressori hanno calcolato probabilità che si discostano soltanto minimamente da quelle calcolate per i punti in superficie. Il regressore CatBoost (Fig. 4.3f) mostra i cambiamenti più marcati.

T04

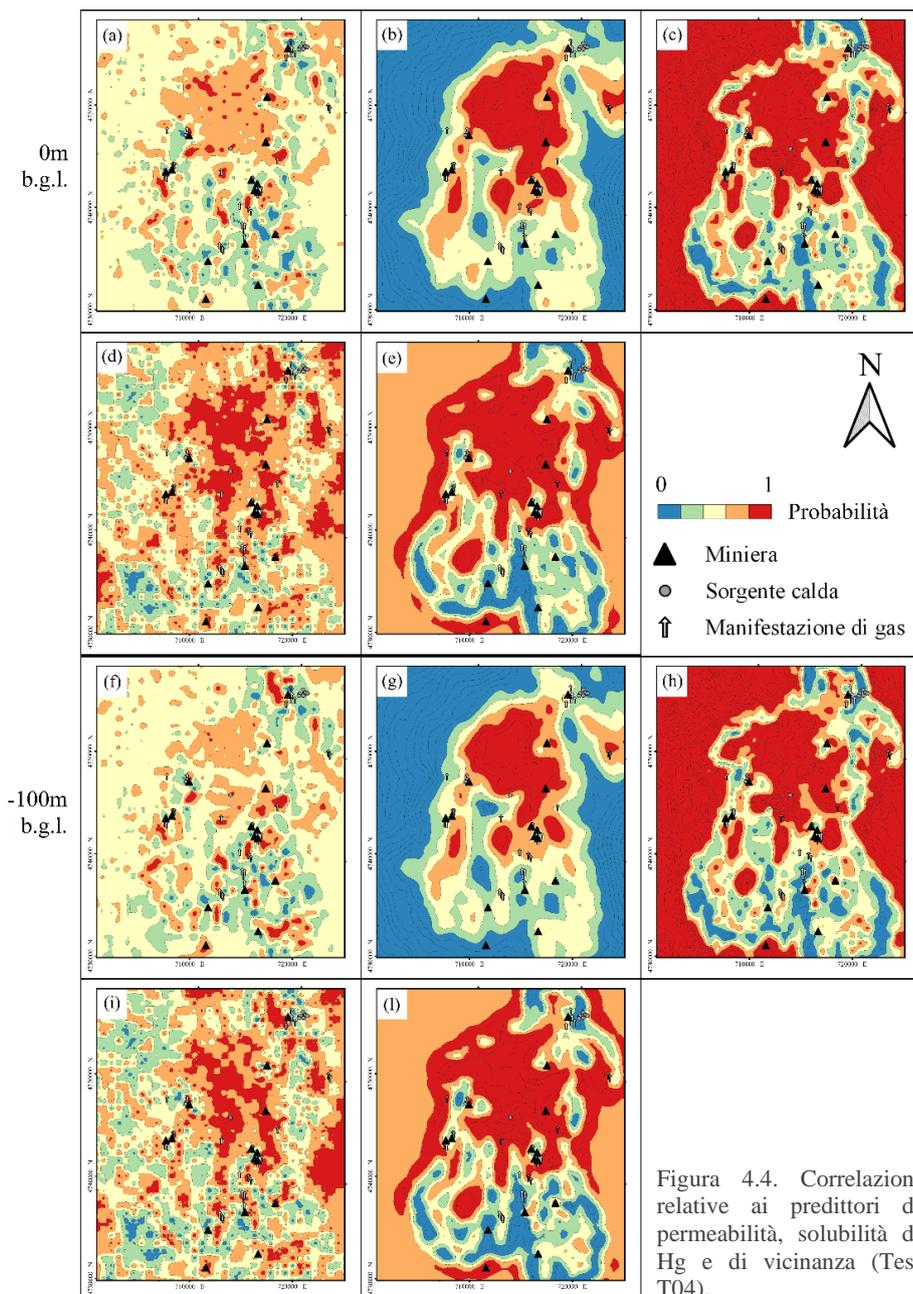


Figura 4.4. Correlazioni relative ai predittori di permeabilità, solubilità di Hg e di vicinanza (Test T04).

Nel test T04 (che combina permeabilità con solubilità e predittori di vicinanza), la distribuzione delle probabilità dei regressori Ordinary Least Square (Fig. 4.4b), Multilayer Perceptron (Fig. 4.4c) e Support Vector (Fig. 4.4e) presenta andamenti simili a quelli del test T03, mentre quelle dei regressori CatBoost (Fig. 4.4a) e Random Forest (Fig. 4.4d) sono più frammentate e discontinue. Nel regressore CatBoost le porzioni più ad est e ovest dell'area di studio presentano valori compresi tra il 40% e il 60% di probabilità, con zone a più bassa probabilità distribuite nei pressi delle principali miniere. Alla profondità di 100m (Fig. 4.4f-l) i regressori CatBoost (f) e Random Forest (i) mostrano valori di probabilità che sono più bassi di quelli calcolati per i punti in superficie, mentre gli altri regressori calcolano valori sostanzialmente simili a quelli calcolati per i punti in superficie.

T05

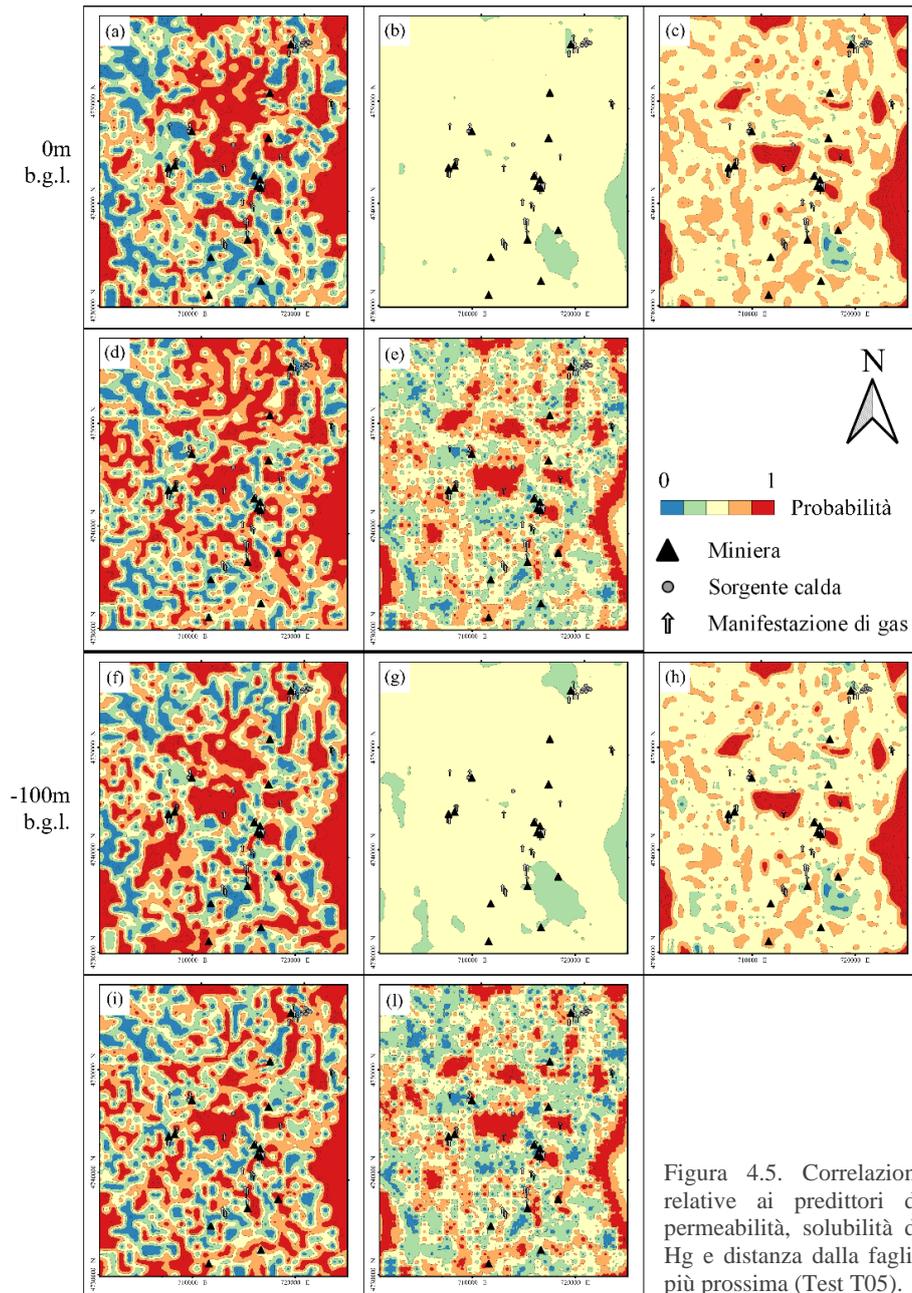


Figura 4.5. Correlazioni relative ai predittori di permeabilità, solubilità di Hg e distanza dalla faglia più prossima (Test T05).

Le distribuzioni dei valori di probabilità dei regressori CatBoost (Fig 4.5a) e Random Forest (Fig 4.5d) del Test 05 (che combina i predittori permeabilità, solubilità del Hg e distanza dalla faglia più prossima) si presentano estremamente discontinue e senza una significativa correlazione con la distribuzione spaziale delle miniere. Il regressore Ordinary Least Square (Fig 4.5b) mostra valori uniformi e compresi tra 40% e 60% in tutta l'area di studio, con aree a probabilità minore in corrispondenza degli affioramenti del Complesso IV e di una piccola fascia nel margine orientale dell'area. I regressori Multilayer Perceptron (Fig 4.5c) e Support Vector (Fig 4.5e) calcolano aree ad alta probabilità che non mostrano correlazione con la distribuzione delle miniere. Alla profondità di 100m (Fig. 4.5f-l), tutti i regressori mostrano distribuzioni di probabilità che sono molto simili a quelle calcolate per i punti in superficie, ma con valori leggermente più bassi.

T06

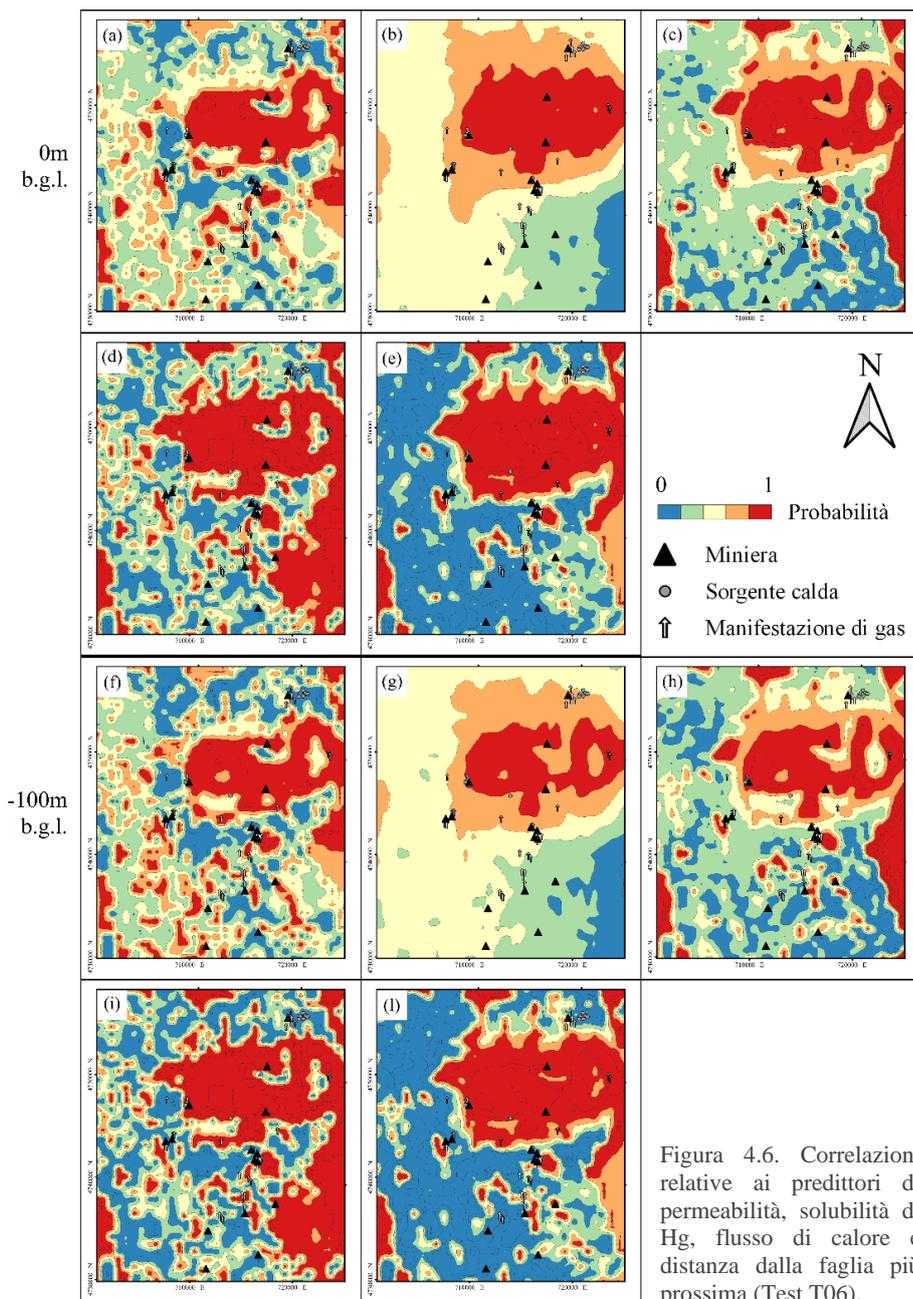
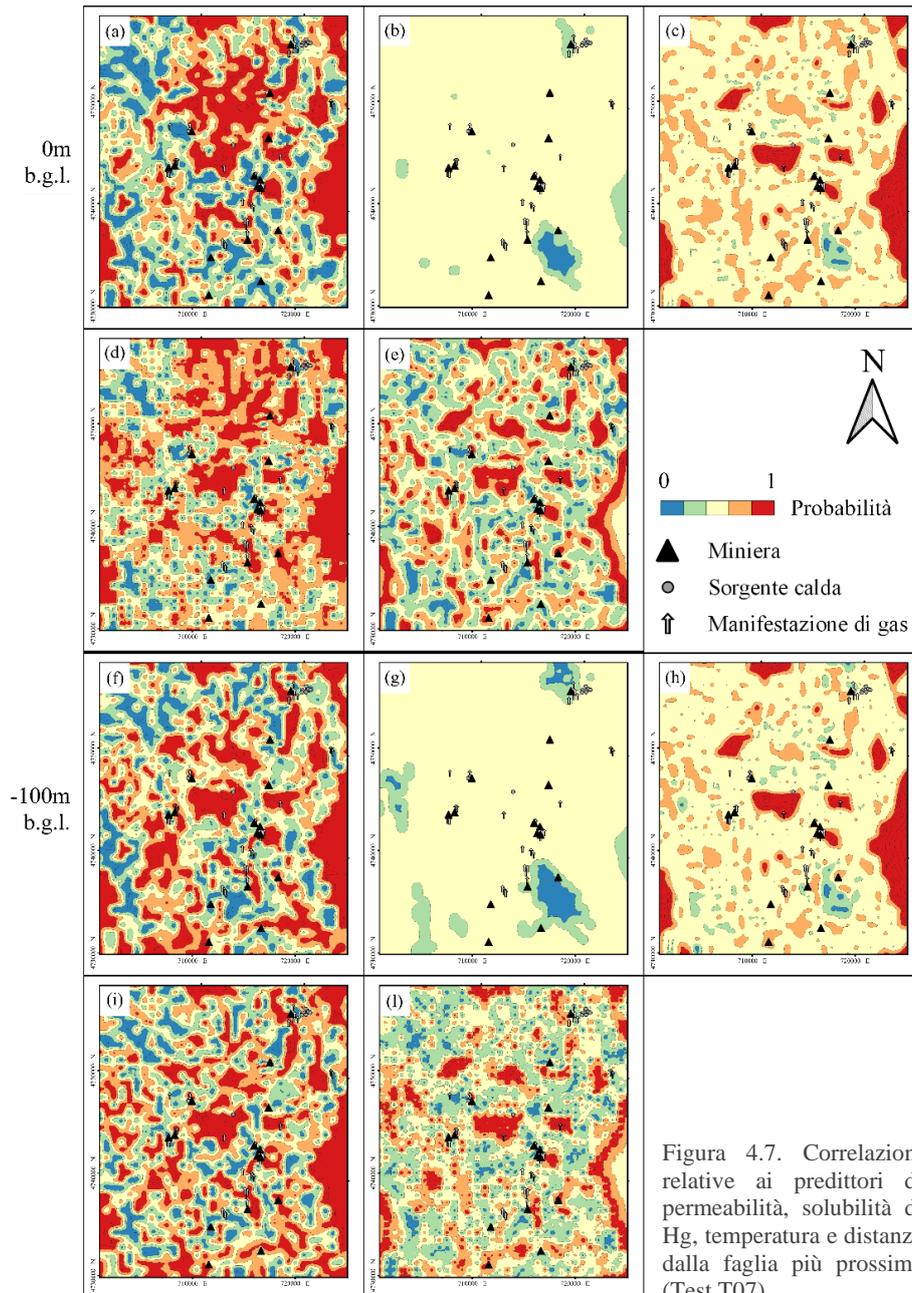


Figura 4.6. Correlazioni relative ai predittori di permeabilità, solubilità di Hg, flusso di calore e distanza dalla faglia più prossima (Test T06).

Le correlazioni risultanti dal test T06 (che considera la permeabilità, la solubilità del Hg, il flusso di calore, e la distanza dalla faglia più prossima) mostrano un'ampia area ad alta probabilità allungata in direzione E-W e corrispondente al Monte Amiata. Quest'area ha un andamento molto simile a quello dei valori di flusso di calore (Fig. 3.14) e d'altra parte mostra una bassa correlazione con la distribuzione delle miniere. I regressori Multilayer Perceptron (Fig. 4.6c), Random Forest (Fig. 4.6d) e Support Vector (Fig. 4.6e) mostrano anche zone ad alta probabilità lungo i margini est e ovest dell'area di studio. Alla profondità di 100m (Fig. 4.6f-l), solamente il regressore Ordinary Least Square (Fig. 4.6g) mostra di valori di probabilità differenti (più bassi) da quelli calcolati per i punti in superficie.

T07



Nel test T07 (che considera la permeabilità, la solubilità del Hg, la temperatura, e la distanza dalla faglia più prossima) soltanto i regressori CatBoost (Fig. 4.7a) e Random Forest (Fig. 4.7d) hanno calcolato una distribuzione di probabilità che approssima quella delle miniere. Queste distribuzioni sono però estremamente discontinue. Il regressore Ordinary Least Square (Fig. 4.7b) mostra una distribuzione uniforme delle probabilità con valori compresi tra 40% e 60%, e valori molto bassi sono presenti in corrispondenza degli affioramenti del Complesso IV nell'area. Gli altri regressori presentano distribuzioni più frammentate, ad eccezione della fascia ad alta probabilità presente nella porzione est dell'area di studio. Alla profondità di 100m (Fig. 4.7f-l), tutti i regressori calcolano valori di probabilità più bassi di quelli calcolati per i punti in superficie.

T08

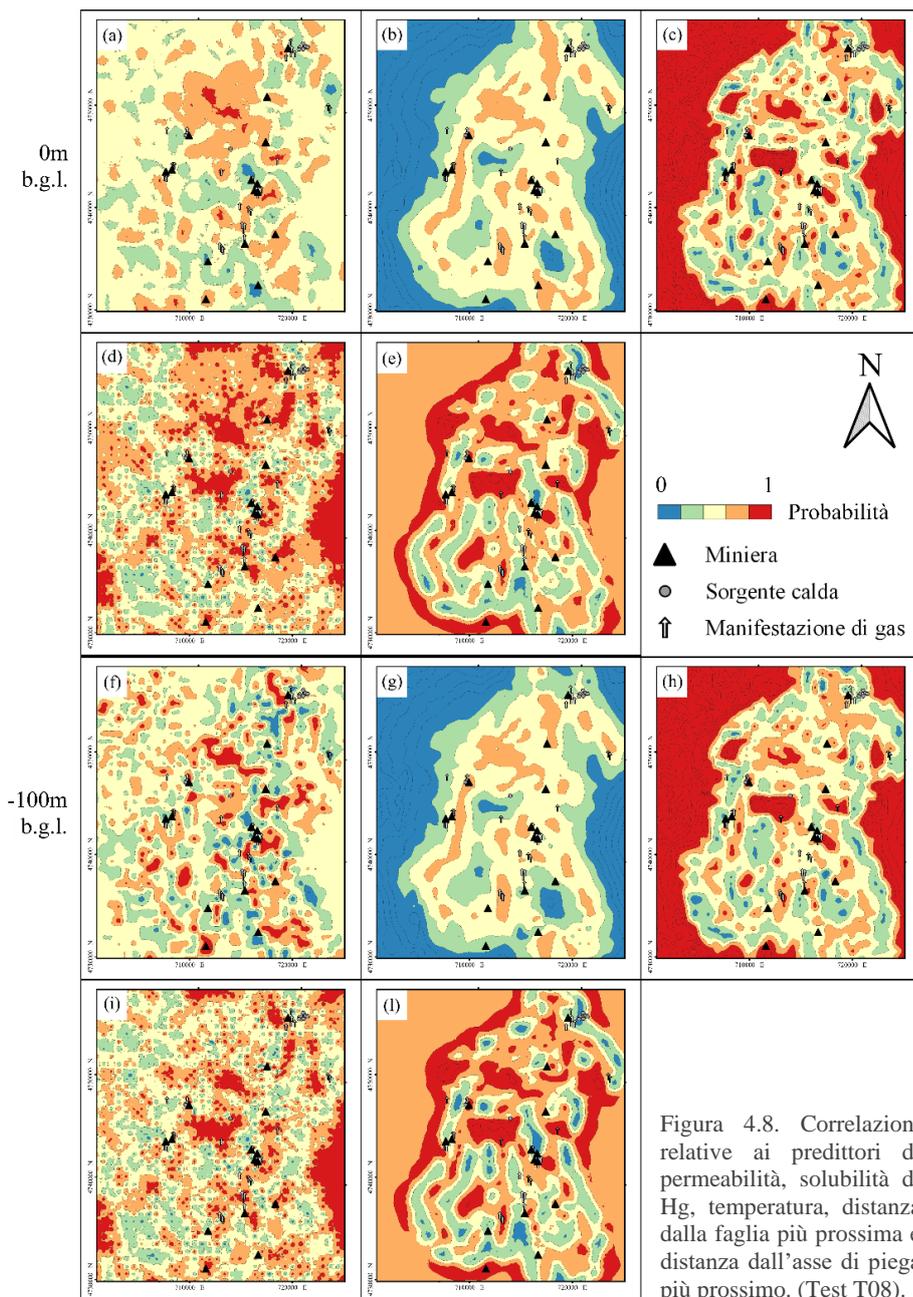


Figura 4.8. Correlazioni relative ai predittori di permeabilità, solubilità di Hg, temperatura, distanza dalla faglia più prossima e distanza dall'asse di piega più prossimo. (Test T08).

I regressori Ordinary Least Square (Fig. 4.8b), Multilayer Perceptron (Fig. 4.8c) e Support Vector (Fig. 4.8e) del test T08 (che combina permeabilità con solubilità del Hg, temperatura, distanze dalla faglia più prossima e distanza dall'asse di piega più prossimo) presentano una distribuzione delle probabilità che segue in modo discontinuo la distribuzione delle miniere. Il regressore CatBoost (Fig. 4.8a) mostra una minore correlazione con le strutture plicative ma maggiore con le faglie, e presenta aree ad alta probabilità localizzate nelle vicinanze di miniere esistenti. La distribuzione delle probabilità del regressore Random Forest (Fig. 4.8d) invece è più discontinua e non mostra evidenti correlazioni con il motivo strutturale. Alla profondità di 100m (Fig. 4.8f-l), tutti i predittori mostrano la stessa geometria di distribuzione di probabilità, ma una generale diminuzione dei valori assoluti.

TA1

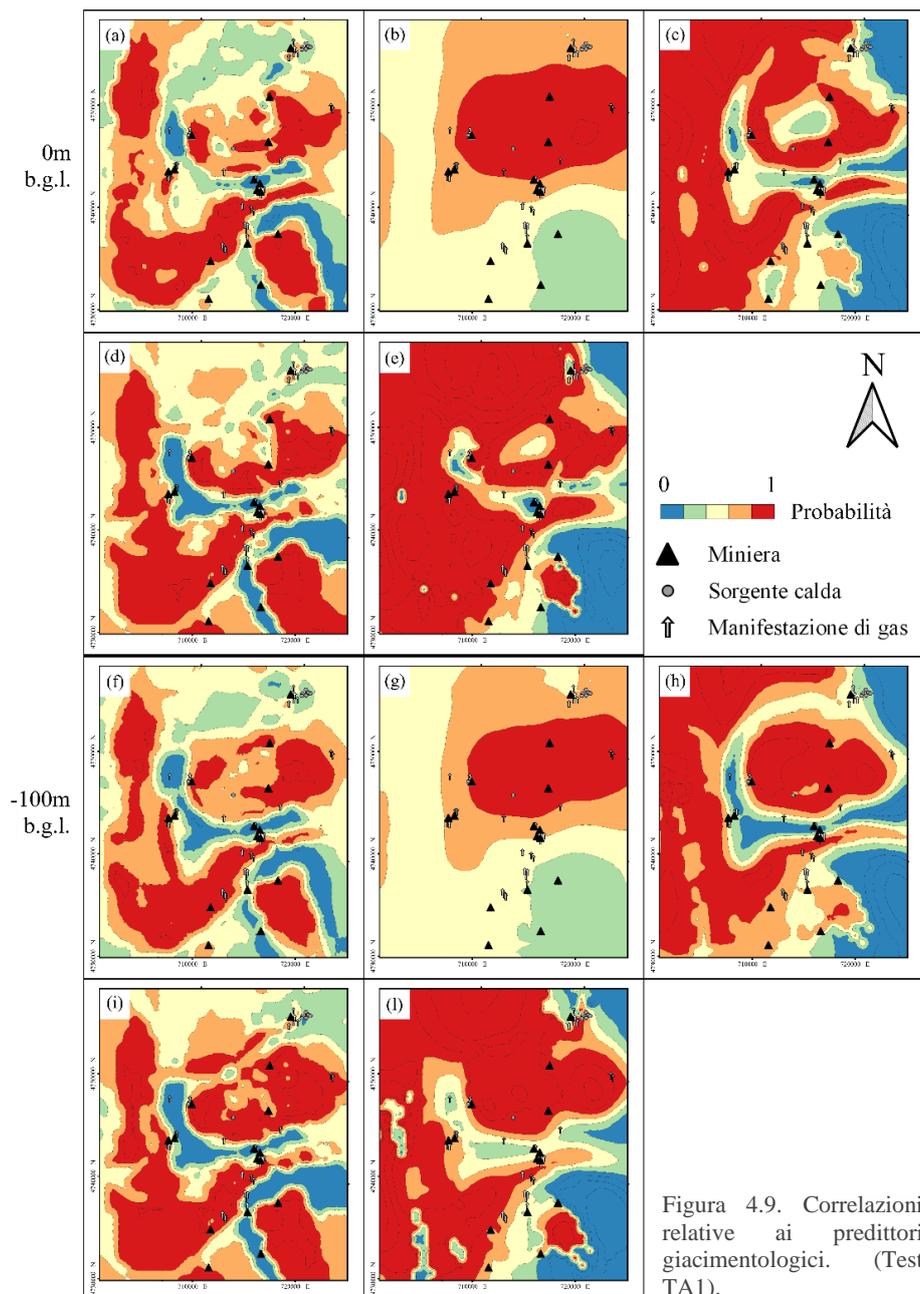


Figura 4.9. Correlazioni relative ai predittori giacimentologici. (Test TA1).

I regressori CatBoost (Fig. 4.9a), Ordinary Least Square (Fig. 4.9b) e Random Forest (Fig. 4.9d) del test TA1 (che utilizza soltanto i predittori giacimentologici come il test T01, ma con un training aumentato di 3003 punti) mostrano distribuzioni dei valori di probabilità simili a quelli del test T06, ma con aree ad alta probabilità di dimensioni maggiori. I regressori Multilayer Perceptron (Fig. 4.9c) e Support Vector (Fig. 4.9e) mostrano invece distribuzioni simili tra loro ma con scarsa correlazione con la distribuzione delle miniere. Alcune aree ad alta probabilità calcolate da diversi regressori sono comparabili, come ad esempio le aree immediatamente a S del Monte Amiata e quelle nei pressi delle miniere del Siele, Solforate e Abetina, nella parte centrale dell'area di studio. Le probabilità calcolate a -100m (Fig. 4.9f-l) sono in generale più estese di quelle calcolate per i punti in superficie tranne per il regressore Ordinary Least Square, in cui è visibile una leggera diminuzione (cf. Fig. 4.9b con Fig. 4.9g).

TA3

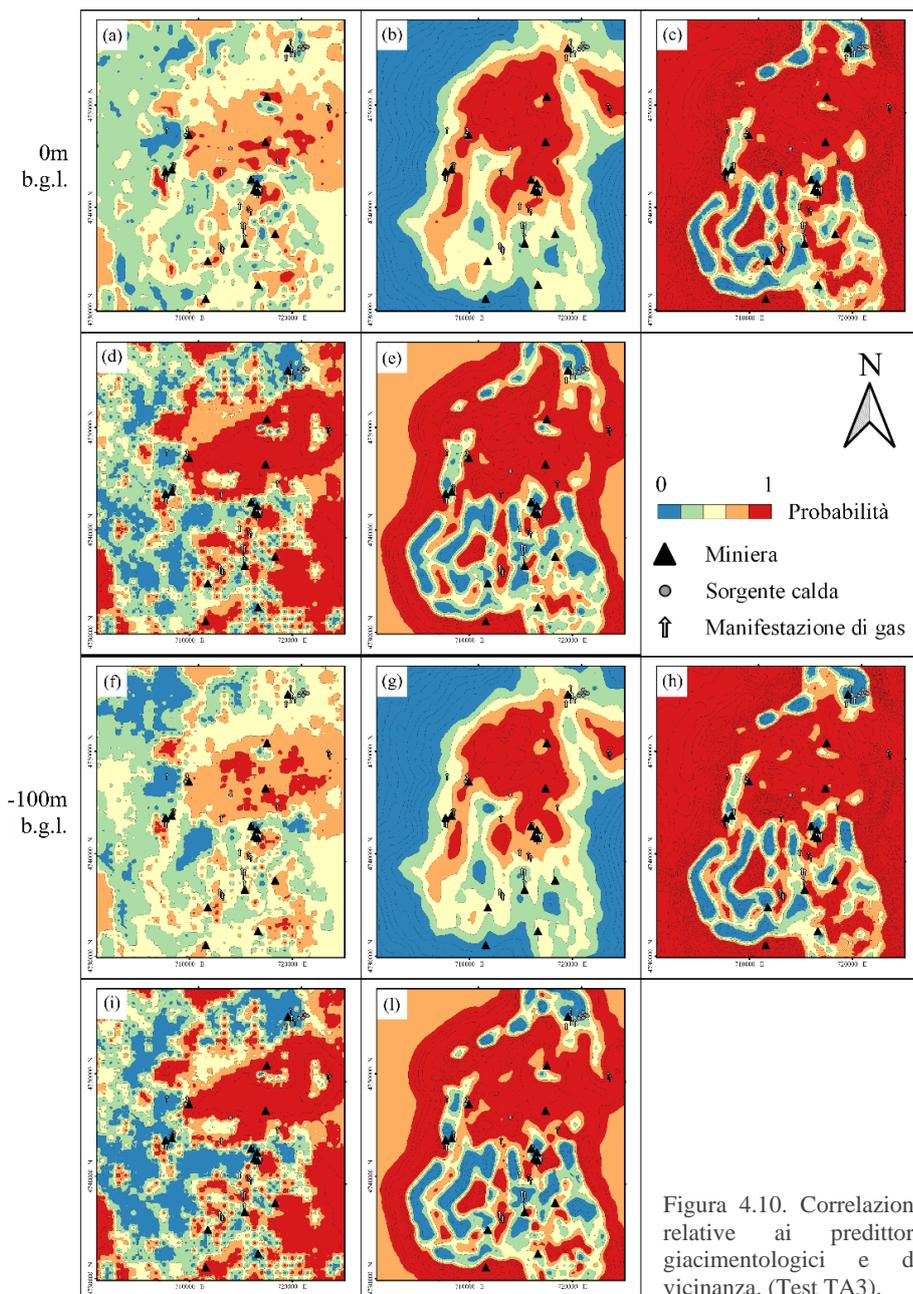


Figura 4.10. Correlazioni relative ai predittori giacimentologici e di vicinanza. (Test TA3).

Il regressore Ordinary Least Square (Fig. 4.10b) del test TA3 (che utilizza tutti i predittori disponibili come il test T03, ma con un training aumentato di 3003 punti) ha generato una distribuzione delle probabilità con andamento N-S che è relativamente continuo e correlabile con la distribuzione delle miniere. Simili distribuzioni orientate N-S sono state generate anche dai regressori Multilayer Perceptron (Fig. 4.10c) e Support Vector (Fig. 4.10e), ma in questi casi le miniere si trovano al di fuori delle aree a maggiore probabilità, oppure le aree ad alta probabilità ricoprono l'intera area di studio. Il regressore Random Forest (Fig. 4.10d) ha generato vaste aree ad alta probabilità che non mostrano correlazioni con la distribuzione delle miniere. L'effetto frastagliato e discontinuo dei punti a maggiore probabilità è dovuto all'interpolatore Kriging. La migliore correlazione è stata ottenuta dal regressore CatBoost (Fig. 4.10a), che presenta una buona correlazione con le miniere e genera una distribuzione con andamento N-S. Le probabilità calcolate a -100m (Fig. 4.10f-l) non mostrano apprezzabili variazioni di probabilità con la profondità.

TA6

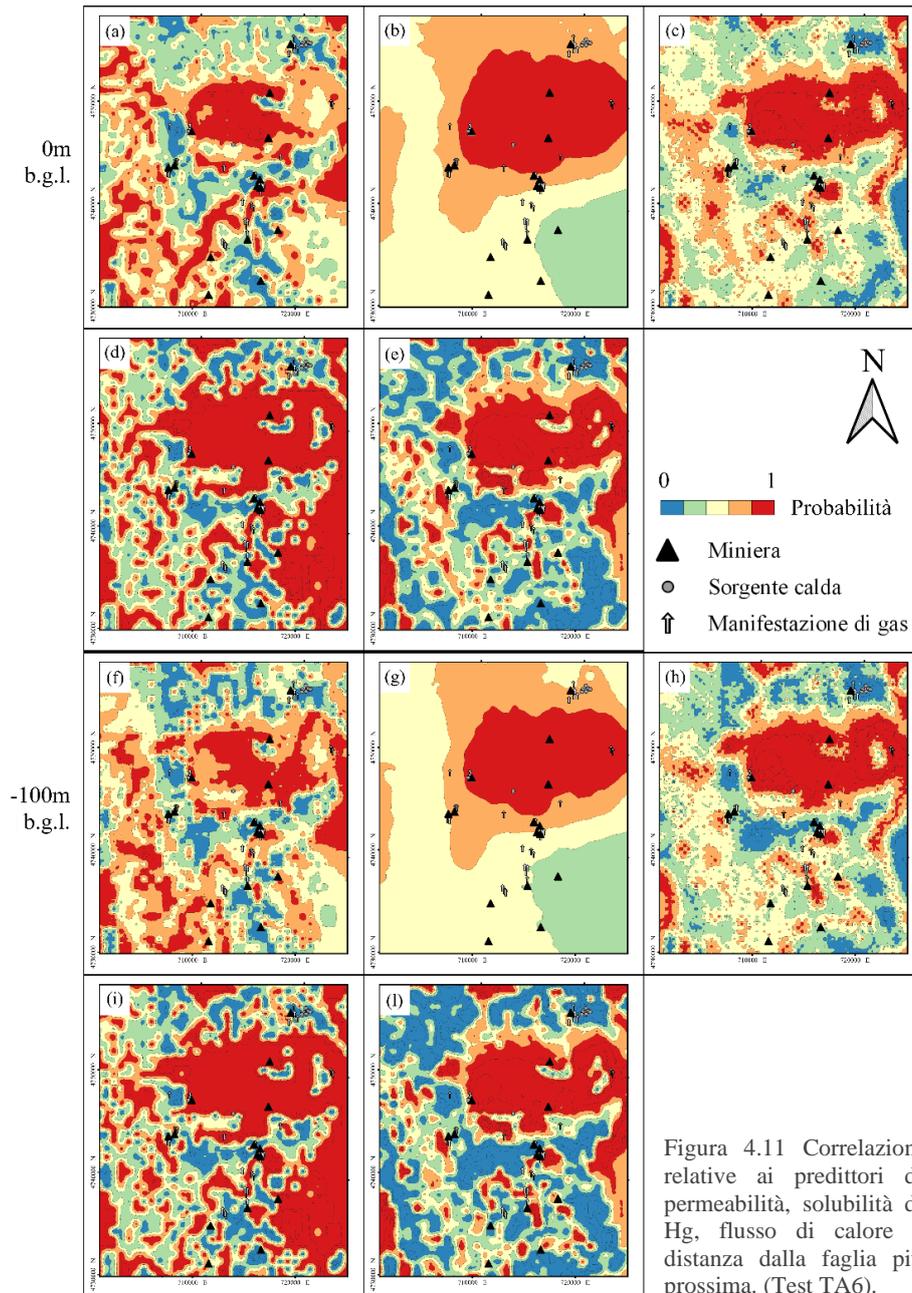


Figura 4.11 Correlazioni relative ai predittori di permeabilità, solubilità di Hg, flusso di calore e distanza dalla faglia più prossima. (Test TA6).

I risultati del regressore Ordinary Least Square (Fig. 4.11b) del test TA6 (che utilizza i predittori permeabilità, solubilità del Hg, flusso di calore, e distanza dalla faglia più prossima come nel test T06, ma con un training aumentato di 3003 punti) mostrano un'unica grande zona ad alta probabilità in corrispondenza del complesso vulcanico del Monte Amiata e nessuna correlazione con la distribuzione delle miniere. Le distribuzioni di probabilità calcolate dagli altri regressori sono piuttosto frammentate, con alti valori sempre localizzati in corrispondenza del Monte Amiata e allungati verso est. Le probabilità calcolate a -100m (Fig. 4.11f-l) non generano variazioni significative, a parte il regressore CatBoost (Fig. 4.11f) che mostra una discreta riduzione areale in corrispondenza della miniera di Bagnore e nella porzione NW dell'area di studio.

TA7

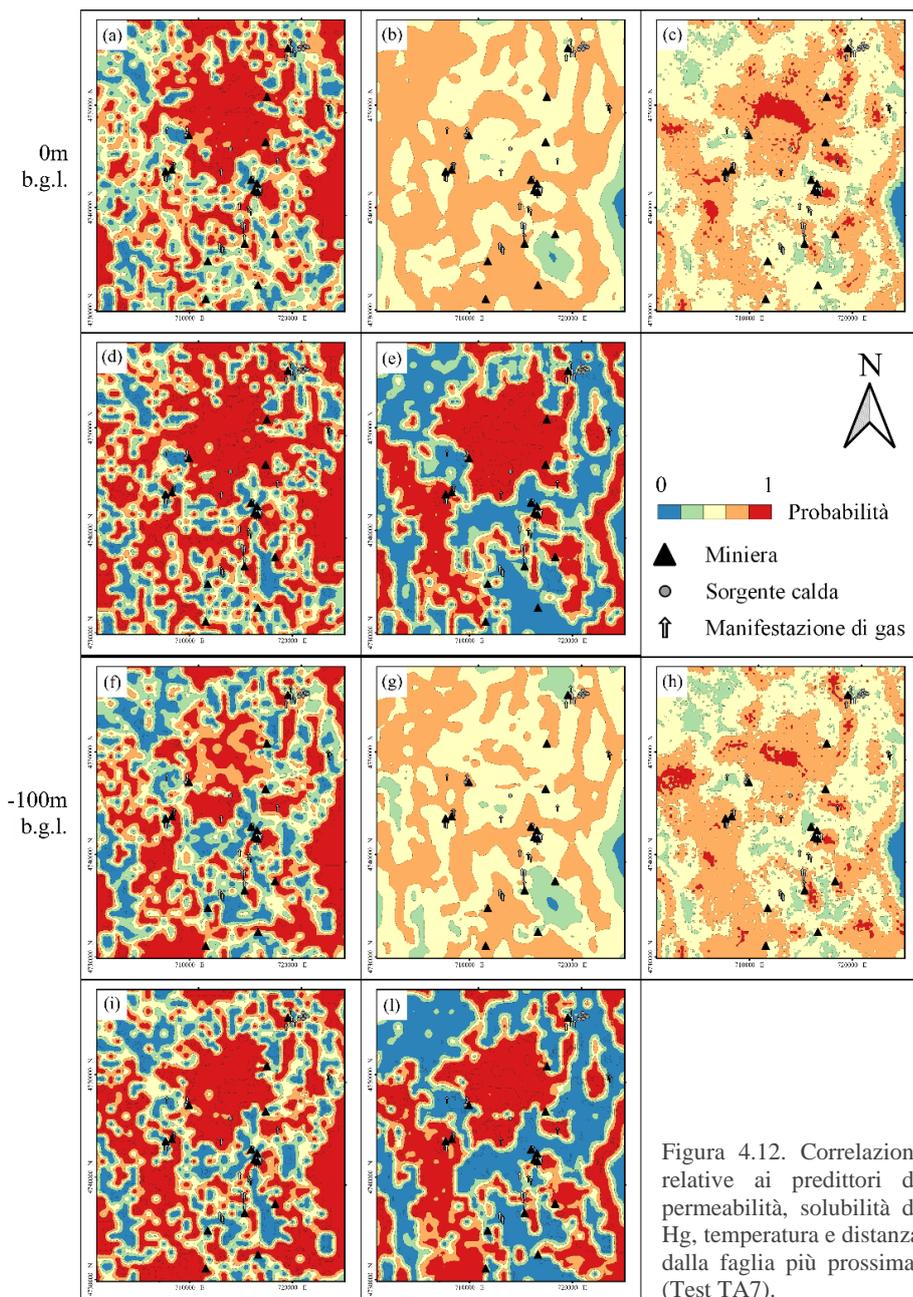
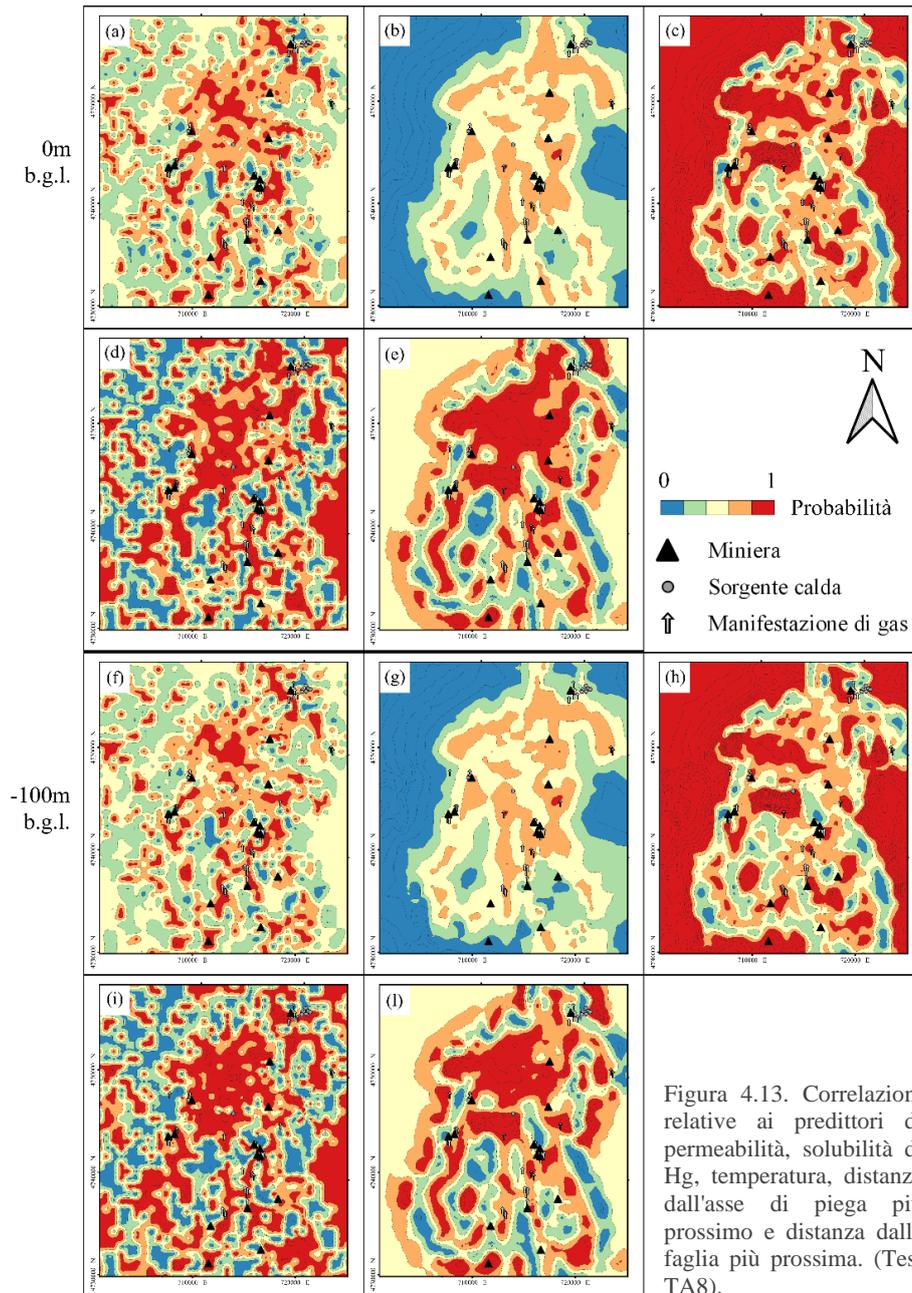


Figura 4.12. Correlazioni relative ai predittori di permeabilità, solubilità di Hg, temperatura e distanza dalla faglia più prossima. (Test TA7).

Nel test TA7 (che fa uso dei predittori permeabilità, solubilità del Hg, temperatura e distanza dalla faglia più prossima come nel test T07, ma con un training aumentato di 3003 punti) i regressori CatBoost (Fig. 4.12a), Random Forest (Fig. 4.12d) e Support Vector (Fig. 4.12e) hanno generato alti valori di probabilità centrati in corrispondenza del complesso vulcanico, ma una discreta correlazione con la distribuzione delle miniere. Il regressore Ordinary Least Square (Fig. 4.12b) mostra valori di probabilità sempre inferiori all'80% mentre il regressore Multilayer Perceptron (Fig. 4.12c) alte probabilità in aree non correlate con le miniere presenti. I regressori (c) ed (e) mostrano una più alta correlazione con il valore di permeabilità legato al Complesso III. Le probabilità calcolate a -100m (Fig. 4.12f-l) mostrano distribuzioni simili a quelle determinate per i punti di superficie, ma tutti i regressori calcolano una generale diminuzione dei valori assoluti.

TA8



I regressori CatBoost (Fig. 4.13a) e Random Forest (Fig. 4.13d) del test TA8 (che combina i predittori permeabilità, solubilità del Hg, distanza dalla faglia più prossima, e distanza dall'asse di piega più prossimo come il test T08, ma con un training aumentato di 3003 punti) mostrano una distribuzione frammentata dei valori di probabilità. I regressori Ordinary Least Square (Fig. 4.13b), Multilayer Perceptron (Fig. 4.13c) e Support Vector (Fig. 4.13e) mostrano invece una distribuzione più continua delle aree ad alta probabilità che segue la distribuzione delle miniere. Il regressore (b) mostra valori di probabilità che superano l'80% nell'area di Piancastagnaio. Il regressore (c) mostra una vasta area con alti valori di probabilità che corrisponde ai valori minimi del regressore (b). Le probabilità calcolate a -100m (Fig. 4.13f-i) mostrano valori simili a quelli calcolati per i punti in superficie con l'eccezione del regressore CatBoost (Fig. 4.13f), che è l'unico a mostrare una leggera riduzione delle aree ad alta probabilità.

TB5

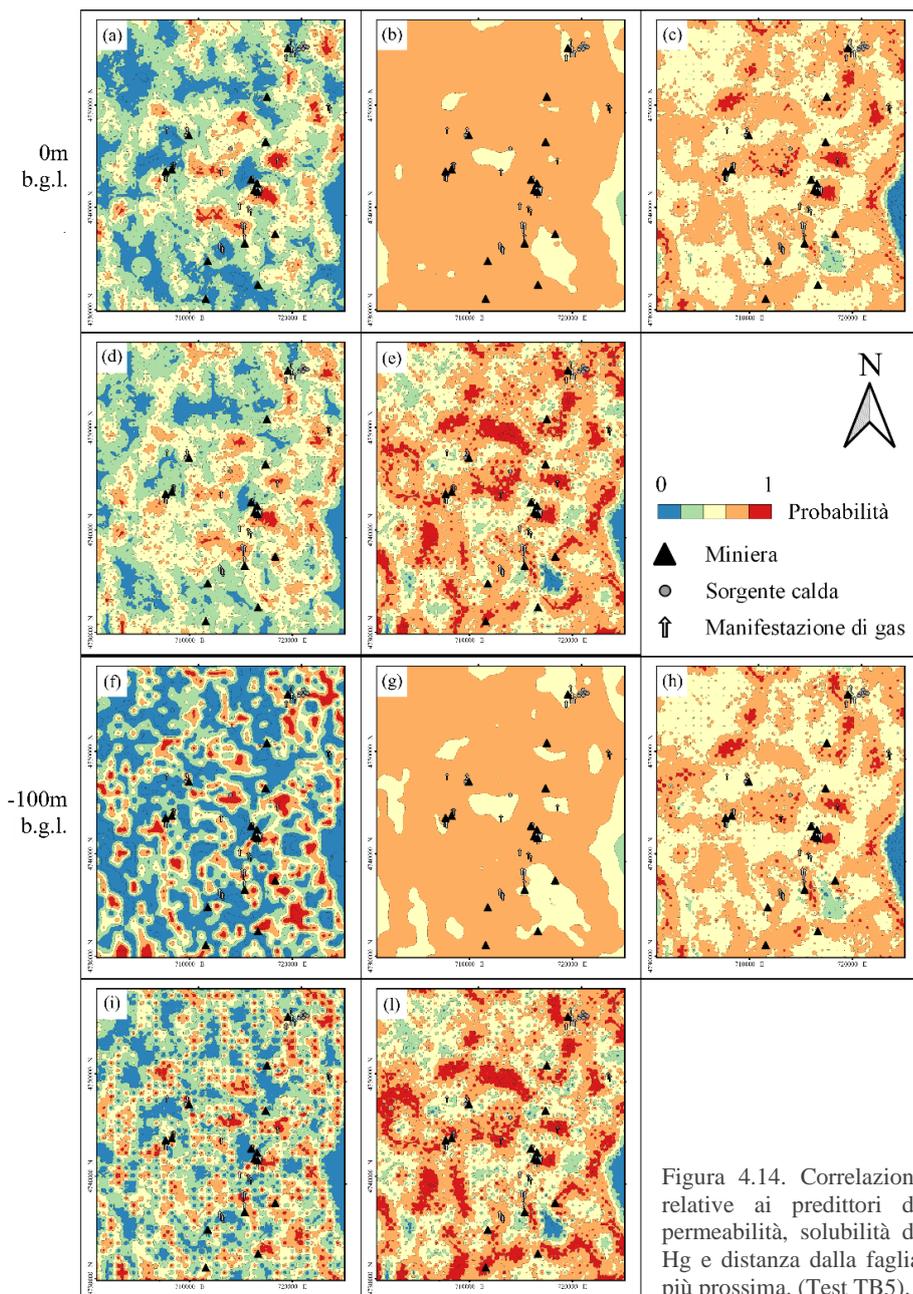


Figura 4.14. Correlazioni relative ai predittori di permeabilità, solubilità di Hg e distanza dalla faglia più prossima. (Test TB5).

I valori risultanti dal test TB5 (che utilizza i predittori permeabilità, solubilità del Hg, distanza dalla faglia più prossima come T05, ma con un training aumentato di 3110 punti) mostrano generalmente probabilità inferiori rispetto ai test precedenti. I regressori CatBoost (Fig. 4.14a) e Random Forest (Fig. 4.14d) mostrano distribuzioni simili tra loro, ma che non riproducono la distribuzione delle miniere. L'Ordinary Least Square (Fig. 4.14b) mostra una distribuzione apparentemente non correlata ad alcun fattore. Anche i regressori Multilayer Perceptron (Fig. 4.14c) e Support Vector (Fig. 4.14e) sono simili tra loro, mostrano valori di probabilità più alti ma poco correlati alla distribuzione delle miniere. Ciò è dato dal fatto che le distribuzioni dei quattro regressori (a), (c), (d) ed (e) mostrano aree ad alte probabilità in prossimità delle miniere dell'area, ma i valori più alti sono determinati sistematicamente nelle immediate vicinanze delle miniere e non in corrispondenza delle stesse. Le probabilità calcolate a -100m (Fig. 4.14f-l) con i regressori CatBoost (Fig. 4.14f), Ordinary Least Square (Fig. 4.14g) e Random Forest (Fig. 4.14i) mostrano lievi riduzioni dei valori di probabilità. Multilayer Perceptron (Fig. 4.14h) non mostra alcuna variazione mentre Support Vector (Fig. 4.14l) ha generato un visibile ma leggero aumento.

TB6

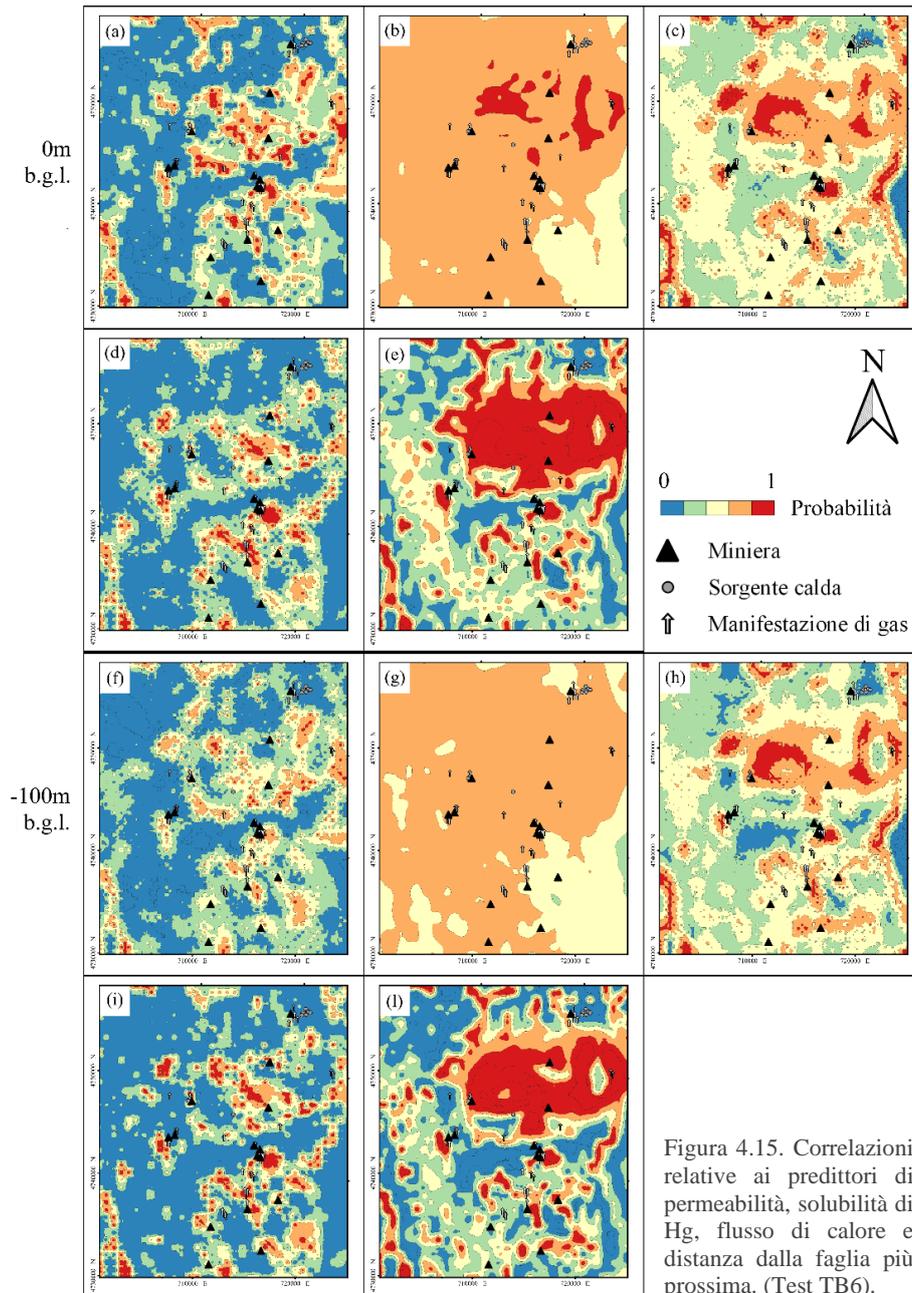


Figura 4.15. Correlazioni relative ai predittori di permeabilità, solubilità di Hg, flusso di calore e distanza dalla faglia più prossima. (Test TB6).

La distribuzione delle predizioni dei regressori CatBoost (Fig. 4.15a) e Random Forest (Fig. 4.15d) sono frammentate (che, come T06, combina i predittori permeabilità, solubilità del Hg, flusso di calore, e distanza dalla faglia più prossima, ma con un training aumentato di 3110 punti) come nei test precedenti a causa del Kriging, ma mostrano una correlazione con la posizione delle miniere che è relativamente buona. Il regressore Ordinary Least Square (Fig. 4.15b) presenta una distribuzione non correlata a nessun predittore. Il regressore Multilayer Perceptron (Fig. 4.15c) mostra una buona correlazione con le miniere ma con aree maggiori rispetto ai regressori CatBoost (a) e Random Forest (d), mentre il regressore Support Vector (Fig. 4.15e) presenta una distribuzione simile a quella del Multilayer Perceptron ma con una grande area in corrispondenza del Monte Amiata e ad est di esso. Le probabilità calcolate a -100m (Fig. 4.15f-l) sono più basse di quelle calcolate per i punti in superficie, e questo è chiaramente visibile per i regressori CatBoost (cf. Fig. 4.15a con Fig. 4.15f) ed Ordinary Least Square (cf. Fig. 4.15b con Fig. 4.15g), mentre gli altri non mostrano variazioni significative.

TB7

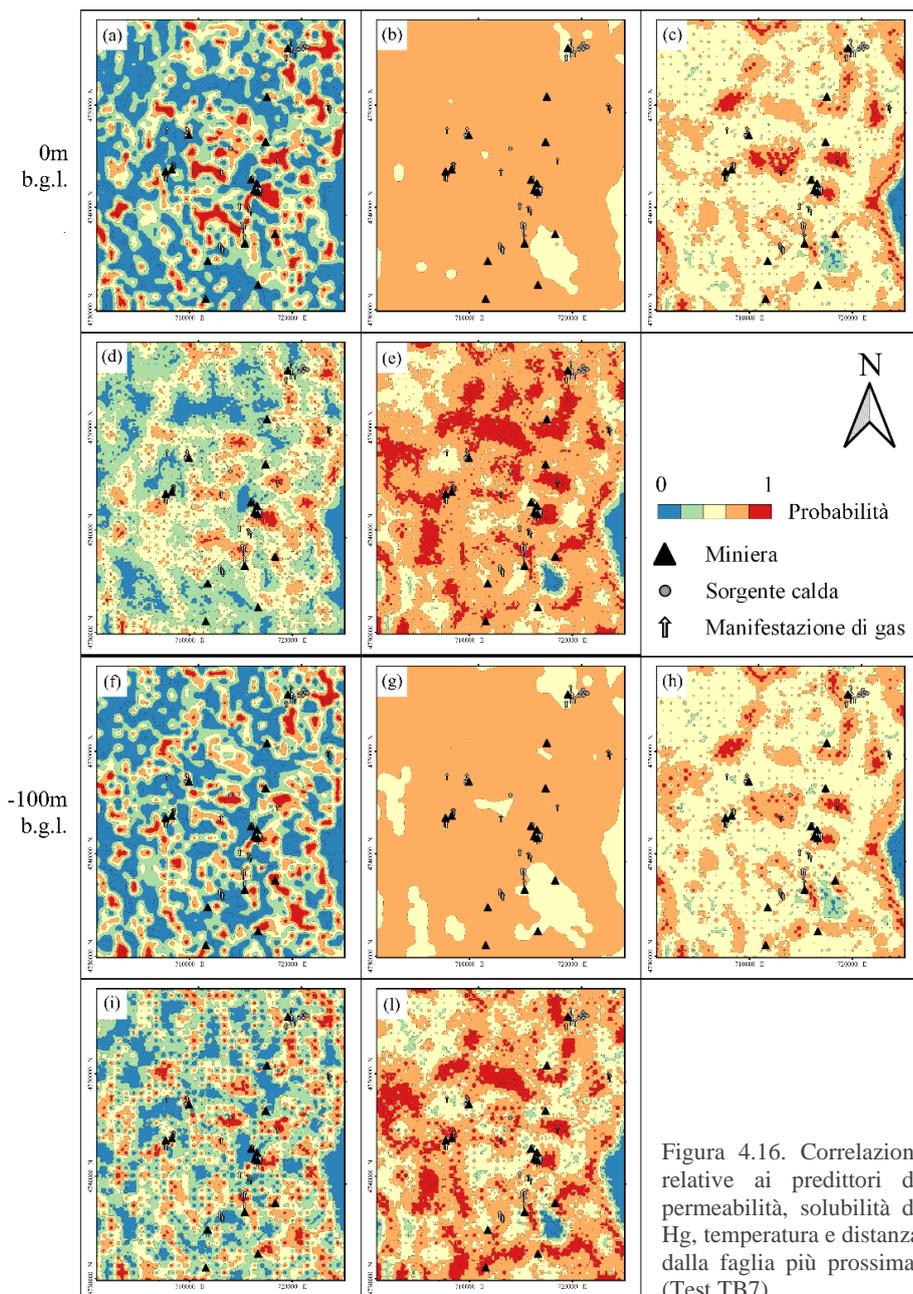


Figura 4.16. Correlazioni relative ai predittori di permeabilità, solubilità di Hg, temperatura e distanza dalla faglia più prossima. (Test TB7).

Il regressore Ordinary Least Square (Fig. 4.16b) del test TB7 (che combina i predittori permeabilità, solubilità del Hg, temperatura, e distanza dalla faglia più prossima come T07, ma con un training aumentato di 3110 punti) non presenta correlazioni con la distribuzione delle miniere. Il regressore CatBoost (Fig. 4.16a) mostra una distribuzione frammentata, non direttamente riconducibile ad alcun predittore. I regressori Multilayer Perceptron (Fig. 4.16c), Random Forest (Fig. 4.16d) e Support Vector (Fig. 4.16e) mostrano le stesse distribuzioni dei medesimi regressori del test TB5. Le probabilità calcolate a -100m (Fig. 4.16f-l) mostrano valori di probabilità più bassi per tutti i regressori, ad eccezione di Random Forest (Fig. 4.16i) che vede aumenti localizzati in corrispondenza dei punti del dataset di test.

TB8

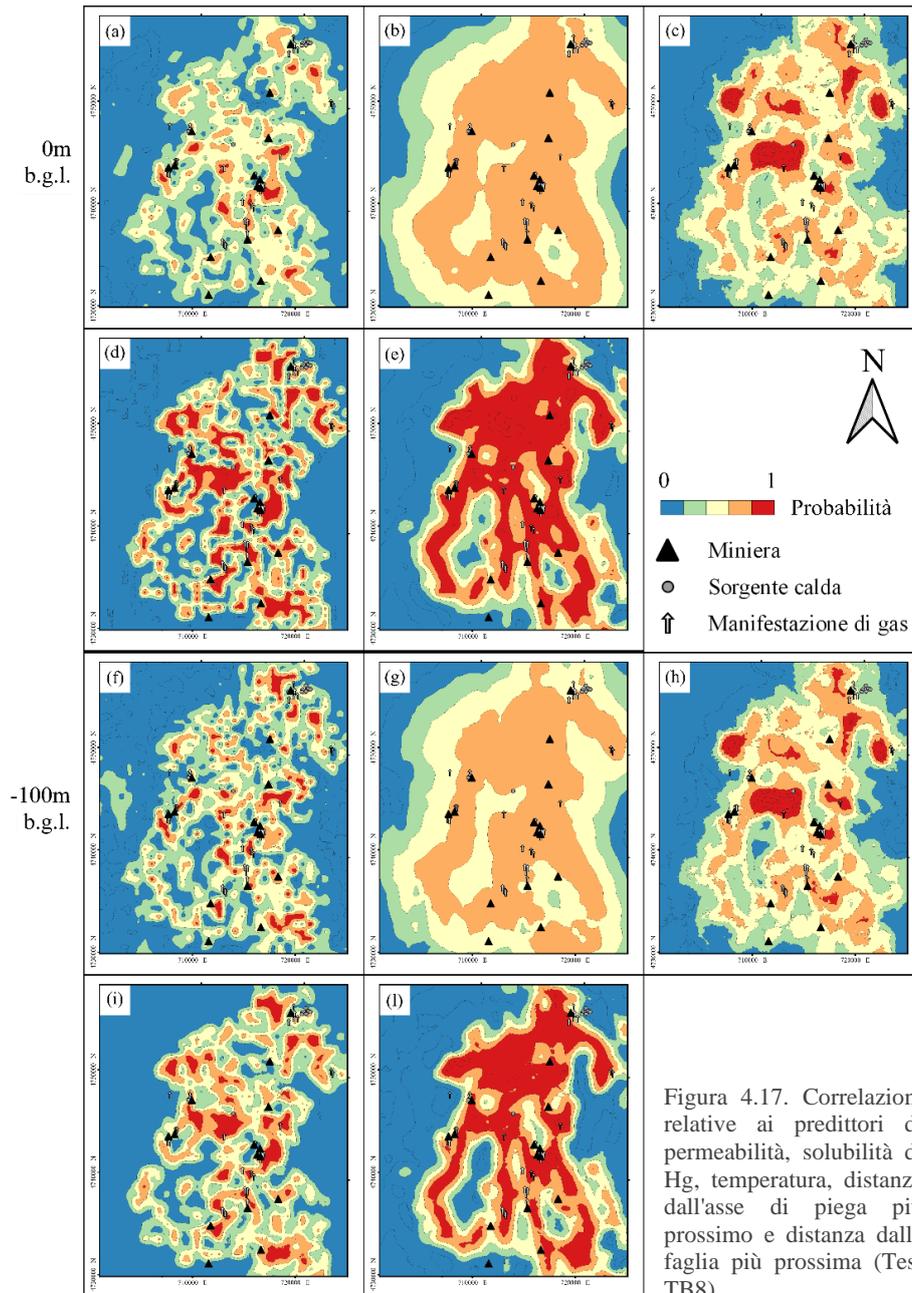


Figura 4.17. Correlazioni relative ai predittori di permeabilità, solubilità di Hg, temperatura, distanza dall'asse di piega più prossimo e distanza dalla faglia più prossima (Test TB8).

L'andamento dei valori di probabilità di tutti i regressori del test TB8 (che, come T08, usa i predittori permeabilità, solubilità del Hg, temperatura, distanza dalla faglia più prossima, e distanza dell'asse di piega più prossimo, ma con un training aumentato di 3110 punti) è influenzato dal contributo del predittore distanza dalla faglia. Il regressore CatBoost (Fig. 4.17a) presenta un'ottima correlazione con le miniere. Il regressore Multilayer Perceptron (Fig. 4.17c) mostra una continuità N-S delle aree ad alta probabilità, ma non ha buona correlazione con le miniere. Random Forest (Fig. 4.17d) presenta una distribuzione frammentata che segue un andamento N-S, ma comunque mostra una buona correlazione con la distribuzione le miniere. I regressori Ordinary Least Square (Fig. 4.17b) e Support Vector (Fig. 4.17e) hanno calcolato ampie zone con alti valori di probabilità ed allungamento N-S che contengono o lambiscono le miniere dell'area. Le probabilità calcolate a -100m (Fig. 4.17f-l) mostrano una generale riduzione dei valori di probabilità, ad eccezione del regressore CatBoost, che presenta un aumento dei valori in zone distali rispetto alle miniere presenti (cf. Fig. 4.17a con Fig. 4.17i).

4.1 Validazione

A partire dai risultati ottenuti tramite l'applicazione degli algoritmi di ML, si è stimata la bontà dei risultati mediante la valutazione dei parametri metrici RMSE, MAE ed R^2 .

I parametri di validazione (Tab. 4.2) quantificano analiticamente i risultati ottenuti per ciascun test e valutabili graficamente nelle Figure 4.1-4.17. In dettaglio, i valori di coefficiente di determinazione (R^2) contrassegnati in verde in tabella determinano in generale una buona correlazione tra i dati originali e quelli predetti, calcolati sulla base del Train-Test split mentre quelli in rosso indicano una bassa correlazione. Esiste una relazione inversa tra i valori di R^2 e quelli di MAE e RMSE, poiché ad alti valori di R^2 corrispondono bassi valori di MAE e RMSE, indicando che maggiore è la correlazione, minore sarà l'errore associato.

Sulla base dei risultati ottenuti, emerge che l'accuratezza dei regressori CatBoost e Random Forest supera gli altri algoritmi nella maggior parte dei test effettuati. In particolare, nei test TB6 e TB8 i regressori CatBoost e Random Forest mostrano valori di $R^2 > 0.95$ e valori di MAE e RMSE < 0.1 .

Tabella 4.1. Parametri numerici di valutazione dell'accuratezza dei diversi regressori utilizzati per ogni test, suddivisi sulla base del dataset di training.

		Training originale								Training aumentato A					Training aumentato B			
		T01	T02	T03	T04	T05	T06	T07	T08	TA1	TA3	TA6	TA7	TA8	TB5	TB6	TB7	TB8
CatBoost	RMSE	0.29	0.13	0.19	0.22	0.21	0.15	0.24	0.23	0.00	0.00	0.00	0.00	0.00	0.14	0.09	0.15	0.08
	MAE	0.15	0.07	0.11	0.15	0.13	0.07	0.16	0.15	0.00	0.00	0.00	0.00	0.00	0.06	0.03	0.06	0.03
	R ²	0.71	0.93	0.86	0.80	0.82	0.91	0.78	0.78	1.00	1.00	1.00	1.00	1.00	0.86	0.94	0.85	0.96
Ordinary Least Square	RMSE	0.52	0.41	0.42	0.42	0.46	0.45	0.50	0.47	0.43	0.40	0.44	0.49	0.50	0.43	0.43	0.43	0.41
	MAE	0.45	0.37	0.37	0.38	0.44	0.40	0.47	0.43	0.38	0.34	0.38	0.45	0.45	0.39	0.38	0.40	0.37
	R ²	-0.10	0.31	0.28	0.29	0.15	0.20	0.01	0.12	0.22	0.34	0.22	0.00	-0.02	-0.28	-0.28	-0.30	-0.16
Multilayer Perceptron	RMSE	0.43	0.41	0.31	0.33	0.43	0.37	0.43	0.38	0.32	0.09	0.23	0.37	0.28	0.40	0.35	0.39	0.32
	MAE	0.36	0.37	0.22	0.27	0.39	0.28	0.38	0.33	0.24	0.06	0.16	0.29	0.20	0.32	0.26	0.31	0.24
	R ²	0.25	0.32	0.62	0.57	0.25	0.46	0.26	0.43	0.57	0.96	0.79	0.42	0.68	-0.10	0.13	-0.06	0.28
Random Forest	RMSE	0.19	0.16	0.15	0.18	0.21	0.18	0.21	0.17	0.00	0.00	0.00	0.00	0.00	0.18	0.06	0.18	0.07
	MAE	0.10	0.09	0.06	0.11	0.12	0.07	0.13	0.09	0.00	0.00	0.00	0.00	0.00	0.07	0.02	0.06	0.02
	R ²	0.85	0.90	0.91	0.87	0.83	0.87	0.82	0.89	1.00	1.00	1.00	1.00	1.00	0.78	0.97	0.78	0.97
Support Vector	RMSE	0.50	0.32	0.33	0.31	0.43	0.37	0.44	0.35	0.34	0.16	0.22	0.35	0.33	0.42	0.36	0.43	0.37
	MAE	0.34	0.22	0.22	0.22	0.32	0.24	0.34	0.24	0.21	0.09	0.14	0.22	0.19	0.27	0.23	0.29	0.23
	R ²	0.01	0.60	0.55	0.61	0.26	0.46	0.22	0.51	0.52	0.89	0.80	0.48	0.54	-0.24	0.08	-0.28	0.06

I valori contrassegnati in verde indicano valori di coefficiente di determinazione R2 compresi tra 0.80 e 0.97, mentre quelli in rosso indicano valori di R2 < 0.35. I valori di R2 = 1 risiedono al di fuori del range di accettabilità dei valori del coefficiente di correlazione perché ciò si verifica quando il valore di R2 è influenzato da overfitting del modello

Discussione

I risultati dei test mostrano che l'utilizzo di differenti combinazioni di predittori e diversi modelli di regressione generano set distinti di valori di probabilità, che corrispondono a correlazioni significativamente diverse di fattori geologici e giacimentologici. Inoltre, i risultati mostrano che l'utilizzo del *data augmentation* genera predizioni di qualità più elevata, come mostrano i valori più bassi di errore medio assoluto (MAE) ed errore quadratico medio (RMSE), e valori significativamente più elevati di coefficiente di determinazione (R^2) calcolati per i test di tipo aumentato.

In totale sono stati eseguiti 24 test ma solo 17 hanno prodotto risultati (i test TA2, TA4, TA5, TB1, TB2, TB3 e TB4 non hanno prodotto risultati) ed il motivo del fallimento di questi test potrebbe essere un errore di lettura dei dati da parte degli algoritmi di ML utilizzati, oppure che non esistono soluzioni per le matrici di calcolo dei regressori. Tuttavia, questa seconda ipotesi dovrebbe essere investigata più nel dettaglio prima di essere identificata con certezza.

Per comprendere le correlazioni oggetto di questa tesi, i test possono essere discussi sia sulla base del modello di regressione che sulla base al dataset di training utilizzato. Con questo approccio è possibile effettuare una valutazione complessiva dei modelli di regressione, nonché una valutazione dell'accuratezza delle predizioni. La valutazione dei modelli di regressione permette di evidenziare vantaggi e svantaggi di ognuno di essi nelle applicazioni ai dati geologici analizzati. Inoltre, l'utilizzo dei parametri di validazione dei risultati per la definizione dell'accuratezza del modello costituisce una parte fondamentale del processo di discussione dei risultati ottenuti.

L'effettiva determinazione di correlazioni tra i motivi geologici-strutturali e quelli giacimentologici nell'area di studio è basata sulla valutazione dell'importanza dei singoli predittori nel calcolo di statistica multivariata. A partire da queste valutazioni, espresse come percentuale di importanza delle caratteristiche, è possibile ottenere un'indicazione sul peso di ogni singola variabile sul calcolo. Anche uno studio della variazione delle predizioni con la probabilità è un ottimo strumento per verificare la bontà delle correlazioni e per individuare eventuali nuove correlazioni che non si apprezzano con un'analisi superficiale dei dati.

5.1 Valutazione dei regressori

La valutazione dei modelli di regressione permette di stabilire quale algoritmo si adatta meglio ai dati di input. Questa valutazione si basa sui parametri metrici RMSE, MAE ed R^2 , e sul confronto grafico dei risultati a parità di predittori utilizzati. Per mostrare in dettaglio come questa valutazione è stata effettuata per i cinque modelli di regressione utilizzati, di seguito si presenta a titolo di esempio la valutazione del test TB6 (Tabella 5.1). Utilizzando il criterio proposto da Lary et al. (2016) – per cui consideriamo buone le correlazioni che esprimono un valore di coefficiente di correlazione R^2 compreso tra 0.80 e 0.97 – dai dati di Tab. 5.1 identifico i regressori CatBoost (a) e Random Forest (d) come quelli che generano i risultati migliori per la combinazione di predittori utilizzata. Questi predittori includono permeabilità, solubilità di Hg, flusso di calore, distanza dalla faglia più prossima e distanza dall'asse di piega più prossimo. Si noti che i regressori (a) e (d) i valori di MAE e RMSE risultano minori di 0.1, e quindi confermano la bontà della predizione.

Tabella 5.1. Valutazione dell'accuratezza dei regressori utilizzati nel test TB6.

		(a)	(b)	(c)	(d)	(e)
		CatBoost	Ordinary Least Squares	Multilayer Perceptron	Random Forest	Support Vector
TB6	RMSE	0.09	0.43	0.35	0.06	0.36
	MAE	0.03	0.38	0.26	0.02	0.23
	R ²	0.94	-0.28	0.13	0.97	0.08

I valori di R² espressi dai regressori Ordinary Least Square (b), Multilayer Perceptron (c) e Support Vector (e) risultano invece molto bassi, in un caso <0, e sono associati a valori di MAE e RMSE piuttosto elevati (sempre >0.23). Nell'insieme, i valori dei parametri di accuratezza per il test TB6 denotano quindi buone correlazioni per i regressori (a) e (d) e scarse correlazioni per i regressori (b), (c) ed (e). Tuttavia, dall'analisi grafica delle interpolazioni prodotte a partire dai predittori scelti (Fig. 4.15a-c), si può affermare che i regressori (a) e (d) non presentano correlazioni significative con le miniere e le manifestazioni geotermiche.

Complessivamente, sulla base dei parametri metrici di accuratezza e sull'analisi delle interpolazioni, si può confermare che i regressori CatBoost e Random Forest basati sugli alberi decisionali sono i regressori che forniscono le migliori correlazioni con i dati di input.

5.2 Valutazione dei risultati

Per stabilire il migliore dei 17 test che hanno prodotto risultati si deve stabilire quale modello di regressione e quale combinazione di predittori fornisce i risultati con migliori correlazioni e parametri di accuratezza. A partire dai parametri di accuratezza, sono stati confrontati i tre migliori risultati di ogni dataset di training (Tabella 5.2).

Tabella 5.2. Valutazione dell'accuratezza dei regressori nel confronto tra i test T04, TA3 e TB8. Si nota come il regressore Ordinary Least Squares (b) presenti bassi valori di R² e alti valori di RMSE e MAE per tutti i test.

		(a)	(b)	(c)	(d)	(e)
		CatBoost	Ordinary Least Squares	Multilayer Perceptron	Random Forest	Support Vector
T04	RMSE	0.22	0.42	0.33	0.18	0.31
	MAE	0.15	0.38	0.27	0.11	0.22
	R ²	0.80	0.29	0.57	0.87	0.61
TA3	RMSE	0.00	0.40	0.09	0.00	0.16
	MAE	0.00	0.34	0.06	0.00	0.09
	R ²	1.00	0.34	0.96	1.00	0.89
TB8	RMSE	0.08	0.41	0.32	0.07	0.37
	MAE	0.03	0.37	0.24	0.02	0.23
	R ²	0.96	-0.16	0.28	0.97	0.06

Dal confronto dei tre test (Fig. 4.4a-e; Fig. 4.10a-e; Fig. 4.17a-e) è chiaro che le correlazioni variano al variare del dataset di training. Complessivamente, i regressori che mostrano le migliori correlazioni sono tutti quelli del test TB8, che utilizza il dataset di training aumentato B (training aumentato con 3110 punti). Per i tre test, i regressori migliori sono CatBoost e Random Forest, anche se per il test TA3 nessuno dei due è in grado di riprodurre il trend di distribuzione orientato N-S delle miniere (cf. Fig. 4.10a e Fig. 4.10e). Si noti che i valori dei parametri di accuratezza del test TA3 risultano massimi per CatBoost e Random Forest (Tabella 5.2), probabilmente a causa dell'*overfitting* del modello generato come conseguenza dell'approccio *copy-paste augmentation* (cf. Cap. 3.5.1). Anche gli altri tre regressori risentono del probabile *overfitting*, che genera un globale aumento dei valori di probabilità.

5.2.1 Analisi di importanza dei predittori

È possibile, attraverso l'analisi di importanza dei predittori, quantificare l'influenza di un singolo predittore sul calcolo effettuato (cf. Cap. 3.5.3).

Tabella 5.3. Analisi di importanza dei predittori. Valori moltiplicati per 100.

	Predittori giacimentologici						Predittori di vicinanza		
	Permeabilità	Solubilità	Flusso di calore	Temperatura	Gradiente termico	Anomalia di Bouguer	Distanza pieghe	Distanza faglie	Distanza tetto C4
T04	2.00	11.00					24.00	37.00	26.00
TA3	0.50	5.00	18.50	6.00	1.00	17.50	6.00	41.00	4.50
TB8	5.00	9.00		9.00			36.00	41.00	

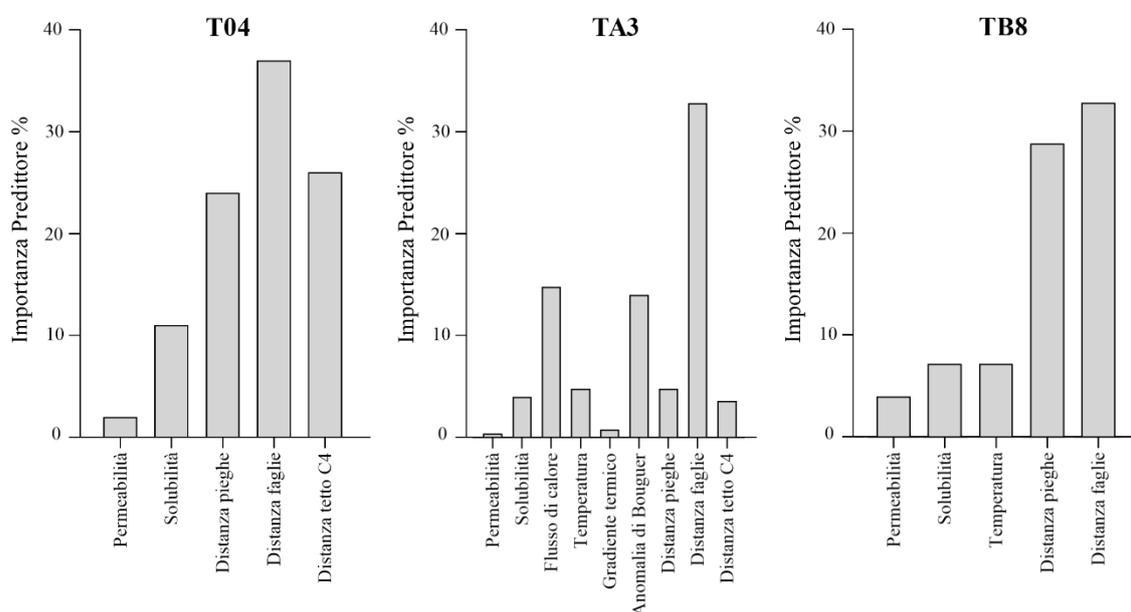


Figura 5.1. Importanza dei predittori sul calcolo. Per i tre predittori analizzati, la variabile che dà il contributo maggiore è il parametro di distanza dalla faglia più prossima.

L'influenza di un predittore sul calcolo determina quanto la correlazione sia dipendente da quel predittore. Dalle informazioni di Tab. 5.2 e dalle Figure 4.4a-e; 4.10a-e, e 4.17a-e si nota come il parametro che ha la maggiore importanza sul calcolo statistico sia il valore di distanza dalla faglia più prossima, seguito dalla distanza dall'asse di piega più prossimo. Gli altri predittori presentano valori variabili, ma sempre inferiori rispetto al contributo dei predittori di vicinanza.

5.2.2 Valutazione a partire dal dataset di training

I tre dataset di training, originale, aumentato A ed aumentato B, possono essere valutati combinando i parametri metrici di accuratezza con l'analisi grafica della correlazione tra distribuzione delle miniere e distribuzione delle aree a maggior probabilità. Confrontare tra loro i risultati delle predizioni effettuate a partire dai tre diversi dataset di training permette di determinare l'influenza della dimensione del dataset iniziale sul calcolo statistico, a patto che tale confronto venga fatto a parità di predittori utilizzati. Per confrontare al meglio i

risultati ottenuti, è utile mettere a confronto i risultati del Test 8 dei regressori CatBoost e Random Forest a parità di predittori utilizzati. È stato scelto il Test 8 perché presenta i migliori valori di parametri di accuratezza. Si ricorda che i predittori utilizzati nel Test 8 sono permeabilità, solubilità di Hg, temperatura, distanza dalla faglia più prossima e distanza dall'asse di piega più prossimo.

Tabella 5.4. Valutazione dell'accuratezza dei regressori riferiti ai test T08, TA8 e TB8.

		(a)	(b)	(c)	(d)	(e)
		CatBoost	Ordinary Least Squares	Multilayer Perceptron	Random Forest	Support Vector
T08	RMSE	0.23	0.47	0.38	0.17	0.35
	MAE	0.15	0.43	0.33	0.09	0.24
	R ²	0.78	0.12	0.43	0.89	0.51
TA8	RMSE	0.00	0.50	0.28	0.00	0.33
	MAE	0.00	0.45	0.20	0.00	0.19
	R ²	1.00	-0.02	0.68	1.00	0.54
TB8	RMSE	0.08	0.41	0.32	0.07	0.37
	MAE	0.03	0.37	0.24	0.02	0.23
	R ²	0.96	-0.16	0.28	0.97	0.06

Dal paragone dei parametri di accuratezza (Tab. 5.4) si deduce che l'utilizzo del dataset di training aumentato B ha comportato un miglioramento significativo nella qualità delle correlazioni prodotte, come indicato dal sostanziale aumento di R² e dalla diminuzione di RMSE e MAE. Anche dal punto di vista delle interpolazioni grafiche si nota un netto miglioramento nelle predizioni in relazione all'aumento del volume di dati contenuti nel dataset di training. Escludendo il test TA8 (Fig. 5.3) a causa dei problemi connessi all'*overfitting*, la scarsa correlazione tra distribuzione spaziale delle miniere e quella delle aree a maggior probabilità per il test T08 (Fig. 5.2) mostra che questo dataset non è adatto a mostrare le correlazioni geologiche-giacimentologiche. Il chiaro miglioramento nella qualità delle predizioni è invece visibile nel test TB8 (Fig. 5.4), che ha generato aree a maggiore probabilità che seguono una direzione N-S e che sono ben correlate la distribuzione spaziale delle miniere.

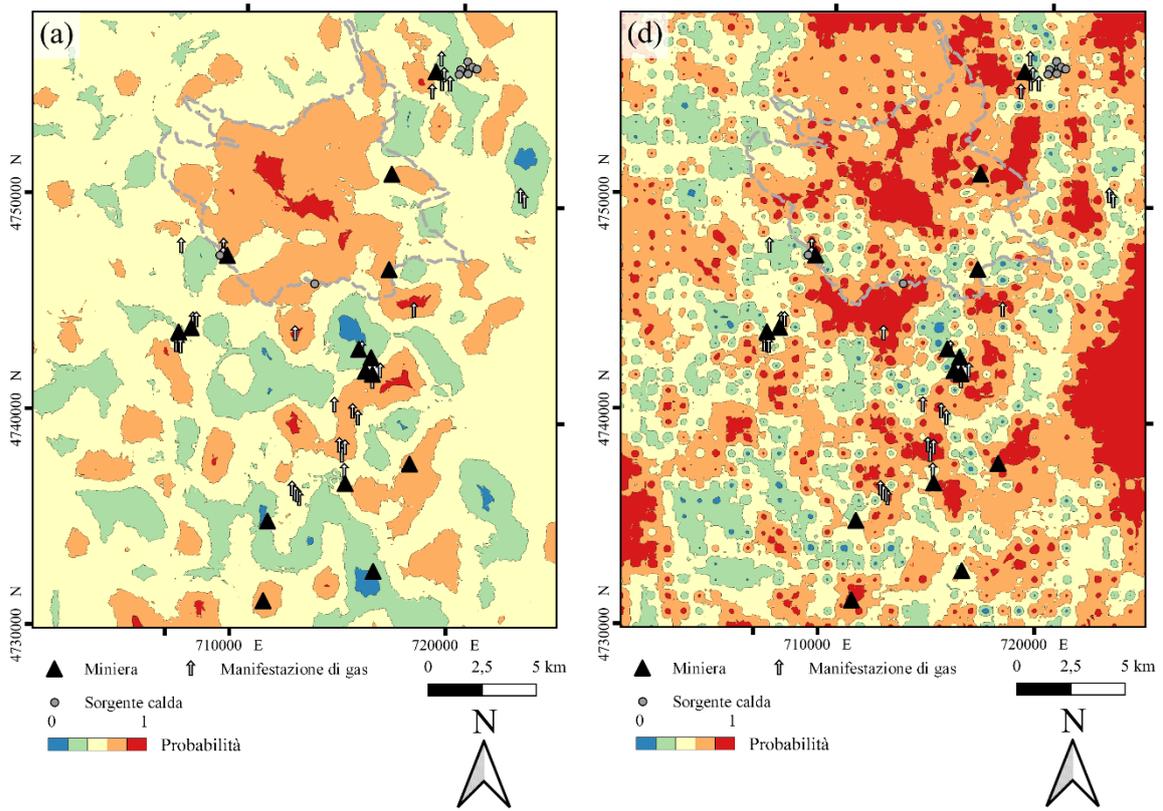


Figura 5.2. Risultati delle predizioni del test T08. (a): regressore CatBoost; (d): regressore Random Forest.

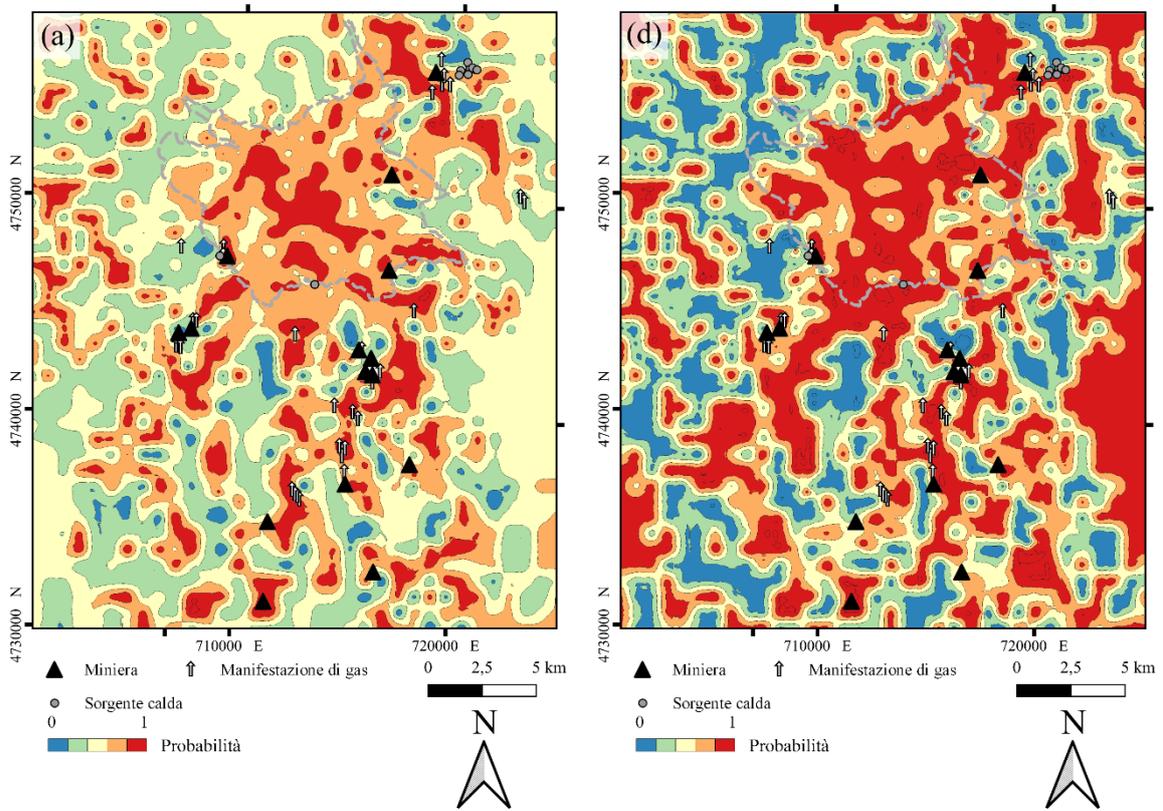


Figura 5.3. Risultati delle predizioni del test TA8. (a): regressore CatBoost; (d): regressore Random Forest.

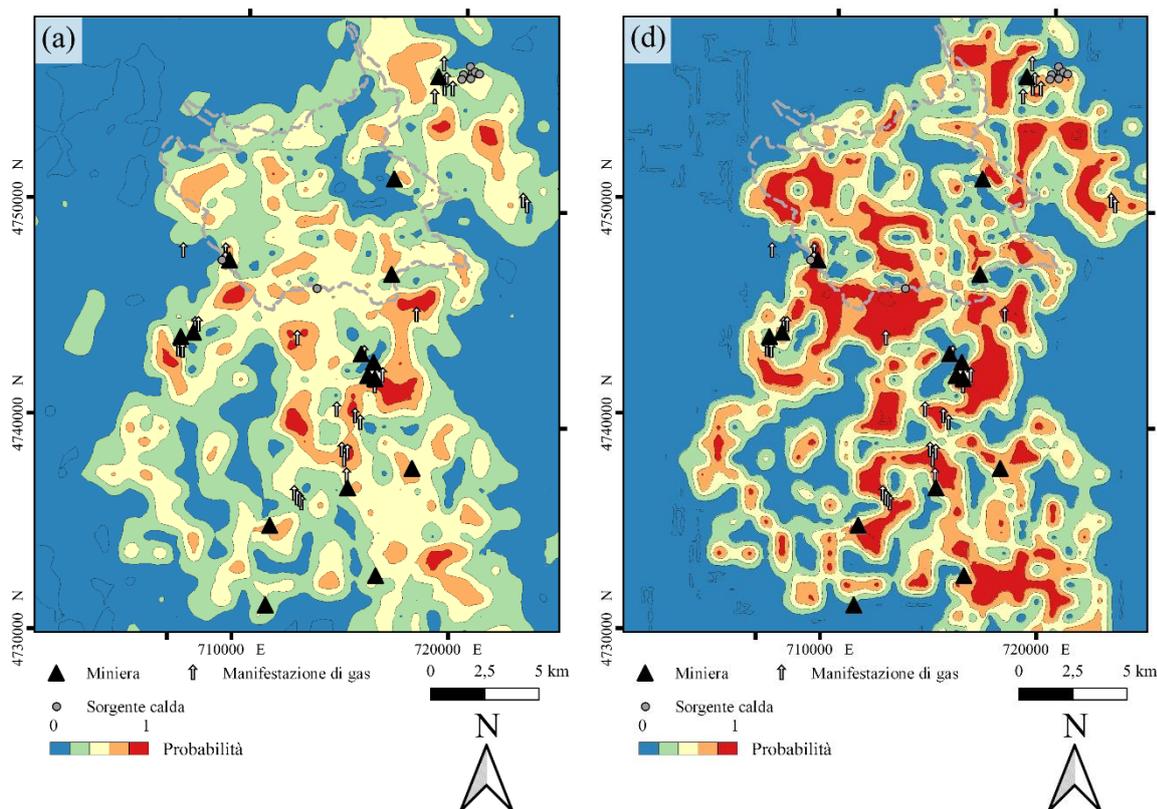


Figura 5.4. Risultati delle predizioni del test TB8. (a): regressore CatBoost; (d): regressore Random Forest.

La valutazione dell'importanza dei predittori sul calcolo è stata effettuata anche per il test 8 (Tab. 5.5) confrontando i risultati di ogni test.

Tabella 5.5. Analisi di importanza dei predittori dei test T08, TA8 e TB8. Valori moltiplicati per 100.

	Predittori geologici			Predittori di vicinanza	
	Permeabilità	Solubilità	Temperatura	Distanza pieghe	Distanza faglie
T08	5.60	6.00	6.40	31.00	51.00
TA8	16.00	6.00	6.00	38.00	34.00
TB8	5.00	9.00	9.00	36.00	41.00

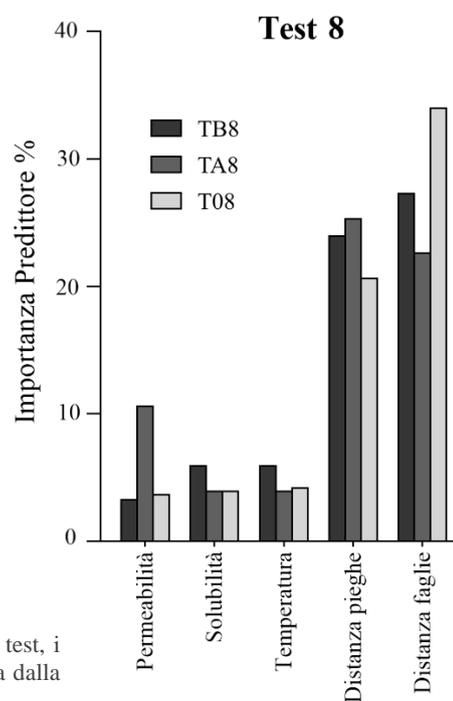


Figura 5.5. Importanza dei predittori sul calcolo. Si nota che, per tutti i test, i predittori che hanno maggiore influenza sulle predizioni sono la distanza dalla faglia più prossima e la distanza dall'asse di piega più prossimo.

Questi dati (Fig. 5.5) mostrano che, indipendentemente dal tipo di test effettuato, i predittori che hanno esercitato maggior peso sono stati la distanza dall'asse di piega e quella dalla faglia più prossima.

5.2.3 Casi particolari

Alcuni test effettuati a partire dal dataset di training originale mostrano scarse correlazioni con la posizione delle mineralizzazioni a cinabro e con l'assetto strutturale, dovute presumibilmente alla scarsità di dati di training. I test mostrano tendenzialmente distribuzioni che seguono l'andamento di alcuni predittori utilizzati, come nel regressore Ordinary Least Square (Fig. 5.6) del test T01 in cui è ipotizzabile un forte contributo da parte del dato di flusso di calore, non mostrando invece correlazioni con l'assetto strutturale o le mineralizzazioni presenti.

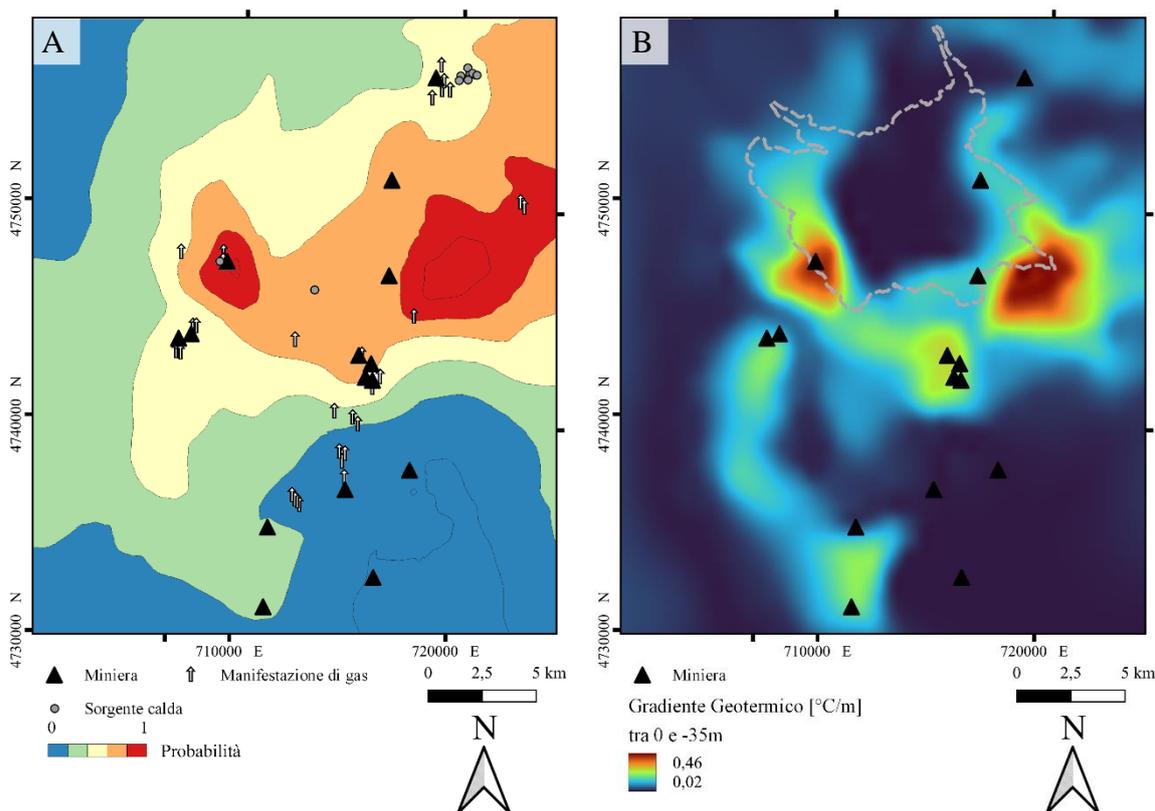


Figura 5.6. Risultato del test T01(b), regressore Ordinary Least Squares. Confronto tra il risultato della predizione effettuata per il test T01 (A) e la carta del gradiente geotermico misurato nei primi 35m di profondità (B). Nel test T01 sono stati utilizzati solo i sei predittori geologici.

È importante sottolineare come non esista sempre una corrispondenza univoca tra parametri di valutazione dell'accuratezza e la qualità delle interpolazioni dei risultati, come mostrato in Tab. 5.6, nel test T02(a) in cui il regressore CatBoost presenta un valore $R^2 = 0.93$ ma una distribuzione delle probabilità in carta (Fig. 5.7) che apparentemente non presenta correlazioni né con le manifestazioni geotermiche e le miniere, né con le faglie (Fig. 5.7a) o con le tracce assiali di piega (Fig. 5.7b).

Tabella 5.6. Valutazione dell'accuratezza dei regressori utilizzati nel test T02.

		(a) CatBoost	(b) Ordinary Least Squares	(c) Multilayer Perceptron	(d) Random Forest	(e) Support Vector
T02	RMSE	0.13	0.41	0.41	0.16	0.32
	MAE	0.07	0.37	0.37	0.09	0.22
	R^2	0.93	0.31	0.32	0.90	0.60

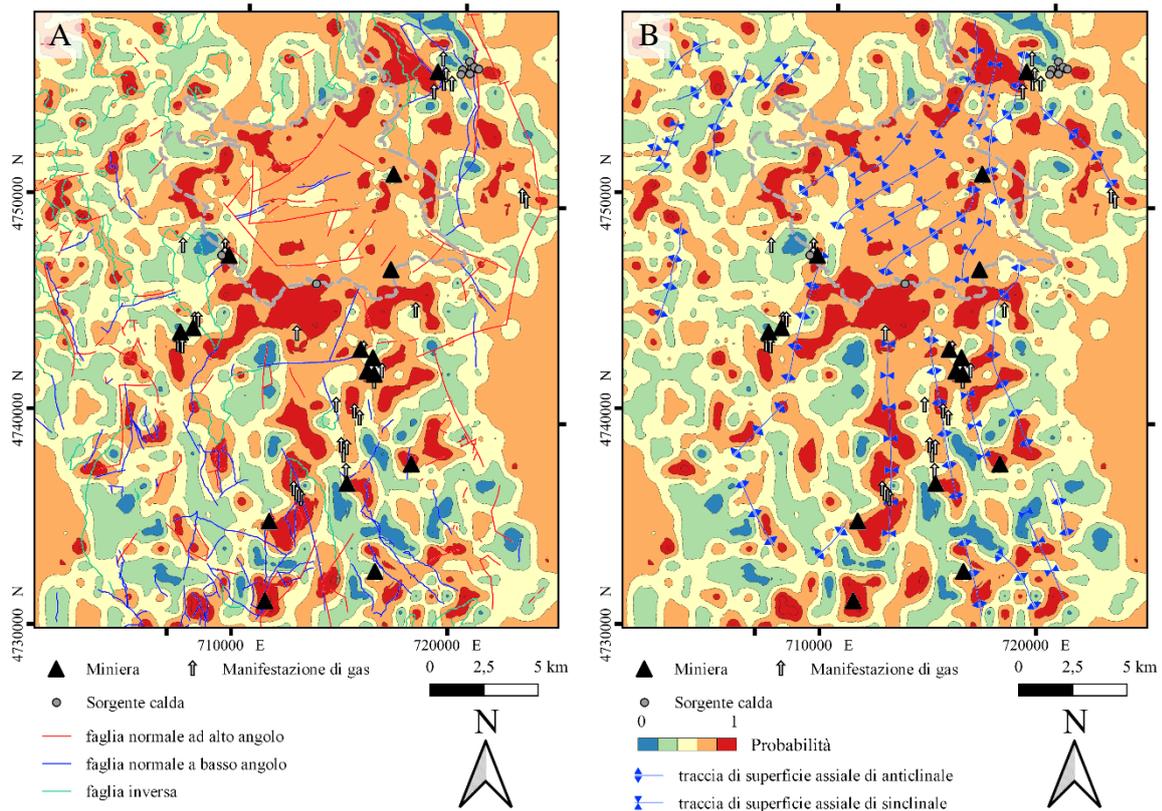


Figura 5.7. Risultati della predizione del test T02(a) – regressore CatBoost. (A): Sovrapposizione delle tracce di faglia ai risultati della predizione; (B): sovrapposizione delle tracce assiali di piega ai risultati della predizione.

Inoltre, in questo test, i predittori utilizzati sono solamente quelli di vicinanza. Per poter discriminare l'importanza dei soli predittori di vicinanza, sarebbe stato funzionale un confronto con i test TA2 e TB2, che però non hanno prodotto risultati.

5.3 Variabilità delle correlazioni con la profondità

I risultati di tutte le predizioni sono stati presentati a 0m dal piano campagna locale e a quota -100m dal piano campagna locale. Per il test migliore (test TB8) è stato effettuato anche uno studio a profondità maggiori, per cercare di comprendere come variano le distribuzioni di probabilità con la profondità.

Lo studio è stato eseguito per i regressori CatBoost (Fig. 5.8-9) e Random Forest (Fig. 5.10-11) del test TB8, elaborando le carte di interpolazione dei dati predetti ogni 100m, fino ad una profondità di -1000m dal piano campagna locale.

La distribuzione delle probabilità superficiale interpolata tramite l'uso di ArcGIS Pro (Fig. 5.8b) è diversa nei valori massimi rispetto a quella interpolata in QGIS (Fig.5.8a). Questo è dovuto all'utilizzo di due diversi interpolatori Kriging, nel primo caso il Kriging Bayesiano Empirico 3D (EBK3D), nel secondo il Kriging ordinario. La differenza risiede proprio nel numero di punti utilizzati per l'interpolazione: il Kriging ordinario interpola 12 punti con un raggio di ricerca fisso di 1000m nel piano bidimensionale, il EBK3D individua autonomamente il numero di punti da interpolare e il raggio di ricerca, che in questo caso è di 26 punti con raggio di ricerca variabile. Il risultato è che le due interpolazioni, pur utilizzando lo stesso dataset ed essendo graficamente simili, differiscono sulla base del tipo di interpolazione.

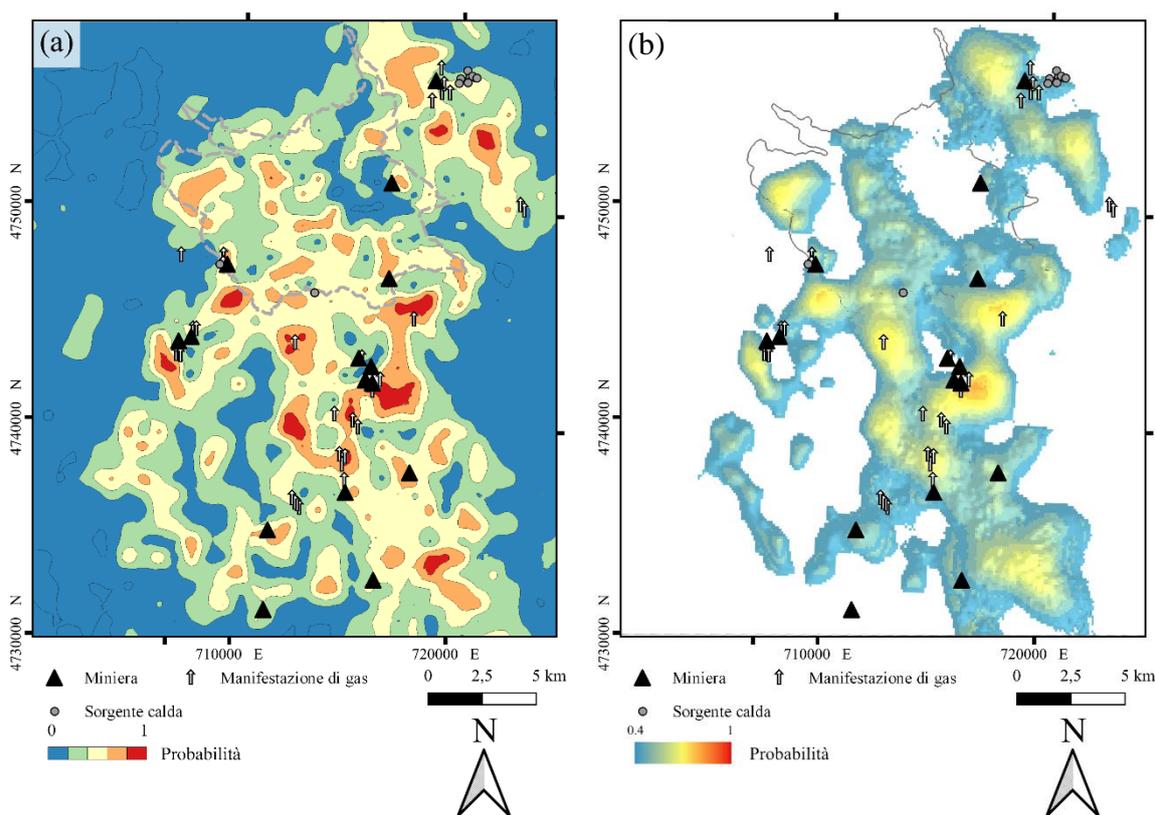


Figura 5.8. Confronto tra l'utilizzo di interpolatori del test TB8(a). (a) Carta delle probabilità interpolata utilizzando il Kriging ordinario bidimensionale. (b) Carta delle probabilità interpolata utilizzando il EBK3D. L'immagine è tridimensionale ma proiettata in vista ortografica (2D).

La Fig. 5.9 mostra una generale diminuzione dei valori di probabilità, che presenta i valori più alti in corrispondenza delle miniere delle Solforate e del Siele, nella porzione centrale dell'area di studio. Questo è in linea con l'evidenza che le mineralizzazioni a cinabro e

stibina si trovano generalmente a profondità che difficilmente superano i 400-500m (Zucchetti, 1964; Brogi et al., 2011; Forconi, 2011), come nel caso della Miniera di Abbadia San Salvatore. La predizione del regressore CatBoost del test TB8 secondo cui a -1000m dal piano campagna si potrebbe ancora trovare la mineralizzazione (probabilità comprese tra 0.4 e 0.5) rappresenta un'idea esplorativa potenzialmente testabile.

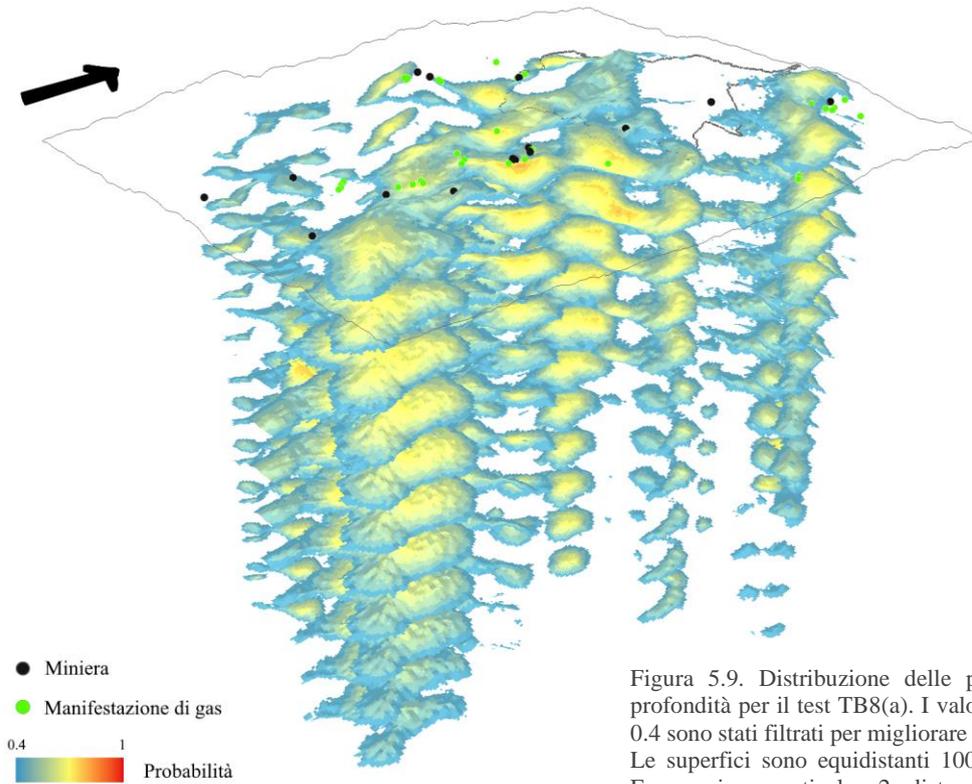


Figura 5.9. Distribuzione delle probabilità con la profondità per il test TB8(a). I valori di probabilità < 0.4 sono stati filtrati per migliorare la visualizzazione. Le superfici sono equidistanti 100m sulla verticale. Esagerazione verticale x2, distanza tra le superfici x10.

Per quanto riguarda il risultato del regressore Random Forest del test TB8 (Fig. 5.10-11), si possono fare le stesse considerazioni fatte per il test appena descritto. Anche in questo caso è visibile una netta differenza di interpolazione delle distribuzioni superficiali di probabilità (Fig. 5.10a-b), data sempre dai differenti metodi utilizzati. La variabilità con la profondità è maggiormente visibile in questo test rispetto al precedente poiché, generalmente, i massimi valori di probabilità diminuiscono fortemente entro i primi 600m dal piano campagna (Fig. 5.11).

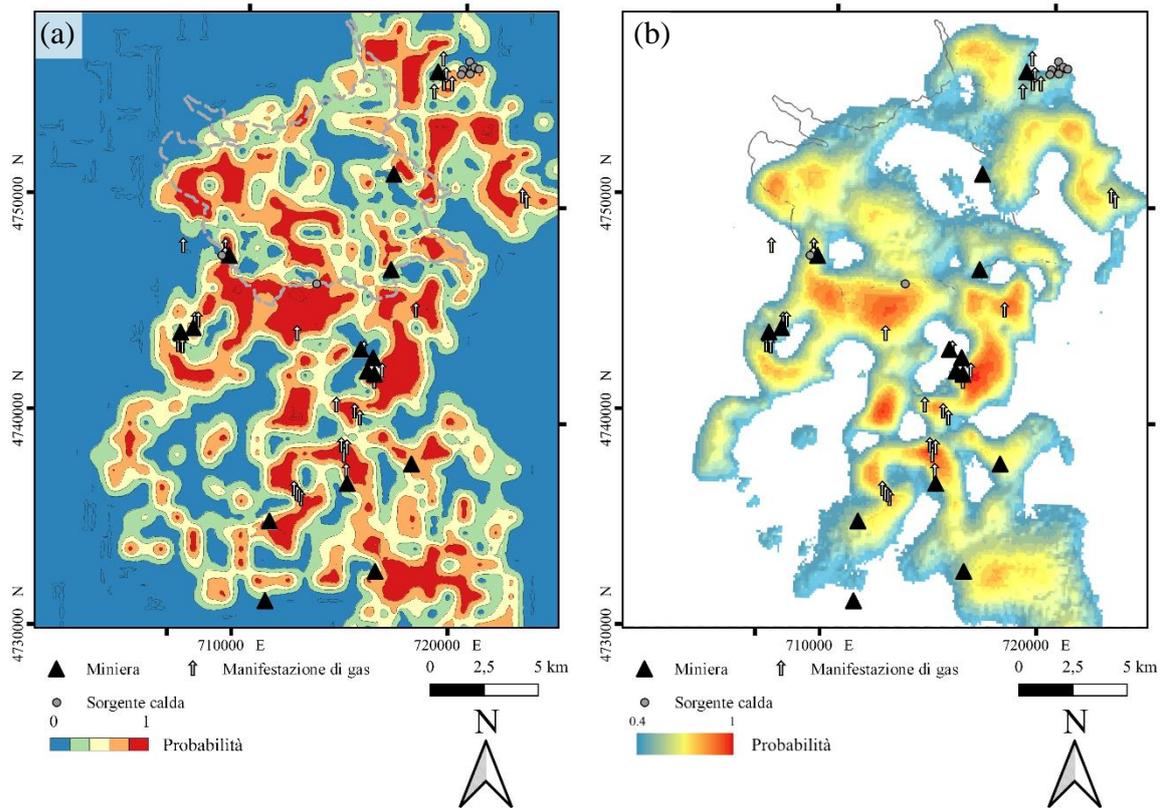


Figura 5.10. Confronto tra l'utilizzo di interpolatori del test TB8(a), regressore CatBoost. (a) Carta delle probabilità interpolata utilizzando il Kriging ordinario bidimensionale. (b) Carta delle probabilità interpolata utilizzando il EBK3D. Con l'utilizzo di EBK3D si ha la perdita di tutte le piccole aree a probabilità massima dovute all'effetto del Kriging ordinario.

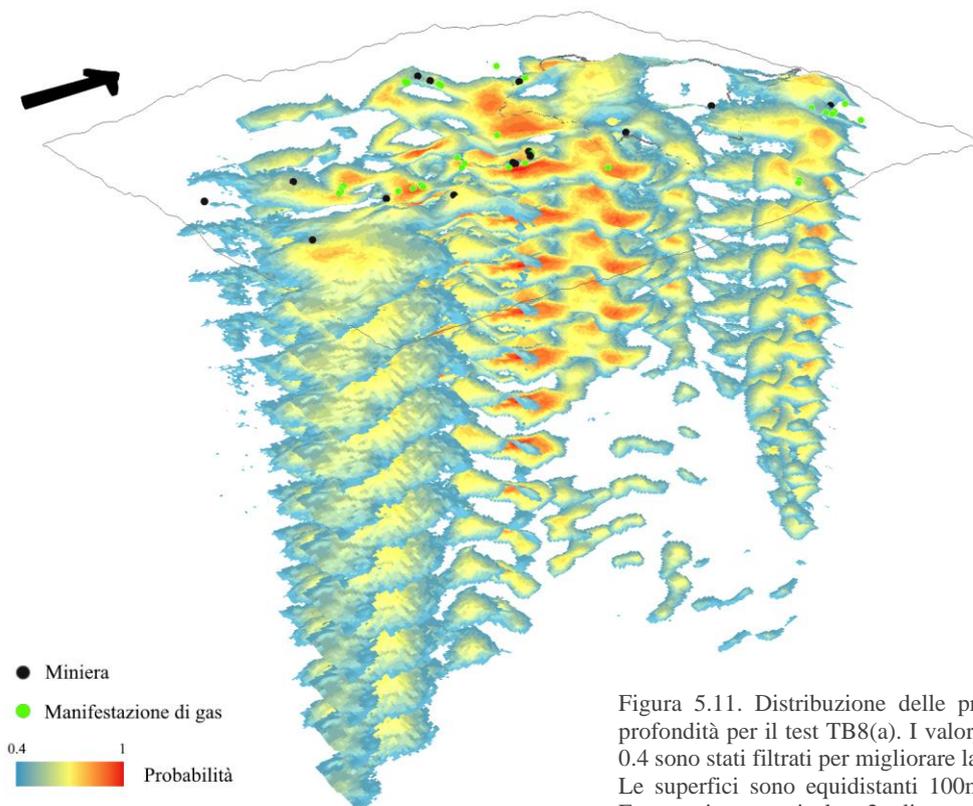


Figura 5.11. Distribuzione delle probabilità con la profondità per il test TB8(a). I valori di probabilità < 0.4 sono stati filtrati per migliorare la visualizzazione. Le superfici sono equidistanti 100m sulla verticale. Esagerazione verticale x2, distanza tra le superfici x10.

Conclusioni

Sulla base dei risultati ottenuti dall'uso di diverse combinazioni di predittori e di diversi modelli di regressione, si può affermare che – nonostante la scarsità di dati geologici e giacimentologici utilizzati in rapporto con la dimensione dell'area di studio – le applicazioni del ML a questa tipologia di indagine conducono a delle conclusioni che si possono riassumere nei seguenti elementi.

In primo luogo, dai risultati emerge che i modelli di regressione basati su alberi decisionali (CatBoost e Random Forest) sono complessivamente più performanti e geologicamente significativi rispetto agli altri modelli, indipendentemente dal tipo di test effettuato (con dataset di training aumentato o non). Questi secondi algoritmi tendono ad essere più performanti quando si ha a che fare con grandi moli di dati, poiché le operazioni di calcolo vengono effettuate correlando il più alto numero possibile di nodi. I modelli basati su alberi decisionali operano invece sul singolo nodo, generando una suddivisione del dataset in base alla scelta presa in ogni nodo; solo al termine del calcolo il modello effettua una stima media di tutte le decisioni prese e genera un risultato. Per questo motivo, gli algoritmi basati su alberi decisionali riescono ad effettuare predizioni sensate anche con un piccolo dataset di informazioni.

In secondo luogo, le tecniche di ML si sono dimostrate utili nell'identificare potenziali correlazioni tra i diversi fattori geologico-strutturali che interessano il sistema geotermico del M.te Amiata. Utilizzare questo approccio in fase di esplorazione potrebbe essere utile perché è economico, può guidare verifiche di campagna nelle fasi successive, e le correlazioni possono essere interpretate attraverso l'utilizzo di software GIS *open source*. Potenzialmente, questo approccio permetterebbe di produrre carte di probabilità con una risoluzione dipendente dalla risoluzione dei dati processati. I dati originali utilizzati per questo lavoro hanno una risoluzione verticale di 10 metri; quindi, il numero di carte che potrebbero essere prodotte è notevolmente superiore a quello mostrato in questo elaborato. Bisogna però tenere presente che il processo di produzione delle carte a partire dai risultati delle operazioni ML richiede molto tempo all'operatore ed anche all'elaborazione software. Quindi, è necessario pianificare la quantità di carte prodotte sulla base della risoluzione finale che si vuole ottenere. In generale, le carte di probabilità si possono produrre con lo scopo di correlarle con il dato geologico di interesse, in modo da vincolare specifici processi.

Nel dettaglio, le combinazioni di predittori utilizzate per il sistema geotermico del M.te Amiata sono state scelte sulla base delle caratteristiche geologico-strutturali dell'area e sulla base dei fattori che controllano efficacemente la precipitazione del cinabro. Dai risultati emerge che la peculiare distribuzione spaziale N-S dei giacimenti del distretto dipende dalla combinazione di un sistema di faglie e di pieghe. È infatti importante notare che – presi singolarmente – questi predittori non hanno generato delle significative correlazioni, suggerendo che soltanto la loro combinazione è geologicamente importante. Questa combinazione di fattori geologici, che si può interpretare come decisiva sulla base dell'importanza statistica di questi due predittori, suggerisce che un sistema interconnesso di clivaggio di piano assiale e di fratture favorevolmente orientate devono aver agito insieme ed in modo concorde nel sistema geotermico per creare una struttura di permeabilità efficiente per il flusso geotermico. Questa struttura deve aver controllato la precipitazione di >117 kt di Hg lungo una fascia relativamente ristretta di territorio, che rappresenta un potenziale target esplorativo per altre mineralizzazioni o altre emergenze geotermiche

(sorgenti calde, venute di gas, ecc). Ciò ha delle evidenti implicazioni anche nel campo della valutazione dei rischi ambientali associati a questo tipo di controllo strutturale.

Un quarto elemento di riflessione che emerge dal confronto dei predittori è che alcuni predittori utilizzati hanno mostrato una scarsa (perfino negativa) influenza sulle correlazioni calcolate. Esempi rilevanti sono l'anomalia di Bouguer, il flusso di calore ed il gradiente geotermico, che in un caso hanno anche forzato il calcolo in maniera evidente. È importante notare che, nonostante i risultati ottenuti in questo lavoro, sarebbe fuorviante escludere a priori il contributo di questi fattori nel processo geotermico, e che i risultati ottenuti dall'analisi di ML permettono di individuare potenziali correlazioni e non vincoli di fatto.

In conclusione, questo lavoro suggerisce che il ML rappresenta uno strumento in grado di suggerire nuove e poco sperimentate relazioni tra elementi geologici-giacimentologici di un'area complessa come un sistema geotermico. Questo strumento è in grado di guidare fasi successive di uno studio geologico complesso ed è probabilmente quello che in futuro sarà in grado di elaborare una mole notevole di dati geologici (*big data*) che sarebbero ingestibili senza modelli di apprendimento automatico.

Bibliografia

- AA.VV. (2020). *Study on the EU's list of critical raw materials (2020) : final report* (pp. 1-158). Publications Office. <https://doi.org/doi/10.2873/904613>
- Accaino, F., Tinivella, U., Rossi, G., & Nicolich, R. (2005). Imaging of CROP-18 deep seismic crustal data. *Bollettino della Societa Geologica Italiana, Supplemento*, 3, 195-204.
- Acocella, V. (2000). Space accommodation by roof lifting during pluton emplacement at Amiata (Italy). *Terra Nova*, 12(4), 149-155. <https://doi.org/10.1046/j.1365-3121.2000.00286.x>
- Anderson, T. R., & Fairley, J. P. (2008). Relating permeability to the structural setting of a fault-controlled hydrothermal system in southeast Oregon, USA. *Journal of Geophysical Research*, 113(B5). <https://doi.org/10.1029/2007jb004962>
- Ardigò. (1983). *Geology of the M.te Amiata mercury district* (pp. 1-40). Rimin.
- Arisi Rota, F., Brondi, A., Dessau, G., & Franzini, M. (1971). *I Giacimenti minerari* (Vol. 27, pp.
- Avanzinelli, R., Lustrino, M., Mattei, M., Melluso, L., & Conticelli, S. (2009). Potassic and ultrapotassic magmatism in the Circum-Tyrrhenian region: Significance of carbonated pelitic vs. pelitic sediment recycling at destructive plate margins. *Lithos*, 113, 213-227. <https://doi.org/10.1016/j.lithos.2009.03.029>
- Awad, M., & Khanna, R. (2015). Support Vector Regression. In *Efficient Learning Machines*. (pp. 67-80). Apress. https://doi.org/10.1007/978-1-4302-5990-9_4
- Bacci, E., Gaggi, C., Lanzillotti, E., Ferrozzi, S., & Valli, L. (2000). Geothermal power plants at Mt. Amiata (Tuscany–Italy): mercury and hydrogen sulphide deposition revealed by vegetation. *Chemosphere*, 40(8), 907-911. [https://doi.org/https://doi.org/10.1016/S0045-6535\(99\)00458-0](https://doi.org/https://doi.org/10.1016/S0045-6535(99)00458-0)
- Baldi, P., Bellani, S., Ceccarelli, A., Fiordelisi, A., Rocchi, G., Squarci, P., & Taffi, L. (1995). Geothermal anomalies and structural features of southern Tuscany. <https://doi.org/10.13140/RG.2.1.2088.5923>
- Baldi, P., Bertini, G., Cameli, G., Decandia, F., Dini, I., Lazzarotto, A., & Liotta, D. (1994). La tettonica distensiva post-collisionale nell'area geotermica di Larderello (Toscana meridionale). In Università di Camerino (Ed.), *Studi Geologici Camerti* (Vol. spec. 1, pp. 183-193).
- Baldi, P., Decandia, F. A., Lazzarotto, A., & Calamai, A. (1974). Studio geologico del substrato della copertura vulcanica laziale nella zona dei laghi di Bolsena, Vico e Bracciano. *Mem. Soc. Geol. It.*, 13(4), 575-606.
- Barbier, E. (1997). Nature and technology of geothermal energy: A review. *Renewable and Sustainable Energy Reviews*, 1(1-2), 1-69. [https://doi.org/10.1016/s1364-0321\(97\)00001-4](https://doi.org/10.1016/s1364-0321(97)00001-4)
- Barchi, M., Landuzzi, A., Minelli, G., & Piali, G. (2001). Outer Northern Apennines. In Vai, G. B. & Martini, I. P. (Eds.), *Anatomy of an Orogen: the Apennines and Adjacent Mediterranean Basins* (pp. 215-253). Springer Netherlands. https://doi.org/10.1007/978-94-015-9829-3_15
- Barelli, A., Bertini, G., Buonasorte, G., Cappetti, G., & Fiordelisi, A. (2000). Recent deep exploration results at the margins of the Larderello Travale geothermal system. *Proceedings World Geothermal Congress*, 965-970.

- Barelli, A., Ceccarelli, A., Dini, I., Fiordelisi, A., Giorgi, N., Lovari, F., & Romagnoli, P. (2010). A Review of the Mt. Amiata Geothermal System (Italy). *Proceedings World Geothermal Congress*
- Barnes, H. L. (2015). Hydrothermal Processes: The development of geochemical concepts in the latter half of the twentieth century. *Geochemical Perspectives*, 4, 1-93. <https://doi.org/10.7185/geochempersp.4.1>
- Barnes, H. L., & Seward, T. M. (1997). *Geochemistry of hydrothermal ore deposits* (pp. 699-736). John Wiley & Sons.
- Batini, F., Brogi, A., Lazzarotto, A., Liotta, D., & Pandeli, E. (2003). Geological features of Larderello-Travale and Mt. Amiata geothermal areas (southern Tuscany, Italy). *Episodes*, 26(3), 239-244. <https://doi.org/10.18814/epiiugs/2003/v26i3/015>
- Beardmore, G. R., & Cull, J. P. (2001). *Crustal heat flow: a guide to measurement and modelling* (pp. Cambridge University Press).
- Bellani, S., & Gherardi, F. (2009). *Thermal Modeling of an Area West of the Mt. Amiata Geothermal Field, Italy* (pp. <https://doi.org/10.13140/RG.2.1.2514.5767>)
- Berger, B. R. (1983). Geologic-Geochemical Features of Hot-Spring. *US Geological Survey Bulletin*(1646), 47.
- Bernabini, M., Favaro, P., & Orlando, L. (1994). Density in Bouguer anomalies and its consequences. *Journal of applied geophysics*, 32(2-3), 187-197.
- Bertani, R. (2010). *Geothermal power generation in the world 2005–2010 update report* (0375-6505). In World Geothermal Congress 2010, Bali, Indonesia, 25-29 April 2010. <https://www.sciencedirect.com/science/article/pii/S0375650511000526>
- Bertini, G., Cappetti, G., Dini, I., & Lovari, F. (1995). Deep drilling results and updating of geothermal knowledge on the Monte Amiata area. *Proceedings World Geothermal Congress*, 2, 1283-1286.
- Billi, B., Cappetti, G., & Luccioli, F. (1986). Enel activity in the research, exploration and exploitation of geothermal energy in Italy. *Geothermics*, 15(5), 765-779. [https://doi.org/https://doi.org/10.1016/0375-6505\(86\)90090-8](https://doi.org/https://doi.org/10.1016/0375-6505(86)90090-8)
- Boncianni, F., Callegari, I., Conti, P., Cornamusini, G., & Carmignani, L. (2005). Neogene post-collisional evolution of the internal Northern Apennines: insights from the upper Fiora and Albegna valleys, (Mt. Amiata geothermal area, southern Tuscany). *Bollettino- Societa Geologica Italiana, Vol Spec 3*, 103-118.
- Bosellini, A. (2005). *Storia Geologica d'Italia: gli ultimi 200 milioni di anni* (pp. 58-61). Zanichelli.
- Branconi, E. (1976). La miniera mercurifera d'Abbadia San Salvatore. *L'Industria Mineraria*, XXVII, 339-353.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification And Regression Trees* (pp. 215-265). Chapman & Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Brogi, A. (2004a). Assetto geologico del nucleo di Falda Toscana affiorante nel settore occidentale del Monte Amiata (Appennino Settentrionale): strutture pre-e sin-collisionali relitte preservate durante lo sviluppo della tettonica distensiva post-collisionale. *Bollettino della Societa Geologica Italiana*, 123(3), 443-462.
- Brogi, A. (2004b). Miocene low-angle detachments and upper crust megaboudinage in the Mt. Amiata geothermal area (Northern Apennines, Italy). *Geodinamica Acta*, 17(6), 375-387. <https://doi.org/10.3166/ga.17.375-387>

- Brogi, A. (2008a). The structure of the Monte Amiata volcano-geothermal area (Northern Apennines, Italy): Neogene-Quaternary compression versus extension. *International Journal of Earth Sciences*, 97, 677-703. <https://doi.org/10.1007/s00531-007-0191-1>
- Brogi, A. (2008b). The Triassic and Palaeozoic successions drilled in the Bagnore Geothermal field and Poggio Nibbio area (Monte Amiata, Northern Apennines, Italy). *Bollettino della Societa Geologica Italiana*, 127, 599-613.
- Brogi, A., & Fabbrini, L. (2009). Extensional and strike-slip tectonics across the Monte Amiata–Monte Cetona transect (Northern Apennines, Italy) and seismotectonic implications. *Tectonophysics*, 476(1-2), 195-209.
- Brogi, A., Fabbrini, L., & Liotta, D. (2011). Sb–Hg ore deposit distribution controlled by brittle structures: The case of the Selvena mining district (Monte Amiata, Tuscany, Italy). *Ore Geology Reviews*, 41(1), 35-48. <https://doi.org/10.1016/j.oregeorev.2011.06.004>
- Brogi, A., Liotta, D., Meccheri, M., & Fabbrini, L. (2010). Transtensional shear zones controlling volcanic eruptions: the Middle Pleistocene Mt Amiata volcano (inner Northern Apennines, Italy). 22(2), 137-146. <https://doi.org/10.1111/j.1365-3121.2010.00927.x>
- Brogi, A., Liotta, D., Ruggieri, G., Capezzuoli, E., Meccheri, M., & Dini, A. (2016). An overview on the characteristics of geothermal carbonate reservoirs in southern Tuscany. *Italian Journal of Geosciences*, 135(1), 17-29. <https://doi.org/10.3301/ijg.2014.41>
- Brown, K. L. (1986). Gold deposition from geothermal discharges in New Zealand. *Economic Geology*, 81(4), 979-983.
- Brown, K. L., & Simmons, S. F. (2003). Precious metals in high-temperature geothermal systems in New Zealand. *Geothermics*, 32(4-6), 619-625. [https://doi.org/10.1016/s0375-6505\(03\)00049-x](https://doi.org/10.1016/s0375-6505(03)00049-x)
- Cadoux, A., & Pinti, D. L. (2009). Hybrid character and pre-eruptive events of Mt. Amiata volcano (Italy) inferred from geochronological, petro-geochemical and isotopic data. *Journal of Volcanology and Geothermal Research*, 179(3-4), 169-190. <https://doi.org/10.1016/j.jvolgeores.2008.10.018>
- Calamai, A., Cataldi, R., Squarci, P., & Taffi, L. (1970). Geology, Geophysics and Hydrogeology of the Monte Amiata Geothermal Fields. *Geothermics, Special Issue I*, 1-9.
- Carmignani, L., Decandia, F. A., Fantozzi, P. L., Lazzarotto, A., Liotta, D., & Meccheri, M. (1994). Tertiary extensional tectonics in Tuscany (northern Apennines, Italy). *Tectonophysics*, 238(1-4), 295-315.
- Cataldi, R. (1967). Remarks on the geothermal research in the region of Monte Amiata (Tuscany — Italy). 30(1), 243-269. <https://doi.org/10.1007/bf02597673>
- Caté, A., Perozzi, L., Gloaguen, E., & Blouin, M. (2017). Machine learning as a tool for geologists. *The Leading Edge*, 36(3), 215-219. <https://doi.org/https://doi.org/10.1190/tle36030215.1>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247-1250. <https://doi.org/10.5194/gmd-7-1247-2014>
- Chawla, N. V. (2005). Data Mining for Imbalanced Datasets: An Overview. In *Data Mining and Knowledge Discovery Handbook* (pp. 853-867). Springer-Verlag. https://doi.org/10.1007/0-387-25465-x_40
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, 1-24.

- Choi, R. Y., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., & Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational Vision Science & Technology*, 9(2), 1-12. <https://doi.org/https://doi.org/10.1167/tvst.9.2.14>
- Clever, H. L., Johnson, S. A., & Derrick, M. E. (1985). The Solubility of Mercury and Some Sparingly Soluble Mercury Salts in Water and Aqueous Electrolyte Solutions. *Journal of Physical and Chemical Reference Data*, 14(3), 631-680. <https://doi.org/10.1063/1.555732>
- Cohen, P. R., & Jensen, D. (1997). Overfitting explained. Sixth International Workshop on Artificial Intelligence and Statistics,
- Conti, P., & Lazzarotto, A. (2004). Geology of Tuscany: evolution of the state-of-knowledge presented by geological maps and the new geological map of Tuscany, 1: 250,000 scale. *The 'Regione Toscana' project of geological mapping, Spec. vol for the 32nd IGC*, 25-50.
- Conticelli, S., Boari, E., Burlamacchi, L., Cifelli, F., Moscardi, F., Laurenzi, M. A., Ferrari Pedraglio, L., Francalanci, L., Benvenuti, M. G., Braschi, E., & Manetti, P. (2015). Geochemistry and Sr-Nd-Pb isotopes of Monte Amiata Volcano, Central Italy: evidence for magma mixing between high-K calc-alkaline and leucititic mantle-derived magmas. *Italian Journal of Geosciences*, 134(2), 266-290. <https://doi.org/10.3301/ijg.2015.12>
- Cox, S. F. (2005). Coupling between deformation, fluid pressures, and fluid flow in ore-producing hydrothermal systems at depth in the crust. *Economic Geology, 100th Anniversary Volume*, 39-75. <https://doi.org/https://doi.org/10.5382/AV100.04>
- Cracknell, M. J., Reading, A. M., & McNeill, A. W. (2014). Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and Self-Organising Maps. *Australian Journal of Earth Sciences*, 61(2), 287-304. <https://doi.org/10.1080/08120099.2014.858081>
- Dallmeyer, R. D., & Liotta, D. (1998). Extension, uplift of rocks and cooling ages in thinned crustal provinces: the Larderello geothermal area (inner Northern Apennines, Italy). *Geological Magazine*, 135(2), 193-202. <https://doi.org/10.1017/s0016756898008309>
- De Castro, C. (1914). Genesi dei giacimenti cinabreriferi del Monte Amiata. In *Memorie Descrittive Carta Geologica d'Italia* (pp. 1-77).
- De Ferrari, P. (1890). *La Miniere di Mercurio del Monte Amiata* (pp. 1-179). Tipografia G. Berbera.
- Dehm, R. M., Klemm, D. D., Müller, C., Wagner, J., & Weber-Diefenbach, K. (1983). Exploration for antimony deposits in southern Tuscany, Italy [Article]. *Mineralium Deposita*, 18(2 Supplement), 423-434. <https://doi.org/10.1007/BF00206490>
- Des Marais, D. J., & Walter, M. R. (2019). Terrestrial Hot Spring Systems: Introduction. *Astrobiology*, 19(12), 1419-1432. <https://doi.org/10.1089/ast.2018.1976>
- Dessau, G. (1952). Antimony deposits of Tuscany. *Economic Geology*, 47(4), 397-413.
- Dessau, G., Duchi, V., & Stea, B. (1972). Geologia e depositi minerari della zona Monti Romani-Monteti. *Mem. Soc. Geol. It.*, 11, 217-260.
- Dini, A. (2017). Miniere e minerali del Distretto mercurifero del Monte Amiata. In *Principe, C., Lavorini, G., Vezzoli, L.; Il Vulcano di Monte Amiata;* (pp. 343-369).
- Dini, I., Ceccarelli, A., Brogi, A., Giorgi, N., Galleni, P., & Rossi, L. (2010). Geological evaluation of the base of the Mt. Amiata volcanic complex (Tuscany, Italy). *Proceedings World Geothermal Congress*,
- Dobak, P. J., Robert, F., Barker, S. L. L., Vaughan, J. R., Eck, D., Sillitoe, R. H., Goldfarb, R. J., Robert, F., & Simmons, S. F. (2020). Chapter 15: Goldstrike Gold System,

- North Carlin Trend, Nevada, USA. In *Geology of the World's Major Gold Deposits and Provinces* (Vol. 23, pp. 313-334). Society of Economic Geologists. <https://doi.org/10.5382/sp.23.15>
- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv pre-print server*. <https://arxiv.org/abs/1810.11363v1>
- Duchi, V. (1976). Determinazione del mercurio mediante spettrofotometria di assorbimento atomico senza fiamma: Dosaggio in travertini dell'Italia centrale. *European Journal of Mineralogy*, 32(1), 243-260.
- Duchi, V., Minissale, A., & Prati, F. (1987). Chemical composition of thermal springs, cold springs, streams, and gas vents in the Mt. Amiata geothermal region (Tuscany, Italy). *Journal of Volcanology and Geothermal Research*, 31, 321-332. [https://doi.org/10.1016/0377-0273\(87\)90075-8](https://doi.org/10.1016/0377-0273(87)90075-8)
- Duplyakin, D., Siler, D. L., Johnston, H., Beckers, K., & Martin, M. (2020). *Using Machine Learning to Predict Future Temperature Outputs in Geothermal Systems*. United States. GRC 2020 Virtual Annual Meeting & Expo,
- Egec Geothermal. (2017). EGEC comments on the European Parliamentary questions on emissions from the geothermal field in Monte Amiata. Bruxelles European Parliament. 1-15.
- Elter, P., & Pertusati, P. (1973). Considerazioni sul limite alpi-appennino e sulle sue relazioni con l'arco delle alpi occidentali. In *Mem. Soc. Geol. Ital.* (Vol. 12, pp. 359-375).
- Esri. (2022a). *How IDW works*. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-idw-works.htm>
- Esri. (2022b). *How Kernel Interpolation With Barriers works*. <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geostatistical-analyst/how-kernel-interpolation-with-barriers-works.htm>
- Esri. (2022c). *How Spline with Barriers works*. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-spline-with-barriers-works.htm>
- Farella, E. M., Özdemir, E., & Remondino, F. (2021). 4D Building Reconstruction with Machine Learning and Historical Maps. *Applied Sciences*, 11(4), 1445. <https://doi.org/10.3390/app11041445>
- Faulds, J. E., Brown, S., Coolbaugh, M., DeAngelo, J., Queen, J. H., Treitel, S., Fehler, M., Mlawsky, E., Glen, J. M., & Lindsey, C. (2020). Preliminary Report on Applications of Machine Learning Techniques to the Nevada Geothermal Play Fairway Analysis. 45th Workshop on Geothermal Reservoir Engineering 2020,
- Ferrari, L., Conticelli, S., Burlamacchi, L., & Manetti, P. (1996). Volcanological evolution of the Monte Amiata, Southern Tuscany: new geological and petrochemical data. *Acta Vulcanologica*, 8, 41-56.
- Ferri, F., Zanolla, C., Porfidia, B., Coren, F., Giori, L., & Cesi, C. (2008). La cartografia gravimetrica digitale d'Italia alla scala 1:250'000. *GNGTS*, 437-438.
- Forconi, S. (2011). *Il cinabro sul Monte Amiata. Giacimenti, impianti di trasformazione ed organizzazione del lavoro* HgS, Ed. (pp. 1-233).
- Forster, M. (2021). *An Introduction to Machine Learning* In CoreSafe Core Solutions & Prospectors Supplies. GeoHUG. https://www.youtube.com/watch?v=R9q_iaCqxcw
- Froncini, F., Caliro, S., Cardellini, C., Chiodini, G., & Morgantini, N. (2009). Carbon dioxide degassing and thermal energy release in the Monte Amiata volcanic-geothermal area (Italy). *Applied Geochemistry*, 24(5), 860-875. <https://doi.org/10.1016/j.apgeochem.2009.01.010>

- Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., & Zoph, B. (2021). Simple copy-paste is a strong data augmentation method for instance segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*,
- Gianelli, G., Puxeddu, M., Batini, F., Bertini, G., Dini, I., Pandeli, E., & Nicolich, R. (1988). Geological model of a young volcano-plutonic system: The geothermal region of Monte Amiata (Tuscany, Italy). *Geothermics*, *17*(5), 719-734. [https://doi.org/https://doi.org/10.1016/0375-6505\(88\)90033-8](https://doi.org/https://doi.org/10.1016/0375-6505(88)90033-8)
- Gueguen, E., Doglioni, C., & Fernandez, M. (1998). On the post-25 Ma geodynamic evolution of the western Mediterranean. *Tectonophysics*, *298*(1-3), 259-269. [https://doi.org/10.1016/s0040-1951\(98\)00189-9](https://doi.org/10.1016/s0040-1951(98)00189-9)
- Guo, L., & Riding, R. (1998). Hot-spring travertine facies and sequences, Late Pleistocene, Rapolano Terme, Italy. *Sedimentology*, *45*(1), 163-180.
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, *7*(1). <https://doi.org/10.1186/s40537-020-00369-8>
- Hardardóttir, V., Hannington, M., & Hedenquist, J. (2013). Metal concentrations and metal deposition in deep geothermal wells at the Reykjanes high-temperature area, Iceland. *Procedia Earth and Planetary Science*, *7*, 338-341.
- Henley, R. W., & Brown, K. L. (1985). The geothermal framework of epithermal deposits. In Berger, B. R. & Bethke, P. (Eds.), *Geology and Geochemistry of Epithermal Systems*. (Vol. 2, pp. 1-24) Reviews in Economic Geology. P.M.
- Hutcheson, G. D. (1999). Ordinary Least-Squares Regression. In Hutcheson, G. D. (Ed.), *The Multivariate Social Scientist* (pp. 56-113). SAGE Publications, Ltd. <https://doi.org/10.4135/9780857028075>
- Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544-1554. <https://doi.org/10.1109/tkde.2018.2861006>
- Klemm, D. D., & Neumann, N. (1984). Ore-Controlling Factors in the Hg-Sb Province of Southern Tuscany, Italy. *Syngeneses and Epigenesis in the Formation of Mineral Deposits*, Berlin, Heidelberg.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*,
- Krivoruchko, K. (2012). Empirical bayesian kriging. *ArcUser Fall*, *6*(10).
- Lars, Louppe, G., Blondel, M., Fabian, Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Alexandre, Ltc, Jaques, Layton, R., Vanderplas, J., Joly, A., Holt, B., & Ga"el. (2013). API design for machine learning software: experiences from the scikit-learn project. *arXiv pre-print server*. <https://arxiv.org/abs/1309.0238>
- Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3-10. <https://doi.org/10.1016/j.gsf.2015.07.003>
- Lattanzi, P. (1999). Epithermal precious metal deposits of Italy-an overview. *34*(5-6), 630-638. <https://doi.org/10.1007/s001260050224>
- Laurenzi, M., Braschi, E., Casalini, M., & Conticelli, S. (2015). New ⁴⁰Ar-³⁹Ar dating and revision of the geochronology of the Monte Amiata Volcano, Central Italy. *Italian Journal of Geosciences*, *134*, 255-265. <https://doi.org/10.3301/IJG.2015.11>
- Lavric, J. T. V., & Spangenberg, J. E. (2003). Stable isotope (C, O, S) systematics of the mercury mineralization at Idrija, Slovenia: constraints on fluid source and alteration processes. *Mineralium Deposita*, *38*(7), 886-899. <https://doi.org/10.1007/s00126-003-0377-9>

- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Lenzi, A., & Caprai, A. (2016). *Emissioni naturali di idrogeno solforato nell'area amiatina parte I: prospezioni* Codice Armonico, Castiglioncello.
- Magi, F., Doveri, M., Menichini, M., Minissale, A., & Vaselli, O. (2019). Groundwater response to local climate variability: hydrogeological and isotopic evidences from the Mt. Amiata volcanic aquifer (Tuscany, central Italy). *Rendiconti Lincei. Scienze Fisiche e Naturali*, 30(1), 125-136. <https://doi.org/10.1007/s12210-019-00779-8>
- Marinelli, G., Barberi, F., & Cioni, R. (1993). Sollevamenti neogenici e intrusioni acide della Toscana e del Lazio settentrionale. *Memorie della Società Geologica Italiana*, 49, 279-288.
- Marroni, M., Molli, G., Ottria, G., & Pandolfi, L. (2001). Tectono-sedimentary evolution of the External Liguride units (Northern Apennines, Italy): Insights in the pre-collisional history of a fossil ocean-continent transition zone. *Geodinamica Acta*, 14(5), 307-320. [https://doi.org/10.1016/S0985-3111\(00\)01050-0](https://doi.org/10.1016/S0985-3111(00)01050-0)
- Marroni, M., Moratti, G., Costantini, A., Conticelli, S., Pandolfi, L., Bonini, M., Cornamusini, G., & Laurenzi, M. A. (2015). Geology of the Monte Amiata region, Southern Tuscany, Central Italy. *134*(2), 171-199. <https://doi.org/10.3301/ijg.2015.13>
- Marroni, M., & Pandolfi, L. (2007). The architecture of an incipient oceanic basin: a tentative reconstruction of the Jurassic Liguria-Piemonte basin along the Northern Apennines–Alpine Corsica transect. *International Journal of Earth Sciences*, 96(6), 1059-1078. <https://doi.org/10.1007/s00531-006-0163-x>
- Martini, I. P., & Saggi, M. (1993). Tectono-sedimentary characteristics of Late Miocene-Quaternary extensional basins of the Northern Apennines, Italy. *Earth-Science Reviews*, 34, 197-233.
- Melo, A., & Li, Y. (2021). Geology differentiation by applying unsupervised machine learning to multiple independent geophysical inversions. *Geophysical Journal International*. <https://doi.org/10.1093/gji/ggab316>
- Meyer, H. M., & Mitchell, A. W. (1947). *Mercury*. In *Minerals Yearbook*. U.S. Geological Survey. <https://search.library.wisc.edu/digital/AOGWXY7NLI77KT8U/pages/AB6R74IQ6VXF628Y>
- Minissale, A. (2004). Origin, transport and discharge of CO₂ in central Italy. *Earth-Science Reviews*, 66(1-2), 89-141. <https://doi.org/10.1016/j.earscirev.2003.09.001>
- Minissale, A., Magro, G., Vaselli, O., Verrucchi, C., & Perticone, I. (1997). Geochemistry of water and gas discharges from the Mt. Amiata silicic complex and surrounding areas (central Italy). *Journal of Volcanology and Geothermal Research*, 79(3-4), 223-251. [https://doi.org/10.1016/s0377-0273\(97\)00028-0](https://doi.org/10.1016/s0377-0273(97)00028-0)
- Ministero dello Sviluppo Economico. (1994). *Inventario delle risorse geotermiche nazionali*. <https://unmig.mise.gov.it/index.php/it/dati/risorse-geotermiche/inventario-delle-risorse-geotermiche-nazionali>
- Morteani, G., Ruggieri, G., Möller, P., & Preinfalk, C. (2011). Geothermal mineralized scales in the pipe system of the geothermal Piancastagnaio power plant (Mt. Amiata geothermal area): a key to understand the stibnite, cinnabarite and gold mineralization of Tuscany (central Italy). *Mineralium Deposita*, 46(2), 197-210. <https://doi.org/10.1007/s00126-010-0316-5>
- Mottana, A., Crespi, R., & Liborio, G. (1998). *Minerali e rocce* (pp. 25).

- Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, 2(5), 183-197. [https://doi.org/https://doi.org/10.1016/0925-2312\(91\)90023-5](https://doi.org/https://doi.org/10.1016/0925-2312(91)90023-5)
- Nisi, B., Vaselli, O., Tassi, F., De Elio, J., Ortega, M., Caballero, J., Rappuoli, D., & Mazadiego, L. F. (2014). Origin of the gases released from the Acqua Passante and Ermeta wells (Mt. Amiata, central Italy) and possible environmental implications for their closure. *Annals of Geophysics*, 57(4), 1-16. <https://doi.org/doi:10.4401/ag-6584>
- Orlando, L., Bernabini, M., Bertini, G., Cameli, G. M., & Dini, I. (1994). Interpretazione preliminare del minimo gravimetrico del Monte Amiata. In Camerino, U. d. (Ed.), *Studi Geologici Camerti* (Vol. Speciale, pp. 175-181). <http://193.204.8.201:8080/jspui/handle/1336/515>
- Peccerillo, A. (2005). *Plio-Quaternary Volcanism in Italy* Springer, Ed. (Vol. 365, pp. 1-370).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Pierotti, L., Cortecchi, G., & Gherardi, F. (2016). Hydrothermal gases in a shallow aquifer at Mt. Amiata, Italy: insights from stable isotopes and geochemical modelling. *Isotopes in Environmental and Health Studies*, 52(4-5), 414-426. <https://doi.org/10.1080/10256016.2015.1113958>
- Pirajno, F. (2020). Subaerial hot springs and near-surface hydrothermal mineral systems past and present, and possible extraterrestrial analogues. *Geoscience Frontiers*, 11(5), 1549-1569. <https://doi.org/10.1016/j.gsf.2020.04.001>
- Principe, C., Vezzoli, L., & La Felice, S. (2017). *Il vulcano di Monte Amiata* E.S.A. - Edizioni Scientifiche e Artistiche, Ed. (pp. 1-399).
- Prokhorenkova, L., Gusev, G., Vorobev, A., Anna, & Gulina, A. (2019). CatBoost: unbiased boosting with categorical features. *arXiv pre-print server*. <https://arxiv.org/abs/1706.09516>
- Regione Toscana. (2009). 99MM020 "Acquifero dell'Amiata". In Caratterizzazione geologica, idrogeologica e idrogeochimica dei Corpi Idrici Sotterranei Significativi della Regione Toscana (CISS). Istituto di Geoscienze e Georisorse - CNR.
- Regione Toscana. (2011). 23FI010 "Acquifero delle vulcaniti di Pitigliano". In Caratterizzazione geologica, idrogeologica e idrogeochimica dei Corpi Idrici Sotterranei Significativi della Regione Toscana (CISS). Istituto di Geoscienze e Georisorse - CNR.
- Regione Toscana. (2018). *Banche dati cartografia geologica*. <https://www.regione.toscana.it/-/banche-dati-cartografia-geologica>
- Rimondi, V., Chiarantini, L., Lattanzi, P., Benvenuti, M., Beutel, M., Venturi, S., Colica, A., Costagliola, P., Gabbani, G., Gray, J. E., Pandeli, E., Pattelli, G., Paolieri, M., & Ruggieri, G. (2015). Metallogeny, exploitation and environmental impact of the Mt. Amiata mercury ore district (Southern Tuscany, Italy). *Italian Journal of Geosciences*, 134(2), 323-336. <https://doi.org/10.3301/ijg.2015.02>
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804-818. <https://doi.org/https://doi.org/10.1016/j.oregeorev.2015.01.001>
- Rong, S., & Bao-Wen, Z. (2018). The research of regression model in machine learning field. *MATEC Web of Conferences*, 176, 01033. <https://doi.org/10.1051/matecconf/201817601033>

- Rytuba, J. J. (1993). Active geothermal systems and gold-mercury deposits in the Sonoma–Clear Lake volcanic fields. *California: SEPM (Society for Sedimentary Geology) Guidebook*, 16, 38-51.
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44(1.2), 206-226.
- Scikit-learn. (2021). *Supervised learning Library*. https://scikit-learn.org/stable/supervised_learning.html#supervised-learning
- Segreto, L. (1991). *Monte Amiata: il mercurio italiano: strategie internazionali e vincoli extraeconomici* Angeli, F., Ed. (pp. 1-355).
- Signorini, R. (1949). Visione odierna della geologia toscana. *Boll. Soc. Geol. It*, 65, 82-90.
- Siler, D. L., Zhang, Y., Spycher, N. F., Dobson, P. F., McClain, J. S., Gasperikova, E., Zierenberg, R. A., Schiffman, P., Ferguson, C., Fowler, A., & Cantwell, C. (2017). Play-fairway analysis for geothermal resources and exploration risk in the Modoc Plateau region. *Geothermics*, 69, 15-33. <https://doi.org/10.1016/j.geothermics.2017.04.003>
- Sillitoe, R. H., & Brogi, A. (2021). Geothermal systems in the Northern Apennines, Italy: Modern analogues of Carlin-style gold deposits. *Economic Geology*, 116(7), 1491-1501. <https://doi.org/10.5382/econgeo.4883>
- Simmons, S. F., & Brown, K. L. (2006). Gold in magmatic hydrothermal solutions and the rapid formation of a giant ore deposit. *Science*, 314(5797), 288-291.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222. <https://doi.org/10.1023/b:stco.0000035301.49549.88>
- Squarci, P. (1981). Geologia e geoidrologia dei sistemi geotermici: l'esempio dei campi geotermici italiani in esercizio: Larderello, Travale, Monte Amiata. In Istituto Nazionale per le Ricerche Geotermiche (Ed.). CNR - Pisa.
- Strappa, O. (1977). Storia delle miniere di mercurio del Monte Amiata. *L'Industria Mineraria*, XXVIII(Parte I-II-III), 252-259, 346-348, 433-439.
- Sun, T., Chen, F., Zhong, L., Liu, W., & Wang, Y. (2019). GIS-based mineral prospectivity mapping using machine learning methods: A case study from Tongling ore district, eastern China. *Ore Geology Reviews*, 109, 26-49. <https://doi.org/10.1016/j.oregeorev.2019.04.003>
- Tanelli, G. (1983). Mineralizzazioni metallifere e minerogenesi della Toscana. *Memorie della Società Geologica Italiana*, 25, 91-109.
- Van Bergen, M. J. (1983). Polyphase metamorphic sedimentary xenoliths from Mt. Amiata volcanics (Central Italy); evidence for a partially disrupted contact aureole. *Geologische Rundschau*, 72(2), 637-662. <https://doi.org/10.1007/bf01822086>
- Van Bergen, M. J. (1985). Common trace-element characteristics of crustal-and mantle-derived K-rich magmas at Mt. Amiata (central Italy). *Chemical Geology*, 48(1-4), 125-135.
- Varekamp, J. C., & Buseck, P. R. (1984). The speciation of mercury in hydrothermal systems, with applications to ore deposition. *Geochimica et Cosmochimica Acta*, 48(1), 177-185. [https://doi.org/https://doi.org/10.1016/0016-7037\(84\)90359-4](https://doi.org/https://doi.org/10.1016/0016-7037(84)90359-4)
- Vaselli, O., Nisi, B., Rappuoli, D., Bianchi, F., Cabassi, J., Venturi, S., Tassi, F., & Raco, B. (2015). Geochemical characterization of the ground waters from the former Hg-mining area of Abbadia San Salvatore (Mt. Amiata, central Italy): criticalities and perspectives for the reclamation process. *Italian Journal of Geosciences*, 134(2), 304-322. <https://doi.org/10.3301/ijg.2015.03>
- White, D. E. (1981). Active geothermal systems and hydrothermal ore deposits. *Society of Economic Geologists, Seventy-Fifth Anniversary Volume*.

- Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79-82. <https://doi.org/10.3354/cr030079>
- Yin, R., Deng, C., Lehmann, B., Sun, G., Lepak, R. F., Hurley, J. P., Zhao, C., Xu, G., Tan, Q., Xie, Z., & Hu, R. (2019). Magmatic-Hydrothermal Origin of Mercury in Carlin-style and Epithermal Gold Deposits in China: Evidence from Mercury Stable Isotopes. *ACS Earth and Space Chemistry*, 3(8), 1631-1639. <https://doi.org/10.1021/acsearthspacechem.9b00111>
- Yu, B. (2020). *Epidode 2: Search [MOOC lecture]* In Malan, D., J., CS50's Introduction to Artificial Intelligence with Python. edX. <https://learning.edx.org/course/course-v1:HarvardX+CS50AI+1T2020/home>
- Zucchetti, S. (1964). I giacimenti cinabreriferi detritici e residuali del distretto amiatino. *L'Industria Mineraria*, 15, 495-507.
- Zucchetti, S. (1965). Confronto fra i giacimenti mercuriferi primari dell'Amiata e della Spagna. *Bollettino della Associazione Mineraria Subalpina*, 2, 1-26.