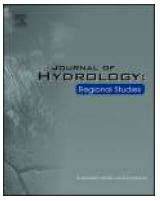
Journal of Hydrology: Regional Studies 58 (2025) 102157

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

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



Deep learning text classification of borehole logs for regional scale modeling of hydrofacies (Po Plain, N Italy)

Alberto Previati [*,1], Valerio Silvestri [2], Giovanni Crosta [3]

DISAT – Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza, 4, Milan 20126, Italy

ARTICLE INFO

Keywords:
Deep learning
Geological text classification
Borehole logs
Hydrofacies
Aquifer heterogeneity
Po Plain

ABSTRACT

Study region: The Po River alluvial plain in northern Italy, spanning 45,700 km^2, stands as one of Europe's most extensive groundwater reservoirs, serving the needs of approximately 15 million residents. It is underlain by a substantial sequence of quaternary fluvio-glacial and alluvial plain deposits originating from the Alpine and Apennine ranges. These deposits form aquifer systems at a regional scale, exhibiting grain size variations that correlate with lithology, proximity to source areas, and depositional age.

Study focus: In this study, we have transformed a considerable volume of qualitative lithotextural data from borehole logs - comprising 39,265 boreholes and 387,297 descriptive intervals - into a semi-quantitative hydrogeological framework. This was achieved through an automated deep learning process that classified geological descriptions into hydrofacies based on the grain size. We employed a long short-term memory (LSTM) recurrent neural network algorithm, which was trained and validated using 86,611 pre-labelled entries encompassing all sediment types within the study region. The word embedding technique enhanced the model accuracy and learning efficiency by quantifying the semantic distances among geological terms. The primary objectives of this research are twofold: (i) to develop a robust deep learning classification model that leverages geological descriptions alongside grain size data, and (ii) to standardize a vast array of sparse and heterogeneous stratigraphic log data for integration into hydrofacies models tailored for hydrogeological applications.

New hydrological insights for the region: The outcome of this work is a novel dataset of semi-quantitative hydrogeological information, boasting a classification model accuracy of 97.4 %. This dataset was incorporated into expansive modeling frameworks, enabling the assignment of hydrogeological parameters based on grain size data, integrating the uncertainty stemming from misclassification. This has markedly increased the spatial density of available information from 0.34 data points/km^2 to 8.7 data points/km^2. The study findings align closely with existing literature maps, offering a robust spatial reconstruction of hydrofacies at different scales. This has significant implications for groundwater research, particularly in the realm of quantitative modeling at a regional scale.

* Corresponding author.
E-mail address: alberto.previati@unimib.it (A. Previati).
[1] 0000–0002-4736–6531
[2] 0009–0008-5423–9670
[3] 0000–0002-3002–3188

https://doi.org/10.1016/j.ejrh.2024.102157
Received 5 August 2024; Received in revised form 9 December 2024; Accepted 26 December 2024
Available online 28 January 2025

1. Introduction

The advent of extensive digital datasets coupled with advancements in artificial intelligence (AI) is revolutionizing our ability to extract meaningful insights from the nonlinear, multidimensional patterns that pervade the natural sciences. Recent advances in machine learning and deep learning algorithms have proven to be exceptionally powerful for lithological classification, leveraging multiple inputs from well logs of hydro-geophysical parameters (Asante-Okyere et al., 2022; Li et al., 2020; Min et al., 2020; Saporetti et al., 2021) or from mapped geophysical properties such as airborne electromagnetic surveys (Harris and Grunsky, 2015; He et al., 2023; Tilahun and Korus, 2023).

The construction of regional-scale hydrogeological and geotechnical models necessitates extensive datasets with adequate spatial resolution. Direct measurements of aquifer characteristics by means of hydrogeological (e.g., pumping tests) or geotechnical tests (e.g., penetration tests) provide quantitative data that can be directly used to generate thematic maps or subsurface models but are often limited by poor spatial density. However, direct measurements of subsurface properties are typically local and proprietary to companies. In contrast, stratigraphic logs from boreholes, typically made public via geological portals under mandates like Italy's Law 646/1984, Germany's Geological Data Act (Geol-DG), and the Swiss Federal Act on Geoinformation, offer qualitative data with a far greater spatial density due to the large number of boreholes typically drilled in densely populated areas. These datasets may differ across countries due to a lack of a common log description standardization, but also within countries due to the period covered by the borehole dataset within which operational and descriptive methods may have changed.

Consequently, Europe is replete with qualitative geological data such as borehole logs, which, while abundant, are not readily assimilated into quantitative hydrogeological and geotechnical models due to the extensive time required to process the written descriptions into operationally significant units like hydrofacies, adhering to the definition established by Anderson (1989). This conversion typically necessitates expert analysis of each report but can be expedited through the application of natural language processing (NLP) techniques rooted in AI (Chowdhary, 2020).

Recently, many examples demonstrated the efficacy of deep learning neural networks in solving NLP tasks such as sentiment analysis, text generation, and text classification (Hirschberg and Manning, 2015; Kowsari et al., 2019; Li et al., 2022). Recurrent neural networks (RNN), for instance, are extensively used to detect patterns in data sequences, including time series and natural language (Sundermeyer et al., 2012), due to their ability to evaluate not only the actual value of an input variable (e.g., a word in a sentence), but also its position in the sequence (e.g., the relationships between words in the sentence).

In addition, NLP techniques like word embedding (e.g., *word2vec* by Mikolov et al., 2013) have demonstrated to be very effective in interpreting geological language (Lawley et al., 2022; Wang et al., 2018) by learning the semantics and meanings of individual terms or sentences within complex textual datasets. These word embeddings, which encode information about a word and its linguistic relationships with other words, have been used (e.g., Lawley et al., 2022; Padarian and Fuentes, 2019), to discriminate between rock types (igneous, metamorphic, and sedimentary) and sediment classes (i.e. gravel, sand, silt, and clay), by evaluating the proximity of embedding vectors in a multidimensional space, often through principal component analysis.

In this context, the targeted classification of textual descriptions, particularly those detailing the granulometry of unconsolidated sediments or the fracturing state of rock masses, combining supervised deep learning and natural language processing (NLP) is a promising method. This approach aims to refine large-scale geological and hydrogeological models by enriching them with increased data volume, as demonstrated by the few and recent works (Fuentes et al., 2020; Lawley et al., 2023).

The objective of this research is to demonstrate the conversion of qualitative geological information from a very large dataset of stratigraphic logs (encompassing 387,297 text descriptions from 39,265 boreholes), into a dataset of semi-quantitative information. This transformation, primed for hydrofacies and hydrogeological modeling, is facilitated by an operational classification system (e.g., grain size based hydrofacies). The system utilizes a deep learning-based NLP algorithm to categorize complex geological and lithostratigraphic text descriptions according to grain size-based hydrofacies. A supervised deep learning text classification algorithm, founded on a Long-Short Term Memory (LSTM) Recurrent Neural Network (RNN) architecture (Hochreiter and Schmidhuber, 1997), was meticulously developed and trained within Matlab®. The algorithm processes textual inputs in the form of geological descriptions from drilling reports, classifying them based on grain size attributes, informed by extensive training on a manually pre-labeled subset of the data. Beyond generating a semi-quantitative hydrogeological dataset from qualitative geological inputs, the study proposes various modeling techniques, including maps, cross-sectional, and three-dimensional spatial representations. In particular, these techniques were designed to spatialize the outcomes inside hydrostratigraphic units known from literature chosen from some sub-domains of the study area. The resulting outcomes, such as maps, cross-sections and 3D model show the heterogeneous distribution of hydrogeological properties, further validated against available quantitative data like grain size analyses, well tests and reference hydrogeological map and cross-sections.

To this aim, the Po plain (northern Italy) serves as the pilot area for this study. The selection of this basin is justified by several factors: 1) the relative homogeneity of the shallow subsurface geology, characterized by glacial, alluvial-fan, and fluvial deposits with grain size variations influenced by the source of the deposits, the distance from the source area and the dynamics of the depositional environment, 2) the high density of geological boreholes due to the high population density of the area and controlled in depth and distribution by urban settlement and geological characteristics, 3) the availability of approximately 22 % of the borehole logs as manually pre-labeled records classified according to grain size characteristics, thanks to prior research (De Caro et al., 2020) within a subset of the study area (3135 km^2). It is important to point out that the intent of this work is not to retrace hydrostratigraphic boundaries extensively delineated by previous studies, but rather to obtain a dataset of quantitative hydrofacies data from boreholes. This dataset will facilitate the derivation of hydrogeologic parameter distributions within the established hydrostratigraphic units.

2. Study area

The Po Plain, encompassing 45,700 km^2 in northern Italy, is a peculiar basin representing both the foreland basin of the Alps (bounding the Plain to the north) and the foredeep basin of the Apennines (bounding the Plain to the south) (Garzanti et al., 2011; Livani et al., 2023). This basin, including the regions of Piemonte, Lombardia, Veneto, and Emilia-Romagna, is characterized by a Neogenic-Quaternary siliciclastic sequence that blankets the external fronts of these mountain chains, composed of syntectonic and recent alluvial sediments from the Po River (Livani et al., 2023; Ori, 1993).

Extensive research over recent decades, driven by hydrocarbon exploration, water supply needs, and stratigraphic studies, has provided a detailed understanding of the basin's subsurface (Amorosi and Pavesi, 2010; Campo et al., 2020; Garzanti et al., 2011; Livani et al., 2023; Ori, 1993; Regione Emilia-Romagna and ENI, 1998; Regione Lombardia and ENI, 2002; Scardia et al., 2006). Notably, clastic bodies originating from Alpine rivers are prevalent, marked by coarse deposits indicative of a robust detrital influx from carbonate, metamorphic, and igneous sources (Ori, 1993). The northern part of the plain features a 'bajada' (Fontana et al., 2014; Guzzetti et al., 1997), a broad coalescence of mountain front fans from the Alps, while the southern portion exhibits smaller Apennine-fed fans interspersed within finer overlying strata, reflecting the source areas lower detrital output and mudrock and flysch abundance (Ori, 1993).

This study analyses data from the shallowest subsurface, primarily hosting Quaternary deposits. These are bounded by unconformities into three main hydrostratigraphic units forming regional- scale aquifer groups (from top to bottom): aquifer group A or 'Upper Po Synthem' (less than 0.45 My), aquifer group B or 'Lower Po Synthem' (less than 0.65 My), and aquifer group C (less than 0.80 My) (Amorosi and Pavesi, 2010; Regione Emilia-Romagna and ENI, 1998; Regione Lombardia and ENI, 2002). The surfaces chosen as hydrostratigraphic boundaries reflects abrupt facies shift from laterally extensive and amalgamated coarse-grained bodies to overlying organic-rich muddy units (Amorosi and Pavesi, 2010). Each hydrostratigraphic unit (or aquifer group) consist of several stratigraphic successions, delineated by transgressive-regressive depositional cycles, consisting of fine-grained deposits (aquitards or aquicludes) and coalesced alluvial fan gravels and sandy fluvial channel belts, which constitute the primary aquifer systems (Amorosi and Pavesi, 2010; Campo et al., 2020). These laterally extensive hydrostratigraphic units (or aquifer groups) embrace different depositional environments (detailed, e.g., in the study by Nichols and Fisher, 2007), from proximal alluvial fans to distal alluvial plains and nearshore, reflecting a very high variability in the grain size and connectivity of the permeable units, which this work aims to clarify.

The geological and hydrogeological settings of the study area, presented in Fig. 1 (redrawn from Servizio Geologico Nazionale, 1998), illustrates the distribution of several key features for a comparative analysis: 1) the spatial distribution of shallow deposit characteristics, such as the grain size, reflects the depositional environment, while the age of deposition helps to distinguish between pre- or post-deglaciation sedimentation phases. Additionally, the size disparity between Alpine and Apennine alluvial fans is

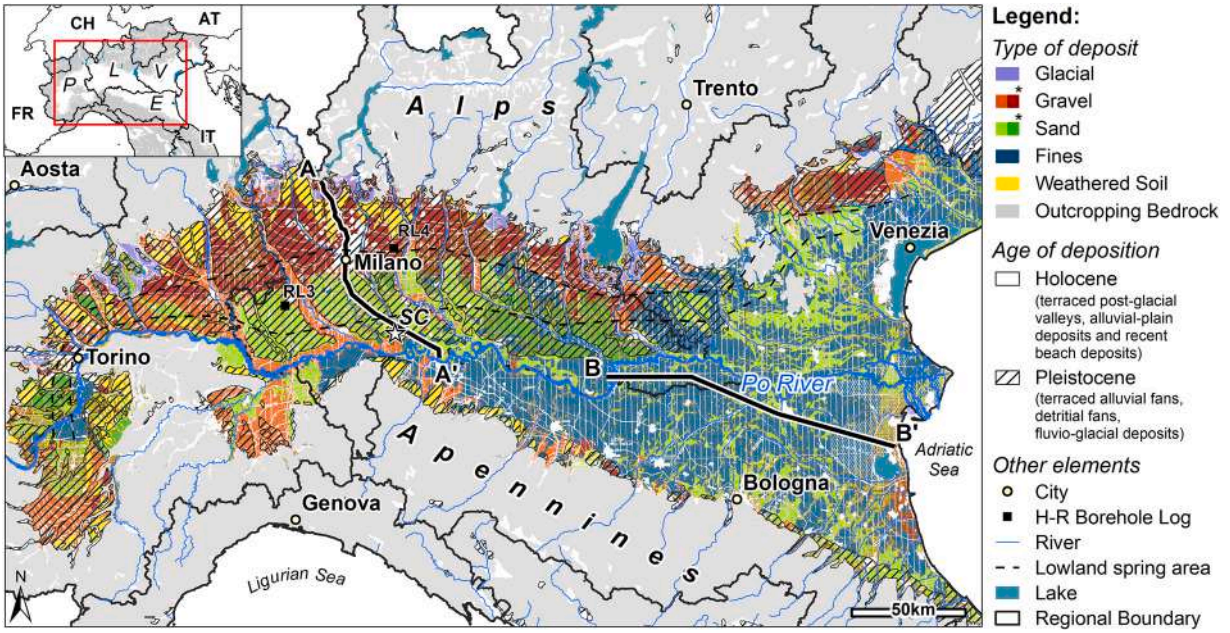


Fig. 1. Hydrogeological map of the study area (redrawn from Servizio Geologico Nazionale, 1998) showing the spatial distribution of the surface deposits. The asterisk symbol in the legend shows the presence of weathered coverage that may influence the permeability. The age of deposition is shown for the Po basin fill only. The trace of the sections A-A' and B-B' are also reported. SC = San Colombano hill. The location of two high-resolution (H-R) borehole logs used later for comparison is shown by black squares. The inset shows the distribution of the four Italian regions covered by this study: P = Piemonte, L = Lombardia, V = Veneto and E = Emilia-Romagna.

noteworthy; 2) in the upper part of the Plain, clastic deposits with high permeability have been locally transformed into conglomeratic units due to $CaCO_3$ dissolution and precipitation processes (e.g., the *'Ceppo'* formation described by Orombelli, 1979); 3) the widespread belt of lowland phreatic springs (known as *'Fontanili'*), underlines the transition in both the slope surface and sediment grain size from the gravel-dominated northern sector of the Plain (*'high plain'*) to the sandier and finer southern sector (*'middle to low plain'*) (De Luca et al., 2014); 4) the *'San Colombano'* (SC) hill, an isolated hill formed by the uplift of a buried anticline-ramp structure during the Pleistocene, is associated to a depositional unconformity (Zuffetti et al., 2018) that will be further discussed in relation to section AA'.

3. Materials and methods

Fig. 2 delineates the workflow adopted for the automated classification of litho-textural descriptive intervals into grain size based hydrofacies. Based on the definition proposed by Anderson (1989), and subsequently widely adopted in subsurface modeling of hydrogeological characteristics using various techniques (Bayer et al., 2011; dell'Arciprete et al., 2012; Marini et al., 2018; Ouellon et al., 2008; Weissmann and Fogg, 1999), this paper uses the term hydrofacies to refer to units characterized by similar, but not necessarily isotropic, hydrogeological properties, primarily determined based on litho-textural grain size descriptions. The training set for the classification model was derived from a designated sub-region within the study area, together with the hydrogeological parameterization of the identified classes. This dataset was derived from the pre-processed data presented in De Caro et al., (2020). Upon the successful training of the classification algorithm, the stratigraphic sequences from the entirety of the available borehole records were subjected to automatic classification. Subsequently, hydrogeological parameters, specifically hydraulic conductivity and porosity, were attributed to each classified unit, drawing upon the empirical relationships discerned within the training set. Thanks to the hydrogeological operative classification of the boreholes, detailed maps and cross-sectional views, that illustrate both the grain size distribution of the sediments and the spatial variation of hydrogeological parameters, were generated. These representations were then compared with 'hard' hydrogeological data derived from direct testing methods, as well as with 'soft' data extrapolated from regional-scale hydrogeological maps and cross-sections. This comparative analysis serves to validate the accuracy of our classification model and its applicability in hydrogeological studies.

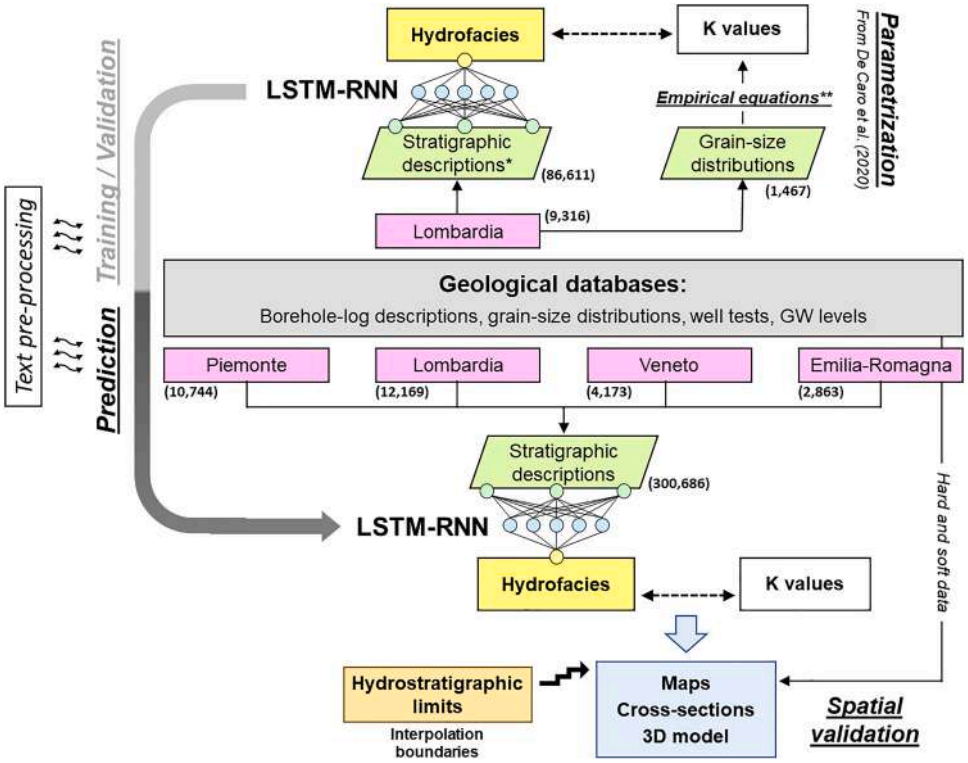


Fig. 2. Workflow for the automatic classification of borehole geological descriptions in grain size based hydrofacies and generation of thematic maps at regional scale. The deep-learning algorithm is based on a long short-term memory network (LSTM-RNN) designed for text classification. The total number of used boreholes (pink boxes), geological text descriptions and grain size data is given in brackets. (*) A quantitative classification of grain size based stratigraphic descriptions is available for training in a subset of the study area (see also Fig. 5). (**) Hydrogeological parameters (K = hydraulic conductivity) of hydrofacies are derived from empirical equations following a previous study (De Caro et al., 2020) in a smaller area included in this study.

3.1. Borehole logs classification with a Deep Learning supervised algorithm

The main objective of this work is to systematically categorize a substantial dataset of geological text descriptions of unconsolidated deposits (mainly) by their grain size attributes into distinct hydrofacies. Borehole logs of unconsolidated deposits, which are primarily based on in situ visual assessments or semi-quantitative field techniques, adhere to the Italian geotechnical classification standard (A.G.I., 1963), yet they often incorporate a diverse array of details such as color, alteration, organic content, and stratification, reflecting the skills of the operators, and the varied focus, and methodologies of the drilling operations (see Section 3.3.1 for the typical structure of the available borehole logs). To address the heterogeneity of the data, advanced deep learning-aided NLP methods were employed to distill grain size information from the textual descriptions, thereby achieving an operational classification into hydrofacies predicated on grain size characteristics (see Section 3.3.2 for the available grain size data and characteristics). The initial phase involved preprocessing the text descriptions by removing punctuation and numbers, and converting to lowercase, and then constructing a dictionary that encapsulated 2782 unique terms extracted from all the records. This dictionary underwent a manual review to correct prominent typographical errors.

Subsequently, based on the common dictionary, word embedding vectors with dimension of 100 were evaluated using the *word2vec* algorithm (Mikolov et al., 2013) to facilitate the learning process. A verification of the semantic learning of this algorithm was done applying a principal component (PCA) analysis with words as observations (rows of the PCA input matrix) and the 100 embedding vectors as variables (columns of the PCA input matrix). Fig. 3 shows the most frequent geological terms in the dataset (translated in English) arrayed along the PC1 axis according to a geological meaning, revealing a spectrum from coarse sediments on the left (lower PC1 values) to finer sediments on the right (higher PC1 values).

Then, a Long Short-Term Memory (LSTM) text classification algorithm was implemented within Matlab® utilizing supervised learning. The LSTM architecture (Hochreiter and Schmidhuber, 1997), a type of RNN, features fully interconnected neuron layers (hidden layers) that collectively process input features to discern patterns. This LSTM network is further equipped with memory units and gates capable of learning long-term dependencies between time steps of sequence data by holding or discarding information on similar patterns that are relevant in the training stage. Additionally, a word embedding layer, included in the network, captures the semantic similarity between terms by mapping the sequences of word indices to the embedding vectors. During training, the word embedding vectors are randomly initialized and corrected when similarities (in the output) are recognized. Finally, a *softmax* layer normalizes the output of the fully connected layers to be used as classification probabilities by the final classification layer.

The architecture of the supervised classification deep learning algorithm proposed in this study is outlined in Fig. 4 and involve the following steps:

1. Text preprocessing: geological descriptions are standardized by removing punctuation, numbers, and converting to lowercase. These preprocessed text strings are transformed in word tokens to form the basis of the dictionary;
2. Document conversion: each document is transformed into a numeric sequence through word encoding, where every word is associated with a unique numeric index from the dictionary. These sequences form the input layer for the neural network;
3. LSTM-RNN network creation: a LSTM-RNN is constructed for text classification. It includes a word embedding layer to capture semantic relationships, a series of fully connected layers for pattern recognition, a *softmax* layer for probability distribution, and a memory cell to retain important sequence information;
4. Network training: the deep learning network is trained using the Adam optimization algorithm (Kingma and Ba, 2015) which efficiently updates network weights and biases based on training data;
5. Text data classification: new text data is classified using the same preprocessing steps of the training stage, ensuring consistency and accuracy in the classification results.

The classification algorithm adopted in this study was trained on a dataset of 69,289 geological descriptions (i.e. text strings) covering all the sediment types. These descriptions exhibit a heterogeneous level of detail, from succinct summaries to elaborate narratives of up to about 100 words. Each record correlates with a grain size class as per the simplified classification system from the Lombardia regional geological database, as delineated by De Caro et al. (2020), encompassing a total of 26 classes (23 for unconsolidated deposits in Table 1, and 3 for cemented deposits or rocks). For the purpose of model validation, an additional 17,322 geological descriptions, along with their corresponding classifications from the same database, were utilized.

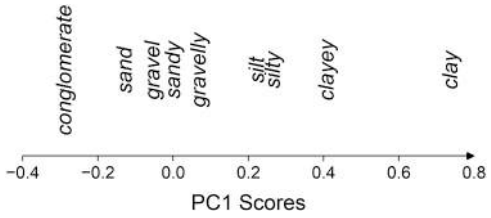


Fig. 3. Visualization of the word embedding vectors for the most frequent litho-textural terms by reducing the dimension of the embedding vectors with PCA. The terms are graded with a geological meaning, having the coarse sediments on the left-hand side of the plot (low values of PC1) gradually becoming finer towards the right-hand side (high values of PC1).

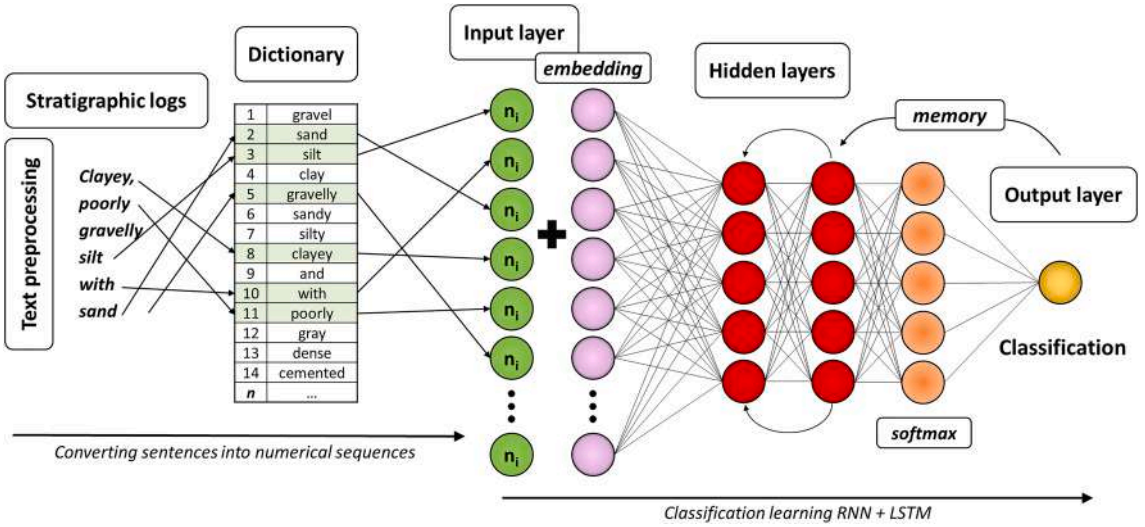


Fig. 4. Schematic representation of the classification learning process from stratigraphic log descriptions (text strings) based on a Recurrent Neural Network (RNN) architecture with a Long Short-Term Memory (LSTM) unit.

After the classification of the geological descriptions into hydrofacies classes, each of them was assigned representative grain size class abundances and hydrogeological parameters as shown in Table 1. The average grain size abundances were derived following the Italian geotechnical classification standard for loose deposits (A.G.I., 1963), serving as a criterion to retrieve the percentage of each grain size class from the hydrofacies classes. Furthermore, values of hydrogeological parameters were assigned to each hydrofacies from the analysis of complete grain size distributions. The methods used to estimate the hydraulic conductivities are listed in Table 1 following the work of De Caro et al. (2020) in a smaller area included in the study area (see also Section 3.3.2). This approach ensures the robustness of the classification model enhancing the granularity and precision of the hydrogeological parameters assigned to each class.

Table 1
Hydrofacies used in this study from the work of De Caro et al. (2020) with the associated mean values of porosity and hydraulic conductivity (K), and average abundances of grain size classes. G = gravel, S = sand, L = silt and C = clay. Method reference for the attribution of the hydraulic conductivity: (1) Alyamani and Şen (1993); (2) Chapuis et al. (2005); (3) Beyer (1964); (4) Harleman et al. (1963); (5) Hazen (1982); (6) Kozeny (1953); (7) Carman (1937); (8) NAVFAC (1974) from Chesnaux et al. (2011); (9) Sauerbrei method from Vuković and Soro (1992), (10) Slichter (1899).

Hydrofacies	Porosity	K (m/s)	Methods	% Gravel	% Sand	% Silt	% Clay
G	0.26	9.0E−02	(1) (4) (5)	100			
GS	0.29	2.4E−03	(1) (4) (5) (10)	70	30		
GL	0.34	6.1E−04	(2) (3) (4)	70		30	
GC	0.38	6.7E−04	(2) (3) (4)	70			30
GSL	0.30	1.5E−05	(1) (4) (5) (10)	60	25	15	
GSC	0.30	1.2E−05	(1) (4) (5) (10)	60	25		15
S	0.34	5.8E−05	(2) (4) (6) (9) (10)		100		
SG	0.31	5.5E−04	(2) (4) (5) (6) (8) (10)	30	70		
SGL	0.30	9.8E−06	(2) (3) (4) (5) (6) (8) (10)	25	60	15	
SGC	0.30	2.8E−04	(2) (3) (4) (5) (6) (8) (10)	25	60		15
SL	0.34	2.2E−05	(2) (3) (4) (5) (6) (7) (8) (9)		70	30	
SLC	0.27	4.3E−06	(2) (3) (4) (5) (6) (7) (8) (9)		60	25	15
SLG	0.33	5.9E−06	(2) (3) (4) (5) (6) (7) (8) (9)	15	60	25	
SC	0.32	1.1E−05	(4) (7) (8) (9)		70		30
SCG	0.33	1.6E−05	(4) (7) (8) (9)	15	60		25
L	0.44	9.5E−06	(4) (7)			100	
LG	0.35	6.9E−06	(4)	30		70	
LS	0.30	5.6E−06	(4)		30	70	
LC	0.40	8.2E−06	(4)			70	30
C	0.50	3.2E−08	(4) (7)				100
CG	0.46	3.4E−06	(4)	30			70
CS	0.43	5.5E−06	(4)		30		70
CL	0.45	1.2E−07	(4)			30	70

3.2. Spatial reconstruction of hydrofacies

In order to outline the potential of a database of geological drilling log descriptions with associated semi-quantitative hydrogeological information, a spatial reconstruction of the hydrofacies has been proposed using the results from the automatic classification algorithm. Following the concepts of the truncated gaussian/multi-gaussian methods (Beucher and Renard, 2016; Mariethoz et al., 2011), one or more gaussian random function simulations (i.e., kriging interpolation) were used to map the spatial distribution of the hydrofacies by thresholding the total domain of variation of the grain size abundances and the derived hydraulic conductivity.

The spatial variability of hydrofacies was reconstructed using two distinct methodologies based on:

1. Grain size abundances: this approach involves ordinary kriging interpolation of the relative abundances of the four primary grain size classes: G (gravel), S (sand), Fi (fines, comprising silt and clay), and Ce (cemented units). The interpolated data result in four thematic layers, from which an unsupervised classification algorithm (k-means clustering) was used to identify homogeneous domains by maximizing the distances within the four-dimensional space between all grid cells. The optimal number of clusters was determined by iterative validation of the predicted aquifer spatial distribution with existing maps and cross-sections from literature. Similarly, a single variable approach was employed by applying the Receiver Operating Characteristic (ROC) method to classify the distribution of coarse units (G and S) into aquifer facies from comparison with the available literature knowledge.
2. Hydraulic conductivity: this method employs ordinary kriging interpolation of the hydraulic conductivity (K) associated with each hydrofacies. The interpolated hydraulic conductivity is then classified using the Gaussian truncation method to delineate the spatial distribution of hydrofacies.

Through these methodologies, hydrogeological maps, cross-sections, and a 3D model were constructed. These products were subsequently validated against available 'soft' hydrogeological data, such as maps and cross-sections from literature, and 'hard' data, including grain size analyses and well tests, ensuring a comprehensive and accurate representation of the hydrogeological characteristics.

3.3. Available data

The comprehensive datasets available for the study area are a testament to the extensive geological explorations conducted for oil and gas research, water supply, and engineering projects. These datasets encompass a wealth of both direct and indirect geological, stratigraphic, hydrogeological, and geotechnical information, including grain size distributions, Lefranc tests, and penetration tests.

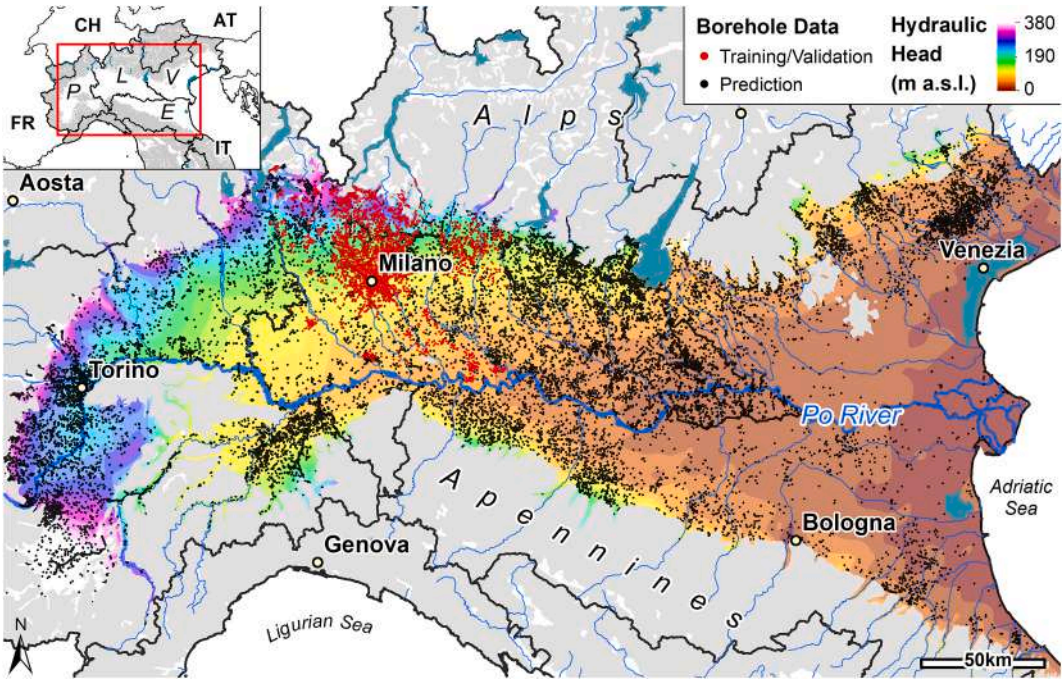


Fig. 5. Map of the study area (Po plain) showing the spatial distribution of the available stratigraphic logs used for training/validation (red) and prediction (black), and the hydraulic head of the shallow aquifer as of the period 2002–2017. The inset shows the distribution of the four Italian regions covered by this study: P = Piemonte, L = Lombardia, V = Veneto and E = Emilia-Romagna. Please note the different spatial density of data between regions P/L, where regional geological databases (in digital format) are available, and regions V/E, where data come only from the national geological database.

Fig. 5 shows the study area and pinpoints the locations of the available borehole logs. The log positions are color-coded: the training (80 %) and validation (20 %) sets (9,316 boreholes containing a total of 86,611 descriptive intervals) are marked in red; the prediction set (29,949 boreholes containing a total of 300,686 descriptive intervals) is marked in black. In the background of Fig. 5 (see also Figure S1 in the supplementary material), the regional scale piezometric level, obtained combining linear (i.e., groundwater head isolines) and pointwise (i.e., groundwater head from well measurements) data from four different sources spanning about 15 years. The specifics regarding the data sources and the years of measurement will be provided subsequently.

The subsurface information used in this study (i.e., stratigraphic logs, grain size analysis, well tests and groundwater data) primarily originates from accessible regional and national geological databases. Additional data are sourced from technical reports derived from private and public projects available to the authors. Table 2 lists the sources of the datasets and specifies the amount of information that will be used in the forthcoming sections.

3.3.1. Stratigraphic logs

Subsurface lithological information (i.e. borehole logs) are available from regional and national geological databases as either scanned sheets or digitalized logs (tables). The latter contain important details such as drilling location, borehole top and bottom elevations, and geological descriptions (digital text) of the crossed stratigraphic units. Only data in digital format were included in this study by retrieving and homogenizing the stratigraphic tables from the available regional (Piemonte and Lombardia) and national public-access online portals. For the Veneto and Emilia-Romagna regions, where only scanned stratigraphic reports are available, the national database provided the necessary digital records (see location of all the stratigraphic logs in Fig. 5). Fig. 6 breaks down the borehole data, presenting the depth intervals reached by the boreholes across the different regions of the study area.

The stratigraphic logs, derived from a variety of sources, such as infrastructure design and water well exploration and development, vary in detail depending on the purpose of the drilling. Usually, the stratigraphic core logging and classification of unconsolidated deposits is performed in situ by visual analysis or semi-quantitative field methods, according to the Italian geotechnical classification guidelines (A.G.I., 1963). Soils are categorized according to the abundances of gravel (G), sand (S), silt (L), and clay (C) separated based on the minimum class diameter (i.e., 2 mm for gravel, 0.06 mm for sand and 0.002 mm for silt).

The classification system describes the soil mixture composition using specific relational terms and abundance ranges for the secondary and following terms. For instance, the term *'with'* indicates a 50–25 % presence (e.g., *'gravel with sand'*), the derived adjective formed with the suffix *'-y'* denotes a 25–10 % presence (e.g., *'sandy gravel'*), and the construct *'poorly -y'* signifies a 10–5 % presence (e.g., *'Poorly sandy gravel'*). Descriptions for very heterogeneous soils may include up to all the four primary terms in descending order and the conjunction *'and'* is used if two classes are equally abundant.

The Lombardia regional database goes a step further, providing a manual classification by a qualified operator that includes up to 40 lithotypes and 12 abundance descriptors, resulting in over 6,000 potential unique combinations. To streamline this complexity, the dataset was condensed following the A.G.I. (1963) guidelines, reclassifying the lithotypes into four primary grain size classes (gravel, sand, silt and clay) and three rock/cemented deposit classes (generic rock, sandstone and conglomerate), while the abundances were simplified into 4 unique values (>50, 50–25, 25–10 and <10 %). After ignoring classes with less than 10 % of abundance, the hierarchical classification by the Lombardia region was aggregated into single terms of up to three single words (G, S, L or C) obtaining 26 classes in total (23 describing loose deposits and 3 for cemented deposits/rocks), echoing the methodology proposed by De Caro et al. (2020) within a smaller subset of the study area (see also Table 1).

Table 3 provides illustrative examples of the borehole log descriptions, the hierarchical classification from the Lombardia dataset, and the simplified classification employed in this study, ensuring a coherent and manageable framework for analysis.

3.3.2. Hard data

In this study, a large amount of quantitative data, including grain size analysis and well tests, was harnessed to parameterize hydrofacies and validate the model outcomes generated after classification and spatialization processes. For the Piemonte and

Table 2
Summary of the data used in this study. The source of data is given in the bottom line of the table for the regional (from 1 to 4) and national (5) geological datasets.

Region	Area (km^2)	Stratigraphic Logs		Direct Tests			Groundwater head	
		Regional	*National*[e]	*Grain sz.*	*Lefranc*	*Well tests*[e]	*Type*	*Year*
Piemonte[a]	9733	8675	2069	3093*	554	1902	Cont.	2002
Lombardia[b]	13,229	18,020	3465	1467**	343**	3100	P (840)	2014
Veneto[c]	11,205	N.A.	4173	N.A.	N.A.	2999	P (279)	2014
Emilia-Romagna[d]	11,523	N.A.	2863	N.A.	N.A.	2400	P (458)	2017

* Grain size data from the Piemonte regional dataset is given as abundance of the four main grain size classes.
** Grain size and Lefranc data for the Lombardia region are retrieved from geological reports from private projects available to the authors. Groundwater head data are available as point measurements (P) or contour lines (Cont.).
[a] https://geoportale.arpa.piemonte.it/
[b] https://www.geoportale.regione.lombardia.it/
[c] https://www.arpa.veneto.it/
[d] https://dati.arpae.it/tl/dataset/acque-sotterranee-2017
[e] https://www.isprambiente.gov.it/it/banche-dati

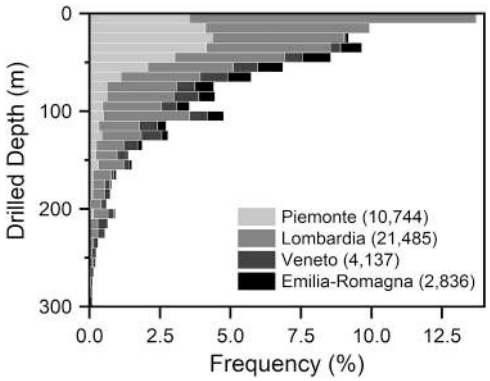


Fig. 6. Number of boreholes reaching a specific depth expressed as frequency (i.e., number of logs in each depth class to the total number of logs).

Table 3
Examples of qualitative soil descriptions according to the Italian standard *(*A.G.I., 1963*)* and the relative intervals of percentage for each grain size class. For the same demonstrative units, it is shown the classification by the regional geological database of Lombardia and the corresponding hydrofacies class as proposed by this study. G = gravel, S = sand, L = silt, and C = clay.

	% of grain size class				*Classification in Lombardia*					
Description	**G** (%)	**S** (%)	**L** (%)	**C** (%)	L1	L2	P2	L3	P3	**Class**
Gravel with sand	> 50	50–25	-	-	**G**	**S**	50–25			**GS**
Gravelly silt with sand	25–10	50–25	> 50	-	**L**	**S**	50–25	**G**	25–10	**LSG**
Sand poorly silty	-	> 50	< 10	-	**S**	**L**	< 10			**S**
Clayey, poorly gravelly silt with sand	10–5	50–25	> 50	25–10	**L**	**S**	50–25	**C**	25–10	**LSC**

Lombardia regions a substantial dataset was available (see Table 2 for the data source) including 1467 full grain size distributions from private projects and available to the authors; 3093 simplified grain size distributions (showing the abundance of the four main grain size classes) from public datasets, and 897 Lefranc well tests from public datasets and private sources. Moreover, 10,401 discharge/drawdown records from well performance tests were retrieved from the national geological database (see also Table 2).

Complete grain size distributions were performed following ASTM standards D6913 and D7928 for sieve analysis and sedimentation analysis, respectively. Each resulting gradation curve was associated to a classified hydrofacies according to the pertaining borehole and the sample depth. Then, following the methodology applied by De Caro et al. (2020) in a smaller area enclosed in this study, characteristic diameters from the grain size curves (e.g., D_{10}, D_{50} and D_{60}) were used to estimate the hydraulic conductivity (K) and porosity by means of empirical equations from the literature (see list in Table 1). Fig. 7 shows the grain size curves used for the

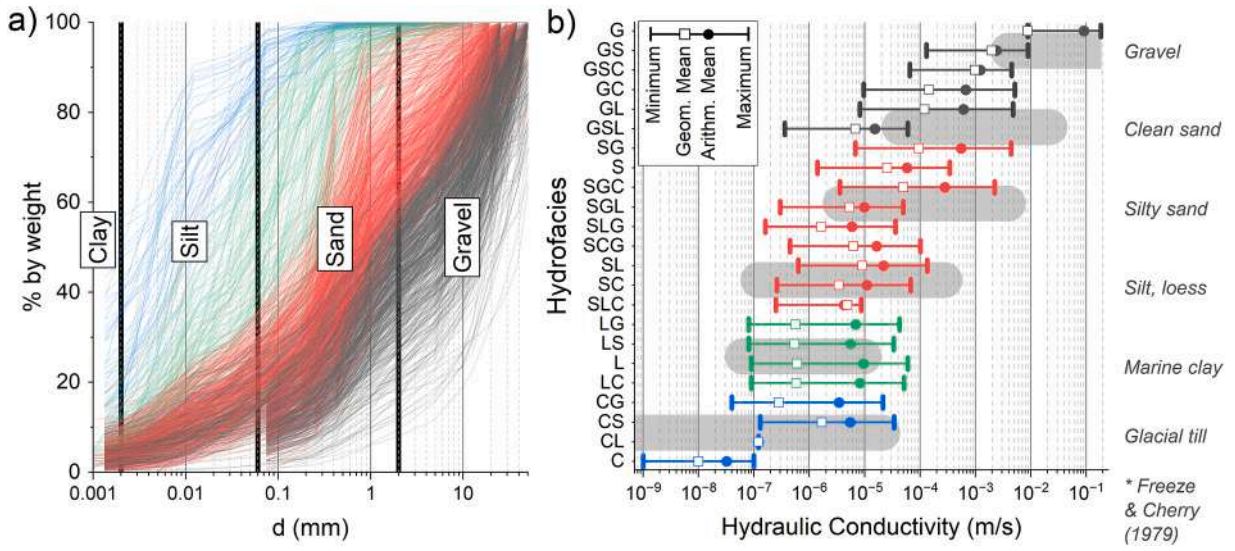


Fig. 7. (a) Grain size distributions available for this study (1467 in total) and (b) estimated values of hydraulic conductivity from empirical correlations for each hydrofacies class in comparison with values suggested by *(*Freeze and Cherry, 1979*)*. G = gravel, S = sand, L = silt, and C = clay.

parametrization (a) and the variability of the estimated K values (b) across various hydrofacies, considering only the mean values for this analysis.

The extensive set of discharge-drawdown tests, including 1077 tests with multiple discharge steps out of the total 10,401, was used to derive the aquifer transmissivity from the ratio between drawdown s (in m) and discharge Q (in m^3/s) applying the Logan's simplification (Eq. 1) to Thiem's equation (Logan, 1964).

$$T = 1.22 * Q/s \tag{1}$$

For tests involving multiple discharge steps, the maximum discharge value was taken into account to ensure the most robust transmissivity estimation.

3.3.3. Soft data

The integration of soft data into this study is pivotal for contextualizing the hydrogeological characteristics of the Po plain. A regional hydrogeological map by the Italian Geological Survey (Servizio Geologico Nazionale, 1998, Fig. 1) containing geomorphological features, and information on the permeability of the surface deposits of the Po plain, was retrieved and compared with the results obtained by this work. No other relevant regional scale hydrogeological maps are known to the authors.

For comparative analysis, hydrogeological cross-sections were obtained from two sources: the regional groundwater characterization and protection plan (PTUA) by Regione Lombardia (2016) and the work of Amorosi and Pavesi (2010). The first NS cross-section, labeled AA' in Fig. 1, crosses the upper part of the plain. It clearly illustrates the transition from proximal Alpine depositional facies, characterized by coarse-grained alluvial fans, to distal facies, mainly sandy alluvial plains and post-glacial deposits. This section also reveals the distribution of both permeable units (shallow and deep aquifers) and low permeability layers, as interpreted from the manual interpretation of available stratigraphic logs along the section. The second EW cross-section, BB' in Fig. 1, is located in the center of the Plain, aligned along the Po channel belt and reaching the Adriatic Sea coastline. This section shows alternances of clayey-silty (aquitards) and coarser sandy (aquifers) deposits mirroring the depositional environment of the distal alluvial plain.

The inclusion of these soft data elements significantly enhances the study by providing insights into the hydrogeological characteristics of the area and validating the accuracy of the results generated through automated categorization and spatial analysis processes.

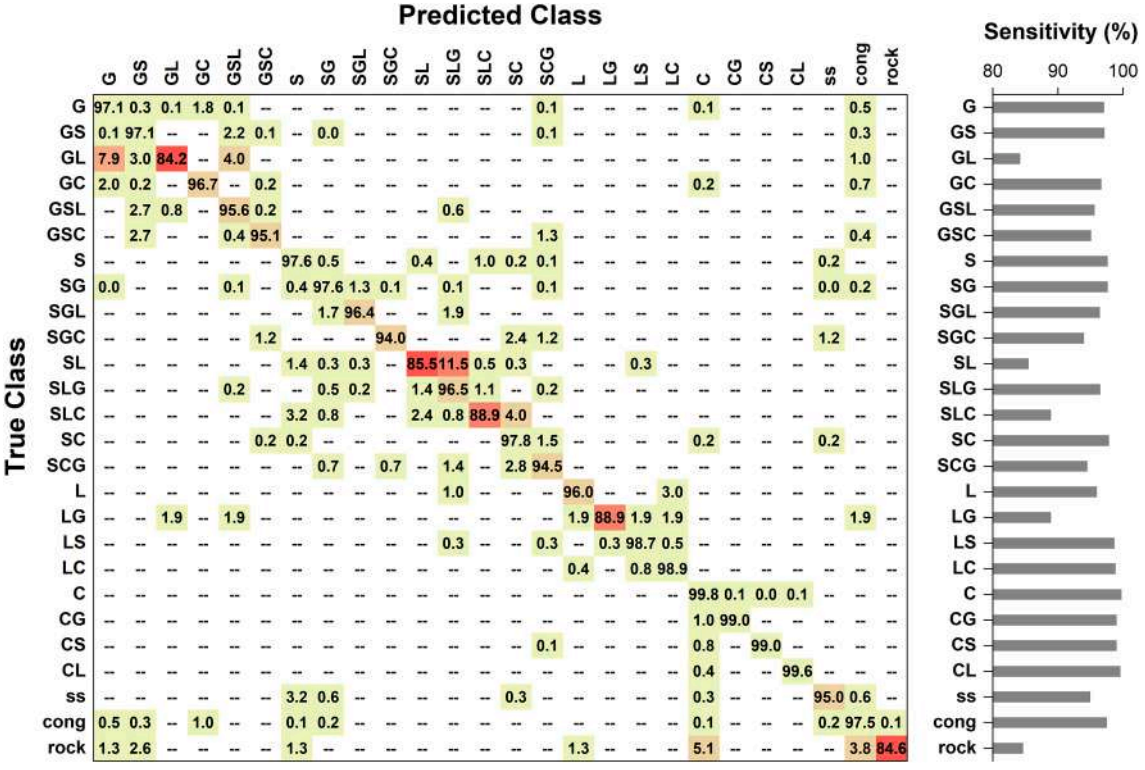


Fig. 8. Confusion matrix evaluated for the proposed classification model on the validation set (17,322 records in total). Values are expressed as a percentage to the total number of observations that has the same true class. G = gravel, S = sand, L = silt, C = clay, ss = sandstone, cong = conglomerate, and rock = generic rock type. The sensitivity of the model is reported in the plot on the right-hand side for each class.

4. Results and discussion

4.1. Classification model

The LSTM classification network underwent training on a single GPU with 6 GB of memory, requiring approximately 10 minutes of computational time. The network accuracy on the validation dataset was determined to be 97.4 % and was computed by dividing the number of accurate predictions by the total predictions made. Given the disparity in sample sizes across different classes, the confusion matrix for the validation dataset is presented in Fig. 8. This matrix has been normalized along the rows, meaning the values are represented as percentages of the total observations within the same true class.

The relevant metrics to evaluate the performance of the classification model are summarized in Table 4 considering the validation set only. Shown for each class are the total number of validation samples, true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), along with calculated accuracy ((TP+TN)/(TP+TN+FP+FN)), precision (TP/(TP+FP)), sensitivity (TP/(TP+FN)), and specificity (1-FP/(FP+TN)). 21 out of 26 classes show a classification sensitivity exceeding 95 %. The lesser sensitivity scores are observed in the 'rock' class and some heterometric classes that contain silt (L). An examination of the confusion matrix reveals that misclassifications within classes of lower sensitivity are generally associated with the most similar classes. For instance, 64 % of the false positives for the SLC class are misclassified as SL, SLG, and SC classes, while the remaining 35 % are classified as S or SG. In the hydrogeological contexts, such misclassifications are of minor significance since most of the incorrectly classified samples often share similar hydraulic parameter values. A probabilistic study on the effects of misclassification on the hydraulic parameterization is presented later in Section 4.3.

Subsequently, the trained LSTM classification network was deployed to classify the hydrofacies of an additional 300,686 unclassified text descriptions of borehole logs available in the Po plain. Fig. 9 illustrates the relative abundance of each class in the prediction set compared to the first occurring grain size-related term in the text descriptions. A notable correspondence exists between the first occurring term and the macro groups of sediment size outlined by arrows in Fig. 9. The proposed classification goes beyond the mere evaluation of first term, finding subclasses (26 hydrofacies) within each main grain size group, each exhibiting significantly different

Table 4
Classification metrics of validation set for the proposed model. For each class it is reported the total number of samples in the validation set, the number of true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), and the accuracy, precision, sensitivity and specificity metrics in percentage. G = gravel, S = sand, L = silt, C = clay, ss = sandstone, cong = conglomerate, and rock = generic rock type.

Class	Samples	TP	FP	FN	TN	Accuracy (TP+TN) / (TP+TN+FP+FN)	Precision TP / (TP+FP)	Sensitivity TP / (TP+FN)	Specificity TN / (TN+FP)
G	1164	1130	31	34	16127	0.996	97.3	97.1	99.8
GS	2342	2275	33	67	14947	0.994	98.6	97.1	99.8
GL	101	85	6	16	17215	0.999	93.4	84.2	100
GC	542	524	37	18	16743	0.997	93.4	96.7	99.8
GSL	479	458	61	21	16782	0.995	88.2	95.6	99.6
GSC	224	213	6	11	17092	0.999	97.3	95.1	100
S	1305	1274	31	31	15986	0.996	97.6	97.6	99.8
SG	2188	2136	24	52	15110	0.996	98.9	97.6	99.8
SGL	360	347	30	13	16932	0.998	92	96.4	99.8
SGC	83	78	4	5	17235	0.999	95.1	94	100
SL	365	312	16	53	16941	0.996	95.1	85.5	99.9
SLG	569	549	60	20	16693	0.995	90.1	96.5	99.6
SLC	126	112	21	14	17175	0.998	84.2	88.9	99.9
SC	548	536	16	12	16758	0.998	97.1	97.8	99.9
SCG	145	137	21	8	17156	0.998	86.7	94.5	99.9
L	99	95	3	4	17220	1.000	96.9	96	100
LG	54	48	1	6	17267	1.000	98	88.9	100
LS	371	366	4	5	16947	0.999	98.9	98.7	100
LC	261	258	6	3	17055	0.999	97.7	98.9	100
C	2527	2521	20	6	14775	0.998	99.2	99.8	99.9
CG	402	398	3	4	16917	1.000	99.3	99	100
CS	708	701	1	7	16613	1.000	99.9	99	100
CL	238	237	2	1	17082	1.000	99.2	99.6	100
ss	317	301	9	16	16996	0.999	97.1	95	99.9
cong	1526	1488	30	38	15766	0.996	98	97.5	99.8
rock	78	66	1	12	17243	0.999	98.5	84.6	100

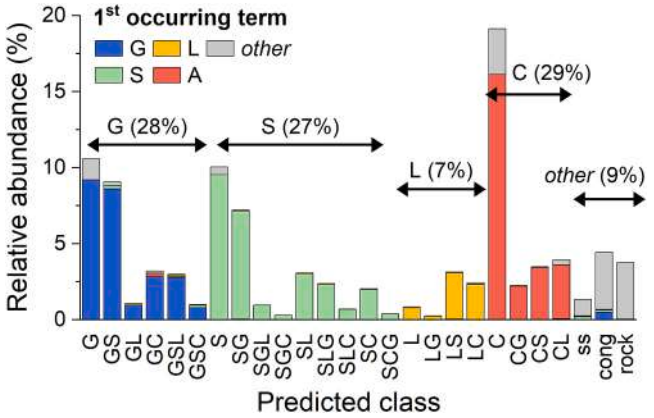


Fig. 9. Bar plot showing the relative abundance of each class in the prediction set (total number of classified samples is 300,686) compared to the first occurring grain size related term in the text descriptions. G = gravel, S = sand, L = silt, C = clay, ss = sandstone, cong = conglomerate, and rock = generic rock typ.

hydrogeological behavior. For instance, among the strings classified within the gravel macro group, which constitutes 28 % of the total, a 37 % represents pure gravel deposits (G), 33 % gravel and sand deposits (GS), and 30 % other hydrofacies that include a mix of gravel and finer materials. These latter categories may possess significantly reduced hydraulic conductivity values, potentially up to two orders of magnitude lower in comparison to G and GS classes, as documented by Freeze and Cherry (1979).

This analysis excludes significant qualitative information beyond the mere sediment type that could enhance the classification and characterization of hydrofacies. For instance, depositional gradation, described using terms such as *fine-*, *medium-*, and *coarse*-grained, is not explicitly incorporated. Although this information is present in the textual descriptions of the training set, it is not quantified in the classification due to the absence of consistent labeling. While a more detailed or multi-level classification could offer improvements, it would also pose challenges in harmonizing descriptions that use gradation terminology with those that omit such details.

4.1.1. Comparison with hard data

To validate the automatic classification of the LSTM network, a comparative analysis was conducted between 2776 laboratory grain size analyses and the corresponding textual descriptions from borehole samples. This dataset consists of 1467 complete grain size distributions and 1309 simplified analyses, which highlighted the prevalence of primary grain size classes. The statistical variability of these measurements, categorized by hydrofacies within sediment size macro groups, is visually represented in Fig. 10a.

The box plots reveal a strong correlation between the grain size classifications inferred from geological descriptions and the actual grain size compositions determined through laboratory analysis. It should be noted that all three classes denoting a very high to absolute clay content (i.e. C: *'clay'*, CL: *'clay and silt'*, CS: *'clay and sand'*) exhibited a higher content of silt over clay in the laboratory analysis. This discrepancy was interpreted as a human bias in visually analyzing very fine sediments, where such materials are frequently labeled as clay rather than silt due to the challenge in making a quantitative distinction. Nonetheless, this interpretation aligns with other established classification systems, such as the ASTM-D2488 standard, where both silt and clay are collectively referred to as fines.

An in-depth comparison between two exemplary high-resolution borehole logs from the study by Regione Lombardia and ENI (2002) and the results of this study is presented in Fig. 10b by evaluating the grain size class abundances every 2 m of depth from all boreholes encompassed by a 2 km radius. The borehole log on the left (RL4 in Fig. 1) shows shallow, highly permeable aquifers (A1, A2 and B) dissected by some minor low permeability horizons reflecting the depositional environment of the high alluvial plain. Downwards, there is an increase in fine sediments reflecting the transition to a deltaic depositional environment. On the right, the borehole log (RL3 in Fig. 1) shows a major sand content locally surpassed by gravel, reflecting the lower energy alluvial plain depositional environment. The alternance of fine and coarse sediments shown in the reference stratigraphic logs is well reproduced by the grain size class abundances predicted by this model, validating the robustness of the approach in predicting the subsurface grain size distribution.

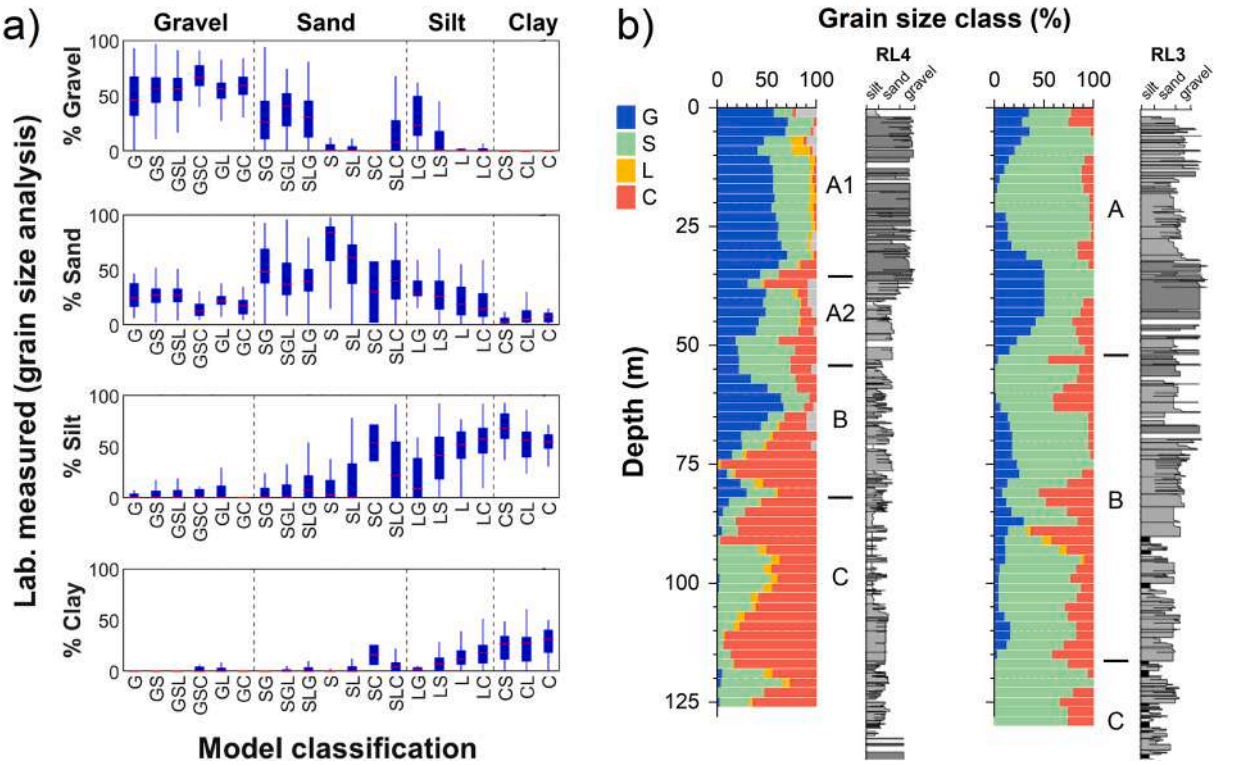


Fig. 10. (a) Boxplots showing the abundances of the four main grain size classes measured by grain size analysis in each class estimated by the classification model. In total 2776 results of grain size analysis with a corresponding classified description are shown. (b) Comparison between two high-resolution borehole logs (grey stratigraphic columns) available from Regione Lombardia and ENI (2002) and the grain size class abundance depth profiles evaluated in the Lombardia high and low plain context (see Fig. 1 for their locations). Grain size class percentages per depth are evaluated from the borehole logs classified in this study, including all boreholes within a 2 km radius from the location of the high-resolution boreholes. The dividing horizons between the three main aquifer groups as described in Section 2 are also reported from Regione Lombardia (2016). G = gravel, S = sand, L = silt, and C = clay.

4.2. *Spatial reconstruction of hydrofacies*

The high spatial density of wells with classified geologic descriptions allowed us to produce very detailed representations of hydrogeological parameters. These are instrumental for conducting regional scale quantitative analyses related to groundwater and subsurface management strategies. For example, the spatial density of data useful for quantitative hydrofacies modeling increased from an average of 0.34 data/km^2, when considering only grain size, Lefranc and discharge-drawdown tests, to 8.7 data/km^2, when including information from the available borehole log descriptions processed by the proposed method. Moreover, the vertical resolution was improved by a quantitative hydrogeological parameterization of all the layers crossed by the borehole logs. This is in contrast with direct measurements, such as well tests, which only provide an equivalent value representative of the total thickness of the screened portion or of the investigated aquifer.

Recent advancements in quantitative stratigraphic modeling techniques, which account for the stratigraphic hierarchy between the units, have been presented with applications to geological and hydrogeological modeling (Giacomelli et al., 2023; Schorpp et al., 2022; Velasco et al., 2012; Zuffetti et al., 2020). These approaches aim to reconstruct the subsurface geometry of stratigraphic bodies but require associating each borehole horizon with a stratigraphic unit based on their characteristics and spatial relationships. To this aim, the borehole data must undergo manual inspection to identify, by expert knowledge, the stratigraphic belonging of each horizon considering a multitude of lithological and stratigraphic criteria (e.g., grain size, sediment gradation, lamination, colors, fossil content) together with the consistency to a common framework of stratigraphic hierarchy. To the authors, this task remains so far unfeasible via an automatic classification algorithm based on common lithological descriptions such as the workflow presented in this study. Therefore, while the current approach permits an evaluation of the subsurface distribution of hydrogeological parameters through the processing of new data sets automatically, it does not facilitate the derivation of precise stratigraphic boundary geometries. Although some grain-size distribution patterns, such as the occurrence of hydrogeological boundaries like fine-grained aquitards, may overlap stratigraphic boundaries, the replication of more complex geometries, such as depositional unconformities and onlap structures, is beyond the scope of this methodology, as it necessitates additional stratigraphic insight. Therefore, known hydrostratigraphic boundaries from the literature have been used as constraints for the proposed interpolations. Thus, the resulting outcomes aim to reconstruct the heterogeneous distribution of the grain-size characteristics and the derived hydrogeological parameters within

hydrostratigraphic units, to support the assessment of large-scale hydrogeological features.

In the next sections, we will explore potential methodologies to incorporate the new semi-quantitative dataset into spatialization frameworks. This integration aims to obtain maps, cross sections and three-dimensional models of hydrogeological variables (i.e. grain-size characteristics, permeability). Hereafter, we will introduce the concept of hydrogeological domains, defined as areas or

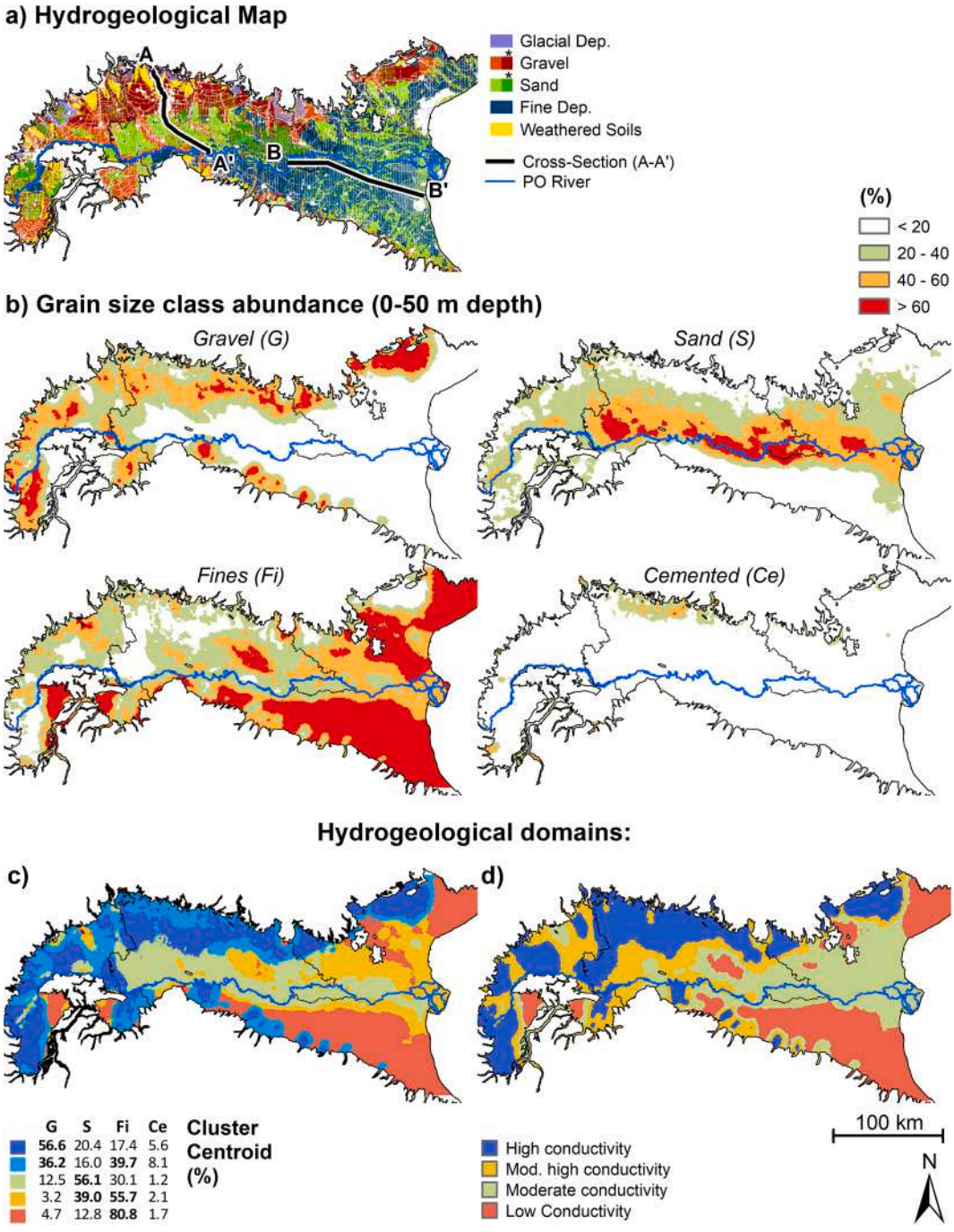


Fig. 11. (a) Hydrogeological map redrawn from (Servizio Geologico Nazionale, 1998) showing the characteristics of the surface deposits in comparison with: (b) Maps showing the percentage of gravel (G), sand (S), fines (silt + clay, Fi) and cemented (Ce) deposits in the shallow subsurface (up to 50 m of depth) predicted by the classification and interpolation models. See the supplementary materials for a comparison of the grain class percentages among specific depth intervals and the kriging error map. (c, d) Hydrogeological domains (map of the hydrofacies) obtained by (c) method #1 (clustering of grain size class abundances, the legend reports the centroids of each cluster highlighting in bold the most relevant class) and (d) method #2 (spatialization and classification of K_{eq}).

volumes of the investigated domain having similar hydrogeological traits. The Kriging interpolation parameters for all the proposed representations are given in Table S1 in the supplementary materials.

4.2.1. Maps of the subsurface hydrogeological domain distribution

The distribution of subsurface hydrogeological features in the Po Plain is reflected by the depositional dynamics occurred during Quaternary (Fontana et al., 2014; Ori, 1993). The northern Alpine sector represents a progressive transition from gravelly to sandy and finer units, reflecting a shift from proximal (alluvial fan) to distal (alluvial plain) locations relative to their sediment sources. This is indicative of the diminishing depositional energy and is mirrored by the decline in topographic gradient. Conversely, the southern Apennine sector is characterized by an abundance of fine deposits, a consequence of mudrocks and flysch presence in the source areas (refer also to Section 2), while coarser deposits are found only at the exits of major valleys, where they form alluvial fans.

These basin geological features are reflected by the grain size characteristics of shallow deposits, which this study aims to predict from borehole logs. Fig. 11a recalls the hydrogeological map of the Po Plain area (Servizio Geologico Nazionale, 1998) previously described in Section 2. Fig. 11b presents the spatial distribution by ordinary Kriging of the depth-averaged grain size class percentages evaluated for each borehole within the initial 50 m of depth. The values were assigned to each unit according to the classified hydrofacies, following Table 1. Boreholes shorter than 45 m were excluded and those longer than 50 m were truncated. For a better representation, fine sediment percentages were aggregated with *'Fines'* in Fig. 11b representing the total content of silt and clay. A clear correspondence is evident between high gravel content areas in Fig. 11b and the coarse-grained, large alluvial fans situated to the north on the Alpine side, as well as the small detrital fans coming out of the Apennine valleys (depicted in Fig. 11a). Areas with a predominance of sandy deposits align with the Po river channel belt and associated remnant meanders in the central plain. Finally, a high content of fine sediments is found in the lower part of the plain, mainly on the Apennine side, due to the predominance of mudrocks in the source area (Ori, 1993).

Fig. 11c and d delineate the main hydrogeological domains reconstructed using the method #1 and #2, respectively (described in Section 3.2). Fig. 11c displays the k-means clustering of the grain size class abundances, with the legend indicating the average content of the grain size classes in each cluster, as determined by the cluster centroid, and the most relevant class highlighted in bold. Fig. 11d portrays the thickness-weighted average hydraulic conductivity (K_{eq}) between 0 and 50 m of depth, after the classification into four main hydrogeological domains, with exact values given in Figure S4 of the supplementary materials. For the Kriging interpolation, the K_{eq} values were log-transformed to reach a Gaussian distribution of the variable that guarantees better stationarity of the variable. The four hydrogeological domains in Fig. 11d are represented by high ($K_{eq} > 3.16*10^{-4}$ m/s), moderately high ($3.16*10^{-4} > K_{eq} \geq 3.16*10^{-5}$ m/s), moderate ($3.16*10^{-5} > K_{eq} \geq 3.16*10^{-6}$ m/s), and low ($K_{eq} < 3.16*10^{-6}$ m/s) hydraulic conductivity values.

A quantitative comparison between the grain size characteristics predicted by this model and the classes depicted in the hydrogeological map is shown by box plots in Fig. 12a. This comparison substantiates the congruence between literature basin-scale data and the model outputs. The goodness of classification using an unsupervised machine-learning classification algorithm (method #1) against the available known classes given in the reference hydrogeological map is given by the confusion matrix in Fig. 12b. From the known classes (Fig. 11a), gravel and sand classes with or without significance amount of weathered coverage have been merged, while 'glacial deposits' and 'weathered soils' have been excluded. From the predicted classes (Fig. 11c), clusters 1–2, and 3–4 have been merged as 'G' and 'S' classes, respectively, while cluster 5 represents the 'Fines' class. Prediction accuracy is 0.77, 0.63 and 0.68 for the 'G', 'S' and 'Fines' classes, respectively. Although some misclassification between 'S' and 'Fines' classes is encountered a promising classification is obtained.

An additional comparison between the predicted hydrogeological domains and the aquifer transmissivity derived from discharge-drawdown well tests by means of Eq. (1) is shown in Figure S6 in the Supplementary Material. A weak correlation was found and attributed to at least two possible reasons: 1) wells that were drilled in areas with a high percentage of fine layers were typically designed to tap into the most conductive layers. These productive layers, despite their limited thickness, can significantly influence the overall transmissivity; 2) interpolation of grain size data is susceptible to inaccuracies, particularly in regions where the subsurface composition is markedly heterogeneous. If the spatial density of data points is insufficient to accurately represent this heterogeneity, local errors in grain size prediction may occur.

4.2.2. Cross-sections

The vertical spatial distribution of the grain-size and hydrogeological properties was reconstructed along with two demonstratives regional scale cross-sections (see Fig. 1 for the trace of the sections AA' and BB'). These sections are selected due to the availability of hydrostratigraphic information used as vertical constraints for the spatialization of the algorithmic outcomes. For each section, codified information from boreholes within a 500-meter distance, on both sides of the section trace were used after a vertically resampling at 0.5 m intervals, yielding 101,058 and 72,911 data points from 788 and 383 boreholes, for section AA' and BB', respectively. Boreholes used in the reference cross-sections from the literature are 36 and 74, respectively.

The first example exploits a north-south cross-section (AA'), north of the Po river, that spans the upper plain. This section captures the transition from proximal Alpine depositional facies (mainly coarse-grained alluvial fan) to distal facies (mainly sandy alluvial plain). A reference hydrogeological cross-section is reported in Fig. 13 from the literature (Regione Lombardia, 2016) delineating both permeable units (shallow and deep aquifers) and low permeability layers. This section was vertically bounded, at the top, by the topographic surface and, at the bottom, by the base of the intermediate aquifer unit adhering to the boundary limit from (De Caro et al., 2020) in agreement with the aquifer bottom surface delineated by the regional groundwater characterization and protection plan (PTUA) by Regione Lombardia (2016). The boundary between the shallow and the intermediate aquifer units (dashed line) does not represent a strong and continuous hydrogeological boundary and, for this reason, the spatialization domain encompass both these

partially connected units. Fig. 13c illustrates the predicted abundances of main grain-size classes along the cross-section, achieved through ordinary Kriging interpolation.

Different classification techniques were compared, including the ROC method for variables 'G', 'S' and 'Fines', and method #1 and #2 described in Section 3.2, validating the predictability of the aquifer facies against the truth given by Fig. 13b. Classification metrics are given in the supplementary material (Table S2). The ensemble of all grain-size variables (method #1) gives the best classification accuracy (0.72) followed by method #2 (0.66). Therefore, the subdivision into hydrogeological domains according to these methods is displayed in Fig. 13d and e.

Fig. 13d was generated by a k-means clustering of the grain size class abundances (method #1), with the legend reporting the average content of the grain size classes for each cluster as determined by the cluster centroid, highlighting the most significant class in bold. Instead, Fig. 13e was generated classifying the interpolated hydraulic conductivity (K) values in three domains (method #2). The results from both methods highlight the north-south transition from more permeable to less permeable domains in the phreatic aquifer as evidenced by the presence of lowland natural springs (see *'Fontanili'* in Section 2). Both methods also delineate a low conductivity horizon (discontinuous aquitard), reflecting the separation between the shallow and intermediate aquifer groups, each exhibiting different degrees of connectivity due to the spatialization method. Method #1, in particular, reveals the presence of areas dominated by cemented units (typically conglomeratic) exemplified by the *'Ceppo'* formation described by Orombelli (1979). The proposed models exhibit limitations around the *'San Colombano'* syndepositional anticline ramp uplift. In this area, the lack of stratigraphic information within the spatialization process, combined with sparse spatial data density, hinders the accurate prediction by this method of complex geometries (as described by Zuffetti and Bersezio, 2021) without the integration of prior stratigraphic knowledge. To avoid unreliable predictions, the modeling domain excludes the area around the *'San Colombano'* structure where late Pleistocene alluvial units rest in unconformity above the marine succession (Zuffetti et al., 2018).

The second reference stratigraphic cross-section develops west-eastward (BB' in Fig. 1, for approximately 100 km) in the center of the Plain up to the Adriatic Sea (Fig. 14). This section, redrawn from Amorosi and Pavesi (2010) in Fig. 14b, portrays a sequence of alternating clayey-silty fines (aquitards) and coarser sandy deposits (aquifers), extending from the topographic surface to the *'Upper Po Synthem'* basal unconformity identified on a seismic basis (450 ka, from Amorosi and Pavesi, 2010). These layers extend vertically for about 50–90 m and reflects the cyclic depositional environment of the distal alluvial plain, characterized by fluvial channel belts, prograding shorelines and deltas. The western part of the section is dominated by thick and laterally continuous permeable units, indicative of four aquifer systems (Fig. 14b). In contrast, the eastern segment is dominated by a series of discrete, relatively thinner, and finer-grained sand bodies, interspersed with thicker silty layers affected by local tectonic deformation towards the basin margins to the east. Amorosi and Pavesi (2010) infer onlapping geometries upon the lower bounding unconformity. Notably, the aquifer systems I to IV as identified in Fig. 14b from Amorosi and Pavesi (2010) belong to the upper hydrostratigraphic unit 'aquifer group A′ identified by Regione Emilia-Romagna and ENI (1998), which, despite small discrepancies, is considered to represent the lateral equivalent south of the Po river of the one identified by Regione Lombardia and ENI (2002), which is portrayed in Fig. 13b from Regione Lombardia (2016).

Fig. 14c illustrates the predicted abundances of the main grain-size classes along the cross-section, achieved through ordinary Kriging interpolation. The 'gravel' and 'cemented' classes are not displayed due to the absence of significant amounts of these materials along the section. For each section shown in Fig. 14, areas with high uncertainty resulting from the spatialization (kriging error variance $>$ 30 %) are hidden by oblique hatching to avoid unreliable predictions.

Fig. 14d shows the hydrofacies predicted by the classification of the grain-size abundances shown in Fig. 14c using the ROC approach (see the classification metrics in the supplementary material). Both 'gravel'+ 'sand' and 'fines' representations were tested showing a good predictability of known aquifers from Fig. 14b, with AUROC values of 0.72 and 0.70, respectively. Looking at the predicted hydrofacies distribution, a correlation is evident between the first three aquifer systems and the predicted aquifers, particularly on the western side of the section where greater thickness and lateral continuity prevail. On the eastern side, despite the lower data density, a reasonable match is still observable and the eastern dipping geometry proposed by the literature is replicated for

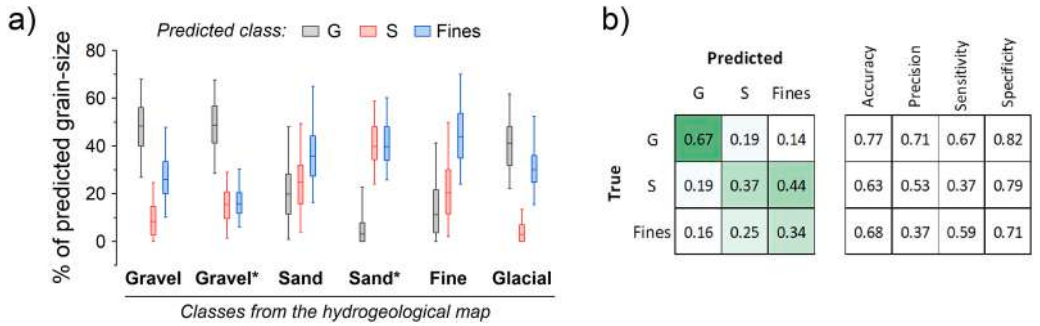


Fig. 12. (a) Boxplots showing the comparison between the grain size abundances predicted by the proposed model and the classes depicted in the hydrogeological map in Fig. 11a (*Deposits with weathered coverage). (b) Confusion matrix and classification metrics of the classification model obtained with method #1 (Fig. 11c) against the reference classification given by the hydrogeological map in Fig. 11a. Values in the confusion matrix are normalized by the total number of samples in each true class.

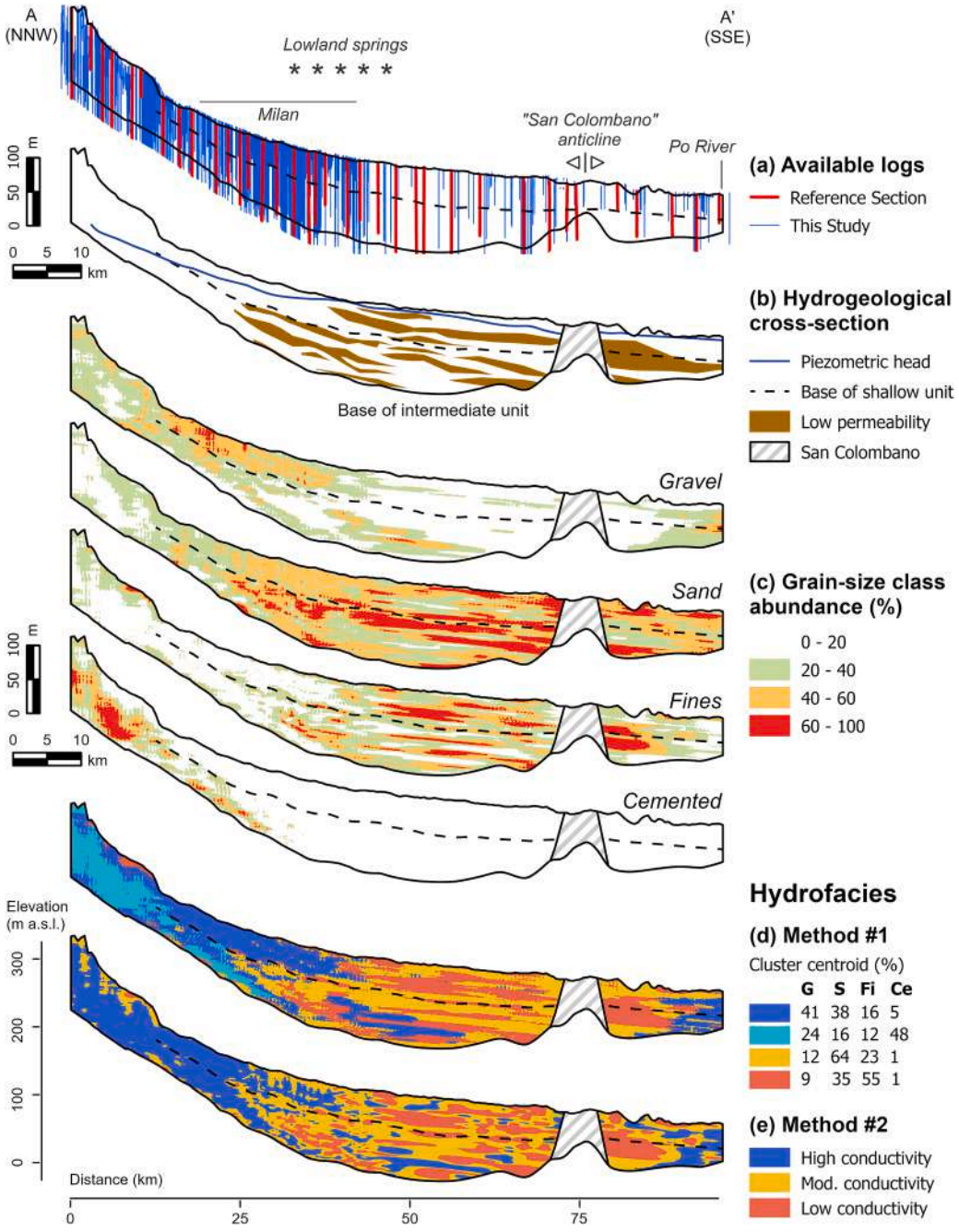


Fig. 13. Hydrogeological cross-section (see Fig. 1 for the trace of the section) redrawn from (Regione Lombardia, 2016) showing (a) the available borehole logs (red, Regione Lombardia, 2016; blue, this study), and (b) the characteristics of the subsurface deposits. The lowland springs ('Fontanili') belt is marked by asterisks, while the 'San Colombano' isolated relief is highlighted by oblique hatching. (c) Cross-sections showing the percentage of (G) gravel, (S) sand, (Fi) fines (silt + clay) and (Ce) cemented deposits predicted by the classification and interpolation models. Spatial distribution of homogenous hydrofacies obtained by: (d) method #1 (clustering of grain size class abundances, the legend reports the centroids of each cluster highlighting in bold the most relevant class) and (e) method #2 (spatialization and classification of the hydraulic conductivity). Vertical scale exaggerated 100 times.

the II and III predicted aquifer units. Finally, the results were validated through the confusion matrix of known aquifers (Fig. 14b) versus predicted aquifers (Fig. 14d), showing an accuracy of 0.67. Due to the higher data density, in the western side of the section, higher values of accuracy, precision and sensitivity (given in the supplementary materials) are obtained reflecting a better predictability of aquifer units and their connectivity. It is important to note that the literature cross-sections are themselves the result of an interpretation process, which could potentially influence the aforementioned validation.

However, the methodology has two main limitations: 1) high spatialization uncertainties, stemming from sparse data distribution, yield poorly reliable geometries (see the central part of section BB' in Fig. 14); 2) the replication of complex stratigraphic geometries is challenging without the integration of stratigraphic rules in the spatialization method. Hence, the proposed method is best suited for predicting the heterogeneous proportions of hydrogeological properties within known subsurface units (e.g., the *'Upper Po Synthem'* of Fig. 14) devoid of any stratigraphic context.

4.2.3. 3D-model

A detailed 3D regional-scale hydrofacies model has been constructed for two subsectors of the Po alluvial plain (Fig. 15). These domains were selected based on the availability of hydrostratigraphic surfaces, which served as physical constraints for the spatial distribution of hydrofacies, enabling a three-dimensional characterization of the heterogeneity in hydrogeological properties. The first 3D domain (Fig. 15a) covers the entire Lombardia region and is developed for the "aquifer group A" *sensu* Regione Lombardia and ENI (2002) for a total volume of $1.54*10^{12}$ m^3, while the second (Fig. 15b) covers the entire Emilia-Romagna region and is developed for the "aquifer systems" A1 and A2 (shallow portion of the "aquifer group A") *sensu* Regione Emilia-Romagna and ENI (1998) for a total volume of $1.39*10^{12}$ m^3. For both domains, the bottom confining hydrostratigraphic surfaces are available online on their respective regional geoportals (https://www.geoportale.regione.lombardia.it/ and https://geoportale.regione.emilia-romagna.it/, respectively).

Grain-size abundances derived from the classified boreholes were systematically resampled at 0.5 m intervals along the borehole axis. After excluding the points outside the selected hydrostratigraphic units, two datasets of about 1.19 and 0.31 million 3D points,

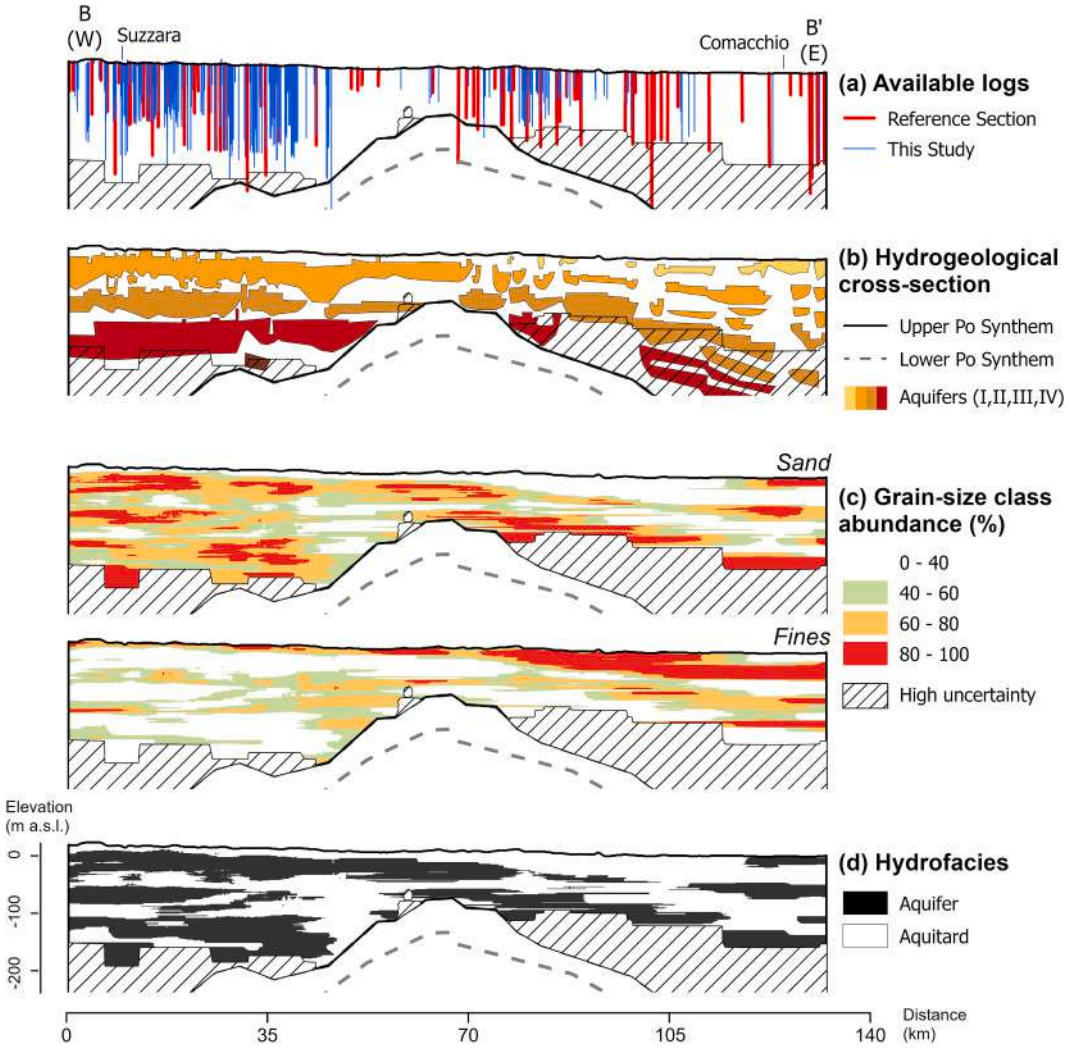


Fig. 14. Hydrostratigraphic cross-section through the central axis of the Plain (see Fig. 1 for the trace of the section) showing: (a) the available borehole logs (red, Amorosi and Pavesi, 2010; blue, this study), and (b) the geometry of the aquifer systems in the 'Upper Po Synthem' redrawn from (Amorosi and Pavesi, 2010). (c) Cross-sections showing the percentage of (S) sand and (Fi) fines (silt + clay) deposits predicted by the classification and interpolation models. (d) Predicted aquifer distribution using method #1 (see text above for details). Oblique hatching shows areas of high spatialization uncertainty (kriging error variance > 30 %). Vertical scale exaggerated 50 times.

distributed across 21,358 and 2836 boreholes and segmented into 185,099 and 24,924 individual descriptive intervals were obtained, respectively for subdomains (a) and (b) in Fig. 15. The interpolations of grain size abundances were performed separately inside the 3D domains enclosed by the hydrostratigraphic structures on two quadratic grids 1000 × 1000 × 2 m spaced, resulting in roughly 770,000 and 694,000 cells, respectively. The models show the predicted abundances of the 'gravel', 'sand' and 'fines' classes by means of ordinary kriging (see the supplementary materials for the model parameters).

Despite the coarse horizontal resolution of the grid, imposed by the computational demands of processing such an extensive dataset, several significant hydrogeological features were identified at the regional scale. These include: (i) a gradual eastward decrease in the grain size towards the Po delta (Fig. 15a and b), (ii) the high gravel content of the upper plain on the Alpine side where large alluvial fans coexist (Fig. 15a), (iii) the presence of smaller gravel-dominated alluvial fans at the outlet of the Apennine valleys, followed by layered alternances of sand and fine deposits (Fig. 15b), (iv) the abundance of sand along the axis of the Po river and the associated channel belts (Fig. 15a). It is noteworthy that the models capture both horizontal and vertical heterogeneities, which might differ according to the scale of analysis and selected variogram parameters but this fall beyond the scope of the current work.

This 3D model has two major limitations: (i) the applicability of this spatialization approach relies on the existence of a priori

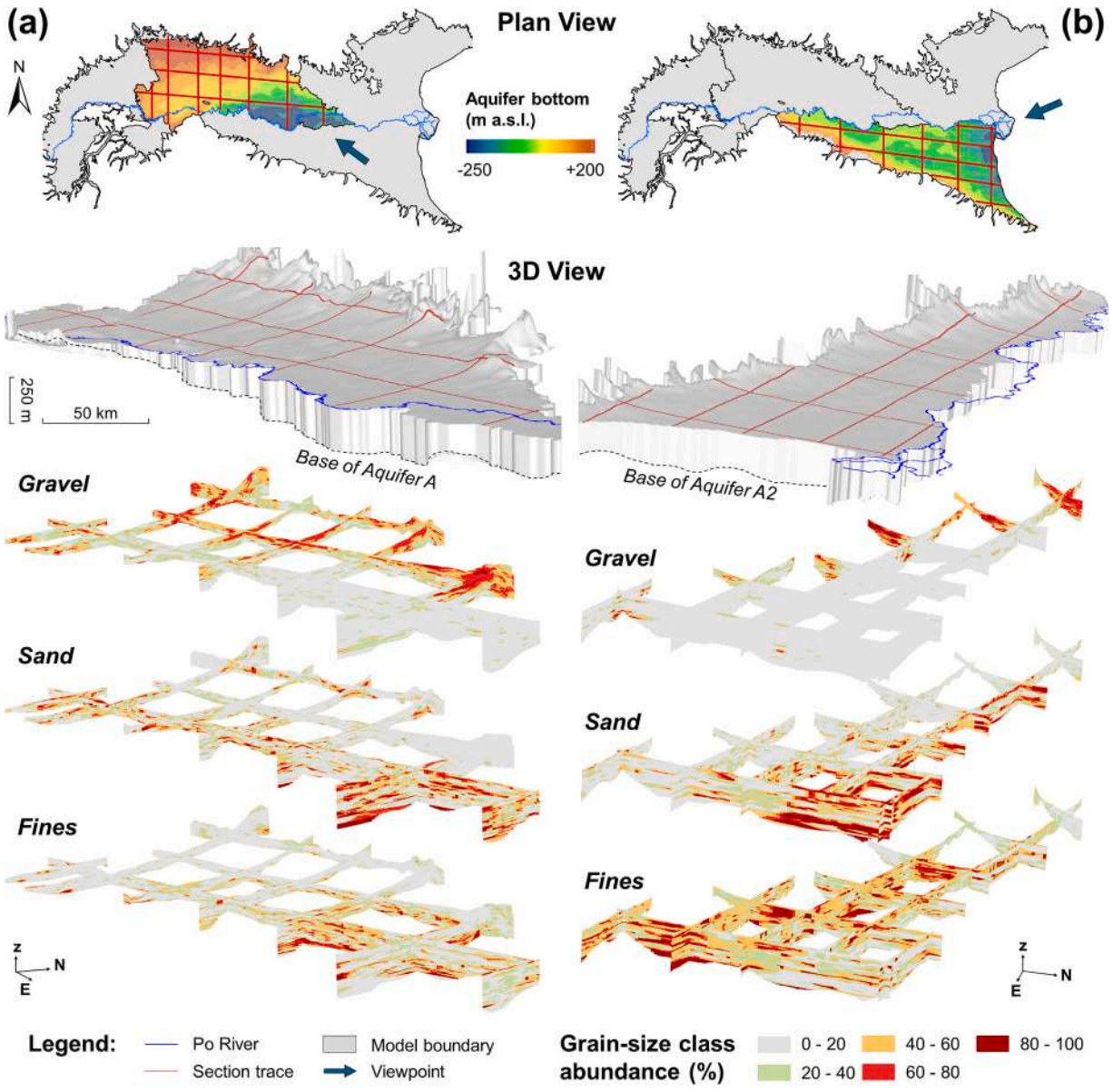


Fig. 15. 3D models for two sub-sectors of the Po plain showing the spatial distribution along regularly spaced cross-sections of the abundances of the three main grain-size classes. Panel (a) shows the 3D model for the Lombardia region inside the "aquifer group A" unit. Panel (b) shows the 3D model for the Emilia-Romagna region inside the "aquifer systems" A1 and A2. Vertical dimension is exaggerated 100 times.

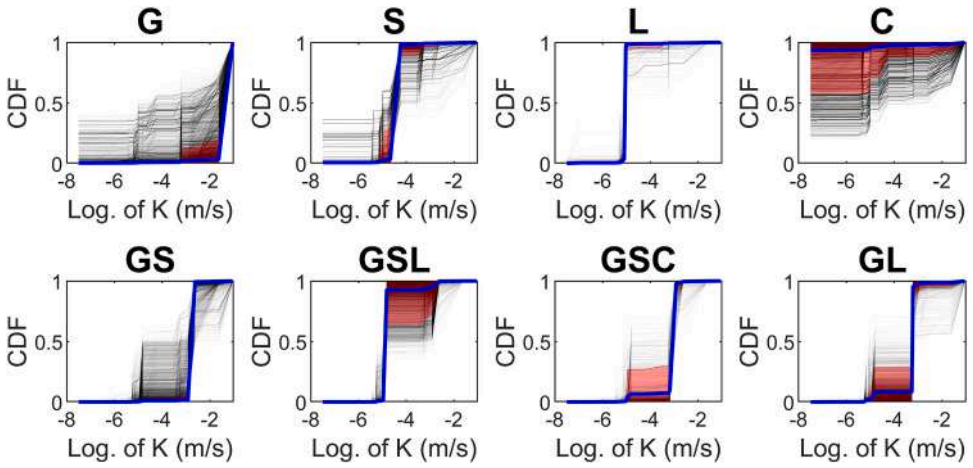


Fig. 16. Cumulative distribution functions (CDF) of the hydraulic conductivity (K) assigned to each classified text string according to the uncertainty of the classifier. The CDFs are grouped by the classification obtained in this analysis for some demonstrative classes. The blue thick line represents the average CDF of each class and the red area represents the 5th to 95th percentile range.

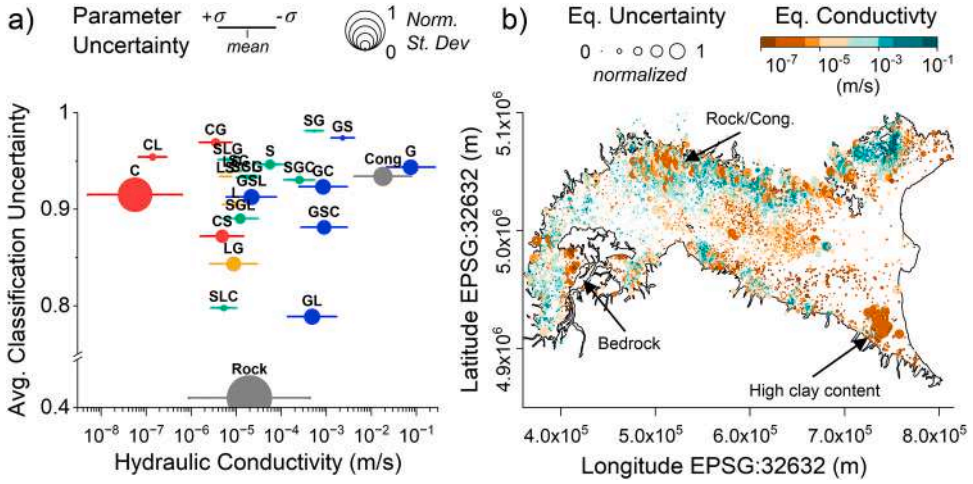


Fig. 17. a) Plot showing the uncertainty for each hydrofacies class expressed as the standard deviation of all the probabilistic estimates of the hydraulic conductivity. The x-axis represents the estimated hydraulic conductivity, the y-axis the average classification uncertainty of each class. Bubbles are sized to the normalized standard deviation of K and the associated horizontal bar represents the ± σ interval range of K. b) Map showing the average horizontal equivalent hydraulic conductivity (color scale) and the associated uncertainty from the normalized standard deviation obtained for each borehole after the probabilistic simulation.

hydrostratigraphic knowledge, which is essential to constrain the hydrofacies reconstructions within hydrostratigraphic coherent bodies, avoiding unreliable prediction at the transition between different hydrogeologically separated stratigraphic units. Unfortunately, a common boundary for the shallow hydrostratigraphic unit is still lacking in the available literature for the entire Po alluvial plain. Nevertheless, this is not a limit of the proposed deep learning hydrofacies classification model itself but of the selected spatialization method which might be optimized according to the scopes and the scale of modeling; (ii) a full cell-by-cell validation of this model remains unfeasible because there is no equivalent product available for comparison.

The recent study by Manzoni et al. (2023) introduces an interesting 3D geological model for the entire Po Plain area, emphasizing the quantification of uncertainty by means of machine learning techniques. They identified the primary sources of uncertainty in the low spatial data density, the occurrence of small-scale patterns, and the delineation of boundaries between major aquifer bodies. While their work provides a robust framework for understanding the spatial distribution of uncertainties, it is posited that the simplification in the classification and parameterization of the hard data, that was not the main focus of their contribution, may lead to oversight of significant variations in the hydrogeological parameters. For instance, the presence of a substantial clay component within a gravelly unit might be overlooked. In contrast, the proposed approach, by means of a deep learning-aided full text analysis of the stratigraphic logs, successfully identified critical non-negligible secondary terms that are fundamental for the hydrogeological classification of heterogeneous sedimentary units. Hence, this capability distinguishes this research by that of Manzoni et al. (2023).

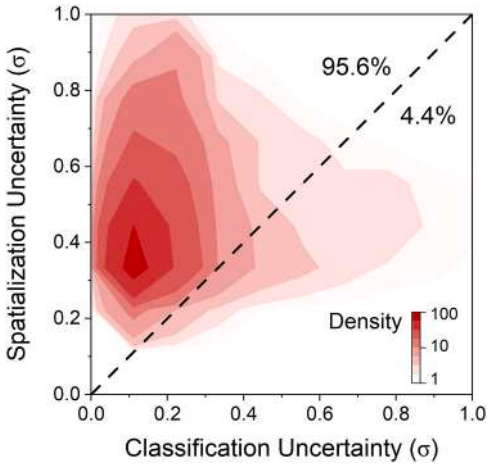


Fig. 18. Density plot comparing the uncertainty (expressed as the standard deviation of the prediction) due to classification (x-axis) and spatialization (y-axis) evaluated over a 1000m-spaced grid across the study area. Due to the large amount of data the plot density was evaluated on a 0.05σ-spaced grid, showing for the majority of the data (95.6 %) a higher uncertainty due to spatialization.

4.3. Classification vs spatialization uncertainty

A workflow to generate large scale hydrofacies/hydrogeological models was presented, integrating qualitative and quantitative stratigraphic and hydrogeological data by means of deep learning text classification and spatialization techniques. Different types of uncertainties in the outputs may arise from:

1) misclassification;
2) parameter estimation;
3) spatialization.

This discussion addresses the uncertainty arising from the automatic classification technique and the spatialization method. The uncertainty associated to the empirical estimation of hydrogeological parameters was not considered opting instead for a single, most representative, value for each hydrofacies class as described in Table 1. This choice was substantiated to avoid excessive analysis in the paper focusing on the novelty of the approach.

The classification uncertainty was retrieved from the normalized output of the *softmax* layer of the LSTM network. This layer consists of a vector whose length corresponds to the number of classification categories, with its values representing the classification probabilities. For each predicted text string, the probability of being classified into each hydrofacies class, was used to build a probability distribution function (PDF) of the hydrogeological parameter by converting the classes into numerical values (i.e., by assigning the value of hydraulic conductivity to each hydrofacies class). Then, by sorting the classes according to the assigned K values, the cumulative distribution function (CDF) of the estimated K was obtained for each text string representing a descriptive interval within a borehole. Fig. 16 shows all the CDFs of hydraulic conductivity (black lines) for some selected classes, highlighting the mean (blue line) and the 5th and 95th percentiles (red area). These examples illustrate how the uncertainty of the classification model impacts the corresponding hydrogeological parametrization. For some classes (e.g., G, S, L, GS) the classifier exhibits a propensity towards a definitive decision resulting in small probabilities of being classified into other classes, thereby receiving a different hydrogeological parameter. In other cases, especially clay (C) and other heterogeneous classes such as, GSL, GSC and GL, the classifier is less decisive and the associated hydrogeological parameters can vary more significantly.

Based on the CDF obtained for each stratigraphic unit, a random sampling technique was adopted to perturb the (optimal) value of conductivity, obtaining a set of 1000 values (realizations) of K for each layer in the borehole. Fig. 17a shows the mean classification uncertainty (y-axis) against the resulting uncertainty from the parameter association (x-axis), expressed as the standard deviation of all K realizations for each hydrofacies class. It is important to note that, while certain classes exhibit low classification uncertainty (less than 0.1), the resulting uncertainty in the associated parameter (K) can vary significantly, depending on the classifier's confidence in assigning alternative classes. For instance, the GS and SG classes are confidently classified and have a low parameter uncertainty because the few indecisions involve classes with similar K values. Conversely, for descriptions falling in the C or G classes, despite being relatively confident, the associated parameter uncertainty is higher due to the markedly different K values of potential alternative classifications.

These realizations were combined to obtain 1000 values of equivalent horizontal hydraulic conductivity (K_{eq}) for each borehole by computing the thickness-weighted average of K for each layer between 0 and 50 m depth. Finally, a probabilistic estimate of K_{eq} for each borehole was obtained considering both the mean and the standard deviation of the 1000 realizations. Fig. 17b shows the equivalent depth-averaged parameter uncertainty for each borehole within the 0–50 m depth interval. The greatest uncertainties were

found at the margins of the plain, where bedrock formations could be shallower than 50 m, or where partially cemented deposits (conglomerates, *'Cong'* class) were typically encountered. Another critical area is the south-eastern sector of the Plain, characterized by shallow deposits with substantial clay content and the use of specific terms to describe units that were not included in the training dataset of the classifier.

Finally, to obtain the hydraulic conductivity map, shown in Fig. 11d and in the supplementary material, an ordinary kriging interpolation was performed introducing a second kind of uncertainty. The kriging algorithm computes prediction error through the standard deviation of the best estimate, which is a function of the spatial separation of the data points and the selected variogram function. Thus, the kriging error map (shown in Figure S5 in the supplementary materials) was instrumental in unraveling the uncertainty attributable to spatialization. This was compared to the uncertainty related to the automatic classification process previously described. Fig. 18 provide a comparative analysis of the uncertainties due to classification (i.e., parameter uncertainty) and spatialization (i.e., interpolation uncertainty), evaluated over a 1000m-spaced grid throughout the study area. Generally, the uncertainty due to spatialization significantly exceed (by more than tenfold in some cases) the uncertainty induced by classification, except for a 4.4 % of the study area. The reader must be aware that the uncertainty due to spatialization depends on the distance between data points and the selected variogram parameters, which are inherently subject to their own degree of uncertainty.

5. Conclusions

This study introduces a method for enhancing the textual geological descriptions of borehole logs, merging the capabilities of a deep learning LSTM-based text classification algorithm, for automatic hydrofacies recognition, with the estimation of hydrogeological parameters informed by grain size data and well drawdown tests. A large dataset of new semi-quantitative hydrogeological data was obtained for the Po alluvial plain sedimentary basin in northern Italy. This markedly increased the spatial density of hydrogeological information at regional scale.

Two approaches were used to reconstruct the spatial distribution of hydrofacies with the new 3D dataset. The results were compared and validated against available qualitative (hydrogeological maps and cross-sections) and quantitative (grain size analyses and well tests) data. The generation of regional scale thematic maps, hydrogeological cross-sections and a 3D model showed good consistency with the existing hydrogeological knowledge documented in the available literature.

The overall high classification accuracy (97.4 %) achieved in the validation can be attributed to several key factors: 1) the extensive size of the training set; 2) the effectiveness of deep learning in classifying natural language text, enhanced by word embedding; 3) the structured grain size information within the stratigraphic descriptions ruled by the Italian geotechnical guidelines; 4) the accuracy of a previously labeled (training) dataset with a specific focus on grain size characteristics. The capability of the algorithm to discern the significance of geological terms, their synonyms, and common typographical errors, related to grain size has been proven, resulting in highly accurate classifications. This was possible even amidst complex borehole log descriptions employing multiple terms for grain size delineation. The major benefit of the proposed approach is the significant increase in the spatial density of data, which is instrumental for generating large-scale hydrofacies/hydrogeological maps and models (from 0.34 data/km^2 on average combining grain size and aquifer tests data, to 8.7 data/km^2 considering also the information from the available borehole log descriptions). This method is replicable, contingent upon the availability of a comprehensive dataset of geological records and a suitably labeled training set encompassing all possible geological categories. However, it is important to acknowledge that descriptive geological data alone are not sufficient for the generation of quantitative subsurface models. Only the integration of qualitative geological textual information with other linkable quantitative data (such as, e.g., grain size analysis) can significantly improve the spatial density of information useful for hydrofacies/hydrogeological modeling.

Moreover, the primary limitation of this approach lies in its reliance on borehole lithological descriptions which are mainly focused on the grain size. Hence, it provides grain-size and hydrogeological information along boreholes without incorporating a stratigraphic perspective. Therefore, while the data generated by this approach can refine the spatialization of hydrogeological properties by different methods, it is not adequate to reconstruct a robust framework of stratigraphic boundaries, which are still necessary as *a priori* information to delineate valid boundaries for interpolation. In fact, the proposed techniques demonstrated high capabilities for predicting hydrofacies proportions inside stratigraphic units rather than delineating their boundaries. At the same time, a supervised approach to the definition of grain size limits and distribution represents a promising prerequisite to support more advanced hydrogeological interpretations. While lacking a hydrostratigraphic perspective, the two cross-sections and the 3D model nonetheless offer valuable insights into the heterogeneity of hydrogeological parameters derived from the distribution of hydrofacies. Further hydrostratigraphic improvement of the borehole logs could be reached by a multi-dimensional training of the classification network introducing other quantitative well logs that are relevant to identify stratigraphic boundaries.

CRediT authorship contribution statement

Alberto Previati: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Valerio Silvestri:** Data curation. **Giovanni Crosta:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

influence the work reported in this paper.

Acknowledgments

This work is an outcome of MUR (Ministry of University and Research) project "URGENT - Urban Geology and Geohazards: Engineering geology for safer, resilieNt and smart ciTies", PRIN 2017HPJLPW.

The authors are deeply indebted to MM s.p.a. and ARUP (Italy) for providing essential data used in this study.

The authors thank also: Prof. **Paolo Frattini** for reviewing the manuscript draft, **Alberto Presta Asciutto** for reviewing and homogenizing the available grain size distributions data used in this study, and **Denise Melis** for reviewing the data availability in different regions of the study area during her Master's thesis.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2024.102157.

Data availability

I have shared the links to access the data used in this work in Table 2

References

A.G.I., 1963. Nomenclatura geotecnica e classifica delle terre (Italian soil classification standard).

Alyamani, M.S., Şen, Z., 1993. Determination of hydraulic conductivity from complete grain-size distribution curves. Groundwater 31, 551–555. https://doi.org/10.1111/j.1745-6584.1993.tb00587.x.

Amorosi, A., Pavesi, M., 2010. Aquifer stratigraphy from the middle-late Pleistocene succession of the Po Basin. Mem. Descr. della Carta Geol. D. 'Ital. XC 7–19.

Anderson, M.P., 1989. Hydrogeologic facies models to delineate large-scale spatial trends in glacial and glaciofluvial sediments. Geol. Soc. Am. Bull. 101, 501–511. https://doi.org/10.1130/0016-7606(1989)101%3C0501:HFMTDL%3E2.3.CO;2.

Asante-Okyere, S., Shen, C., Osei, H., 2022. Enhanced machine learning tree classifiers for lithology identification using Bayesian optimization. Appl. Comput. Geosci. 16, 100100. https://doi.org/10.1016/j.acags.2022.100100.

Bayer, P., Huggenberger, P., Renard, P., Comunian, A., 2011. Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. J. Hydrol. (Amst.) 405, 1–9. https://doi.org/10.1016/j.jhydrol.2011.03.038.

Beucher, H., Renard, D., 2016. Truncated Gaussian and derived methods. Comptes Rendus - Geosci. 348, 510–519. https://doi.org/10.1016/j.crte.2015.10.004.

Beyer, W., 1964. Zur bestimmung der wasserdurchlässigkeit von kiesen und sanden aus der kornverteilungskurve (In German). Wasser Wasser 14, 165–168.

Campo, B., Bohacs, K.M., Amorosi, A., 2020. Late Quaternary sequence stratigraphy as a tool for groundwater exploration: lessons from the Po River Basin (northern Italy). Am. Assoc. Pet. Geol. Bull. 104, 681–710. https://doi.org/10.1306/06121918116.

Carman, P.C., 1937. Fluid flow through granular beds. Trans. Inst. Chem. Eng. Lond. 15, 150–156.

Chapuis, R.P., Dallaire, V., Marcotte, D., Chouteau, M., Acevedo, N., Gagnon, F., 2005. Evaluating the hydraulic conductivity at three different scales within an unconfined sand aquifer at Lachenaie, Quebec. Can. Geotech. J. 42, 1212–1220.

Chesnaux, R., Baudement, C., Hay, M., 2011. Assessing and comparing the hydraulic properties of granular aquifers on three different scales. Proceedings of Geohydro Conference, Quebec City (Quebec), Canada 1–10.

Chowdhary, K.R., 2020. Fundamentals of artificial intelligence. Fundam. Artif. Intell. https://doi.org/10.1007/978-81-322-3972-7.

De Caro, M., Perico, R., Crosta, G.B., Frattini, P., Volpi, G., 2020. A regional-scale conceptual and numerical groundwater flow model in fluvio-glacial sediments for the Milan Metropolitan area (Northern Italy). J. Hydrol. Reg. Stud. 29, 100683. https://doi.org/10.1016/j.ejrh.2020.100683.

De Luca, D.A., Destefanis, E., Forno, M.G., Lasagna, M., Masciocco, L., 2014. The genesis and the hydrogeological features of the Turin Po Plain fontanili, typical lowland springs in Northern Italy. Bull. Eng. Geol. Environ. 73, 409–427. https://doi.org/10.1007/s10064-013-0527-y.

dell'Arciprete, D., Bersezio, R., Felletti, F., Giudici, M., Comunian, A., Renard, P., 2012. Comparison of three geostatistical methods for hydrofacies simulation: a test on alluvial sediments. Hydrogeol. J. 20, 299–311. https://doi.org/10.1007/s10040-011-0808-0.

Fontana, A., Mozzi, P., Marchetti, M., 2014. Alluvial fans and megafans along the southern side of the Alps. Sediment Geol. 301, 150–171. https://doi.org/10.1016/j.sedgeo.2013.09.003.

Freeze, R.A., Cherry, J.A., 1979. Groundwater. Prentice-Hall, Englewood Cliffs, NJ.

Fuentes, I., Padarian, J., Iwanaga, T., Willem Vervoort, R., 2020. 3D lithological mapping of borehole descriptions using word embeddings. Comput. Geosci. 141. https://doi.org/10.1016/j.cageo.2020.104516.

Garzanti, E., Vezzoli, G., Andò, S., 2011. Paleogeographic and paleodrainage changes during Pleistocene glaciations (Po Plain, northern Italy). Earth Sci. Rev. 105, 25–48.

Giacomelli, S., Zuccarini, A., Amorosi, A., Bruno, L., Di Paola, G., Martini, A., Severi, P., Berti, M., 2023. 3D geological modelling of the Bologna urban area (Italy). Eng. Geol. 324. https://doi.org/10.1016/j.enggeo.2023.107242.

Guzzetti, F., Marchetti, M., Reichenbach, P., 1997. Large alluvial fans in the north-central Po Plain (Northern Italy). Geomorphology 18, 119–136. https://doi.org/10.1016/S0169-555X(96)00015-3.

Harleman, D.R.F., Mehlhorn, P.F., Rumer Jr, R.R., 1963. Dispersion-permeability correlation in porous media. J. Hydraul. Div. 89, 67–85.

Harris, J.R., Grunsky, E.C., 2015. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. Comput. Geosci. 80, 9–25. https://doi.org/10.1016/j.cageo.2015.03.013.

Hazen, A., 1982. Some Physical Properties of Sands and Gravels, with Special Reference to Their Use in Filtration. In: State Sanitation: A Review of the Work of the Massachusetts State Board of Health, II. Harvard University Press.

He, M., Zhou, J., Li, P., Yang, B., Wang, H., Wang, J., 2023. Novel approach to predicting the spatial distribution of the hydraulic conductivity of a rock mass using convolutional neural networks. Q. J. Eng. Geol. Hydrogeol. 56. https://doi.org/10.1144/qjegh2021-169.

Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. Science 349 (1979), 261–266.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings 1–15.

Kowsari, K., Meimandi, K.J., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D., 2019. Text classification algorithms: A survey. Inf. (Switz.) 10, 1–68. https://doi.org/10.3390/info10040150.
Kozeny, J., 1953. Das wasser im boden. Grundwasserbewegung. Hydraul.: ihre Grundl. und Prakt. Anwend. 380–445.
Lawley, C.J.M., Raimondo, S., Chen, T., Brin, L., Zakharov, A., Kur, D., Hui, J., Newton, G., Burgoyne, S.L., Marquis, G., 2022. Geoscience language models and their intrinsic evaluation. Appl. Comput. Geosci. 14. https://doi.org/10.1016/j.acags.2022.100084.
Lawley, C.J.M., Gadd, M.G., Parsa, M., Lederer, G.W., Graham, G.E., Ford, A., 2023. Applications of Natural Language Processing to Geoscience Text Data and Prospectivity Modeling. Nat. Resour. Res. 32, 1503–1527. https://doi.org/10.1007/s11053-023-10216-1.
Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P.S., He, L., 2022. A Survey on Text Classification: From Traditional to Deep Learning. ACM Trans. Intell. Syst. Technol. 13. https://doi.org/10.1145/3495162.
Li, Z., Kang, Y., Feng, D., Wang, X.M., Lv, W., Chang, J., Zheng, W.X., 2020. Semi-supervised learning for lithology identification using Laplacian support vector machine. J. Pet. Sci. Eng. 195, 107510. https://doi.org/10.1016/j.petrol.2020.107510.
Livani, M., Petracchini, L., Benetatos, C., Marzano, F., Billi, A., Carminati, E., Doglioni, C., Petricca, P., Maffucci, R., Codegone, G., Rocca, V., Verga, F., Antoncecchi, I., 2023. Subsurface geological and geophysical data from the Po Plain and the northern Adriatic Sea (north Italy). Earth Syst. Sci. Data 15, 4261–4293. https://doi.org/10.5194/essd-15-4261-2023.
Logan, J., 1964. Estimating Transmissibility from Routine Production Tests of Water Wells. Ground Water 2, 35–37. https://doi.org/10.1111/j.1745-6584.1964.tb01744.x.
Manzoni, A., Porta, G.M., Guadagnini, L., Guadagnini, A., Riva, M., 2023. Probabilistic reconstruction via machine-learning of the Po watershed aquifer system (Italy). Hydrogeol. J. https://doi.org/10.1007/s10040-023-02677-8.
Mariethoz, G., Renard, P., Cornaton, F., Jaquet, O., 2011. High-resolution truncated plurigaussian simulations for the characterization of heterogeneous formations. https://doi.org/10.1111/j.1745-6584.2008.00489.x.
Marini, M., Felletti, F., Beretta, G.Pietro, Terrenghi, J., 2018. Three geostatistical methods for hydrofacies simulation ranked using a large borehole lithology dataset from the venice hinterland (NE Italy). Water (Switz.) 10. https://doi.org/10.3390/w10070844.
Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings 1–12.
Min, X., Pengbo, Q., Fengwei, Z., 2020. Research and application of logging lithology identification for igneous reservoirs based on deep learning. J. Appl. Geophy 173, 103929. https://doi.org/10.1016/j.jappgeo.2019.103929.
NAVFAC, 1974. Design manual-soil mechanics, foundations, and earth structures. US Government Printing Office, Washington, DC.
Nichols, G.J., Fisher, J.A., 2007. Processes, facies and architecture of fluvial distributary system deposits. Sediment Geol. 195, 75–90. https://doi.org/10.1016/j.sedgeo.2006.07.004.
Ori, G.G., 1993. Continental depositional systems of the Quaternary of the Po Plain (northern Italy). Sediment. Geol.
Orombelli, G., 1979. Il Ceppo dell'Adda: revisione stratigrafica. Riv. Ital. Paleontol. STRATIGR 85.
Ouellon, T., Lefebvre, R., Marcotte, D., Boutin, A., Blais, V., Parent, M., 2008. Hydraulic conductivity heterogeneity of a local deltaic aquifer system from the kriged 3D distribution of hydrofacies from borehole logs, Valcartier, Canada. J. Hydrol. (Amst.) 351, 71–86. https://doi.org/10.1016/j.jhydrol.2007.11.040.
Padarian, J., Fuentes, I., 2019. Word embeddings for application in geosciences: development, evaluation, and examples of soil-related concepts. Soil 5, 177–187. https://doi.org/10.5194/soil-5-177-2019.
Regione Emilia-Romagna, ENI, 1998. Riserve idriche sotterranee della Regione Emilia-Romagna. Selca, Firenze.
Regione Lombardia, 2016. PTUA 2016. Elaborato 2: caratterizzazione, monitoraggio e classificazione dei corpi idrici sotterranei [WWW Document]. URL ⟨https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/servizi-e-informazioni/Enti-e-Operatori/territorio/governo-delle-acque/piano-tutela-acque-pta/piano-tutela-acque-pta⟩.
Regione Lombardia, ENI, 2002. Geologia degli acquiferi padani della Regione Lombardia. SELCA, Firenze.
Saporetti, C.M., Goliatt, L., Pereira, E., 2021. Neural network boosted with differential evolution for lithology identification based on well logs information. Earth Sci. Inf. 14, 133–140. https://doi.org/10.1007/s12145-020-00533-x.
Scardia, G., Muttoni, G., Sciunnach, D., 2006. Subsurface magnetostratigraphy of Pleistocene sediments from the Po Plain (Italy): constraints on rates of sedimentation and rock uplift. Bull. Geol. Soc. Am. 118, 1299–1312. https://doi.org/10.1130/B25869.1.
Schorpp, L., Straubhaar, J., Renard, P., 2022. Automated hierarchical 3D modeling of quaternary aquifers: The ArchPy approach. Front Earth Sci. 10, 1–17. https://doi.org/10.3389/feart.2022.884075.
Servizio Geologico Nazionale, 1998. Carta idrogeologica regionale (Tavola II fuori testo), in: Memorie Descrittive Della Carta Geologica d'Italia. Vol. 56.
Slichter, C.S., 1899. Theoretical investigation of the motion of ground waters. The 19th Ann. Rep. US Geophys Survey. 304–319.
Sundermeyer, M., Schlüter, R., Ney, H., 2012. LSTM neural networks for language modeling. 13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012 1, 194–197. https://doi.org/10.21437/interspeech.2012-65.
Tilahun, T., Korus, J., 2023. 3D hydrostratigraphic and hydraulic conductivity modelling using supervised machine learning. Appl. Comput. Geosci. 19, 100122. https://doi.org/10.1016/j.acags.2023.100122.
Velasco, V., Cabello, P., Vázquez-Suñé, E., López-Blanco, M., Ramos, E., Tubau, I., 2012. A sequence stratigraphic based geological model for constraining hydrogeological modeling in the urbanized area of the Quaternary Besòs delta (NW Mediterranean coast, Spain). Geol. Acta 10, 373–393. https://doi.org/10.1344/105.000001757.
Vuković, M., Soro, A., 1992. Determination of hydraulic conductivity of porous media from grain-size composition. Water Resources Publications, Littleton, Colorado.
Wang, C., Ma, X., Chen, Jianguo, Chen, Jingwen, 2018. Information extraction and knowledge graph construction from geoscience literature. Comput. Geosci. 112, 112–120. https://doi.org/10.1016/j.cageo.2017.12.007.
Weissmann, G.S., Fogg, G.E., 1999. Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. J. Hydrol. (Amst.) 226, 48–65.
Zuffetti, C., Bersezio, R., 2021. Space–time geological model of the Quaternary syntectonic fill of a foreland basin (Po basin, Northern Italy). Sediment Geol. 421, 105945. https://doi.org/10.1016/j.sedgeo.2021.105945.
Zuffetti, C., Bersezio, R., Contini, D., Petrizzo, M.R., 2018. Geology of the San Colombano hill, a quaternary isolated tectonic relief in the Po Plain of Lombardy (Northern Italy). J. Maps 14, 199–211. https://doi.org/10.1080/17445647.2018.1443166.
Zuffetti, C., Comunian, A., Bersezio, R., Renard, P., 2020. A new perspective to model subsurface stratigraphy in alluvial hydrogeological basins, introducing geological hierarchy and relative chronology. Comput. Geosci. 140, 104506. https://doi.org/10.1016/j.cageo.2020.104506.