

Article

A K-Nearest Neighbors Algorithm in Python for Visualizing the 3D Stratigraphic Architecture of the Llobregat River Delta in NE Spain

Manuel Bullejos ¹ , David Cabezas ² , Manuel Martín-Martín ^{3,*}  and Francisco Javier Alcalá ^{4,5} 

¹ Departamento de Álgebra, University of Granada, 18010 Granada, Spain; bullejos@ugr.es

² Departamento de Análisis Matemático, University of Granada, 18010 Granada, Spain; dcabezas@ugr.es

³ Departamento de Ciencias de la Tierra y Medio Ambiente, University of Alicante, 03080 Alicante, Spain

⁴ Departamento de Desertificación y Geo-Ecología, Estación Experimental de Zonas Áridas (EEZA-CSIC), 04120 Almería, Spain; fjalcala@eeza.csic.es

⁵ Instituto de Ciencias Químicas Aplicadas, Facultad de Ingeniería, Universidad Autónoma de Chile, Santiago 7500138, Chile

* Correspondence: manuel.martin.m3@gmail.com

Abstract: The k-nearest neighbors (KNN) algorithm is a non-parametric supervised machine learning classifier; which uses proximity and similarity to make classifications or predictions about the grouping of an individual data point. This ability makes the KNN algorithm ideal for classifying datasets of geological variables and parameters prior to 3D visualization. This paper introduces a machine learning KNN algorithm and Python libraries for visualizing the 3D stratigraphic architecture of sedimentary porous media in the Quaternary onshore Llobregat River Delta (LRD) in northeastern Spain. A first HTML model showed a consecutive 5 m-equispaced set of horizontal sections of the granulometry classes created with the KNN algorithm from 0 to 120 m below sea level in the onshore LRD. A second HTML model showed the 3D mapping of the main Quaternary gravel and coarse sand sedimentary bodies (lithosomes) and the basement (Pliocene and older rocks) top surface created with Python libraries. These results reproduce well the complex sedimentary structure of the LRD reported in recent scientific publications and proves the suitability of the KNN algorithm and Python libraries for visualizing the 3D stratigraphic structure of sedimentary porous media, which is a crucial stage in making decisions in different environmental and economic geology disciplines.

Keywords: Llobregat River Delta; 3D stratigraphic architecture; granulometry; KNN algorithm; Python libraries



Citation: Bullejos, M.; Cabezas, D.; Martín-Martín, M.; Alcalá, F.J.

A K-Nearest Neighbors Algorithm in Python for Visualizing the 3D Stratigraphic Architecture of the Llobregat River Delta in NE Spain.

J. Mar. Sci. Eng. **2022**, *10*, 986.

<https://doi.org/10.3390/jmse10070986>

jmse10070986

Academic Editor: Niki Evelpidou

Received: 23 June 2022

Accepted: 17 July 2022

Published: 19 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Numerical modeling has expanded the classic qualitative geological visualizations toward strategies for delineating essential stratigraphic elements of sedimentary basins under quantitative criteria. Nowadays, numerical modeling tools allow interactive 3D visualizations while measuring the magnitude and uncertainty of the modeled geological variables and parameters [1–4]. As a result, modern 3D modeling has the ability to assimilate new data into the models as they are generated, thus allowing more real, accurate, and interactive visualizations than classic 2D representations. These features make the numerical modeling of great interest to take decisions in different environmental and economic geology disciplines.

There is a variety of applications for 3D visualization based on different interpolation algorithms and programming languages, including commercial software such as MOVE (Petroleum Experts Ltd., Edinburgh, UK), 3D Geomodeler (Intrepid Geophysics), Autocad Civil (Autodesk, Inc.), Gocad (Emerson Paradigm Roxar), ArcGis, PETREL (Geology and Modeling from Schlumberger), VOXI (Earth Modeling from Geosoft), and Geoscene 3D (I-GIS), as well as open source tools such as Gempy [5] and OSGeo [6]. In general, commercial

applications have friendly environments and technical support for users, but they are expensive. The advantage of open source applications is the zero cost and adaptability (modifying or extending the sources), but the absence of technical support for users and sometimes low reliability are its negative parts. The open source Python libraries [5–7] and posts listing libraries [8–10] devoted to Geographic Information Systems (GIS) and mapping are of special interest for visualizing geological structures and stratigraphic elements in the fields of mining, engineering, and hydrogeology. Some scientific documents develop or apply computer tools to different fields of geology [11–14], and some social media channels post data analytics and machine learning educational applications focused on geology [15].

The experience of the researchers' team of this paper with Python libraries for geological data handling and 3D visualization [16] was decisive in developing a new application for classifying and visualizing the 3D stratigraphic architecture of sedimentary porous media based on a machine learning KNN algorithm and Python libraries. This application uses (i) the KNN algorithm to create a consecutive 5 m-equispaced set of horizontal sections of the granulometry classes grouped as an interactive 3D HTML model; and (ii) different Python libraries to create an interactive 3D HTML model of the essential stratigraphic elements (coarse lithosomes and basement top surface) of the sedimentary basin. On the basis of the high density of boreholes and the subsequent geological knowledge gained during the last six decades, the Quaternary onshore Llobregat River Delta (LRD) near Barcelona city in northeastern Spain (Figure 1) was selected to show the application.



Figure 1. Geological sketch map of the onshore LRD area (green line contour), modified and simplified from Custodio [17], Medialdea and Solé-Sabarís [18], Alonso et al. [19], and Gámez et al. [20].

This paper uses the public granulometry dataset prepared by the Water Authority of Catalonia (Agència Catalana de l'Aigua, ACA) in the LRD region, which is available on request. A Jupyter notebook describing the data classification and 3D visualization, as well as an operative version of the Python code, can be downloaded from the GitHub repository described in Supplementary Materials. The HTML files do not require any additional tools apart from a web browser to view different perspectives, hide elements, enlarge or focus on specific areas or elements and take snapshots of a particular view. These files are included in Supplementary Materials.

2. Study Area

The LRD is a densely populated coastal plain of 98 km², forming the southwestern sector of the metropolitan area of Barcelona city in the Catalonia region in northeastern Spain (Figure 1). This area includes other minor-order cities such as El Prat de Llobregat, L'Hospitalet de Llobregat, Cornellà, Sant Boi, Viladecans, Gavà and Castelldefels. The water abundance and its strategic location near Barcelona city have favored the development of an important industrial activity since the XIX century. The high groundwater exploitation rates to supply to the increasing population and industrial activity have produced negative consequences on groundwater quantity and quality, including seawater intrusion into aquifers and high levels of pollution [17,21–24].

Other modern development milestones occurred in Barcelona city, and its metropolitan area, such as the Olympic Games in 1992 and the Llobregat Delta Infrastructure and Environment Plan (Pla d'Infraestructures i Medi Ambient del Delta del Llobregat, PDL) started in 1994 [25] have modified the LRD land use. The PDL included large civil infrastructures with variable underground development affecting groundwater bodies and posing at risk the water provision for different uses. In response, in 2004, ACA created the Technical Unit of the Llobregat Aquifers (Mesa Tècnica dels Aqüífers del Llobregat, METALL) to compile and homogenize the huge geological and hydrogeological information in order to prepare groundwater flow numerical models aimed at assessing the cumulative impact of the large civil works on the groundwater resource [26]. This public geological and hydrogeological database is available on request and has been used in this paper.

From a geological point of view, the LRD is regionally fed by the Llobregat River and its tributaries arriving from the Pre-Pyrenean Range and locally from the Llobregat River lower valley reliefs (Garraf and Collserola massifs) belonging to the Catalanian Coastal Range [23,24]. This range is a NE–SW-oriented mountain chain that gives pass downward to the Mediterranean coast (Figure 1). The LRD is also bounded toward the NE by the Montjuïc relief. The geological studies in the area started at the end of the XIX century by Almera [27]. In the previous century, several studies [18–20,28,29] allowed proposing geological maps and 2D cross-sections aimed to support hydrogeological studies. Sedimentological studies performed in the 1970s and 1980s [28–30] allowed clarifying the geology of the LRD. In the 1980s, the prodeltaic bodies of the emerged delta, dated as Holocene, were studied in detail [31] and the geological characterization of the continental shelf with support of 2D marine seismic reflection took place [32–34]. These studies, with a strong sedimentological component, allowed the sequential division of the LRD and the arrangement of the Quaternary materials. In the 1990s, this huge geological background was combined with the former geological information compiled from dozens of boreholes to make modern groundwater evaluations [35]. Coinciding with the PDL development at the beginning of the XXI century, new geological data allowed fine research [20,36–39] aimed to detail the 2D stratigraphic architecture of the LRD, thus giving a response to the Pliocene–Quaternary boundary, the confident definition of those coarse detritic levels officially cataloged as productive, high-yielding aquifers and the detailed identification of those interconnected stratigraphic structures (of sedimentary and tectonic origin) through which seawater intrusion into aquifers and mobilization of pollutants take place [22,23,26].

This area represents a Neogene rifted margin associated with the opening of the Valencia Trough and is affected by several fault families, probably active and mainly oriented NE–SW (Morrot and Tibidabo fault families) and NW–SE (Llobregat fault family), that conditioned the reliefs and the location of the Llobregat River outlet [18–30]. The Catalanian Coastal Range includes rocks of Paleozoic (granites and slates) and Mesozoic (Triassic conglomerates, sandstones, and pelites; Jurassic dolostones and limestones; Cretaceous marly limestones), whereas the Montjuïc Mount is made of Miocene calcarenites and marly limestones (Figure 1). Pliocene and older rocks are considered the basement, which is separated from the Quaternary formations by an important unconformity surface [20,38,39] (Figure 2). The Pliocene basement is made of estuarine marls, silts, and clays [18–20,38,39].

The Quaternary record was divided into two depositional sequences, the Pleistocene and Holocene ages [18–20,32,33,37,39]. The terms upper detrital complex and lower detrital complex have also been adopted in the scientific literature [20] for the same depositional sequences (Figure 2). According to geophysical studies in the offshore delta in the marine platform [32–34], the lower detrital complex can be divided into three parasequences [20]. In general terms, the lower detrital complex is made of conglomerate bodies (locally with sand) with intercalated silt- and clay-rich intervals [30–34]. The upper detrital complex, from bottom to top, is made of a sand layer, a silt bed, gravel (locally with sand), and upper silt and clay cover forming the current alluvial plain and the associated coastal wetlands and marshes [30,39]. The LRD coastal plain is also modeled by different streams coming from the neighbor reliefs [37–39] and the regional littoral drift, which distributes the shoreline sedimentation towards the SW [40,41].

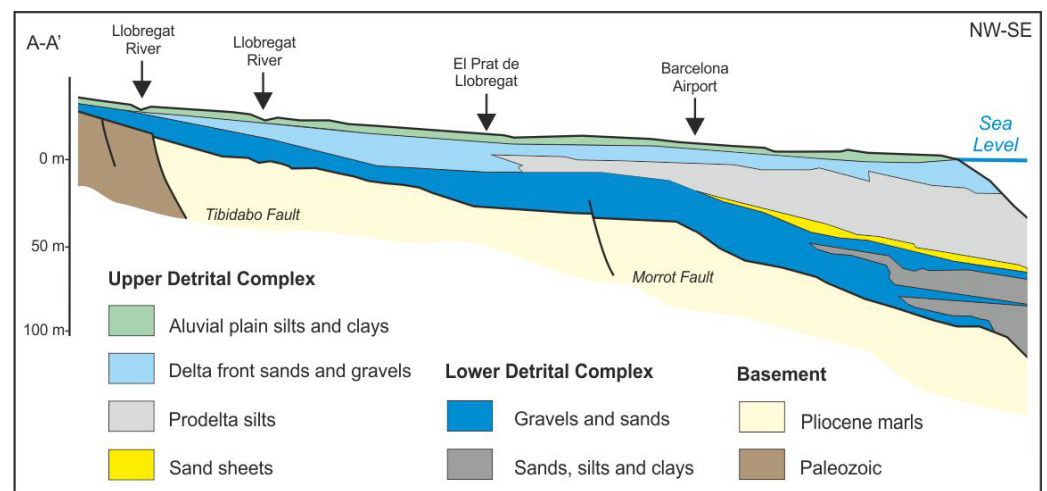


Figure 2. Geological sketch cross-section A-A' (NW-SE oriented) of the LRD located in Figure 1, modified and simplified from Medialdea and Solé-Sabarís [18], Marqués [30] and Simó et al. [39].

3. Methodology

3.1. Data Compilation

The public geological and hydrogeological database prepared by ACA, which is available on request, was consulted, and the granulometry dataset generated by METALL from both Lab test values and proxy values after visual recognition of the predominant lithologies in the LRD were compiled. The lithological records of 433 boreholes in the onshore LRD [42] and their granulometry records were checked in order to detect possible outliers. The detected outliers were suppressed or corrected when possible, thus providing the granulometry dataset used in this paper. The dataset consisted of georeferenced XLS (Excel) files with meter-by-meter granulometry values. The boreholes' location (coordinates x and y) and prospecting depth (coordinate z) lead to a georeferenced array of granulometry data associated with the prospected lithologies over space and depth. The georeferenced data were clustered into four main granulometry classes: clay–silt (<1 mm), coarse sand (1–5 mm), gravel (>5 mm), and basement. This working flow is synthesized in Figure 3A.

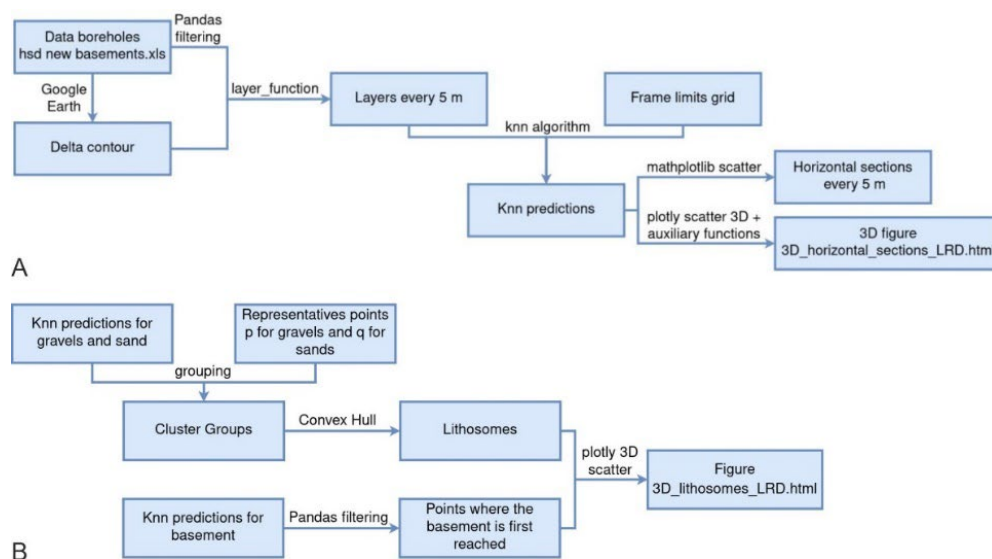


Figure 3. Flow diagram showing the methodological stages, including data compilation and software implementation for the 3D mapping of (A) the consecutive 5 m-equispaced set of horizontal sections of the granulometry classes created with the KNN algorithm from 0 to 120 m below sea level in the onshore LRD and (B) the stratigraphic architecture of the onshore LRD (coarse lithosomes and the basement top surface) created with Python libraries.

3.2. Python Programming Language

Python is an open-source language widely used in many environmental topics, including geological ones. The object-oriented, high-level Python programming language [43] was used for analyzing the granulometry data and visualizing the 3D HTML models of essential stratigraphic elements. From the many packages or modules that use this programming language for releasing a wide variety of problems, the Python packages used here were (i) NumPy [44] for data computing, (ii) Pandas [45] for data analysis and processing, (iii) Plotly [46] as a graphing library, (iv) Scipy [47] for interpolation and render algorithms, and (v) Scikit-learn [48] for the machine learning KNN algorithm. In addition, an inverse distance weighted interpolation algorithm obtained from the 3D Terrain Modelling in Python viewer from the GEODOSE block [49] was also used. This working flow is synthesized in Figure 3. The Jupyter notebooks describing the Python code and its explanatory instructions, as well as the operative HTML version of the code, can be downloaded from the GitHub repository described in Supplementary Materials. This HTML version can be opened and analyzed by using a web browser only, without the need to install Python. The Jupyter notebooks also explain step by step how to proceed to achieve our goals (a figure, a prediction, a cluster, etc.). If the readers want to adapt the code to their data or particular aims, they will need to know Python and might have to modify our auxiliary functions accordingly.

3.3. KNN Algorithm

The k-nearest neighbors (KNN) algorithm is a non-parametric supervised machine learning classifier that uses proximity and similarity of data to make classifications or predictions about the grouping of an individual data point [50–53]. This ability makes the KNN algorithm well suited for regression and classification problems of geological variables and parameters [51,52]. The KNN algorithm is used to classify the granulometry dataset prior to 3D visualization by assuming that the predicted value at a point is mostly dependent on the neighboring data. Under this same assumption of similarity, the KNN algorithm was used to infer data from a collection of measuring to a wider space of individuals, in this case, the nodes of a grid in the LRD.

Before using the KNN algorithm, the strategy for predicting new data attending to the environmental forces controlling the granulometry spatial distribution must be analyzed. The onshore LRD is formed of sub-horizontal layers of sedimentary material, thus determining more accurate 2D predictions than 3D ones. Since the target is 3D visualization, a consecutive 5 m-equispaced set of horizontal layers was adopted. In each horizontal layer, the KNN algorithm searches the K-nearest granulometry data and adopts the predominant value. A weight inversely proportional to the distance was assigned to each neighbor's data in order to prioritize the nearest ones. The rationale for choosing the parameter K considered that smaller K values might produce unrealistic polygonal regions while larger K values lead to more natural and smoother regions although they may ignore isolated data. This procedure follows the geological logic, thus avoiding problems with edge/boundary effects commonly seen in KNN. Once the granulometry dataset was carefully checked for possible outliers, we consider that every piece of data must be considered. So, we shall use $K = 1$, i.e., the KNN algorithm will only check the nearest neighbor to every point. This working flow is synthesized in Figure 3A. A Jupyter notebooks hosted in the GitHub repository described in Supplementary Materials include the code.

3.4. The 3D Mapping of the Granulometry Horizontal Sections

The Python module Pandas was used to read and process the boreholes' granulometry data XLS file. We used Google Earth to draw the LRD contour and the Python geometry function Polygon to create the LRD contour polygon. Then, we defined (i) the X and Y bounds and the grid size where the KNN algorithm will run, (ii) the exploring depth limit defined by the first prospection of the basement top surface, (iii) the adopted 5 m exploring depth interval, and (iv) the function 'layer_function' to classify the granulometry data attending to the above defined three classes and its depth.

Suitability of the 5 m equidistance of the horizontal KNN maps was based on (i) the LRD depth vs. length ratio since the Quaternary onshore LRD is about 120 m depth and its horizontal extension is about 15 km length; (ii) the borehole granulometry source data are arranged meter to meter; (iii) the sedimentary bodies interesting for us must have a certain mappable entity, for instance of 5 m thick at least; and (iv) there is a dense net of boreholes, but those are not closer of the order of the ten meters in the best of the cases.

Next, the KNN algorithm was executed, and the matplotlib function scatter used to create the 2D horizontal layers with predictions bounded within polygonal regions. The Python Plotly function scatter3d was used to integrate the 2D predictions into an interactive 3D HTML model ([3D_Horizontal_Sections_LRD.html](#) (accessed on 12 July 2022)). Some auxiliary functions to arrange data in a specific format were used. This working flow is synthesized in Figure 3A. The Jupyter notebooks hosted in the GitHub repository as Supplementary Materials include the code.

3.5. The 3D Mapping of the Stratigraphic Architecture and Basement Top Surface

This section describes the two steps followed to create the interactive 3D HTML model ([3D_Lithosomes_LRD.html](#) (accessed on 12 July 2022)) that allows a complete view of coarse lithosomes (gravel and coarse sand) of the onshore Quaternary LRD and the basement top surface.

Firstly, the spatial info about coarse granulometry classes clustered in every equispaced horizontal layer created with the above KNN algorithm was used to define the volume of coarse lithosomes in the incipient 3D HTML model. To this end, the function 'grouping' was the recursive cluster procedure used to group the points around a given start point selected in each cluster of points. The recursive nature of this function 'grouping' implies repeating the same calculation many times, but its definition is simple, and it does well with the nucleation strategy. Once the granulometry data of a lithosome were clustered, the Convex Hull algorithm developed by the SciPy community [54] was implemented. The 3D

convex hull of a georeferenced dataset is the smallest polyhedron that wraps them all. The convex hull function must be applied to each obtained group. The function 'lithosome' was used to calculate the corresponding Convex Hull. The output of the function 'lithosome' is a list of four elements (points, vertices, simplices, and name) defining the computed convex hull. At this stage, an overall checking was needed to ensure output is geologically suitable by removing points distorting the shape of groups or adding others with specific granulometry values in sparse data areas. Attending to the general decreasing data density with depth, the radio to get proper clusters was redefined, and the 'grid' parameter was decreased from 150 to 50 to reduce the computation time. Once all grid points are classified by their granulometry by using the KNN algorithm, they are grouped by lithosomes by means of the function 'grouping'. All clusters should be wrapped together by using their corresponding convex wrappings.

In a second step, the function `data_lithosome` we defined used the function `Mesh3d` by `plotly.graph_objects`. This allowed to shape the data in the proper drawing format, and the basement top surface was added to complete the interactive 3D HTML model. To this end, the basement KNN predictions were used, and the grid size decreased again up to 50 to reduce the computation time. The Pandas package used the KNN predicted points to calculate where the basement is firstly prospected (defining the coordinate z) and saved the output data as a CSV file. An Inverse Distance Weighting (IDW) interpolation algorithm was used to map the basement top surface. The mapped basement top surface was processed again to remove possible outliers prior to being merged with the above grouped coarse lithosomes. This working flow is synthesized in Figure 3B. Jupyter notebooks hosted in the GitHub repository as Supplementary Materials include the code.

4. Results

4.1. The 3D Mapping of the Granulometry Horizontal Sections

As described in Section 3.1, the public granulometry dataset produced by the ACA from 433 boreholes in the onshore LRD was clustered into the clay-silt (<1 mm), coarse sand (1–5 mm), gravel (>5 mm), and basement classes. The KNN algorithm was used to create the consecutive 5 m-equispaced set of horizontal layers of the granulometry classes from 0 to 120 m b.s.l. These horizontal layers used a regular 300 m × 300 m grid over the entire onshore LRD surface. The granulometry dataset at a given depth was used to produce the granulometry of the nodal grid points at that depth. Figure 4 shows six representative horizontal sections of the granulometry classes at different depths, in which the location of boreholes with granulometry data is also displayed. As shown, the data density (the number of boreholes) decreases with depth, so the accuracy of the KNN predictions also decreased with depth. However, how the mapping uncertainty of the spatial KNN predictions increases attending to the decreasing data density with depth was not evaluated because this task is out of the scope of this paper and the subject of ongoing research.

Figure 5 shows the interactive 3D HTML model created from the consecutive 5 m-equispaced set of horizontal layers of the granulometry classes created with the KNN algorithm from 0 to 120 m below sea level. An interactive 3D HTML version of this model ([3D_Horizontal_Sections_LRD.html](#) (accessed on 12 July 2022)) is provided in Supplementary Materials. The interactive 3D HTML model can be opened with any browser and allows observing different views, zooming, rotating, and moving around, as well as hiding elements by clicking in the legend to focus on details.

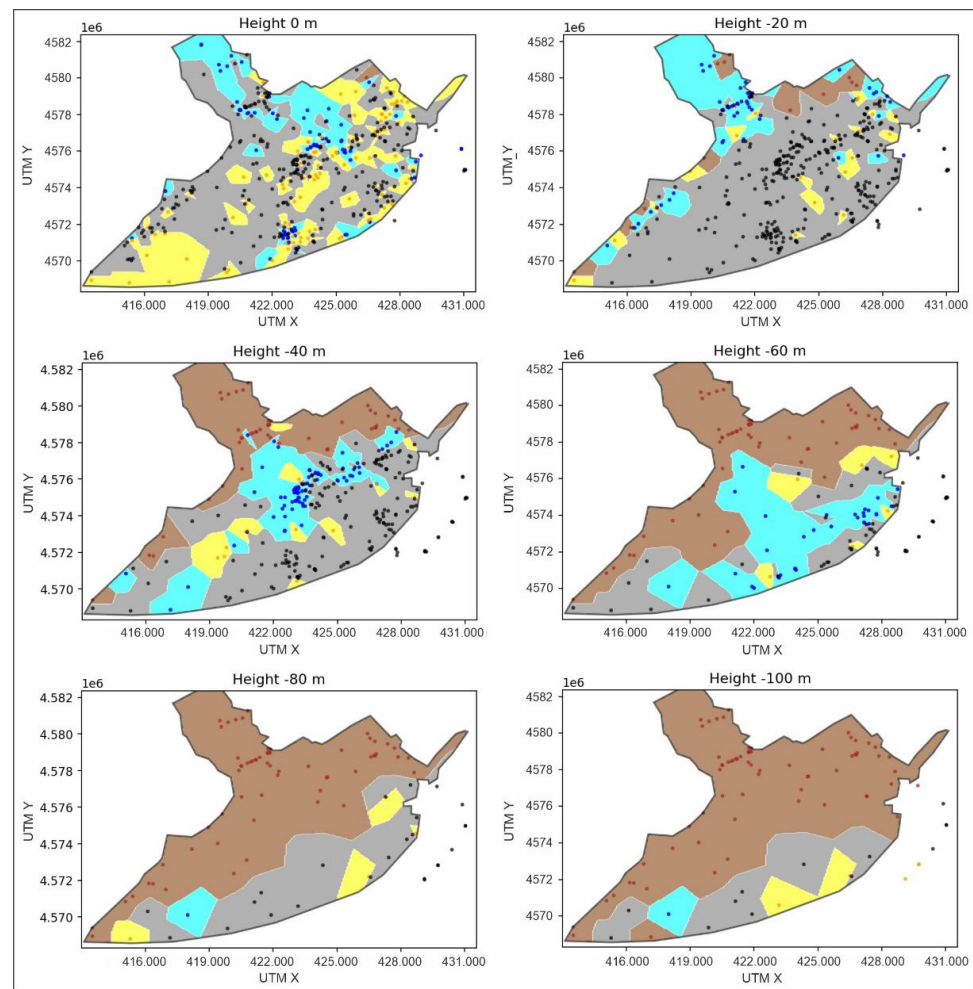


Figure 4. Some horizontal sections of the granulometry classes created with the KNN algorithm at 0, 20, 40, 60, 80, and 100 m below sea level in the onshore Quaternary LRD. The granulometry data were clustered into the gravel (>5 mm, cyan), coarse sand (1–5 mm, yellow), clay-silt (<1 mm, gray), and basement (brown) classes. The points display the boreholes with granulometry data at the modeled depth.

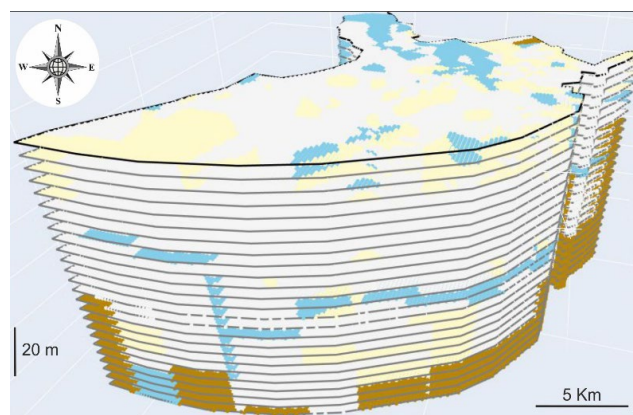


Figure 5. The consecutive 5 m-equispaced set of horizontal layers of the granulometry classes created with the KNN algorithm from 0 to 120 m b.s.l. in the onshore LRD. The color assigned to each granulometry class is cyan for gravel, yellow for coarse sand, light-grey for clay-silt and brown for the basement. An interactive 3D HTML version of this model is included in Supplementary Material ([3D_Horizontal_Sections_LRD.html](#) (accessed on 12 July 2022)).

4.2. The 3D Mapping of the Stratigraphic Architecture and Basement Top Surface

Figure 6 shows the interactive 3D HTML model created for visualizing the stratigraphic architecture (coarse lithosomes and the basement top surface) of the onshore Quaternary LRD, including some partial views relative to (i) coarse lithosomes and basement top surface (Figure 6A), (ii) gravel lithosome and basement top surface (Figure 6B), (iii) coarse sand lithosome and basement top surface (Figure 6C), and (iv) basement top surface only (Figure 6D). These partial views help to interpret the spatial distribution of lithosomes and the shape of the basement top surface. An interactive 3D HTML version of this model ([3D_Lithosomes_LRD.html](#) (accessed on 12 July 2022)) is included in Supplementary Materials. This interactive 3D HTML model can also be opened with any browser and allows observing different views, zooming, rotating, and moving around, as well as hiding elements by clicking in the legend to focus on details.

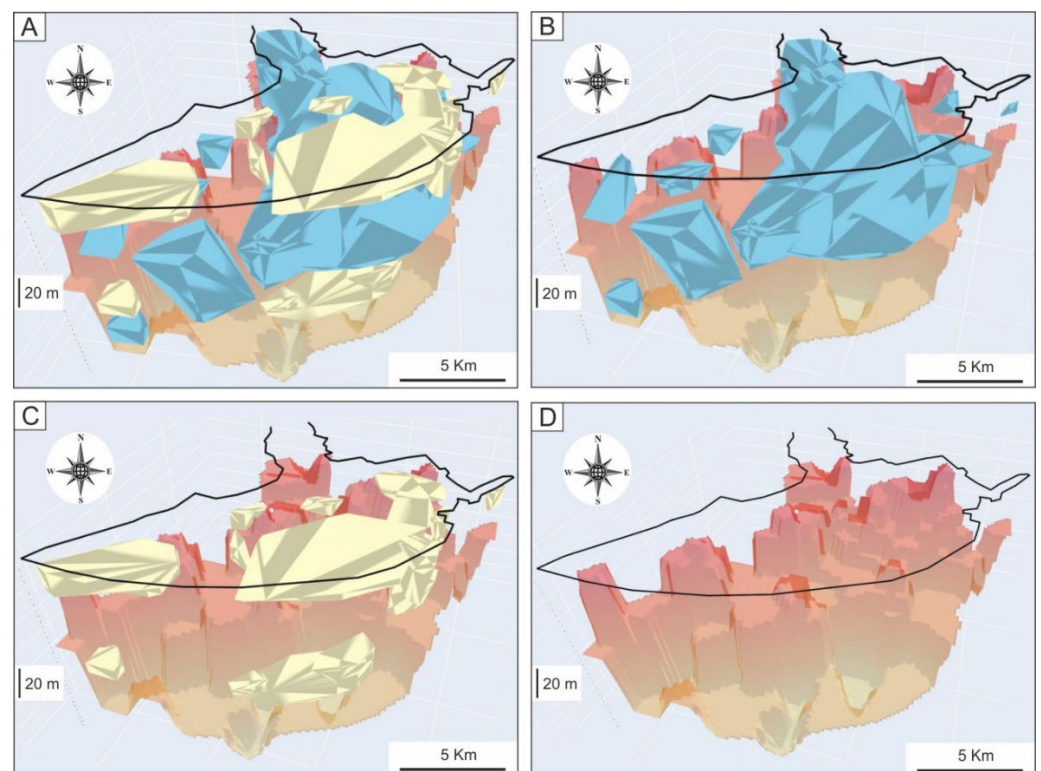


Figure 6. The 3D stratigraphic architecture (coarse lithosomes and the basement top surface (BTS)) of the onshore LRD. (A) Gravel and coarse sand lithosomes and BTS. (B) Gravel lithosomes and BTS. (C) Coarse sand lithosomes and BTS. (D) Basement top surface. The color assigned to each granulometry class is cyan for gravel, yellow for coarse sand, and reddish-brownish for the basement. An interactive 3D HTML version of this model is included in Supplementary Materials ([3D_Lithosomes_LRD.html](#) (accessed on 12 July 2022)).

There is a large, continuous, in time, gravel lithosome near the current Llobregat River course and other minor ones at different depths in the SW sector of the onshore Quaternary LRD (Figure 6B). There are also two extensive coarse sand lithosomes at different depths, the shallowest one being more important (Figure 6C). The basement top surface shows a general steeped shape deepening toward the marine platform with an over-imposed horst-graben structure, probably due to faulting (Figure 6C). When the sub-square raised and sunken sectors structure of the basement is compared with the geological sketch map from Figure 1, a clear correlation between the horst-graben boundaries and the main Tibidabo, Llobregat, and Morrot fault families and their associated minor-order faults is observed. These results reproduce well the complex sedimentary structure of the onshore LRD reported in recent scientific publications [20,37–39,55,56].

5. Discussion and Conclusions

This paper shows the suitability of the combined use of the KNN algorithm and Python libraries for classifying geological data and visualizing the 3D stratigraphic structure of sedimentary basins. Visualizing first the 3D stratigraphic architecture of sedimentary basins is crucial for making decisions in different environmental and economic geology disciplines, with clear applications in groundwater, oil, engineering, and mineral exploration, as well as in surveying a variety of physical and mechanical ground properties. In the onshore LRD, coarse (gravel and coarse sand) lithosomes storing groundwater are what matter, so this paper focused on them. However, this target may change in other areas with other lithosomes and other resources or physical and mechanical ground properties to be explored.

This paper also successfully meets the challenge of evolving from qualitative geological features subjected to interpretation toward quantitative geological data associated with physical parameters. This is a challenge in applied geology, which this research addresses by using parameter granulometry instead of lithological descriptions. The use of quantitative data allows successive numerical modeling exercises with assimilation of new data as they are generated to refine predictions progressively. The public granulometry dataset produced by ACA from 433 boreholes [10] in the onshore LRD was essential in conducting this research.

With regards to the KNN algorithm, it was used for classification purposes under the assumption that the granulometry value in a point must be similar to the measured magnitude in nearby locations. The KNN algorithm predicted the granulometry classes in a consecutive 5 m-equispaced set of horizontal sections created from 0 to 120 m below sea level in the onshore LRD. The created interactive 3D HTML model is included in Supplementary Materials ([3D_Horizontal_Sections_LRD.html](#) (accessed on 12 July 2022)). On the other hand, Python has a wide variety of libraries for data handling (such as Pandas and Numpy) and visualization (such as Matplotlib), including a friendly KNN interface available through the machine learning library scikit-learn to infer the magnitude of a given parameter or variable in points where data could not be collected. As deduced, Python is a suitable programming language for the modeling of geological parameters and variables, as the authors of this paper already proposed in previous research [16]. The created interactive 3D HTML model for visualizing the 3D stratigraphic architecture (coarse lithosomes and the basement top surface) of the LRD is included in Supplementary Materials ([3D_Lithosomes_LRD.html](#) (accessed on 12 July 2022)). In general, the results reproduce well the complex sedimentary structure of the LRD reported in recent scientific publications [20,37–39,55,56], thus proving the suitability of the open-source KNN algorithm and Python libraries for visualizing the 3D stratigraphic structure of sedimentary basins. Other proves of the 3D model reliability are as follows. We recently used an algorithm different from KNN to create other 3D models of the LRD; the results being quite similar to these [16]. Moreover, we have determined the horizontal and vertical reliability of our 3D model. At the LRD scale, the lower reliability of the 3D model was tentatively evaluated as 25% of the estimated value, whereas reliability at a small scale of observation was a function of distance and magnitude of granulometry data among nearby points reaching $\pm 50\%$ in the worst cases. The specific methodology to evaluate the 3D model reliability will be the subject of a future publication.

Although there are many commercial and open-source libraries and software devoted to geological visualization, its use requires money in the case of the commercial ones and effort to learn how they work. The open-source KNN algorithm and Python libraries have proven efficient, accessible, and friendly for the purpose of visualizing the 3D stratigraphic architecture of sedimentary basins. Currently, additional research to develop routines aimed at assessing the mapping uncertainty is ongoing.

The introduced Python code can be implemented in other areas with different geological features to model other variables and parameters. Here, the introduced application uses granulometry for visualizing coarse lithosomes of the LRD, which is a sedimentary porous media with huge information about this parameter. In this and other cases, the

accuracy of predictions depends on available data density, as expected. In other sparse data areas, the readers can use widely-known empirical relationships to indirectly infer granulometry from physical parameters deduced from hydraulic tests such as pumping tests, geotechnical tests such as Lefranc, Digital Image Analysis of ground textures, or pore-water findings from geophysical surveys such as Electrical Resistivity Tomography and Ground Penetrating Radar techniques [57–63]. In these cases, a representative enough granulometry dataset from each lithological class is needed to ensure the statistical representativeness of the measured vs. deduced data pairs experimental relationships. In this paper, we used granulometry as a physical parameter, but other physical and chemical variables and parameters such as density, porosity, hydraulic conductivity, thermal conductivity, distribution of a pollutant, and occurrence of minerals, among most others, can also be used. This widens the scope of the application introduced in this paper to many environmental and/or economic interests.

Supplementary Materials: The two interactive 3D HTML models [3D_Horizontal_Sections_LRD.html](#) (accessed on 12 July 2022) and [3D_Lithosomes_LRD.html](#) (accessed on 12 July 2022) can be downloaded at: <https://www.mdpi.com/article/10.3390/jmse10070986/s1>. The Python code, the KNN algorithm, and the detailed instructions on how to download and run the code are hosted in a GitHub repository and can be downloaded at: <https://github.com/dcabezas98/knn-stratigraphic-visualization> (accessed on 12 July 2022).

Author Contributions: M.B., D.C., M.M.-M. and F.J.A. contributed to the conceptualization, methodology, formal analysis, data curation, writing, and review of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: Research Project PID2020-114381GB-I00 of the Spanish Ministry of Science and Innovation, Research Groups and Projects of the Generalitat Valenciana from the University of Alicante (CTMAGA), and Research Groups FQM-343 and RNM-188 of the Junta de Andalucía.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors are grateful to the administrative and technical staff of the Water Authority of Catalonia for accessing the public borehole and granulometry databases from the Llobregat River Delta. Two anonymous reviewers are also acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jessell, M. Three-dimensional geological modelling of potential-field data. *Comput. Geosci.* **2001**, *27*, 455–465. [CrossRef]
2. Wycisk, P.; Hubert, T.; Gossel, W.; Neumann, C. High-resolution 3D spatial modelling of complex geological structures for an environmental risk assessment of abundant mining and industrial megasites. *Comput. Geosci.* **2009**, *35*, 165–182. [CrossRef]
3. Ford, J.; Mathers, S.; Royse, K.; Aldiss, D.; Morgan, D.J.R. Geological 3D modelling: Scientific discovery and enhanced understanding of the subsurface, with examples from the UK. *Z. Dtsch. Ges. Geowiss.* **2010**, *161*, 205–218. [CrossRef]
4. Rohmer, O.; Bertrand, E.; Mercerat, E.D.; Régnier, J.; Pernoud, M.; Langlaude, P.; Alvarez, M. Combining borehole log-stratigraphies and ambient vibration data to build a 3D Model of the Lower Var Valley, Nice (France). *Eng. Geol.* **2020**, *270*, 105588. [CrossRef]
5. GemPy: Open-Source 3D Geological Modeling. Available online: <https://www.gempy.org> (accessed on 9 June 2022).
6. OSGeo: The Open Source Geospatial Foundation. Available online: <https://www.osgeo.org/> (accessed on 9 June 2022).
7. GeoPandas. Available online: <https://geopandas.org/en/stable> (accessed on 9 June 2022).
8. Albion: 3D Geological Models in QGIS. Available online: <https://gitlab.com/Oslandia/albion> (accessed on 9 June 2022).
9. GISgeography. 15 Python Libraries for GIS and Mapping. Available online: <https://gisgeography.com/python-libraries-gis-mapping> (accessed on 9 June 2022).
10. Parpoil, B. Open Source and Geology. Available online: <https://oslandia.com/en/2020/07/09/geologie-open-source> (accessed on 9 June 2022).
11. Hobona, G.; James, P.; Fairbairn, D. Web-based visualization of 3D geospatial data using Java3D. *IEEE Comput. Graph. Appl.* **2006**, *26*, 28–33. Available online: <https://ieeexplore.ieee.org/document/1652923> (accessed on 12 July 2022). [CrossRef]

12. Evangelidis, K.; Papadopoulos, T.; Papatheodorou, K.; Mastorokostas, P.; Hilas, C. 3D geospatial visualizations: Animation and motion effects on spatial objects. *Comput. Geosci.* **2018**, *111*, 200–212. [[CrossRef](#)]
13. Semmo, A.; Trapp, M.; Jobst, M.; Doellner, J. Cartography-oriented design of 3D geospatial information visualization—overview and techniques. *Cartogr. J.* **2015**, *52*, 95–106. [[CrossRef](#)]
14. Miao, R.; Song, J.; Zhu, Y. 3D Geographic Scenes Visualization Based on WebGL. In Proceedings of the 6th International Conference on Agro-Geoinformatics, Fairfax, VA, USA, 7–10 August 2017; IEEE: Fairfax, VA, USA, 2017; Volume 1, pp. 1–6. Available online: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8046999> (accessed on 9 June 2022).
15. Pycrz, M. GeostatsGuy Lectures. Available online: <https://www.youtube.com/c/GeostatsGuyLectures> (accessed on 9 June 2022).
16. Bullejos, M.; Cabezas, D.; Martín-Martín, M.; Alcalá, F.J. A Python Application for Visualizing the 3D Stratigraphic Architecture of the Onshore Llobregat River Delta in NE Spain. *Water* **2022**, *14*, 1882. Available online: <https://www.mdpi.com/2073-4441/14/12/1882> (accessed on 12 July 2022). [[CrossRef](#)]
17. Custodio, E. Seawater intrusion in the Llobregat Delta near Barcelona (Catalonia, Spain). In *Groundwater Problems in the Coastal Areas, Studies and Reports in Hydrology*; UNESCO: Paris, France, 1987; Volume 45, pp. 436–463.
18. Medialdea, J.; Solé-Sabaris, L. Geological Map of Spain, Scale 1:50,000, Sheet n° 448. In *El Prat de Llobregat, Memory and Maps*; Geological Survey of Spain: Madrid, Spain, 1991. Available online: <http://info.igme.es/cartografiadigital/geologica/Magna50Hoja.aspx?language=es&id=448> (accessed on 18 April 2022).
19. Alonso, F.; Peón, A.; Rosell, J.; Arrufat, J.; Obrador, A. Geological Map of Spain, Scale 1:50,000, Sheet n° 421. In *Barcelona, Memory and Maps*; Geological Survey of Spain: Madrid, Spain, 1974. Available online: <http://info.igme.es/cartografiadigital/geologica/Magna50Hoja.aspx?language=es&id=421> (accessed on 18 April 2022).
20. Gámez, D.; Simó, J.A.; Lobo, F.J.; Barnolas, A.; Carrera, J.; Vázquez-Suñé, E. Onshore–offshore correlation of the Llobregat deltaic system, Spain: Development of deltaic geometries under different relative sea-level and growth fault influences. *Sediment. Geol.* **2009**, *217*, 65–84. [[CrossRef](#)]
21. Almera, J. Mapa Geológico y Topográfico De La Provincia De Barcelona: Región Primera o De Contornos de la Capital Detallada, Scale 1:40,000, Memory and Maps, Diputación de Barcelona, Barcelona. 1891. Available online: <https://cartotecadigital.icgc.cat/digital/collection/catalunya/id/2174> (accessed on 18 April 2022).
22. Abarca, E.; Vázquez-Suñé, E.; Carrera, J.; Capino, B.; Gámez, D.; Batlle, F. Optimal design of measures to correct seawater intrusion. *Water Resour. Res.* **2006**, *42*, W09415. [[CrossRef](#)]
23. Vázquez-Suñé, E.; Abarca, E.; Carrera, J.; Capino, B.; Gámez, D.; Pool, M.; Simó, T.; Batlle, F.; Niñerola, J.M.; Ibáñez, X. Groundwater modelling as a tool for the European Water Framework Directive (WFD) application: The Llobregat case. *Phys. Chem. Earth* **2006**, *31*, 1015–1029. [[CrossRef](#)]
24. Postigo, C.; Ginebreda, A.; Barbieri, M.V.; Barceló, D.; Martín-Alonso, J.; de la Cal, A.; Boleda, M.R.; Otero, N.; Carrey, R.; Solà, V.; et al. Investigative monitoring of pesticide and nitrogen pollution sources in a complex multi-stressed catchment: The lower Llobregat River basin case study (Barcelona, Spain). *Sci. Total Environ.* **2021**, *755*, 142377. [[CrossRef](#)] [[PubMed](#)]
25. Resolution 12956/1994. Cooperation Agreement on Infrastructure and Environment in the Llobregat Delta. In *Official Journal of Spain*; Ministry of Public Works, Transports and Environment: Madrid, Spain; Government of Spain: Madrid, Spain, 1994. Available online: https://www.boe.es/diario_boe/txt.php?id=BOE-A-1994-12956 (accessed on 18 April 2022).
26. Official Statement. The Water Authority of Catalonia Creates the Technical Unit of the Llobregat Aquifers. In *Official Journal of Catalonia*; Department of the Environment and Housing, Government of Catalonia: Barcelona, Spain, 2004. Available online: <https://govern.cat/salaprensa/notes-premsa/68710/agencia-catalana-aigua-crea-mesa-tecnica-dels-aqueifers-del-llobregat> (accessed on 18 April 2022).
27. Medialdea, J.; Solé-Sabaris, L. Geological Map of Spain, Scale 1:50,000, Sheet n° 420. In *Hospitalet de Llobregat, Memory and Maps*; Geological Survey of Spain: Madrid, Spain, 1973. Available online: <http://info.igme.es/cartografiadigital/geologica/Magna50Hoja.aspx?language=es&id=420> (accessed on 18 April 2022).
28. Llopis, N. Tectomorfológia del Macizo del Tibidabo y valle inferior del Llobregat. *Estud. Geogr.* **1942**, *3*, 321–383.
29. Solé-Sabaris, L. Ensayo de interpretación del Cuaternario Barcelonés. *Misc. Barcinonensia* **1963**, *2*, 7–54.
30. Marqués, M.A. *Les Formacions Quaternàries del Delta del Llobregat*; Institut d’Estudis Catalans: Barcelona, Spain, 1984.
31. Manzano, M. Estudio Sedimentológico del Prodelta Holoceno del Llobregat. Master’s Thesis, University of Barcelona, Barcelona, Spain, 1986.
32. IGME. Geological Map of the Spanish Continental Shelf and Adjacent Areas, Scale 1:200,000, Sheet n° 42E. In *Barcelona, Memory and Maps*; Geological Survey of Spain: Madrid, Spain, 1989. Available online: <https://info.igme.es/cartografiadigital/tematica/Fomar200Hoja.aspx?language=es&id=42E> (accessed on 18 April 2022).
33. IGME. Geological Map of the Spanish Continental Shelf and Adjacent Areas, Scale 1:200,000, Sheet n° 42. In *Tarragona, Memory and Maps*; Geological Survey of Spain: Madrid, Spain, 1986. Available online: <https://info.igme.es/cartografiadigital/tematica/Fomar200Hoja.aspx?language=es&id=42> (accessed on 18 April 2022).
34. Serra, J.; Verdaguer, A. La Plataforma Holocena en el Prodelta del Llobregat. In *X Congreso Nacional de Sedimentología*; Obrador, A., Ed.; University of Barcelona: Barcelona, Spain, 1983; Volume 2, pp. 49–51.
35. Iribar, V.; Carrera, J.; Custodio, E.; Medina, A. Inverse modelling of seawater intrusion in the Llobregat delta deep aquifer. *J. Hydrol.* **1997**, *198*, 226–247. [[CrossRef](#)]

36. Alcalá-García, F.J.; Miró, J.; García-Ruz, A. Sobre la intrusión marina en el sector oriental del acuífero profundo del delta del Llobregat (Barcelona, España). Breve descripción histórica y evolución actual. *Bol. Real Soc. Española Hist. Nat.* **2002**, *97*, 42–49.
37. Alcalá-García, F.J.; Miró, J.; Rodríguez, P.; Rojas-Martín, I.; Martín-Martín, M. Actualización Geológica del Delta del Llobregat (Barcelona, España). Implicaciones Geológicas e Hidrogeológicas. In *Tecnología de la Intrusión de Agua de Mar en Acuíferos Costeros: Países Mediterráneos*; López-Geta, J.A., de la Orden, J.A., Gómez, J.D., Ramos, G., Mejías, M., Rodríguez, L., Eds.; Geological Survey of Spain: Madrid, Spain, 2003; Volume 1, pp. 45–52.
38. Alcalá-García, F.J.; Miró, J.; Rodríguez, P.; Rojas-Martín, I.; Martín-Martín, M. Características estructurales y estratigráficas del substrato Plioceno del Delta de Llobregat (Barcelona, España)—Aplicación a los estudios hidrogeológicos. *Rev. Geotemas* **2003**, *5*, 23–26.
39. Simó, J.A.; Gámez, D.; Salvany, J.M.; Vázquez-Suñé, E.; Carrera, J.; Barnolas, A.; Alcalá, F.J. Arquitectura de facies de los deltas cuaternarios del río Llobregat, Barcelona, España. *Geogaceta* **2005**, *38*, 171–174.
40. Font, J.; Julia, A.; Rovira, J.; Salat, J.; Sanchez-Pardo, J. Circulación marina en la plataforma continental del Ebro determinada a partir de la distribución de masas de agua y los microcontaminantes orgánicos en el sedimento. *Acta Geol. Hisp.* **1987**, *21*, 483–489.
41. Chiocci, F.L.; Ercilla, G.; Torres, J. Stratal architecture of Western Mediterranean Margins as the result of the stacking of Quaternary lowstand deposits below ‘glacio-eustatic fluctuation base-level’. *Sediment. Geol.* **1997**, *112*, 195–217. [[CrossRef](#)]
42. Alcalá, F.J.; Martín-Martín, M.; García-Ruz, A. A lithology database from historical 457 boreholes in the Llobregat River Delta aquifers in northeastern Spain. *Figshare Dataset* **2020**. [[CrossRef](#)]
43. Python Programming Language. Available online: <https://www.python.org> (accessed on 9 June 2022).
44. Numpy. Available online: <https://numpy.org> (accessed on 13 June 2022).
45. Pandas. Available online: <https://pandas.pydata.org/> (accessed on 13 June 2022).
46. Plotly. Available online: <https://plotly.com> (accessed on 9 June 2022).
47. Scipy. Available online: <https://scipy.org> (accessed on 13 June 2022).
48. Scikit-learn. Available online: <https://scikit-learn.org/stable/install.html#installation-instructions> (accessed on 13 June 2022).
49. GEODOSE. Available online: <https://www.geodose.com/2019/09/3d-terrain-modelling-in-python.html> (accessed on 13 June 2022).
50. Gou, J.; Ma, H.; Ou, W.; Zeng, S.; Rao, Y.; Yang, H. A generalized mean distance-based k-nearest neighbor classifier. *Expert Syst. Appl.* **2019**, *115*, 356–372. [[CrossRef](#)]
51. Pratama, H. Machine Learning: Using Optimized KNN (K-Nearest Neighbors) to Predict the Facies Classifications. In Proceedings of the 13th SEGJ International Symposium, Tokyo, Japan, 12–14 November 2018; Society of Exploration Geophysicists of Japan: Tokyo, Japan, 2018; Volume 1, pp. 538–541. [[CrossRef](#)]
52. Wang, X.; Yang, S.; Zhao, Y.; Wang, Y. Lithology identification using an optimized KNN clustering method based on entropy-weighted co-sine distance in Mesozoic strata of Gaoqing field, Jiyang depression. *J. Pet. Sci. Eng.* **2018**, *166*, 157–174. [[CrossRef](#)]
53. Huang, S.; Huang, M.; Lyu, Y. An Improved KNN-Based Slope Stability Prediction Model. *Adv. Civ. Eng.* **2020**, *2020*, 8894109. [[CrossRef](#)]
54. Convex Hull Algorithm. Available online: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.ConvexHull.html> (accessed on 9 June 2022).
55. Parcerisa, D.; Gámez, D.; Gómez-Gras, D.; Usera, J.; Simó, J.A.; Carrera, J. Estratigrafía y petrología del subsuelo precuaternario del sector SW de la depresión de Barcelona (Cadenas Costeras Catalanas, NE de Iberia). *Rev. Soc. Geol. España* **2008**, *21*, 93–109.
56. Salvany, J.M.; Aguirre, J. The Neogene and Quaternary deposits of the Barcelona city through the high-speed train line. *Geol. Acta* **2020**, *18*, 1–19. [[CrossRef](#)]
57. Payton, R.L.; Chiarella, D.; Kingdon, A. The influence of grain shape and size on the relationship between porosity and permeability in sandstone: A digital approach. *Sci. Rep.* **2022**, *12*, 7531. [[CrossRef](#)]
58. Boadu, F.K. Hydraulic conductivity of soils from grain-size distribution: New models. *J. Geotech. Geoenviron. Eng.* **2000**, *126*, 739–746. [[CrossRef](#)]
59. Torskaya, T.; Shabro, V.; Torres-Verdín, C.; Salazar-Tio, R.; Revil, A. Grain shape effects on permeability, formation factor, and capillary pressure from pore-scale modeling. *Transp. Porous Media* **2014**, *102*, 71–90. [[CrossRef](#)]
60. Nabawy, B.S. Estimating porosity and permeability using Digital Image Analysis (DIA) technique for highly porous sandstones. *Arab. J. Geosci.* **2014**, *7*, 889–898. [[CrossRef](#)]
61. De Lima, O.A.; Sri, N. Estimation of hydraulic parameters of shaly sandstone aquifers from geoelectrical measurements. *J. Hydrol.* **2000**, *235*, 12–26. [[CrossRef](#)]
62. Paz, C.; Alcalá, F.J.; Carvalho, J.M.; Ribeiro, L. Current uses of ground penetrating radar in groundwater-dependent ecosystems research. *Sci. Total Environ.* **2017**, *595*, 868–885. [[CrossRef](#)] [[PubMed](#)]
63. Paz, C.; Alcalá, F.J.; Ribeiro, L. Ground penetrating radar attenuation expressions in shallow groundwater research. *J. Environ. Eng. Geophys.* **2020**, *25*, 153–160. [[CrossRef](#)]