Article

Confidence of a k-Nearest Neighbors Python Algorithm for the 3D Visualization of Sedimentary Porous Media

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
2 Departamento de Análisis Matemático, University of Granada, 18010 Granada, Spain
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 Almeria, Spain
5 Instituto de Ciencias Químicas Aplicadas, Facultad de Ingeniería, Universidad Autónoma de Chile, Santiago 7500138, Chile
* Correspondence: manuel.martin@ua.es

Abstract: In a previous paper, the authors implemented a machine learning k-nearest neighbors (KNN) algorithm and Python libraries to create two 3D interactive models of the stratigraphic architecture of the Quaternary onshore Llobregat River Delta (NE Spain) for groundwater exploration purposes. The main limitation of this previous paper was its lack of routines for evaluating the confidence of the 3D models. Building from the previous paper, this paper refines the programming code and introduces an additional algorithm to evaluate the confidence of the KNN predictions. A variant of the Similarity Ratio method was used to quantify the KNN prediction confidence. This variant used weights that were inversely proportional to the distance between each grain-size class and the inferred point to work out a value that played the role of similarity. While the KNN algorithm and Python libraries demonstrated their efficacy for obtaining 3D models of the stratigraphic arrangement of sedimentary porous media, the KNN prediction confidence verified the certainty of the 3D models. In the Llobregat River Delta, the KNN prediction confidence at each prospecting depth was a function of the available data density at that depth. As expected, the KNN prediction confidence decreased according to the decreasing data density at lower depths. The obtained average-weighted confidence was in the $0.44-0.53$ range for gravel bodies at prospecting depths in the $12.7-72.4$ m b.s.l. range and was in the $0.42-0.55$ range for coarse sand bodies at prospecting depths in the $4.6-83.9$ m b.s.l. range. In a couple of cases, spurious average-weighted confidences of 0.29 in one gravel body and 0.30 in one coarse sand body were obtained. These figures were interpreted as the result of the quite different weights of neighbors from different grain-size classes at short distances. The KNN algorithm confidence has proven its suitability for identifying these anomalous results in the supposedly well-depurated grain-size database used in this study. The introduced KNN algorithm confidence quantifies the reliability of the 3D interactive models, which is a necessary stage to make decisions in economic and environmental geology. In the Llobregat River Delta, this quantification clearly improves groundwater exploration predictability.

Keywords: KNN algorithm; confidence degree; data classification; python libraries; 3D stratigraphic architecture; Llobregat River Delta

1. Introduction

Numerical modeling is increasingly replacing the classic qualitative geological methods of data representation and mapping. Quantitative criteria are increasingly used for the 3D mapping of stratigraphic bodies of sedimentary basins. In fact, numerical tools allow interactive 3D visualizations while providing a measure of the accuracy of the mapped parameters and variables [1–4]. The result is more real, accurate interactive visualizations than classic 2D mapping since modern 3D visualizations can assimilate new data into the

prototypes during the modeling stages. These qualities make numerical modeling very interesting for decision making in diverse applied geology functionalities.

Currently, a wide number of modeling tools for 3D visualizations exist. They use diverse programming languages and interpolation algorithms, including open-source software such as Gempy [5] and OSGeo [6] as well as commercial tools such as the 3D Geomodeller (Intrepid Geophysics), MOVE (Petroleum Experts Ltd., Edinburg, UK), Autocad Civil (Autodesk, Inc., San Rafael, CA, USA), Gocad (Emerson Paradigm Roxar), ArcGis, VOXI (Earth Modeling from Geosoft), PETREL (Geology and Modeling from Schlumberger) and Geoscene 3D (I-GIS). Usually, commercial tools have friendly environments and technical support for users, but they can be expensive. While the adaptability (extending or updating the source) and zero cost of open-source tools make them advantageous, their disadvantages include their lack of technical support for users and the fact that they are sometimes unreliable. 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 diverse applied geology fields, such as those devoted to groundwater, mining and geotechnical exploration. The Python programming language is increasingly being used in a wide range of scientific documents to create computer tools that are applicable to different fields of geology [11–15]. The scientific relevance is high, as demonstrated by the existence of some social media channels devoted to post-machine learning educational routines aimed at geological purposes [16].

The experience gained by the researchers responsible for this work with Python libraries for 3D visualization and geological data handling was essential for developing new applications devoted to data classification and the 3D visualization of the stratigraphic architecture of sedimentary bodies (essentially porous media) [17,18]. In the latest application, Bullejos et al. [18] used (i) a machine learning KNN algorithm to produce an interactive 3D model based on a set of horizontal (five-meter-equispaced) sections of the grain-size classes and (ii) used some Python libraries to produce interactive 3D models of the stratigraphic arrangement of the essential sedimentary bodies. Because of the high density of boreholes and the subsequent geological knowledge gained during the last six decades, the Quaternary onshore Llobregat River Delta (LRD) groundwater body near Barcelona city in northeastern Spain was selected to run the application. To design and run these models, the public grain-size database created by the Water Authority of Catalonia (Agència Catalana de l'Aigua) for groundwater purposes in the study area, which is available on request, was used. Grain size (or granulometry) is a physical parameter ideal for classification purposes, which in turn determines the hydraulic and geotechnical behavior of sedimentary formations cataloged as aquifers.

The application of Bullejos et al. [18] included Jupyter notebooks describing the methodology and a version of the Python code, which is downloadable from the GitHub repository (https://github.com/dcabezas98/knn-stratigraphic-visualization, accessed on 17 November 2022). A web browser can be used to open the created 3D models, take snapshots of a particular view, hide elements, view different perspectives, and enlarge or focus on a single element or specific area. The main limitation of this application was that it did not include specific routines for quantifying the confidence of the interactive 3D models. This limitation affects the reliability of the KNN predictions and should be amended for better decision making in applied geology. This paper introduces a metric inspired by the Similarity Ratio [19–21] to quantify the confidence of the interactive 3D models created by Bullejos et al. [18]. The Similarity Ratio is a common confidence metric in the machine-learning KNN literature [19,20]. The implemented confidence metric assigns weights instead of similarity since the similarity condition is equivalent to assigning weights inversely proportional to the distance between each grain-size class.

2. Study Area

The Llobregat River Delta (LRD) (Figure 1) is a coastal plain of 98 km^2 in the SW sector of the metropolitan area of Barcelona (Catalonia region, NE Spain) (Figure 1A). This populated area includes other smaller cities such (L'Hospitalet de Llobregat, Cornellà, Viladecans, Sant Boi, Gavà, Castelldefels and El Prat de Llobregat). Since the XIX century, Barcelona city's location and its abundance of water resources have made it favorable for the implantation of many industries in the LRD. As a consequence, groundwater has been exploited highly to satisfy the agriculture, industry and population water demands. This exploitation has reduced the groundwater quantity and damaged the quality due to seawater intrusion and contaminant leachates [22–25]. LRD land use has also been subjected to noticeable changes over the last few decades, such as those brought about by the Barcelona Olympic Games and the Llobregat Delta Infrastructure and Environment Plan (Pla d'Infraestructures i Medi Ambient del Delta del Llobregat: LRD Infrastructure Plan) [26]. In the case of the LRD Infrastructure Plan, the construction of large civil infrastructures with different underground development added stress to the groundwater resource. As a result, the Water Authority of Catalonia created the Technical Unit of the Llobregat Aquifers (Mesa Tècnica dels Aqüífers del Llobregat) in 2004 to gather and systematize the widely available hydrogeological and geological information [27]. The purpose was to evaluate the hydrogeological risk of these civil works. This public geological and hydrogeological database was considered in this work.

The LRD is mainly nourished by the Llobregat River and from its tributaries to a lesser extent. Sediments mostly come from the Pre-Pyrenean Range. Other supplies come from the Garraf and Collserola massifs (Catalonian Coastal Range) in the lower part of the Llobregat River valley [28,29]. The LRD is NW bounded by the Catalonian Coastal Range, SE bounded by the Mediterranean Sea, and NE bounded by the Montjuïc relief (Figure 1A). At the end of the XIX century, the first geological studies in LRD started [30]. Geological studies in the second half of the XX century [28,29,31–33] defined the geological basis of the first hydrogeological evaluations. A number of sedimentological studies aimed at clarifying the stratigraphic–sedimentological architecture of the LRD were also made during this period [32–34]. Additionally, the prodeltaic bodies of the emerged LRD delta were classified as Holocene bodies [35], marine seismic reflections were used to elucidate the continental shelf geology [36–38], and the internal division and sequential structure of the Quaternary LRD were defined. Modern groundwater evaluations [39] were performed in the nineties by combining the geological information generated in the LRD. In the first decade of the XXI century, new geological data in the framework of the LRD Infrastructure Plan revealed detailed findings [40–44] about the Pliocene–Quaternary boundary, the geometry of coarse-grained bodies defining the productive aquifers, and the sedimentary and tectonic structures that enable seawater intrusion and the transport of contaminants [23,24,40].

The above geological background revealed how this area consists of a Neogene rifted margin contemporaneous to the Valencia Trough opening, where a number of probably active fault families exist. The most important faults are the NE–SW oriented Tibidabo and Morrot faults and the NW–SE oriented Llobregat fault. These fault families control the position of the Llobregat River outlet and the local orography [28–40]. Paleozoic granites and slates, Triassic conglomerates, sandstones and pelites, Jurassic dolostones and limestones, and Cretaceous marly limestones form the Catalonian Coastal Range. Miocene calcarenites and marly limestones make up Montjuïc Mountain (Figure 1A). Estuarine marls, silts and clays were formed during the Pliocene [28,29,31,40–44]. An important unconformity marks the Pliocene–Quaternary boundary (Pliocene rocks and older rocks constitute the basement) [40–44] (Figure 1B). The Quaternary sedimentation was divided into two depositional sequences from the Pleistocene and Holocene [28,29,31,36,37,40,42,44], namely the so-called upper and lower detrital complex, respectively [40] (Figure 1B). After geophysical surveys in the offshore delta [36–38], the lower detrital complex was in turn divided into three minor depositional sequences [31]. Commonly, the lower detrital complex consists of conglomerate bodies and local sand with fine terrigenous material

intercalated [34–38]. The case of the upper detrital complex is more diverse, including bottom to top sands, silts, gravels (with sands) and the uppermost silt-clay thin level associated with coastal marshes and wetlands [34,44]. Diverse streams coming from the neighbor reliefs contribute to determining the present landscape [42–44]. Finally, the regional littoral drift towards the SW distributes the shoreline sedimentation in that direction [45,46].
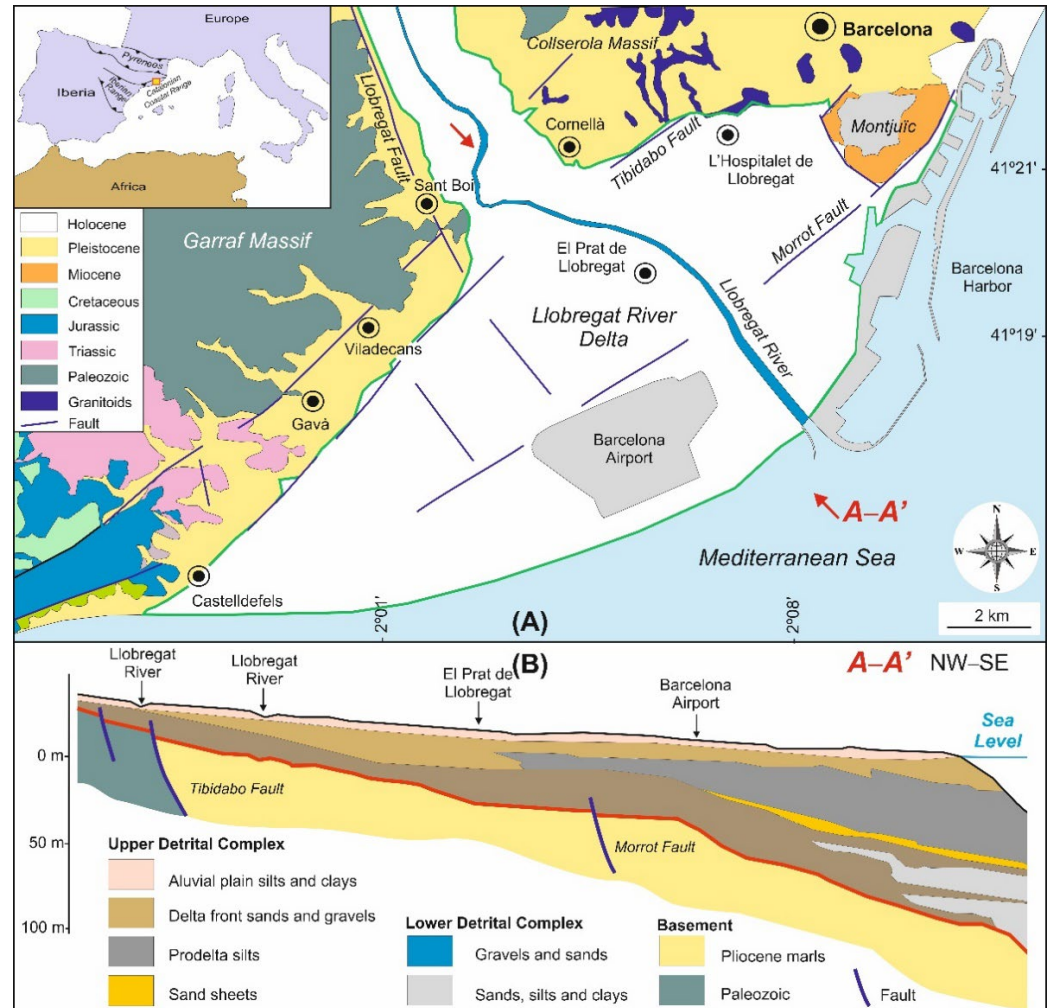

Figure 1. (**A**) Simplified geological map of the onshore LRD area (green line) updated and simplified from Medialdea and Solé-Sabarís [28,29], Almera [30] and Alonso et al. [31]. (**B**) Simplified geological cross-section A−A' (with NW–SE orientation) of the onshore LRD; see location in (**A**).

3. Methodology

3.1. Data Compilation

As introduced, the Technical Unit of the Llobregat Aquifers [27] prepared a grain-size database for groundwater purposes in the LRD. This database included accurate values from lab tests and proxy values after the visual surveys of the sediments identified in 433 onshore boreholes LRD [47]. The borehole lithologies and their grain-size records were revised to notice outliers. The detected outliers were suppressed or corrected when possible, thus providing the grain-size database used in this paper. The database was an XLS (Excel) file with meter-by-meter grain-size georeferenced values. The borehole's location (coordinates x and y) and prospecting depth (coordinate z) lead to a georeferenced array of grain-size data over space and depth. The georeferenced data were classified into four classes attending to proper grain-size limits for feasible groundwater mechanical exploitation through pumping wells, such as (i) clay–silt (<1 mm) for very low-to-low-

yielding fine materials, coarse sand (1−5 mm) for moderate-to-high-yielding medium-grain-sized materials, gravel (>5 mm) for very high-yielding coarse materials and basement, as used by Bullejos et al. [18].

The names and limits of this groundwater-oriented grain-size classification do not coincide with the formal names and limits established in official sedimentological and geotechnical grain-size classifications. For a reliable comparison of results, this paper has implemented the same data clustering and subsequent input files. Sections 3.2–3.4 are devoted to describing the methodological stages developed by Bullejos et al. [18] for KNN predictions (Figure 2), whereas Section 3.5 describes the methodology developed in this paper to quantify the confidence of the KNN predictions.
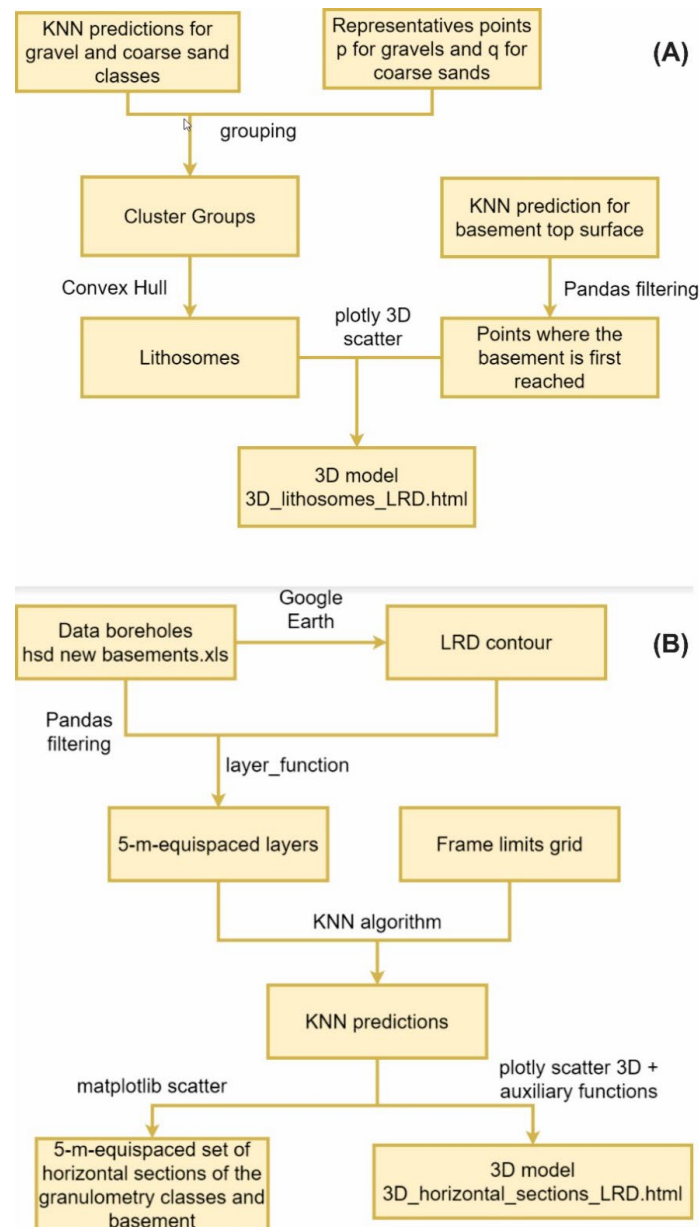

Figure 2. Flow diagram showing the methodological stages developed by Bullejos et al. [18] to create interactive 3D models of essential stratigraphic elements in the onshore LRD. The stages include data compilation and software implementation for the 3D mapping of (**A**) consecutive horizontal sections (5 m equispaced) of the grain-size classes generated by the KNN algorithm from 0 to 120 m b.s.l., and (**B**) coarse sedimentary bodies (lithosomes) and the basement top surface (the stratigraphic architecture) generated by Python libraries.

3.2. Python Programming Language

Python [48] is a popular open-source, object-oriented programming language. It has been used in a wide variety of environmental interests, including geological ones. In this work, we used this high-level programming language to examine the grain-size data and visualize the 3D models of essential stratigraphic elements of hydrogeological interest. The Python packages used here were (i) NumPy [49] for data computing, (ii) Pandas [50] for analyzing and processing data, (iii) Plotly [51] as a graphing library, (iv) Scipy [52] for interpolating and rendering algorithms, and (v) Scikit-learn [53] for the KNN classifier. In addition, an inverse distance weighted interpolation algorithm obtained from the 3D Terrain Modeling in the Python viewer from the GEODOSE block [54] was also utilized. This methodological flow is summarized in Figure 2. The Jupyter notebooks containing the Python code and its explanatory instructions can be accessed in the GitHub repository cited in Bullejos et al.'s Supplementary Materials [18]. This version of the code can be read using a web browser, so installing the Python kernel is not necessary.

3.3. KNN Algorithm

The k-nearest neighbors (KNN) algorithm is a robust machine learning classifier, which takes the proximity and similarity of the data into account to resolve classification and regression problems [55–58]. Bullejos et al. [18] used this capability to classify the grain-size database prior to 3D visualization. This operation assumed that the inferred class at a given point is mostly dependent on the nearest data. Under this assumption of similarity, Bullejos et al. [18] used the KNN algorithm to estimate data from a set of measurements to a wider space of individuals (in this case, the nodes of a 300 m × 300 m regular grid).

3.4. The 3D Mapping of the Essential Stratigraphic Elements

Figure 2 shows the working flow of the five-meter equispaced sets of horizontal sections generated by Bullejos et al. [18]. In each horizontal section, the KNN examined the K-nearest grain-size classes and selected the most frequent one. The routine assigned a weight inversely proportional to the distance between each neighbor in order to favor the closest ones. The choice of a suitable value for the parameter K is a significant issue. Smaller values of K may generate unrealistic polygonal regions, whereas larger values of K favor more smother, natural regions although they may disregard scattered data. The geological logic was used to conduct this process, hence avoiding problems with edge/boundary effects typically viewed in KNN. After the grain-size database was revised to include outliers, every supervised georeferenced datum was considered. Then, K = 1 was imposed, i.e., the KNN algorithm only searches for the neighbors that are nearest to every point. A Jupyter notebook in the GitHub repository mentioned in Bullejos et al. [18] includes the code.

Bullejos et al. [18] also utilized the Python library Pandas to load and process the XLS file containing the grain-size data (Figure 2). Google Earth and the Python Object from the Python Geometry package were used to define and generate the LRD contour, respectively. The next step involved defining (i) the grid (with defined X and Y bounds and nodal spacing) to make the nodal KNN predictions, (ii) the maximum modeling depth determined by the first basement top surface prospection, (iii) the 5 m depth exploration and (iv) the function 'layer_function' to classify the grain-size data according to the three defined grain-size classes and their depths.

The convenience of the 5 m equispaced set of horizontal sections of the grain-size classes (2D KNN predictions) was set as follows: (i) the LRD depth vs. length ratio since the Quaternary onshore LRD is approximately 120 m deep and 15 km long (horizontal length); (ii) the grain-size data arrangement meter to meter; (iii) the minimum dimension of the mappable sedimentary bodies of hydrogeological interest (which were at least 5 m thick in this case); and (iv) the distances between the data.

Next, the KNN algorithm was executed and the Matplotlib function scatter was used to generate the 2D horizontal sections with predictions delimited by polygonal region borders [18]. The plotly package was employed to create the interactive 3D models with the essential stratigraphic elements, such as the successive five-meter equispaced set of horizontal sections of the grain-size classes, the 3D visualization of the gravel and coarse sand sedimentary bodies, and the basement top surface delineation.

The mapped sedimentary bodies can be regarded as 'lithosomes' since they are a volume of rock of uniform character. Lithosomes are characterized by a grain size, composition and internal structure that is clearly differentiable from the neighboring rocks. As described in Section 3.1, the grain-size definition used in this paper was based on the Wentworth classification [59].

3.5. The 3D Models as HTML Files

The 3D models introduced by Bullejos et al. [18] were created as HTML (Hyper-text Markup Language) files. This paper uses this same file type to save the newly created 3D models. The HTML system is a standardized one for tagging text files in order to achieve font, color, graphic and hyperlink effects on World Wide Web pages, and it is the standard markup language for creating web pages [60–62]. The advantage of saving our 3D models as HTML files is that no additional tools are required (apart from a web browser to open them). So, once a model is opened with the use of a browser, it is possible to see different perspectives of the model or hide some elements to focus on others. Furthermore, it allows us to zoom in on a particular area, thereby providing us with a more precise view of it.

3.6. The 3D Mapping of the Confidence of the Essential Stratigraphic Elements

As stated in Section 1, the main limitation of the methodology developed by Bullejos et al. [18] was its lack of a routine to measure the confidence of the interactive 3D models of the essential stratigraphic elements of hydrogeological interest. Inspired by the Similarity Ratio described in the machine-learning KNN literature [19,20], this section introduces a metric to quantify the confidence of the KNN predictions. This metric assigns weights inversely proportional to the distance between each grain-size class (Figure 3) and takes the one with the highest weight at the target modal point. This weighing assignation satisfies the properties of the grain-size classification since it is based on the similarity (like or unlike) of neighboring data (all the data are determined by the physically based controls acting over this parameter).

In each nodal point of the same regular grid used by Bullejos et al. [18] for the KNN predictions, the weight of the nearest neighbor used to infer grain size was determined. The operation of dividing this weight by the weight of the nearest neighbor of a different grain-size class in the Similarity Ratio formulation was replaced by two sequential steps. First, the nearest data point from each of the four grain-size classes was determined, where one of them was the neighbor used for prediction. Later, the weights of the four grain-size classes were added, and the value was placed in the denominator. As a result, the normalized confidence metric for each nodal point was expressed as follows:

$$C_N(t) = \frac{W_{nn}}{\sum_{GC} W_{ndG}} \tag{1}$$

C_N is the normalized confidence associated with the prediction at t; W_{nn} is the weight of t's nearest neighbor; $\sum$ is the summation of the grain-size classes; (GC) is clay–silt, coarse sand, gravel and basement; and W_{ndG} is the weight of the data point of the grain-size class G that is closest to t among the data points in the grain-size class G.
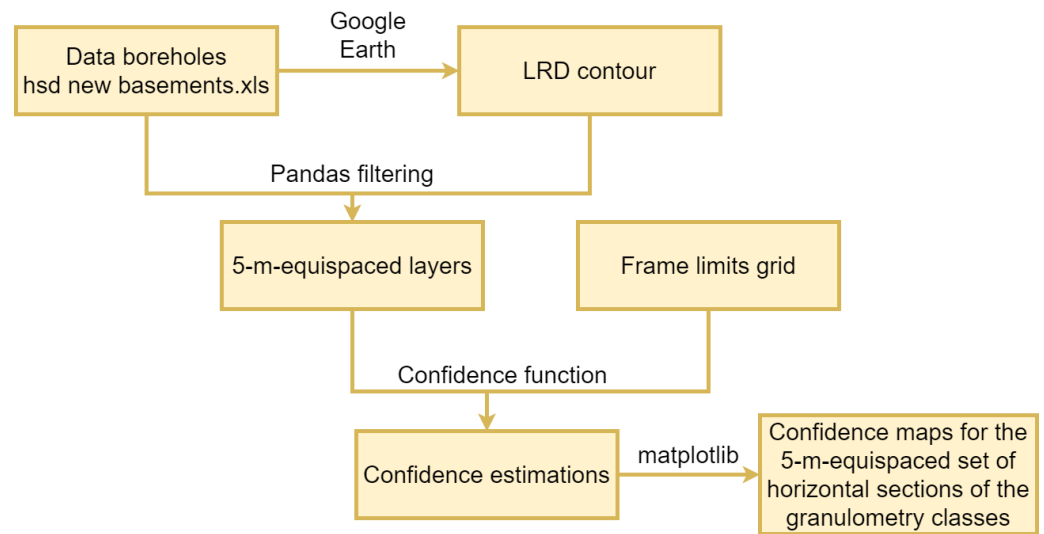
Figure 3. Methodological stages for mapping the confidence of the interactive 3D models of the essential stratigraphic elements in the onshore LRD, such as the horizontal 2D sections (5 m equispaced) of the grain-size (or granulometry) classes generated by the KNN algorithm from 0 to 120 m b.s.l., and the 3D stratigraphic architecture (coarse sedimentary bodies and the basement top surface) generated by Python libraries.

Equation (1) is a normalization of weights adding up to 1, so the result can be interpreted as a probability in the 0.25–1 range since the confidence of all grain-size classes (we predict the class with the highest confidence) add up to 1. The expression takes the value of 0.25 in the worst case in which all classes are equally likely, takes the value of 0.33 if three classes are equally likely and the fourth is very unlikely, and so on. A 0.6 confidence should not be treated as a 60% probability that the prediction will be correct. This would be a naïve interpretation since transforming classifier scores into probabilities is a significant issue. This metric allows us to estimate how confident the classifier for each KNN prediction is. Areas with high data density sharing a sole grain-size class lead to higher confidence values, whereas limits between areas with two or more different KNN predictions lead to lower confidence values. Furthermore, confidence is a continuous function that varies over a horizontal section (2D data field) at a given prospecting depth to the interval [0.25,1]. Points close to each other have similar confidence values.

Additionally, the function 'predict_proba' in the Scikit-learn implementation of KNN was implemented to estimate the approximate goodness of the confidence prediction [63]. For a given point, this function estimates the probability of belonging to each class attending to the weight of the neighbors of such class among the K nearest neighbors. However, this method cannot be performed in the case K = 1 since we would always obtain 100%.

The basement top surface confidence mapping was a particular case since below this stratigraphic element, no grain-size classes associated with any other cataloged geological material of hydrogeological interest will be found. This class deserved special treatment. The confidence formulation was modified to favor the basement when several basement data close to each other and apart from other non-basement data points exist as follows: (i) the normalized confidence was worked out as explained above, and (ii) the nearest neighbor (which is a basement) was ignored. When basement data still remained, a bonus was applied to the heaviest data. This confidence bonus was computed in a similar way to the normalized confidence (Equation (1)), in this case including the weight differences between the (new) nearest basement and the highest non-basement data in the numerator (as shown below).

$$b_C = \frac{W_{(n)nb} - W_{nnb}}{\sum_{GC} W_{ndG}} \qquad (2)$$

More precisely, b_C is the normalized confidence bonus; $W_{(n)nb}$ is the weight of the (new) nearest basement; W_{nnb} is the weight of the nearest non-basement; $\sum$ is the summation of the grain-size classes; (GC) is clay–silt, coarse sand, gravel and basement (ignoring the previous basement points); and W_{ndG} is the weight of the nearest data of the grain-size class G.

Finally, actual (or corrected) confidence is obtained as follows:

$$C_A = C_N + b_C(1 - C_N) \tag{3}$$

This function is continuous and adopts values in the $0.25-1$ range. This process will be repeated iteratively; meanwhile, new basement data with greater weights than other non-basement data are found, thus ignoring the closest, the two closest and the three closest basement data. Since no significant improvement with many iterations was noticed, this process was limited to considering the four closest basement data only.

4. Results

4.1. The Mapping of the Grain-Size Horizontal Sections: KNN Predictions and Confidences

For a reliable integration of KNN predictions and their confidences, the same successive horizontal layers (5 m equispaced) of the grain-size classes, the same regular grid in each layer over the entire onshore LRD surface, and the same grain-size classes (clay–silt (<1 mm), coarse sand (1–5 mm), gravel (>5 mm) and basement) implemented by Bullejos et al. [18] for KNN predictions were used for KNN prediction confidences (Figure 4). Four illustrative KKN predictions and their KNN confidence counterparts at 0, 20, 50 and 100 m b.s.l. are included in Figure 4A−H, respectively. In both cases, the boreholes' locations with grain-size data are presented. Figure 4A−D is an extract of the KNN predictions over the 120 m prospected depth in the LRD that can be found in the Jupyter notebook created by Bullejos et al. [18] (https://github.com/dcabezas98/knn-stratigraphic-visualization, accessed on 17 November 2022). Figure 4E−H is the equivalent extract for the KNN prediction confidences over the 120 m prospected depth in the LRD.

As described in Section 3.5, confidence is expressed in the dimensionless $0.25-1$ range from a dark purple color for the lowest value to a red color for the highest one (Figure 4E−H). The following intermediate confidence levels exist: low ($0.25-0.35$), satisfactory ($0.35-0.50$), high ($0.50-0.70$) and very high ($0.70-1$).

Regarding the confidence evolution with prospecting depth (Figure 4), we first focus on the shallowest 2D sections (0, 20 and 50 m b.s.l.). Close to the land surface (0 m), satisfactory to high confidences in the $0.40-0.70$ range were found. At 20 m b.s.l., confidence reaches the $0.60-0.90$ range in the coastal fringe and central sectors of the LRD, while the other sectors show values similar to those quantified at 0 m depth. At 50 m b.s.l., the area in the coastal fringe with confidences in the $0.60-0.90$ range decreases, while the other sectors of the LRD show lower values. At 50 m b.s.l., the high to very high confidences in the inland northern sector are associated with the basement's first appearance. In this inland northern sector, the very high confidences in the $0.70-1$ range are associated with the basement top surface's first appearance. The confidence increases and stabilizes from 50 m b.s.l. to 100 m b.s.l. because the basement classification does not cause divergences regarding the grain-size classes of the Quaternary sedimentation. At 100 m b.s.l., the central and coastal areas occupied by Quaternary sedimentation show moderate to high confidence levels. In general, the confidence level of the modeled Quaternary sedimentation tends to decrease with depth according to the decreasing data density from 0 m to 100 m b.s.l., whereas the basement in the $50-100$ m b.s.l. range is typically well-classified and shows higher confidence levels.
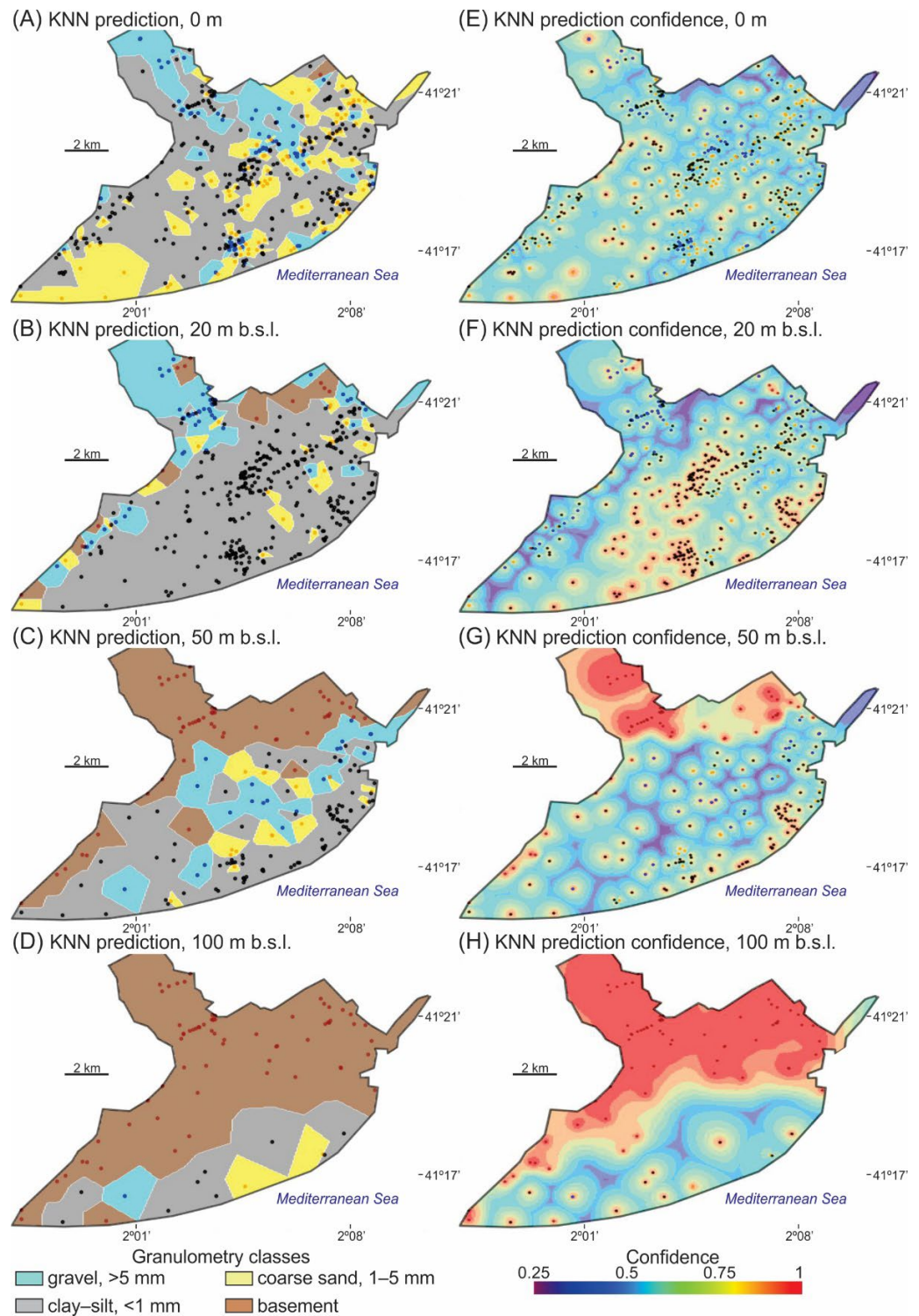
Figure 4. Horizontal sections at 0, 20, 50 100 m b.s.l. of the grain-size (or granulometry) classes generated by the KNN algorithm in the onshore Quaternary LRD [18]. (**A**−**D**) The KNN predictions, with grain-size data classified into the gravel (>5 mm, cyan), coarse sand (1–5 mm, yellow), clay–silt (<1 mm, gray) and basement (brown) classes. (**E**−**H**) The corresponding KNN prediction confidences are expressed in the dimensionless 0.25–1 range. Points A through H represent the boreholes with grain-size data at the mapped depth.

Figure 5A shows the interactive 3D model generated by the KNN algorithm from the consecutive set of horizontal layers (5 m equispaced) of the grain-size classes from 0 to

120 m b.s.l., as in Bullejos et al. [18]. This interactive 3D model (LRD_Classes_Layers.html) is provided in Supplementary Materials. Figure 5B–D show the confidence of the corresponding 5 m equispaced set of horizontal layers of the grain-size classes from 0 to 120 m b.s.l. For a better visual representation and description, the confidence associated with the basement (Figure 5B) and the gravel (Figure 5C) and coarse sand (Figure 5D) grain-size classes were secluded and mapped separately. The interactive 3D models for the basement confidence (Basement_Confidence_Layers.html), gravel sedimentary body (or lithosome) confidence (Gravel_Lithosomes_Confidence_Layers.html) and coarse sand sedimentary body confidence (Sand_Lithosomes_Confidence_Layers.html) are provided in Supplementary Materials. Figure 5 has been gathered from the 3D figures created by Bullejos et al. [18] (https://github.com/dcabezas98/knn-stratigraphic-visualization, accessed on 17 November 2022). These and the subsequent interactive 3D models can be loaded by using any browser. When loaded, the user can observe different views and zoom, rotate, and capture panoramic views. Additionally, the elements can be hidden to focus on a specific detail by simply clicking on the legend.
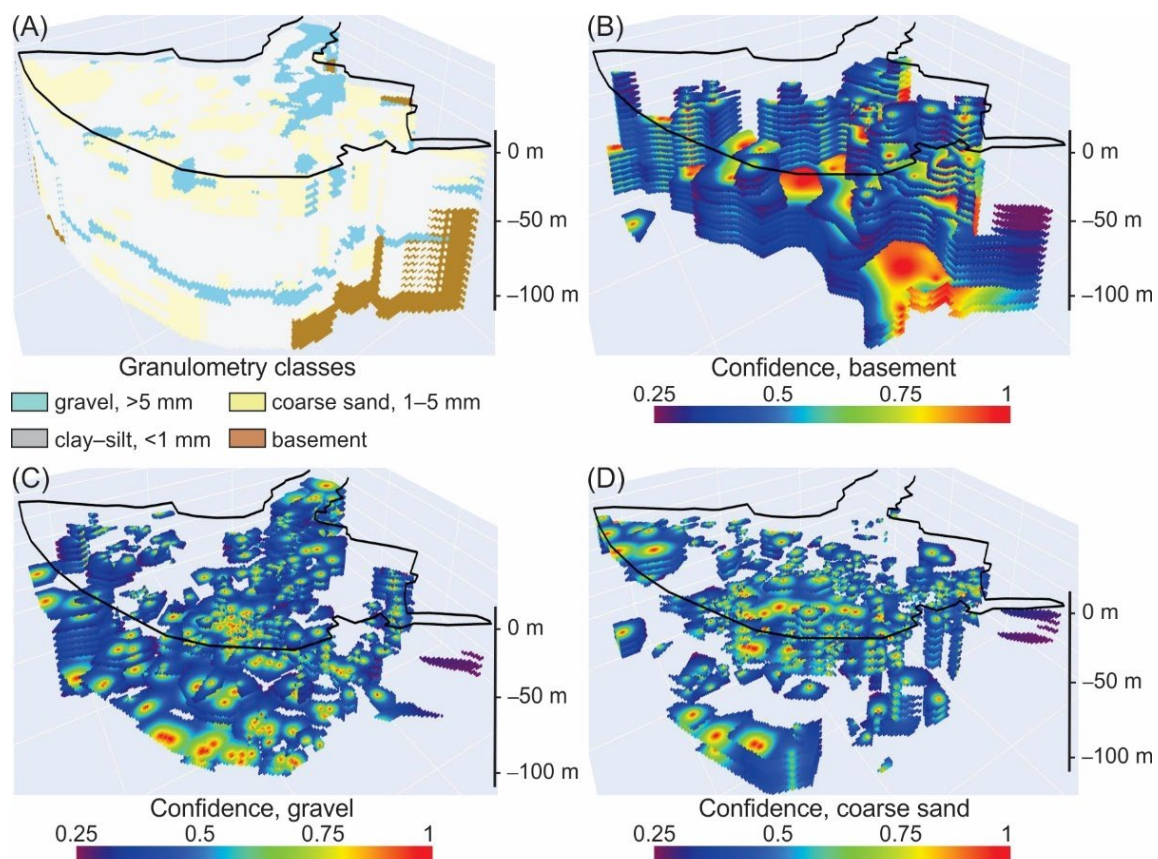

Figure 5. The consecutive horizontal layers (5 m equispaced) of the grain-size (or granulometry) classes generated by the KNN algorithm from 0 to 100 m b.s.l. in the onshore LRD [18] and their corresponding confidences. (**A**) The KNN predictions, with grain-size data classified into the gravel (>5 mm, cyan), coarse sand (1−5 mm, yellow), clay–silt (<1 mm, gray) and basement (brown) classes. An interactive 3D version of this model (LRD_Classes_Layers.html) appears in Supplementary Materials. (**B**−**D**) The corresponding KNN prediction confidences in the 0.25−1 range for the basement top surface (**B**) and the secluded areas concerning the gravel (**C**) and coarse sand (**D**) sedimentary bodies (or lithosomes). Interactive 3D versions of the confidence models for the basement top surface (Basement_Confidence_Layers.html), gravel sedimentary bodies (Gravel_Lithosomes_Confidence_Layers.html) and coarse sand sedimentary bodies (Sand_Lithosomes_Confidence_Layers.html) are included in the Supplementary Materials.

4.2. The 3D Mapping of the Stratigraphic Architecture and Basement Top Surface: KNN Predictions and Confidences

As shown in Figure 4A−D, the number of boreholes (data density) noticeably decreased with depth, so the shape of the sedimentary bodies and their confidences deteriorated and decreased with depth, respectively. With these horizontal sections, an interactive 3D model of the stratigraphic architecture (essential stratigraphic elements of hydrogeological interest) of the Quaternary onshore LRD was created (Figure 5). The ways in which the KNN prediction confidence varies attending to the data density and prospecting depth are described below.

Regarding the KNN prediction, Bullejos et al. [17,18] used a grouping procedure and defined 14 gravel sedimentary bodies and 17 coarse sand ones reaching a prospecting depth of 120 m b.s.l., which represents the entire onshore Quaternary sedimentation space in the LRD (Figure 6A−C). An interactive 3D model of the KNN prediction (3D_Lithosomes_And_Basement_LRD.html) is included in Supplementary Materials. Figure 6 presents partial views of the above-mentioned 3D interactive model, including the spatial distribution of coarse sedimentary bodies (Figure 6A,B) and the shape of the basement top surface (Figure 6C). The 3D visualization represents a gravel sedimentary body that is very continuous over time, close to the present Llobregat River course. In the SW sector, other minor gravel sedimentary bodies are also displayed at different depths (Figure 6A). There are also two big coarse sand bodies at different depths: the most important, shallowest one and others of little relevance (Figure 6B). The basement top surface reveals a general staggered shape moving deeper and deeper into the sea. In this stagger, an over-imposed horst-graben structure is visible (probably related to faulting) (Figure 6C). When the above-mentioned structure of the basement was compared with the geological map from Figure 1, a clear relationship between the horst-graben structures and the Tibidabo, Llobregat, and Morrot fault families can be observed. The 3D models effectively reproduced the sedimentary complexity of the area recently reported in the scientific literature [42–44,64,65].

Regarding the KNN prediction confidence (Figure 6D−F), the prospecting depth was restricted to 100 m b.s.l. So, the number of gravel sedimentary bodies was reduced to 13 ($n = 13$), while the same 17 ($n = 17$) coarse sand sedimentary bodies were maintained (Table 1). The KKN prediction confidence of the gravel sedimentary bodies (Figure 6D), coarse sand sedimentary bodies (Figure 6E) and basement top surface (Figure 6F) reflect the decreasing number of boreholes (data density) with depth, i.e., the upper part of the sedimentary bodies has higher confidences than the lower one. This means that the average-weighed confidence of a given lithosome integrates the high to very high confidences from the upper part and the low-to-moderate confidences from the lower part. The average-weighed confidence and other metrics and statistics for the gravel sedimentary bodies, coarse sand sedimentary bodies and the basement top surface are in Table 1. The interactive 3D models for the KNN prediction confidences for the gravel sedimentary bodies (3D_Gravel_Lithosomes_Confidence.html), coarse sand sedimentary bodies (3D_Sand_Lithosomes_Confidence.html) and basement top surface (3D_Basement_Confidence.html).are included in the Supplementary Materials.
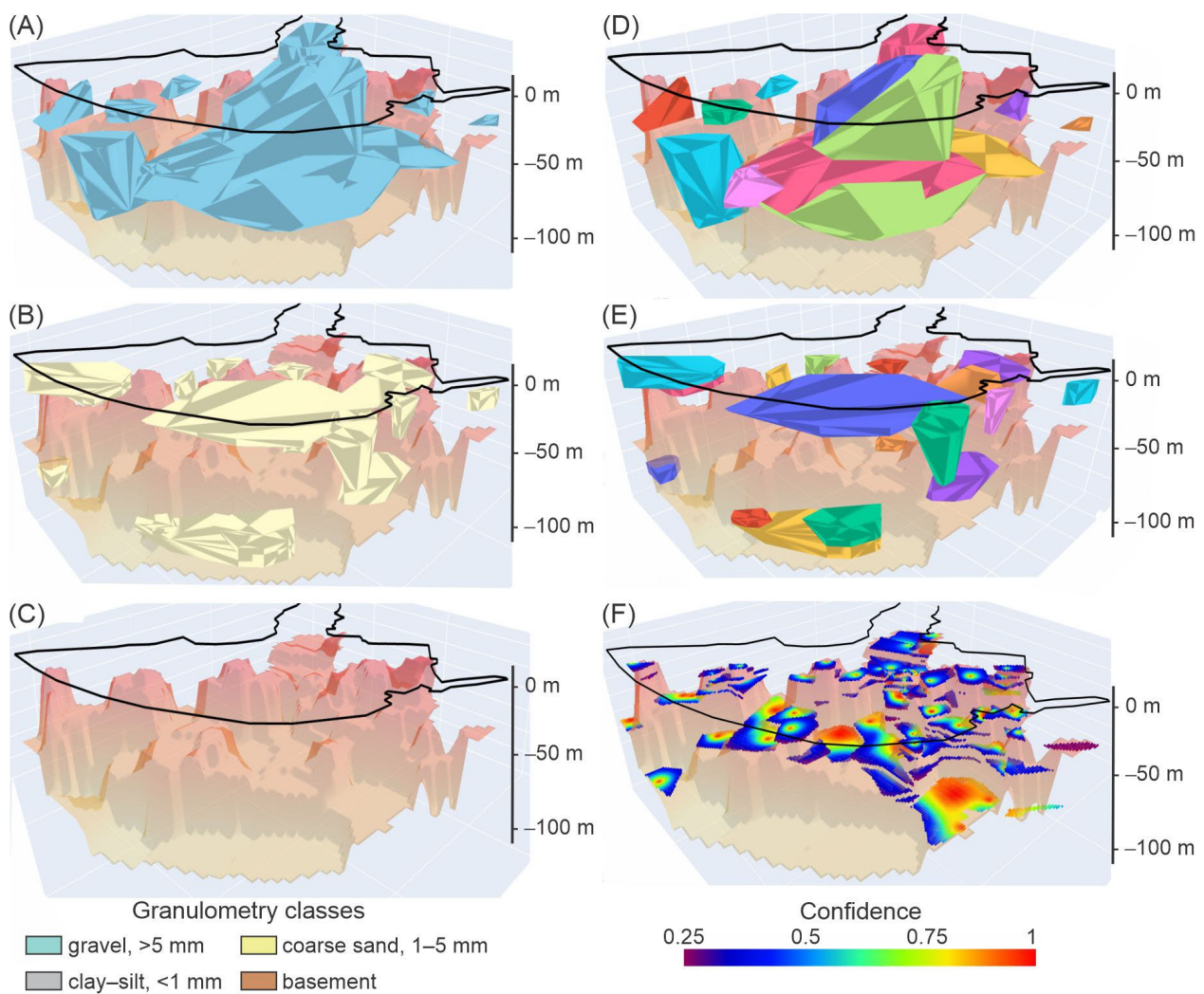
Figure 6. The 3D stratigraphic architecture (essential stratigraphic elements of hydrogeological interest and their confidences) in the LRD. (**A**−**C**) The KNN algorithm predictions [18] for gravel sedimentary bodies (or lithosomes) (>5 mm, cyan) and the basement's top surface (BTS) (**A**), coarse sand sedimentary bodies (1–5 mm, yellow), the only BTS (**B**) and BTS (brown) (**C**). In Supplementary Materials, an interactive 3D version of this model can be found (3D_Lithosomes_And_Basement_LRD.html). (**D**−**F**) The corresponding KNN prediction confidences, expressed in the dimensionless 0.25–1 range, for gravel sedimentary bodies (**D**), coarse sand sedimentary bodies (**E**) and only BTS (**F**). The interactive 3D models for the gravel sedimentary bodies' confidence values (3D_Gravel_Lithosomes_Confidence.html), coarse sand sedimentary bodies' confidence values (3D_Sand_Lithosomes_Confidence.html), and the BTS confidence values (3D_Basement_Confidence.html) are included in Supplementary Materials.

As described in the previous section, confidence is expressed in the 0.25−1 dimensionless range, including the following intermediate levels: low (0.25−0.35), satisfactory (0.35−0.50), high (0.50−0.70) and very high (0.70−1) (Figure 6D−F). In the LRD, the average-weighed confidences varied in the 0.44−0.53 and 0.42−0.55 ranges for prospecting depths in the 12.7−72.4 m b.s.l. and 4.6−83.9 m b.s.l. ranges for gravel (*n* = 12) and coarse sand (*n* = 16) sedimentary bodies, respectively (Table 1). Low average-weighed confidences in one gravel body of 0.29 and in one coarse sand one of 0.30 were found in the SW sector, probably associated with the poor classification in this sector where very different grain-size classes exist closely among them.

Table 1. Basic metrics and statistics of the KNN prediction confidence of the observed (and defined) gravel sedimentary bodies, coarse sand sedimentary bodies and the basement top surface in the onshore LRD.

Gravel Sedimentary Bodies	AW Confidence [1]	AW Depth [2]	Depth Min [2]	Depth Max [2]	LRD Interval
grlit1	0.50	−72.4	−40	−100	Lower
grlit2	0.51	−51.4	−40	−60	Lower
grlit3	0.52	−55.5	−45	−70	Lower
grlit4	0.49	−56.6	−50	−65	Lower
grlit5	0.51	−45.7	−35	−50	Middle
grlit6	0.50	−28.3	0	−40	Middle to Lower
grlit7	0.48	−29.8	−10	−40	Middle to Lower
grlit8	0.44	−29.5	−25	−35	Middle
grlit9	0.53	−19.7	−10	−25	Middle to Lower
grlit10	0.29	−20.0	−15	−25	Middle to Lower
grlit11	0.45	−17.5	−15	−20	Lower
grlit12	0.49	−12.7	0	−25	Lower
grlit13	0.50	−25.0	0	−40	Middle to Lower
Median	0.50	−29.5	−15	−40	
Average	0.48	−35.7	−21.9	−45.8	
Standard Deviation (±1σ)	0.06	18.6	18.2	22.9	
CV [3]	0.13	−0.52	−0.83	−0.50	

Coarse sand sedimentary bodies	AW confidence [1]	AW depth [2]	Depth min [2]	Depth max [2]	LRD interval
snlit1	0.51	−83.9	0	−100	Upper to Lower
snlit2	0.54	−35.0	0	−90	Upper to Lower
snlit3	0.54	−25.5	0	−90	Upper to Lower
snlit4	0.53	−39.7	0	−100	Upper to Lower
snlit5	0.42	−65.9	0	−80	Upper to Lower
snlit6	0.54	−12.7	0	−60	Upper to Lower
snlit7	0.54	−24.4	0	−90	Upper to Lower
snlit8	0.54	−8.8	0	−40	Middle to Lower
snlit9	0.51	−10.5	−5	−20	Lower
snlit10	0.46	−49.1	−5	−80	Upper to Lower
snlit11	0.46	−17.1	0	−40	Middle to Lower
snlit12	0.55	−10.1	−5	−55	Upper to Lower
snlit13	0.43	−40.6	0	−60	Upper to Lower
snlit14	0.53	−16.3	0	−60	Upper to Lower
snlit15	0.54	−4.6	0	−10	Lower
snlit16	0.53	−9.2	0	−60	Upper to Lower
snlit17	0.30	−6.6	0	−15	Lower
Median	0.53	−12.7	0	−60	
Average	0.50	−21.2	−1.2	−51.5	
Standard Deviation (±1σ)	0.07	18.9	2.2	25.4	
CV [3]	0.13	−0.89	−1.90	−0.49	

Basement top surface	AW confidence [1]	AW depth [2]	Depth min [2]	Depth max [2]	LRD interval
basement	0.78	−70.4	0	−100	Upper to Lower

[1] AW confidence—average-weighted KNN confidence in the dimensionless 0.25−1 range. [2] AW—average-weighted prospecting depth, Depth min—minimum prospecting depth, and Depth max—maximum prospecting depth in m b.s.l. [3] CV—coefficient of variation (standard deviation (±1σ) to average value ratio) as a dimensionless fraction. LRD stratigraphic interval as defined in Bullejos et al. [17].

In general, the 0.48 ± 0.06 and 0.50 ± 0.07 average-weighted confidences for the gravel and coarse sand sedimentary bodies, respectively, mean that the KNN algorithm has a satisfactory to high ability to predict these grain-size classes regarding the adjacent ones. The average-weighted confidence for the basement top surface is 0.78, thus showing the

ability of the KNN algorithm to predict this class regarding the adjacent coarse classes of the onshore Quaternary sedimentation.

5. Discussion and Conclusions

In a previous paper, Bullejos et al. [18] introduced a KNN Python algorithm [48–58] for the 3D visualization of essential stratigraphic elements of hydrogeological interest in sedimentary porous media [22–24]. Based on the proximity and similarity to make classifications or predictions about the grouping of an individual data point, the application used the grain-size database prepared by METALL in the onshore Quaternary LRD groundwater body [28–37] for groundwater purposes to generate two interactive 3D models [60–62]. The first model was the consecutive horizontal sections (5 m equispaced) of the grain-size classes generated by the KNN algorithm from 0 to 120 m b.s.l. The second model was the 3D visualization of the main gravel and coarse sand sedimentary bodies forming aquifers and the basement top surface generated by Python libraries. This paper solves an important limitation concerning the confidence quantification [55–58] of the two above interactive 3D models. The quantification of the KNN prediction confidence uses a variant of the Similarity Ratio based on weights that decrease with the distance between each grain-size class instead of similarity.

The rationale for using this metric departs from the adopted K = 1 in the KNN algorithm and the defined four grain-size classes on the basis of proper limits for feasible groundwater mechanical exploitation through pumping wells (clay–silt, coarse sand, gravel and basement). This means that only the nearest neighbor of each grain-size class must represent that class when the grain size at the point to be classified is inferred, i.e., the closest grain-size class is assigned to that point. The accuracy of the KNN prediction decreases when the weight of the neighbors from different classes is very similar since it is influenced by a close call. The KNN prediction confidence increases when the weight of neighbors from different classes is quite different. At a given 2D data field at a given prospecting depth (horizontal section), the KNN prediction confidence is therefore a function of the data density. At a 3D data field, the confidence also reflects the common decreasing data density (number of boreholes) with prospecting depth. The divergences regarding this general behavior are closely associated with very different grain-size classes.

For the consecutive horizontal sections (which are five-meter-equispaced sections) of the grain-size classes, the overall high confidences associated with the higher data density in the shallower horizontal sections (Figure 4E−H) tend to decrease in some sectors when the widely varied points of different classes mingled and were close to each other. At 20 m b.s.l., the clay–silt class predominates in the central and coastal sectors of the LRD. The abundance of data matching in class makes the algorithm confident that clay–silt is found in such places. For deeper prospecting depths (50 m b.s.l. and deeper), this matching decreases due to the decreasing data density, although the appearance of gravel bodies adjacent to coarse sand and clay–silt bodies indicates a similar decrease in confidence values (Figure 4E−H).

In the LRD, the average-weighed confidence varied in the 0.44−0.53 and 0.42−0.55 ranges for prospecting depths in the 12.68−72.39 m b.s.l. and 4.55−83.94 m b.s.l. ranges for the gravel (n = 12) and coarse sand (n = 16) sedimentary bodies, respectively (Table 1). The KKN prediction confidence for the gravel sedimentary bodies (Figure 6D), coarse sand sedimentary bodies (Figure 6E) and basement top surface (Figure 6F) are mostly subjected to the decreasing data density with depth (Figure 4E−H), i.e., the upper part of the sedimentary bodies has higher confidence values than the lower one. This means that the average-weighed confidence of a given lithesome integrates the high to very high confidences from the upper part (the shallower horizontal sections) and the low-to-moderate confidences from the lower part (the deeper horizontal sections). This matching produced spurious average-weighed confidences of 0.29 in one small gravel lithesome and 0.30 in one small coarse sand lithesome (Table 1; Figure 7). This additional ability of the KNN prediction confidence calculation is of special interest because it secludes these kinds

of disparities due to the close points of different classes in supposedly well-depurated grain-size databases, such as those used in this research paper. The average-weighted confidence for the basement top surface is 0.78 (Table 1), thus showing the ability of the KNN algorithm to predict this class regarding the adjacent grain-size classes of the onshore Quaternary deposits.

The visualization of the 3D stratigraphic architecture (essential sedimentary bodies of hydrogeological interest) and its confidence must be managed together for predictable interpretations. This is a prerequisite for decision making in applied (mainly environmental and economic) geology. The 3D modeling seeks to improve the groundwater management and governance, optimization of groundwater monitoring networks, drilling of proper pumping wells in the LRD and similar groundwater bodies, as well as the evaluation of the impact of the large civil works included in the LRD Infrastructure Plan on the groundwater resource. The spatial distribution of the grain size (and the subsequent definition of sedimentary bodies) may also be of assistance in designing compensatory measures for aquifer protection and recovery, including the choice of suitable sites for managed (artificial) aquifer recharge and specific measures to control the advance of seawater intrusion and the mobilization of contaminants.

Figure 7. Range and average-weighed prospecting depth vs. average-weighed KNN prediction confidence for the 13 gravel (**A**) and 17 coarse sand (**B**) sedimentary bodies, as shown in Table 1. LRD stratigraphic interval as defined in Bullejos et al. [17] and achieved confidence levels, namely low (0.25−0.35), satisfactory (0.35−0.50) and high (0.50−0.70), are also indicated.

The introduced application also quantifies the confidence of the interactive 3D models. In the onshore LRD, the overall confidence of the interactive 3D models was in the 0.4−0.6 range. These figures enable rational geological interpretations prior to explorations. The KNN prediction confidence can also be used to depurate input data as a further technique aimed at improving the quality of predictions. This ability has interesting implications for geological modeling strategies based on progressive data assimilation. Similar to the grain size, the introduced KNN prediction confidence calculation can also be used to quantify the confidence of other mapped geological elements and features based on other variables and parameters with similar classifiable behavior.

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. Der Dtsch. Ges. Fur 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 17 November 2022). [CrossRef] [PubMed]

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; Volume 1, pp. 1–6. Available online: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8046999 (accessed on 9 June 2022).

15.   Husillos, C. cesarhusrod/sarai_piezo_precip: Mejora de la Documentación (v1.0.1). Zenodo. 2022. Available online: https://doi.org/10.5281/zenodo.7197288 (accessed on 17 November 2022).

16.   Pyrcz, M. GeostatsGuy Lectures. Available online: https://www.youtube.com/c/GeostatsGuyLectures (accessed on 9 June 2022).

17.   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. [CrossRef]

18.   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. [CrossRef]

19.   Delany, S.J.; Cunningham, P.; Doyle, D. Generating estimates of classification confidence for a case-based spam filter. In *International Conference on Case-Based Reasoning*; 3620 of LNAI; Springer: Berlin/Heidelberg, Germany, 2005; pp. 170–190.

20.   Hu, R.; Delany, S.J.; Mac Namee, B. Sampling with confidence: Using k-nn confidence measures in active learning. In Proceedings of the 8th International Conference on Case-based Reasoning, ICCBR, Seattle, WA, USA, 20−23 July 2009; Volume 9, pp. 181–192.

21.   Murphy, A.; Redfern, S. Confidence Measures in Multiclass Speech Emotion Recognition using Ensemble Learning to Catch Blunders. *Int. J. Sci. Technol. Eng.* **2015**, *2*, 118–122. Available online: https://ijste.org/Article.php?manuscript=IJSTEV2I3013 (accessed on 17 November 2022).

22.   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.

23.   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]

24.   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]

25.   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]

26.   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; 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).

27.   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/salapremsa/notes-premsa/68710/agencia-catalana-aigua-crea-mesa-tecnica-dels-aqueifers-del-llobregat (accessed on 18 April 2022).

28.   Medialdea, J.; Solé-Sabarís, L. *Geological Map of Spain, Scale 1:50,000, Sheet n° 420*; 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).

29.   Medialdea, J.; Solé-Sabarís, L. *Geological Map of Spain, Scale 1:50,000, Sheet n° 448*; 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).

30.   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).

31.   Alonso, F.; Peón, A.; Rosell, J.; Arrufat, J.; Obrador, A. *Geological Map of Spain, Scale 1:50,000, Sheet n° 421*; 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).

32. Llopis, N. Tectomorfología del Macizo del Tibidabo y valle inferior del Llobregat. *Estud. Geográficos* **1942**, *3*, 321–383.
33. Solé-Sabarís, L. Ensayo de interpretación del Cuaternario Barcelonés. *Misc. Barcinonensia* **1963**, *2*, 7–54.
34. Marqués, M.A. *Les Formacions Quaternàries del Delta del Llobregat*; Institut d'Estudis Catalans: Barcelona, Spain, 1984.
35. Manzano, M. Estudio Sedimentológico del Prodelta Holoceno del Llobregat. Master's Thesis, University of Barcelona, Barcelona, Spain, 1986.
36. IGME. *Geological Map of the Spanish Continental Shelf and Adjacent Areas, Scale 1:200,000, Sheet nº 42E.*; 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).
37. IGME. *Geological Map of the Spanish Continental Shelf and Adjacent Areas, Scale 1:200,000, Sheet nº 42*; 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).
38. 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.
39. 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]
40. 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]
41. 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. *Boletín Real Soc. Española Hist. Nat.* **2002**, *97*, 42–49.
42. 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.
43. 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. *Geo-Temas* **2003**, *5*, 23–26.
44. Simó, J.A.; Gàmez, D.; Salvany, J.M.; Vàzquez-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.
45. 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.
46. 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]
47. 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]
48. Python Programming Language. Available online: https://www.python.org (accessed on 9 June 2022).
49. Numpy. Available online: https://numpy.org (accessed on 13 June 2022).
50. Pandas. Available online: https://pandas.pydata.org (accessed on 13 June 2022).
51. Plotly. Available online: https://plotly.com (accessed on 9 June 2022).
52. Scipy. Available online: https://scipy.org (accessed on 13 June 2022).
53. Scikit-learn. Available online: https://scikit-learn.org/stable/install.html#installation-instructions (accessed on 13 June 2022).
54. GEODOSE. Available online: https://www.geodose.com/2019/09/3d-terrain-modelling-in-python.html (accessed on 13 June 2022).
55. 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]
56. 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]
57. Wang, X.; Yang, S.; Zhao, Y.; Wang, Y. Lithology identification using an optimized KNN clustering method based on entropy-weighed co-sine distance in Mesozoic strata of Gaoqing field, Jiyang depression. *J. Pet. Sci. Eng.* **2018**, *166*, 157–174. [CrossRef]
58. Huang, S.; Huang, M.; Lyu, Y. An Improved KNN-Based Slope Stability Prediction Model. *Adv. Civ. Eng.* **2020**, 8894109. [CrossRef]
59. Wentworth, C.K. A Scale of Grade and Class Terms for Clastic Sediments. *J. Geol.* **1922**, *30*, 377–392. [CrossRef]
60. Leifeld, P. texreg: Conversion of Statistical Model Output in R to LATEX and HTML Tables. *J. Stat. Softw.* **2013**, *55*, 1–24. [CrossRef]
61. Li, Y.; Yang, Z.; Chen, X.; Yuan, H.; Liu, W. A stacking model using URL and HTML features for phishing webpage detection. *Future Gener. Comput. Syst.* **2019**, *94*, 27–39. [CrossRef]
62. Gur, I.; Nachum, O.; Miao, Y.; Safdari, M.; Huang, A.; Chowdhery, A.; Narang, S.; Fiedel, N.; Faust, A. Understanding HTML with Large Language Models. *arXiv* **2022**, arXiv:2210.03945v1. [CrossRef]

63. Scikit-learn: KNeighborsClassifier, Predict_Proba. Available online: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html#sklearn.neighbors.KNeighborsClassifier.predict_proba (accessed on 9 September 2022).

64. Salvany, J.M.; Aguirre, J. The Neogene and Quaternary deposits of the Barcelona city through the high-speed train line. *Geologica Acta* **2020**, *18*, 1–19. [CrossRef]

65. 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. Geológica España* **2008**, *21*, 93–109.