ISSN 2712-8008
Volume / Том 25 №2 2025
Journal Homepage: http://vestnik.pstu.ru/geo/
Perm Journal of Petroleum and Mining Engineering

UDC 622+519.688: 550.3: 553.9
Article / Статья

Automation of interwell correlation using dynamic time scale transformation algorithm: application experience

Ivan S. Tunev, Vadim A. Bariakh

LUKOIL-Engineering LLC (3a Permskaya st., Perm, 614015, Russian Federation)

Автоматизация межскважинной корреляции с использованием алгоритма динамической трансформации временной шкалы: опыт применения

И.С. Тунев, В.А. Барях

ООО «ЛУКОЙЛ-Инжиниринг» (Российская Федерация, 614015, г. Пермь, ул. Пермская, 3а)

Received / Получена: 24.10.2024. Accepted / Принята: 05.12.2024. Published / Опубликована: 01.04.2025

Keywords:
interwell correlation, automatic well correlation, autocorrelation, geophysical well logging, logging, geophysics, dynamic time scale transformation algorithm, algorithm, programming, clustering, data analysis, statistics, data processing.

The article presents an approach to automation of interwell correlation based on the Dynamic Time Warping algorithm for dynamic time scale transformation. This method is an effective tool for analyzing and comparing time series (in this case, well logging curves). The key stages of implementing this approach are described, starting with preliminary data preparation: selecting the most informative well logging curves, processing gaps, smoothing and standardizing the data. Then, clustering methods (for example, k-means using PCA) are applied to determine reference wells that cover all possible behavior options for well logging curves from the available data. At the next stage, the Dynamic Time Warping algorithm is used to calculate the similarity between reference and target wells, which allows finding the optimal match between them. Finally, the position of formation intersections in target wells is predicted based on expert marks in reference wells.
Experience shown high consistency of the algorithm results with expert estimates in 85 % of cases, confirming its effectiveness. Discrepancies were noted, which in some cases were caused by subjective factors in expert interpretation, while the Dynamic Time Warping algorithm demonstrated effectiveness in correctly recognizing the assigned behavior patterns of well logging curves.
The proposed approach based on Dynamic Time Warping not only improves the accuracy and objectivity of data interpretation, but also serves as a tool for identifying and correcting subjective errors associated with the human factor. This is especially important when working with large volumes of data, where traditional analysis methods become extremely labor-intensive and vulnerable to errors.
Thus, the use of Dynamic Time Warping opens up new opportunities for automating the processes of data analysis and correlation in geophysical research, contributing to the increased efficiency of specialists in this field.

Ключевые слова:
межскважинная корреляция, автоматическая корреляция скважин, автокорреляция, ГИС, каротаж, геофизика, алгоритм динамической трансформации временной шкалы, алгоритм, программирование, кластеризация, анализ данных, статистика, обработка данных.

Представлен подход к автоматизации межскважинной корреляции, основанный на алгоритме динамической трансформации временной шкалы Dynamic Time Warping. Данный метод является эффективным инструментом для анализа и сравнения временных рядов (в данном случае каротажных кривых). Описаны ключевые этапы реализации данного подхода, начиная с предварительной подготовки данных: выбор наиболее информативных кривых геофизических исследований скважин, обработка пропусков, сглаживание и стандартизация данных. Затем применяются методы кластеризации (например, k-means с использованием PCA) для определения эталонных скважин, которые охватывают все возможные варианты поведения кривых геофизических исследований скважин из имеющихся данных. На следующем этапе используется алгоритм Dynamic Time Warping для вычисления сходства между эталонными и целевыми скважинами, что позволяет найти оптимальное соответствие между ними. Наконец, выполняется прогнозирование положения пластопересечений в целевых скважинах на основе экспертных отметок в эталонных скважинах.
Опыт применения показал высокую согласованность результатов алгоритма с экспертными оценками в 85 % случаев, подтверждая его эффективность. Отмечены расхождения, которые в ряде случаев были вызваны субъективными факторами при экспертной интерпретации, в то время как алгоритм Dynamic Time Warping продемонстрировал эффективность в правильном распознавании заданных ему шаблонов поведения кривых геофизических исследований скважин.
Предложенный подход не только повышает точность и объективность интерпретации данных, но и служит инструментом для выявления и коррекции субъективных ошибок, связанных с человеческим фактором. Это особенно важно при работе с большими объемами данных, где традиционные методы анализа становятся крайне трудоемкими и уязвимыми для ошибок.
Таким образом, использование Dynamic Time Warping открывает новые возможности для автоматизации процессов анализа и корреляции данных в геофизических исследованиях, способствуя повышению эффективности работы специалистов в данной области.

Ivan S. Tunev (ORCID: 0009-0003-5075-242X) – 1st Category Geologist at Department of Geological Structure Monitoring (tel.: +007 (952) 659 16 39, e-mail: Ivan.Tunev@lukoil.com). The contact person for correspondence.
Vadim A. Baryakh (Author ID in Scopus: 57205619691, ORCID: 0009-0006-6453-5313) – Head of Department of Geological Structure Monitoring (tel.: +007 (902) 794 48 72, e-mail: Vadim.Baryakh@lukoil.com).

Тунев Иван Сергеевич (ORCID: 0009-0003-5075-242X) – геолог 1-й категории отдела мониторинга геологического строения (тел.: +007 (952) 659 16 39, e-mail: Ivan.Tunev@lukoil.com). Контактное лицо для переписки.
Барях Вадим Александрович (ORCID: 0009-0006-6453-5313) – начальник отдела мониторинга геологического строения (тел.: +007 (902) 794 48 72, e-mail: Vadim.Baryakh@lukoil.com).

Please cite this article in English as:
Tunev I.S., Bariakh V.A. Automation of interwell correlation using dynamic time scale transformation algorithm: application experience. *Perm Journal of Petroleum and Mining Engineering*, 2025, vol.25, no.2, pp. 66-74. DOI: 10.15593/2712-8008/2025.2.2

Просьба ссылаться на эту статью в русскоязычных источниках следующим образом:
Тунев, И.С. Автоматизация межскважинной корреляции с использованием алгоритма динамической трансформации временной шкалы: опыт применения / И.С. Тунев, В.А. Барях // Недропользование. – 2025. – Т.25, №2. – С. 66–74. DOI: 10.15593/2712-8008/2025.2.2

Introduction

The data volume is rapidly growing with the development of modern technologies, challenging researchers and specialists in various fields, including the oil industry. Traditional data analysis methods, especially for large information volumes, as well log (WL) data, are often time-consuming. Correlating a large number of wells manually becomes an extremely labor-intensive and complex task.

Geology experts must compare a great deal of curves with one another, which is costly in time and can lead to subjective conclusions. It necessitates implementing automated analytical processes to improve the efficiency and accuracy in decision-making.

The paper presents the approach to automating interwell correlation, based on the Dynamic Time Warping (DTW) algorithm and implemented by using the Python programming language [1–14]. The following main stages are outlined in this study:

1) well data preparation and pre-processing, specifically the selection of the most representative well logging curves, processing of missing data, smoothing and standardization;

2) using clustering methods (K-means, with PCA application) to determine the minimum sample of reference wells that will contain all possible variations and features of the well logging curves from the available data;

3) implementation of the DTW algorithm to calculate the similarity between the reference and target wells using a distance matrix. During the algorithm execution, each reference well is compared to each target well, and the optimal correspondence is subsequently determined on the minimum DTW distance;

4) determination of layer intersection marks in target wells based on expert (previously known) marks in reference wells using optimal DTW distance between the wells.

Preparation of input data

In the process of well data analysis, particularly in the context of geophysical studies, the first stage is defined as the data preparation and pre-processing. It includes several key steps to ensure the quality and reliability of the subsequent analysis.

First, the most informative logging curves are selected, which clearly reflect the changes in reservoir properties. Redundant and non-informative curves are excluded from further analysis, which allows focusing on the most relevant data. To test the algorithm, Middle Devonian terrigenous deposits (D2) are singled out. Using neutron gamma logging (NGK) and gamma logging (GK) is sufficient to monitor their properties [15–18].

The next step is to process the missing data. Missing values can significantly influence the results of the analysis, so their processing is extremely important. Gaps can be filled with mean values or medians, and interpolated by neighbor data, or removed from the analysis [19–21].

To reduce the effect of noise, the well logging curves need to be smoothed. One of the commonly used smoothing methods is the moving average method [22, 23], which formula is as follows:

$$MA_t = \frac{1}{n}\sum_{i=0}^{n-1} y_{t-i}, \quad (1)$$

where MA_t is the value of the moving average at time t; y_{t-i} is the value of the initial series at time $t-i$; n is the smoothing width.

This method helps to reduce random fluctuations and identify general trends, improving data interpretation and increasing the accuracy of analysis.

Selecting the smoothing window width (n) is an important step. A window that is too small may not eliminate enough noise, while a window that is too large may lead to excessive smoothing and loss of important details. Choosing the smoothing window width for well logging curves requires a compromise between noise reduction and useful data preservation.

On the tested field, the geological cross-section is characterized by interbedding of fine layers, which can be significant for correlation. Considering the data sampling interval of 0.1 m, a smoothing value of $n = 10$ was chosen, corresponding to a smoothing window of 1 m. It helps to reduce the influence of noise and high-frequency oscillations on the data, while maintaining a sufficient detail to identify minor features of the cross-section.

To facilitate comparison between different well logging curves, the data need to be standardized. The z-score normalization method is used for standardization to transform the data so that it has a mean of 0 and a standard deviation of 1 [24, 25]. The formula for the z-score is as follows:

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (2)$$

where z_i is a standardized value for the point i; x_i is the initial value for the point i; μ is an average value of the initial data; σ is a standard deviation of initial data.

After standardization, all well logging curves have the same statistical characteristics, which facilitates their comparison and analysis.

The performed well logging data preprocessing steps result in smoothed and standardized data, which enhances interpretation and increases analysis accuracy, as well as facilitates comparison between different well logging curves (Fig. 1).

Determination of minimum reference wells samples

It is important to create a reference sample for predictive model construction of the reservoir characteristics based on well research data. To obtain the high-quality and representative reference sample, it is reasonable to use clustering methods. Clustering allows the existing wells to be divided into homogeneous groups (clusters) based on characteristic features. It helps to form a minimum set of reference wells, covering all the selected clusters.

In this study, the overall DTW distance [26–37] between two time series, represented by well logging curves, is proposed as the key clustering parameter. The total DTW distance serves as similarity magnitude between time series.

DTW is used to construct a distance matrix between all possible pairs of points in two time series. An optimal path through this matrix is then determined, which minimizes the total distance between the corresponding points (Fig. 2).

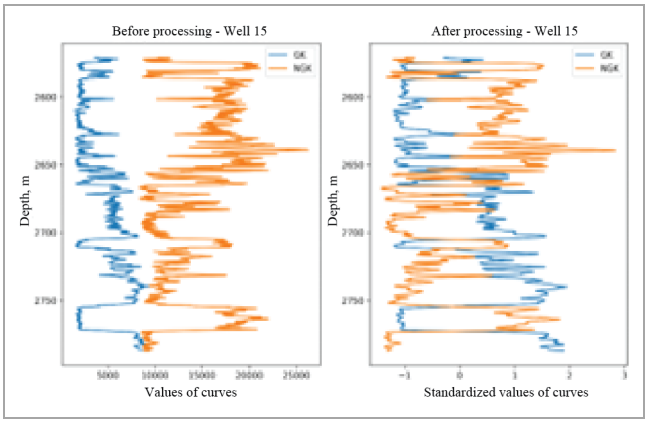

Fig. 1. Well log curves before and after processing

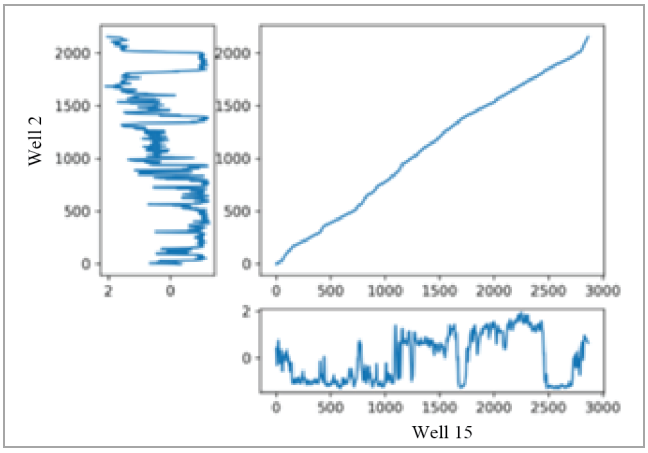

Fig. 2. Distance matrix between well pairs using gamma logging

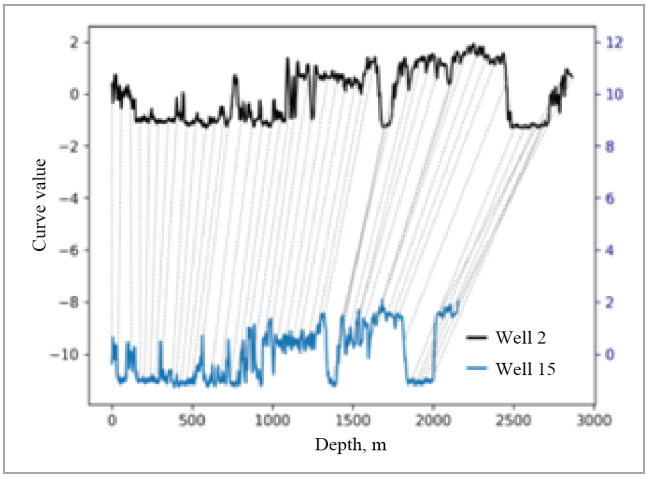

Fig. 3. Correspondence graph

In general, the DTW distance matrix is calculated as follows:

$$DTW(X,Y) = \min_{\gamma}\left(\sum_{(i,j)\in\gamma} d(x_i, y_j)\right), \qquad (3)$$

where $DTW(X,Y)$ is the overall distance between two time series X,Y; i,j are indices that indicate the position of an element in a time series $X\,u\,Y$ respectively; γ is the set of all possible paths representing a sequence of index pairs (i,j); $d(x_i,y_j)$ is the distance function between the elements x_i, of time series X and elements y_j, of time series Y; $\min_{\gamma}$ is identified minimum value among all possible paths γ.

A graphical model of the two time series comparison is shown in the curve correspondence graph (Fig. 3), where the optimal DTW path is shown as a line that connect corresponding points on the well logging curves.

When analyzing multiple time series for each well, in this case, the neutron gamma logging and gamma logging, the distance matrix represents the sum of the distances for each pair of well logging curves.

Two wells – A and B are considered as an example of algorithm's operation. For each well, a data matrix w is formed, where each row represents the sampling frequency (depth), and each column corresponds to the value of the respective well logging curve:

$$w_A = \begin{bmatrix} NGK_{A1} & GK_{A1} \\ NGK_{A2} & GK_{A2} \\ \vdots & \vdots \\ NGK_{An} & GK_{An} \end{bmatrix} \quad w_B = \begin{bmatrix} NGK_{B1} & GK_{B1} \\ NGK_{B2} & GK_{B2} \\ \vdots & \vdots \\ NGK_{Bm} & GK_{Bm} \end{bmatrix}$$

Next, a distance matrix D_{AB} of $(n+1)\times(m+1)$ size is created, where $n\,u\,m$ is the number of values in the matrices $w_A\,u\,w_B$ respectively.

Each (i,j) element in the matrix D_{AB} (except the first column and first row) is calculated by a recurrent relation that includes the spatial distance between points and the accumulated distance from previous paths:

$$D_{\mathrm{AB}}(i,j) = d\left(\mathrm{NGK}_{\mathrm{A}}^{i}, \mathrm{NGK}_{\mathrm{B}}^{j}\right) + d\left(\mathrm{GK}_{\mathrm{A}}^{i}, \mathrm{GK}_{\mathrm{B}}^{j}\right) +$$

$$+ \min\left(\sum_{(p,q)\in\Gamma} w_{(p,q)} \cdot D_{\mathrm{AB}}\,(i-\mathrm{p}, j-q)\right). \qquad (4)$$

The formula components are the following:

1) $D_{AB}(i,j)$ is an element of the distance matrix D_{AB}, the cumulative distance between two wells A and B in positions i and j respectively;

2) $d\left(NGK_A^i, NGK_B^j\right)$ is a spatial distance between parameter NGK for well A in position i and for well B in position j;

3) $d\left(GK_A^i, GK_B^j\right)$ is a spatial distance between parameter GK for well A in position i and for well B in position j;

4) $\min\left(\sum_{(\mathrm{p},q)\in\Gamma} w_{(\mathrm{p},q)} \cdot D_{\mathrm{AB}}\,(i-p, j-q)\right)$ is the sum minimum of weighted cumulative distances for all valid transitions (p,q) from the set Γ, where Γ is the set of valid transitions in the step template; (p,q) is a pair of indices defining a specific transition in the step template; $w_{(p,q)}$ is a transition weight (p,q) in the step template; $D_{AB}(i-p, j-q)$ are the values from previous steps in the distance matrix D_{AB}.

The step pattern determines how the algorithm moves through the distance matrix. To find the optimal alignment between two series, it specifies the allowed transitions between time series elements and their weights. Patterns can be defined individually according to the specific tasks and data characteristics. To solve the problem of well logging curve matching, the Rabiner–Juang template was applied [38]. Although the template is standard, it can be modified and adapted to fit specific tasks and data characteristics. In Python, this template is defined by the rabinerJuangStepPattern (type, slope.weighting) function, where the first argument is the step pattern type, that

determines the valid transitions between time series elements and their weights. The second argument (slope.weighting) is a parameter that determines how the transitions will be weighted. The classification of Rabiner–Juang step patterns includes seven main templates, each having four subtypes of weights. Thus, we have a total of 28 potential transition templates, however, not every combination is meaningful in the context of a particular task.

After the distance matrix is calculated, the optimal alignment step is determined. This step is a sequence of index pairs (i, j), which minimizes the total cumulative distance between two time series:

$$(i-p, j-q) = arg \min_{(p,q)\in\Gamma} \left(w_{(p,q)} \cdot D_{AB}(i-p, j-q) \right). \quad (5)$$

Thus, finding the optimal step is a process of backtracking through the distance matrix, where at each step the next element is selected that minimizes the sum of the distances with given weights and previous steps. This process continues until the initial element of the matrix is reached.

The final step is to find the total DTW distance, which is the cumulative distance between wells A and B. This value is found in the last element of the distance matrix D_{AB}, specifically in $D_{AB}(n, m)$, where n, m are the lengths of time series in wells A and B, respectively.

The formula for calculating the total distance in DTW is the same as for the matrix elements, but in this case the final element of the matrix is taken:

$$D_{AB}(n, m) = d(NGK_A^n, NGK_B^m) + d(GK_A^n, GK_B^m) +$$

$$+\min \left(\sum_{(p,q)\in\Gamma} w_{(p,q)} \cdot D_{AB}\,(n-p, m-q) \right), \quad (6)$$

where $D_{AB}(n, m)$ is the total cumulative distance between the time series of wells A and B.

The distance matrix using multiple time series is a square matrix where each element contains a measure of the distance between corresponding pairs of time series (Fig. 4).

The step pattern used in this work is defined by the function of rabinerJuangStepPattern (4, "c") (Fig. 5). Type 4 is focused on anomaly detection, while subtype “c” includes weighting functions that can account for global trends and anomalies. This combination allows for anomaly analysis considering global patterns and identifying complex relationships, which is crucial in the context of interwell correlation.

After calculating the DTW distance for all well pairs, a square matrix of size $N \times N$ is created, where N is the number of wells. The matrix elements are initialized with values corresponding to the DTW distances between well pairs (Fig. 6). This matrix will be used later for clustering the wells on their similarity.

It is worth noting that using the DTW algorithm for clustering and for best matching of well logging curves has different approaches:

Clustering:

– objective: well grouping based on minimum DTW distance;

– method: the DTW distance matrix is used to divide wells into clusters using the K-means algorithm;

– results: each well receives a cluster label, which is determined by its overall similarity to other wells in the cluster.

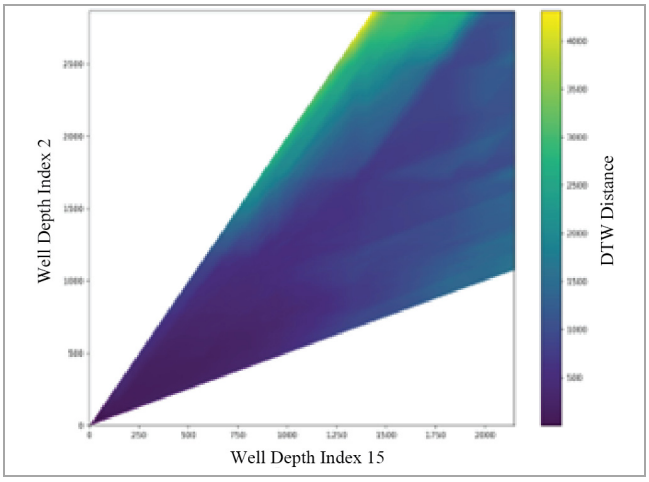

Fig. 4. Distance matrix between a pair of wells using gamma and neutron gamma logging

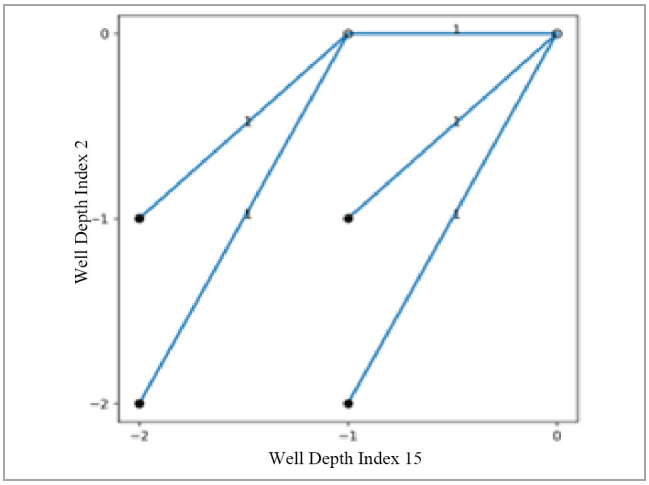

Fig. 5. RabinerJuangStepPattern (4, "c"). By selecting the step pattern (number) and weights (letter), different combinations of traversal can be specified, depending on the configuration of the compared time series

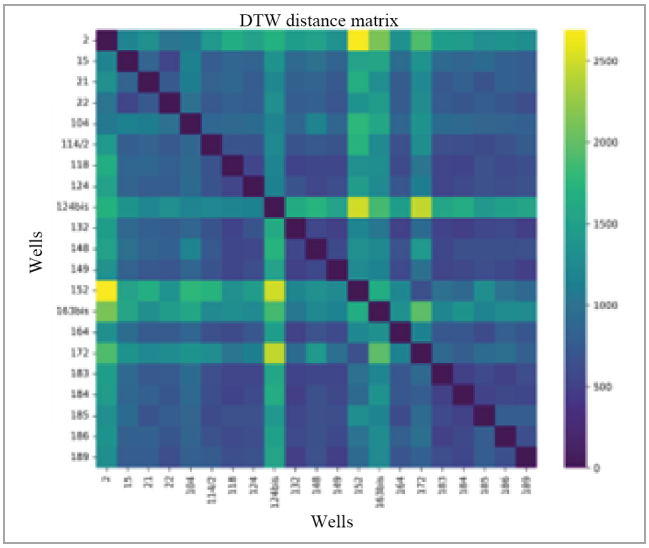

Fig. 6. DTW distance matrix for all pairs of wells

Selecting the best match between reference and target wells:

– objective: to find a reference well that is most similar to the target well;

– method: comparison of each target well with each reference well;

– results: the best well is selected on the minimum DTW distance, regardless of the cluster label.

To summarize, DTW is used in clustering to analyze global patterns for grouping wells by similarity. Whereas local similarity between two specific wells is assessed to select the best well, regardless of their cluster.

To simplify visualization and reduce computational complexity, the DTW distance matrix is transformed using the Principal Component Analysis (PCA). PCA is a statistical method for transformation of the original variables into new ones called principal components, that are uncorrelated with one another and ordered by their variance. It allows the most important data to be preserved in a lower-dimensional space [39, 40].

Formally, PCA is performed with multiplying the original data matrix X by the weight matrix W, obtained from the eigenvectors of the covariance matrix X. The result is a transformed data matrix Z:

$$Z = XW. \quad (7)$$

One of the key aspects for applying the Principal Component Analysis is determining the optimal number of principal components specified in the process of calculating the covariance matrix eigenvectors. In our case, we retain two principal components, as the primary goal is to visually interpret and analyze well clustering, which is much easier to represent in two-dimensional space. In addition, in this paper, two principal components explain a significant part of the total data variance, specifically 73 %, which compromise the information retention and data simplification. The proportion of explained variance for the first component is 48 %, and for the second component is 25 %.

Before applying clustering, it is necessary to determine the optimal number of clusters. The elbow and the silhouette method are proposed to solve the task [41–43]. The elbow method is based on the sum of squared distances within the clusters (SSE), while the silhouette method assess the clustering quality for each object. Using both methods allows for a more accurate determination of the optimal clusters number and improves clustering results.

Formula for SSE calculation is as follows:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} (x - \mu_i)^2, \quad (8)$$

where k is a number of clusters; C_i is the i-th cluster; x is a data point in the cluster C_i; μ_i is the cluster centroid C_i.

The elbow method algorithm is as follows: for each k from 1 to a specified maximum value (in our case, 10 clusters) clustering is performed, SSE is calculated for each k, a dependence graph of SSE from k is constructed. The optimal number of clusters is determined by the point at which the curve begins to level off (Fig. 7, *a*).

The silhouette coefficient $s(i)$ for the object i is calculated with the formula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}, \quad (9)$$

where $a(i)$ is the average distance from the object i to all other objects within the cluster; $b(i)$ is the average distance from the object i to all objects in the nearest neighbor cluster.

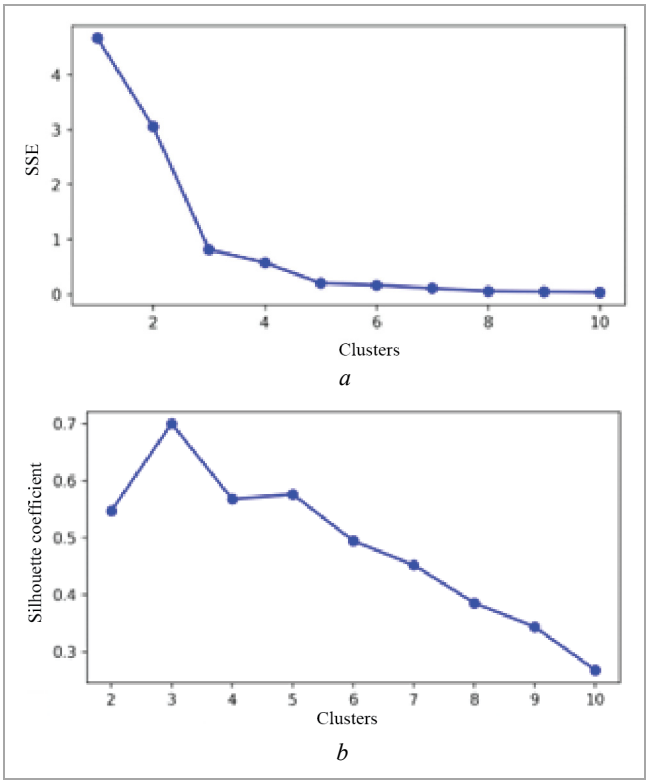

Fig. 7. Graphs: *a* – elbow method; *b* – silhouette method

The silhouette method algorithm consists of the following steps: for each k from 2 to a specified maximum value, clustering is performed, then the average silhouette coefficient is calculated for all objects at the given k. The graph of the average silhouette coefficient as a function of k is constructed, and the optimal number of clusters is determined by the maximum value of the average silhouette coefficient (Fig. 7, *b*).

As a result of applying the elbow and the silhouette method, it is established that three clusters are optimal, as further increasing the number of clusters does not significantly improve the clustering quality. However, the graphs show a certain response to the value of five clusters, which may indicate a more complex data structure related to geological features as stratigraphic unconformities, faults, and changes in sedimentation conditions. These factors can introduce uncertainty into the interwell correlation. In this case, using five clusters may help to form a more representative sample of training wells that will cover a broader range of well log behavior, and, thereby, increasing the accuracy of the correlation process.

Next, clustering is performed using the K-means method [44]. k centroids are randomly chosen, and then each data point is assigned to the nearest centroid, formally expressed as:

$$c_i = arg \min_j \|x_i - \mu_j\|^2, \quad (10)$$

where c_i is the cluster index for the point x_i; μ_j is the cluster centroid of j.

Thereafter, the centroids are recalculated as the average value of all points in the corresponding cluster:

$$\mu_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i, \quad (11)$$

where C_j is a set of points assigned to a cluster j; $C_j \vee$ is the number of points in this cluster.

The steps of assigning points to clusters and recalculating centroids are repeated until convergence, that is, until the centroids no longer change (Fig. 8).

As a result of clustering, most wells formed two compact clusters, indicating a similar pattern of the well logging curves within these groups. However, several wells were separated into clusters, significantly different from the bulk of the data.

Such separation may be caused by various factors. Firstly, geological features such as tectonics or sedimentation conditions may have led to the formation of atypical well logging curves in these wells. Secondly, it is possible that the quality of well logging data in these wells is unsatisfactory, that may also have led to "outliers". The identified unique well groups require more careful study and individual analysis.

To assess the quality of clustering, the Davies – Bouldin Index and the Calinski – Harabasz Index are used [45].

The Davis – Boldin Index (DBI) is calculated as the average ratio value of intra-cluster distance to inter-cluster distance:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{S_i + S_j}{d(c_i, c_j)}\right), \quad (12)$$

where k is the number of clusters; S_i and S_j is intra-cluster distance for clusters i and j respectively; $d(c_i, c_j)$ is the inter-cluster distance between cluster centroids i and j.

The Calinski – Harabasz index (CHI) is calculated as the ratio of inter-cluster variance to intra-cluster variance:

$$CHI = \frac{BGSS}{WGSS} \cdot \frac{n-k}{k-1}, \quad (13)$$

where $BGSS$ is inter-cluster sum of squares; $WGSS$ is intra-cluster sum of squares; k is a number of clusters.

As a result of the data clustering using the K-means method, the Davies – Bouldin index was 0.37, indicating low intra-cluster dispersion and high inter-cluster distance, which is a sign of good clustering quality. The Calinski – Harabasz index is 88.90, which also indicates good separability of clusters.

After determining the minimum sample of wells, the best ratio between the "reference well – target well" pairs is calculated by the optimal step on the distance matrix using the algorithm described earlier. Then a pair of wells within the minimum DTW distance is selected.

The determination of layer intersections in target wells is based on previously known expert points in reference wells. For each point of layer intersection in the reference well, the corresponding point in the target well is found with indices from the DTW alignment matrix.

First, the nearest point to the specified depth of d_k layer intersection is found for the reference well. It is performed by calculating the minimum difference between a given depth and all depths in a reference well:

$$i_{\min} = \text{argmin}|D_1 - d_k|, \quad (14)$$

where D_1 is the array of depth values for the reference well; $i_{\min}$ is the index of the nearest point in the array of reference well depths to the specified depth.

Then the $DTW(i,j)$ matrix is used. The matrix contains the information of the indices in the reference well corresponded to the indices in the target well.

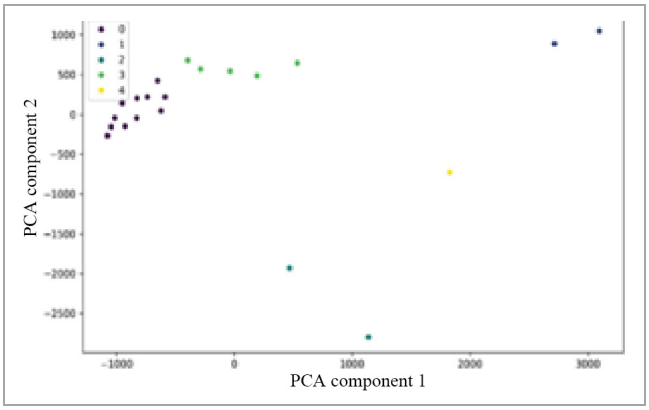

Fig. 8. Well clustering with PCA

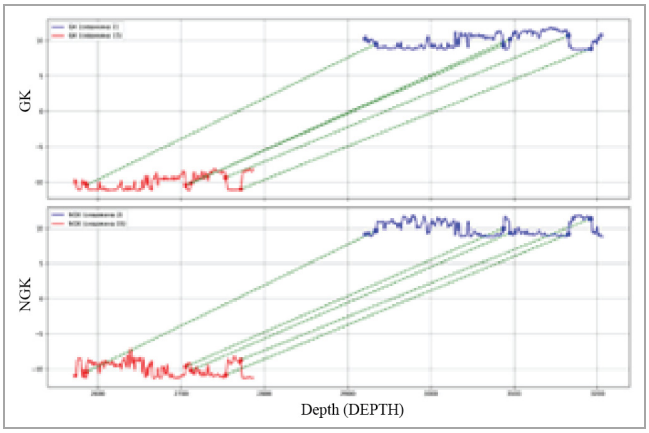

Fig. 9. Correspondence of layer intersection points

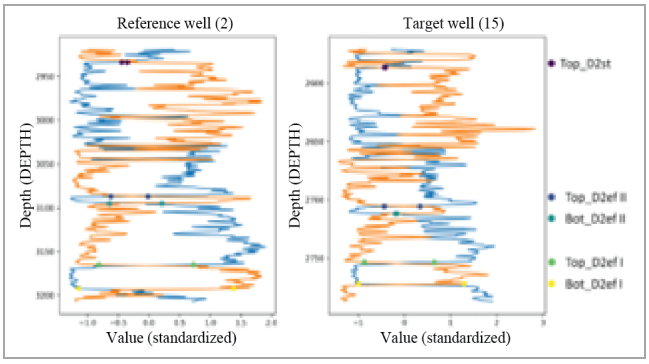

Fig. 10. Correlation results for wells 2 and 15

The corresponding point in the target well is found using the $i_{\min}$ index:

$$j_{\min} = \text{DTW}(i_{\min}, j), \quad (15)$$

where $j_{\min}$ is index of the corresponding point in the target well.

Thus, the search layer intersection in the target well is at the depth of $D_2[j_{\min}]$, where D_2 is an array of depth values for the target well (Fig. 9).

This algorithm is performed for each pair of wells, resulting in calculated layer intersections for the target wells (Fig. 10).

Analysis of results

To assess the results, a comparative analysis was performed between expert and algorithm-calculated layer intersections. The total number of layer intersections for

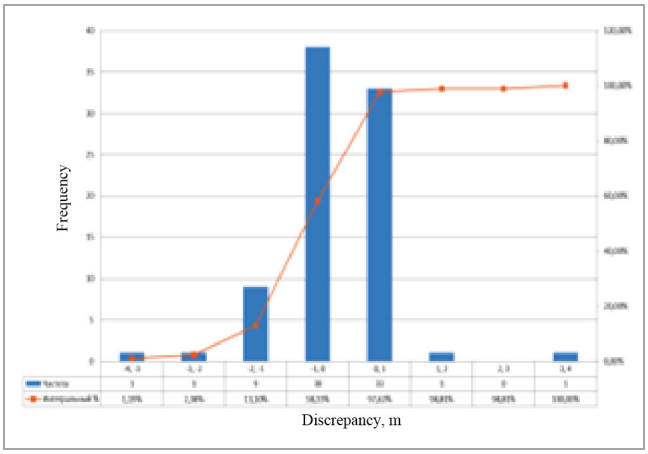

Fig. 11. Discrepancies between expert and calculated layer intersections

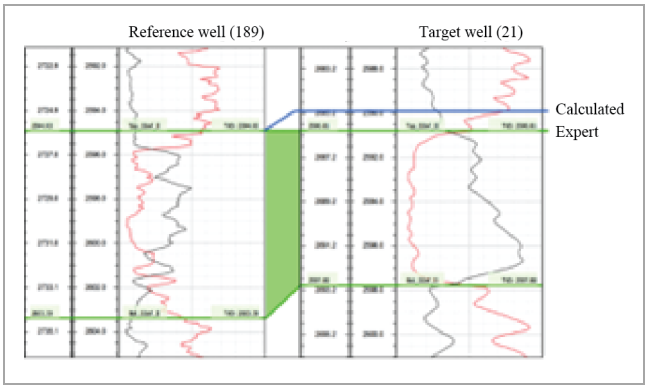

Fig. 12. Correlation scheme of the D2ef-II formation (wells 189 and 21)

which the algorithm was tested on amounts to 84. The distribution of discrepancies across ranges showed that the majority of the results (84.53 %) are concentrated within the range of –1…1, indicating a high consistency with expert marks in most cases (Fig. 11).

However, there are significant discrepancies. A detailed analysis of the wells with substantial discrepancies (>1 m) revealed that not in all cases the errors can be attributed to the algorithm's performance.

As an example, a pair of wells 189 and 21 is considered (Fig. 12). For well 21, the discrepancy between the calculated and expert depths of the D2ef-II formation top is 3.1 m.

In the reference and target wells, the expert applied different methods of identifying the top of the D2ef-II formation, which led to discrepancies in the results. In contrast, the DTW algorithm demonstrated its effectiveness by correctly identifying the points according to the training template, regardless of subjective factors. This case emphasizes that the algorithm can identify errors being unnoticed by experts. Thus, the DTW algorithm improves the accuracy and objectivity of data interpretation, and serves as a tool for identifying and correcting subjective human errors.

Conclusion

The approach based on the dynamic time warping algorithm, demonstrates high efficiency and accuracy in inter-well correlation performed automatically. This approach significantly reduces time costs, increases the objectivity of results and identifies human errors in expert interpretation, making it promising for widespread practical application in geological and geophysical fields, especially for large data volumes.

References

1. Keogh E., Pazzani M. Scaling up Dynamic Time Warping for Datamining Applications. *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, pp. 285-289. DOI: 10.1145/347090.347153
2. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 2011, no. 12, pp. 2825-2830. DOI: 10.48550/arXiv.1201.0490
3. McKinney W. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 2010, pp. 1-6. DOI: 10.25080/Majora-92bf1922-00a
4. Hunter J.D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 2007, no. 9 (3), pp. 90-95.
5. Semenov I.A., Shishkin M.D. Ispol'zovanie vozmozhnostei biblioteki Matplotlib dlia vizualizatsii dannykh v rabote inzhenerov i issledovatelei [Using the potential of the Matplotlib library for data visualization in the work of engineers and researchers]. *Sovremennye tekhnologii i nauchno-tekhnicheskii progress*, 2024, no. 11, pp. 169-170.
6. Tarshkhoeva Zh.T. Iazyk programmirovaniia Python. Biblioteki Python [Python programming language. Python libraries]. *Molodoi uchenyi*, 2021, no. 5 (347), pp. 20-21.
7. Galimullin N.R. Python: ispol'zovanie Python dlia avtomatizatsii povsednevnykh zadach [Python: using Python to automate everyday tasks]. *Ekonomika i upravlenie: problemy, resheniia*, 2024, vol. 9, no. 9 (150), pp. 69-76. DOI: 10.36871/ek.up.p.r.2024.09.09.010
8. Efremov E.A. Razrabotka graficheskikh interfeisov s pomoshch'iu graficheskoi biblioteki Tkinter [Developing GUIs with the Tkinter GUI Library]. *Molodezhnyi nauchno-tekhnicheskii vestnik*, 2017, no. 6, 44 p.
9. Driankova D.A. Obrabotka propushchennykh znachenii i dublikatov v dannykh s pomoshch'iu biblioteki Pandas dlia iazyka programmirovaniia Python [Processing of missing values and duplicates in data using the Pandas library for the Python programming language]. *Dnevnik nauki*, 2023, no. 6 (78). DOI: 10.51691/2541-8327_2023_6_11
10. Van der Walt S., Colbert S.C., Varoquaux G. The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering*, 2011, no. 13 (2), pp. 22-30. DOI: 10.1109/MCSE.2011.37
11. Harris C.R., Millman K.J., Van der Walt S.J. et al. Array programming with NumPy. *Nature*, 2020, no. 585 (7825), pp. 357-362. DOI: 10.1038/s41586-020-2649-2
12. Li Ya., Shao Yu., Fu Ch. Exploring Indexing and Slicing in NumPy Arrays. *Journal of Physics: Conference Series*, 2024, vol. 2832, no. 1, 012019 p. DOI: 10.1088/1742-6596/2832/1/012019
13. Waskom M. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 2021, vol. 6, no. 60, 3021 p. DOI: 10.21105/joss.03021
14. Xu Ch., Fu L., Lin T., Li W., Ma Sh. Machine learning in petrophysics: Advantages and limitations. *Artificial Intelligence in Geosciences*, 2022, no. 3. DOI: 10.1016/j.aiig.2022.11.004
15. Kostikov D.V., Lariukhin A.I., Lialin V.E. Sravnenie metodov umen'sheniia obuchaiushchei vyborki pri interpretatsii dannykh geofizicheskikh issledovanii skvazhin s pomoshch'iu neironnoi seti [Comparison of methods for reducing the training sample in the interpretation of well logging data using a neural network]. *Trudy mezhdunarodnogo simpoziuma "Nadezhnost' i kachestvo"*, 2007, vol. 1, C. 118–123.
16. Gulin A.B. Primenenie metodov mashinnogo obucheniia dlia interpretatsii dannykh geofizicheskikh issledovanii skvazhin [Application of machine learning methods to interpret well logging data]. *Problemy razrabotki mestorozhdenii uglevodorodnykh i rudnykh poleznykh iskopaemykh*, 2022, vol. 1, pp. 41-45.
17. Kovalevskii E.V., Gogonenkov G.N., Perepechkin M.V. Utochnenie geologicheskikh modelei posredstvom ispol'zovaniia avtomaticheskoi korreliatsii skvazhin [Refinement of geological models using automatic well correlation]. *Nedropol'zovanie XXI vek*, 2007, no. 4, pp. 28-31.
18. Lineman D.J., Mendelson J.D., Toksos M.N. Well to well log correlation using knowledge-based systems and dynamic depth warping. *SPWLA 28th Annual Logging Symposium "High-impact palynology in petroleum geology: applications from Venezuela (Northern South America)" (June 29 – July 2, 1987)*. AAPG Bull., 1987, vol. 86, pp. 421-453.
19. Azur M.J., Stuart E.A., Frangakis C., Leaf P.J. Multiple Imputation by Chained Equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, 2011, vol. 20, no. 1, pp. 36-49. DOI: 10.1002/mpr.329
20. Siddique J., Belin T.R. A Comparison of Six Methods for Missing Data Imputation. *Journal of Biometrics & Biostatistics*, 2008, vol. 3, no. 2, pp. 1-15. DOI: 10.4172/2155-6180.1000224

21. Wu R., Hamshaw S.D., Yang L. et al. Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data. *IEEE Sensors Journal*, 2022, vol. 22, no. 11, pp. 10671-10683. DOI: 10.1109/jsen.2022.3166643
22. Chen Y., Li D., Li Y., Ma X., Wei J. Use Moving Average Filter to Reduce Noises in Wearable PPG During Continuous Monitoring. *Lecture Notes of the Institute for Computer Sciences*, 2016, no. 181, pp. 193-203. DOI: 10.1007/978-3-319-49655-9_26
23. Soltanian M., Ghaemmaghami Sh. Recognition of Visual Events using Spatio-Temporal Information of the Video Signal. *Signal and Data Processing*, 2021, vol. 18, no. 1, pp. 134-119. DOI: 10.52547/jsdp.18.1.134
24. Al-Faiz M.Z., Ibrahim A.A., Hadi S.M. The Effect of Z-Score Standardization (Normalization) on Binary Input Due the Speed of Learning in Back-Propagation Neural Network. *Ira. J. Inf. Commun. Technol.*, 2018, vol. 1, pp. 42-48. DOI: 10.31987/ijict.1.3.41
25. Al Shalabi L., Shaaban Z. Normalization as a preprocessing engine for data mining and the approach of preference matrix. *International Conference on Dependability of Computer Systems, 25-27 May 2006*. Szklarska Poreba, 2006, pp. 207-214. DOI: 10.1109/DEPCOS-RELCOMEX.2006.38
26. Salvador S., Chan P. Fast DTW: Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 2004, no. 11 (5), pp. 70-80.
27. Kate RJ. Using dynamic time warping distances as features for improved time series classification. *Data Mining and Knowledge Discovery*, 2016, vol. 30, no. 2, pp. 283-312. DOI: 10.1007/s10618-015-0418-x
28. Sylvester Z. Automated multi-well stratigraphic correlation and model building using relative geologic time. *Basin Research*, 2023, vol. 35, no. 5, pp. 1961-1984. DOI: 10.1111/bre.12787
29. Sharafutdinov T.R., Shaibakov R.A. Aprobatsiia algoritma avtokorreliatsii dannykh geofizicheskikh issledovanii skvazhin na primere otlozhenii achimovskoi tolshchi Pravdinskogo mestorozhdeniia [Testing an algorithm of the well log data autocorrelation on the example of Achimov sequence of Pravdinskoye field]. *Nauchno-tekhnicheskii vestnik OAO "NK "Rosneft'"*, 2012, no. 1, pp. 18-22.
30. Butorin A.V., Sevost'ianov A.I., Stulikov S.K., Timirgalin A.A. Avtomatizirovannaia uviazka seismicheskikh i skvazhinnykh dannykh s ispol'zovaniem algoritma Dynamic Time Warping [DTW based automated seismic–well tie]. *Neftianoe khoziaistvo*, 2019, no. 12, pp. 30-32. DOI: 10.24887/0028-2448-2019-12-30-32
31. Sakoe H., Chiba S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1978, vol. 26, iss. 1, pp. 43-49. DOI: 10.1109/TASSP.1978.1163055
32. Zakrevskii K.E., Gazizov R.K., Ryzhikov E.A., Freidin K.V. Tekhnologiia otsenki soglasovannosti rezul'tatov avtomaticheskoi korreliatsii skvazhin po dannym geofizicheskikh issledovanii [Consistency evaluation technology for automatic well-log correlation using well logging data]. *Neftianoe khoziaistvo*, 2021, no. 8, pp. 22-26. DOI: 10.24887/0028-2448-2021-8-22-26
33. Latypov I.D., Shagimardanova L.R., Markov A.V., Evgrafov N.A. Razrabotka podkhodov k avtomatizirovannoi vnutriplastovoi korreliatsii po dannym geofizicheskikh issledovanii skvazhin s primeneniem mashinnogo obucheniia [Development of approaches to automated correlation from well log data using machine learning]. *Ekspozitsiia Neft' Gaz*, 2024, no. 4 (105), pp. 47-51. DOI: 10.24412/2076-6785-2024-4-47-51
34. Jiang Y., Qi Y., Wang W.K. et al. Event DTW: An improved Dynamic Time Warping algorithm for aligning biomedical signals of nonuniform sampling frequencies. *Sensor*, 2020, vol. 20 (9), pp. 1-13. DOI: 10.3390/s20092700
35. Cai W., Ding J., Li Zh. et al. Study on Rock Mechanics Parameter Prediction Method Based on DTW Similarity and Machine-Learning Algorithms. *Petrophysics*, 2024, vol. 65, no. 1, pp. 128-144. DOI: 10.30632/pjv65n1-2024a7
36. Hladil J., Cejchan P., Koptikova L. et al. The dynamic time-warping approach to comparison of magnetic-susceptibility logs and application to lower devonian calciturbidites (Prague Synform, Bohemian Massif). *Geologica Belgica*, 2010, vol. 13, no. 4, pp. 385-406.
37. Romanenko A.A. Vyravnivanie vremennykh riadov: prognozirovanie s ispol'zovaniem DTW [Time Series Alignment: Forecasting Using DTW]. *Mashinnoe obuchenie i analiz dannykh*, 2011, vol. 1, no. 1, pp. 77-85.
38. Rabiner L., Juang B. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 1986, vol. 3, no. 1, pp. 4-16. DOI: 10.1109/MASSP.1986.1165342
39. Denisenko A.A. Issledovanie metodov sokrashcheniia razmernosti v zadache raspoznavaniia obrazov [Research of dimensionality reduction methods in the pattern recognition problem]. *Mezhdunarodnyi zhurnal prikladnykh i fundamental'nykh issledovanii*, 2020, no. 12, pp. 65-69. DOI: 10.17513/mjpfi.13161
40. Basna H., Adnan A. Journal of Soft Computing and Data Mining A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Applied Soft Computing*, 2021, no. 2, pp. 20-30. DOI: 10.30880/jscdm.2021.02.01.003
41. Bova V.V., Kravchenko Iu.A., Rodzin S.I. Metody i algoritmy klasterizatsii tekstovykh dannykh (obzor) [Methods and algorithms for text data clustering (review)]. *Izvestiia IuFU. Tekhnicheskie nauki*, 2022, no. 4 (228), pp. 122-143. DOI: 10.18522/2311-3103-2022-4-122-143
42. Orlov V.I., Kazakovtsev L.A., Masich I.S. Primenenie kriteriia silueta v algoritme avtomaticheskoi gruppirovki elektroradioizdelii kosmicheskogo primeneniia [Silhouette criterion for automatic grouping algorithm of spaceship electronic components]. *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta imeni akademika M.F. Reshetneva*, 2016, vol. 17, no. 4, pp. 883-890.
43. Umargono E., Suseno J., Gunawan S.K. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. *The 2nd International Seminar on Science and Technology*, 2019. DOI: 10.2991/assehr.k.201010.019
44. Hartigan J.A., Wong M.A. A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, vol. 28, iss. 1, pp. 100-108. DOI: 10.2307/2346830
45. Firman Ashari I., Nugroho E., Baraku R., Yanda I., Liwardana R. Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta. *Journal of Applied Informatics and Computing*, 2023, no. 7, pp. 89-97. DOI: 10.30871/jaic.v7i1.4947

Библиографический список

1. Keogh, E. Scaling up Dynamic Time Warping for Datamining Applications / E. Keogh, M. Pazzani // Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. – 2002. – P. 285–289. DOI: 10.1145/347090.347153
2. Scikit-learn: Machine learning in Python / F. Pedregosa, G. Varoquaux, A. Gramfort [et al.]// Journal of Machine Learning Research, – 2011. – No. 12. – P. 2825–2830. DOI: 10.48550/arXiv.1201.0490
3. McKinnev. W. Data structures for statistical computing in Python / W. McKinney // Proceedings of the 9th Python in Science Conference, – 2010. – P. 1–6. DOI: 10.25080/Majora-92bf1922-00a
4. Hunter, J.D. Matplotlib: A 2D graphics environment / J.D. Hunter // Computing in Science & Engineering. – 2007. – No. 9 (3). – P. 90–95.
5. Семенов, И.А. Использование возможностей библиотеки Matplotlib для визуализации данных в работе инженеров и исследователей / И.А. Семенов, М.Д. Шишкин // Современные технологии и научно-технический прогресс. – 2024.– № 11. – С. 169–170.
6. Таршхоева, Ж.Т. Язык программирования Python. Библиотеки Python / Ж.Т. Таршхоева // Молодой ученый. – 2021. – № 5 (347). – С. 20–21.
7. Галимуллин, Н.Р. Python: использование Python для автоматизации повседневных задач / Н.Р. Галимуллин // Экономика и управление: проблемы, решения. – 2024. – Т. 9, № 9 (150). – С. 69–76. DOI: 10.36871/ek.up.p.r.2024.09.09.010
8. Ефремов, Е.А. Разработка графических интерфейсов с помощью графической библиотеки Tkinter / Е.А. Ефремов // Молодежный научно-технический вестник. – 2017. – № 6. – С. 44.
9. Дрянкова, Д.А. Обработка пропущенных значений и дубликатов в данных с помощью библиотеки Pandas для языка программирования Python / Д.А. Дрянкова // Дневник науки. – 2023. – № 6 (78). DOI: 10.51691/2541-8327_2023_6_11
10. Van der Walt, S. The NumPy array: A structure for efficient numerical computation / S. Van der Walt, S.C. Colbert, G. Varoquaux // Computing in Science & Engineering. – 2011. – No. 13(2). – P. 22–30. DOI: 10.1109/MCSE.2011.37
11. Array programming with NumPy / C.R. Harris, K.J. Millman, S.J. Van der Walt [et al.] // Nature. – 2020. – No. 585 (7825). – P. 357–362. DOI: 10.1038/s41586-020-2649-2
12. Li, Ya. Exploring Indexing and Slicing in NumPy Arrays / Ya. Li, Yu. Shao, Ch. Fu // Journal of Physics: Conference Series. – 2024. – Vol. 2832, no. 1. – P. 012019. DOI: 10.1088/1742-6596/2832/1/012019
13. Waskom, M. Seaborn: Statistical data visualization / M. Waskom // Journal of Open Source Software. – 2021. – Vol. 6, no. 60. – P. 3021. DOI: 10.21105/joss.03021
14. Machine learning in petrophysics: Advantages and limitations / Ch. Xu, L. Fu, T. Lin, W. Li, Sh. Ma // Artificial Intelligence in Geosciences. – 2022. – No. 3. DOI: 10.1016/j.aiig.2022.11.004
15. Костиков, Д.В. Сравнение методов уменьшения обучающей выборки при интерпретации данных геофизических исследований скважин с помощью нейронной сети / Д.В. Костиков, А.И. Ларюхин, В.Е. Лялин // Труды международного симпозиума «Надежность и качество». – 2007. – Т. 1. – С. 118–123.
16. Гулин, А.Б. Применение методов машинного обучения для интерпретации данных геофизических исследований скважин / А.Б. Гулин // Проблемы разработки месторождений углеводородных и рудных полезных ископаемых. – 2022. – Т. 1. – С. 41–45.
17. Ковалевский, Е.В. Уточнение геологических моделей посредством использования автоматической корреляции скважин / Е.В. Ковалевский, Г.Н. Гогоненков, М.В. Перепечкин // Недропользование XXI век. – 2007. – № 4. – С. 28–31.
18. Lineman, D.J. Well to well log correlation using knowledge-based systems and dynamic depth warping / D.J. Lineman, J.D. Mendelson, M.N. Toksos // SPWLA 28th Annual Logging Symposium "High-impact palynology in petroleum geology: applications from Venezuela (Northern South America)" (June 29 – July 2, 1987). – AAPG Bull., 1987. – Vol. 86. – P. 421–453.
19. Multiple Imputation by Chained Equations: What is it and how does it work? / M.J. Azur, E.A. Stuart, C. Frangakis, P.J. Leaf // International Journal of Methods in Psychiatric Research. – 2011. – Vol. 20, no. 1. – P. 36–49. DOI: 10.1002/mpr.329

20. Siddique, J. A Comparison of Six Methods for Missing Data Imputation / J. Siddique, T.R. Belin // Journal of Biometrics & Biostatistics. – 2008. – Vol. 3, no. 2. – P. 1–15. DOI: 10.4172/2155-6180.1000224
21. Data Imputation for Multivariate Time Series Sensor Data With Large Gaps of Missing Data / R. Wu, S.D. Hamshaw, L. Yang [et al.] // IEEE Sensors Journal. – 2022. – Vol. 22, no. 11. – P. 10671–10683. DOI: 10.1109/jsen.2022.3166643
22. Use Moving Average Filter to Reduce Noises in Wearable PPG During Continuous Monitoring / Y. Chen, D. Li, Y. Li, X. Ma, J. Wei // Lecture Notes of the Institute for Computer Sciences. – 2016. – No. 181. – P. 193–203. DOI: 10.1007/978-3-319-49655-9_26
23. Soltanian, M. Recognition of Visual Events using Spatio-Temporal Information of the Video Signal / M. Soltanian, Sh. Ghaemmaghami // Signal and Data Processing. – 2021. – Vol. 18, no. 1. – P. 134–119. DOI: 10.52547/jsdp.18.1.134
24. Al-Faiz, M.Z. The Effect of Z-Score Standardization (Normalization) on Binary Input Due the Speed of Learning in Back-Propagation Neural Network / M.Z. Al-Faiz, A.A. Ibrahim, S.M. Hadi // Ira. J. Inf. Commun. Technol. – 2018. – Vol. 1. – P. 42–48. DOI: 10.31987/ijict.1.3.41
25. Al Shalabi, L. Normalization as a preprocessing engine for data mining and the approach of preference matrix / L. Al Shalabi, Z. Shaaban // International Conference on Dependability of Computer Systems, 25–27 May 2006. – Szklarska Poreba, 2006. – P. 207–214. DOI: 10.1109/DEPCOS-RELCOMEX.2006.38
26. Salvador, S. Fast DTW: Toward accurate dynamic time warping in linear time and space / S. Salvador, P. Chan // Intelligent Data Analysis. – 2004. – No. 11 (5). – P. 70–80.
27. Kate, RJ. Using dynamic time warping distances as features for improved time series classification / RJ. Kate // Data Mining and Knowledge Discovery. – 2016. – Vol. 30, no. 2. – P. 283–312. DOI: 10.1007/s10618-015-0418-x
28. Sylvester, Z. Automated multi-well stratigraphic correlation and model building using relative geologic time / Z. Sylvester // Basin Research. – 2023. – Vol. 35, no. 5. – P. 1961–1984. DOI: 10.1111/bre.12787
29. Шарафутдинов, Т.Р. Апробация алгоритма автокорреляции данных геофизических исследований скважин на примере отложений ачимовской толщи Правдинского месторождения / Т.Р. Шарафутдинов, Р.А. Шайбаков // Научно-технический вестник ОАО «НК «Роснефть». – 2012. – № 1. – С. 18–22.
30. Автоматизированная увязка сейсмических и скважинных данных с использованием алгоритма Dynamic Time Warping / А.В. Буторин, А.И. Севостьянов, С.К. Стуликов, А.А. Тимиргалин // Нефтяное хозяйство. – 2019. – № 12. – С. 30–32. – DOI: 10.24887/0028-2448-2019-12-30-32
31. Sakoe, H. Dynamic programming algorithm optimization for spoken word recognition / H. Sakoe, S. Chiba // IEEE Transactions on Acoustics, Speech and Signal Processing. – 1978. – Vol. 26, iss.1. – P. 43–49. DOI: 10.1109/TASSP.1978.1163055
32. Технология оценки согласованности результатов автоматической корреляции скважин по данным геофизических исследований / К.Е. Закревский, Р.К. Газизов, Е.А. Рыжиков, К.В. Фрейдин // Нефтяное хозяйство. – 2021. – № 8. – С. 22–26. DOI: 10.24887/0028-2448-2021-8-22-26
33. Разработка подходов к автоматизированной внутрипластовой корреляции по данным геофизических исследований скважин с применением машинного обучения / И.Д. Латыпов, Л.Р. Шагимарданова, А.В. Марков, Н.А. Евграфов // Экспозиция Нефть Газ. – 2024. – № 4 (105). – С. 47–51. DOI: 10.24412/2076-6785-2024-4-47-51
34. Event DTW: An improved Dynamic Time Warping algorithm for aligning biomedical signals of nonuniform sampling frequencies / Y. Jiang, Y. Qi, W.K. Wang [et al.] // Sensor. – 2020. – Vol. 20 (9). – P. 1–13. DOI: 10.3390/s20092700
35. Study on Rock Mechanics Parameter Prediction Method Based on DTW Similarity and Machine-Learning Algorithms / W. Cai, J. Ding, Zh. Li [et al.] // Petrophysics. – 2024. – Vol. 65, no. 1. – P. 128–144. DOI: 10.30632/pjv65n1-2024a7
36. The dynamic time-warping approach to comparison of magnetic-susceptibility logs and application to lower devonian calciturbidites (Prague Synform, Bohemian Massif) / J. Hladil, P. Cejchan, L. Koptikova [et al.] // Geologica Belgica. – 2010. – Vol. 13, no. 4. – P. 385–406.
37. Романенко, А.А. Выравнивание временных рядов: прогнозирование с использованием DTW / А.А. Романенко // Машинное обучение и анализ данных. – 2011. – Т. 1, № 1. – С. 77–85.
38. Rabiner, L. An introduction to hidden Markov models / L. Rabiner, B. Juang // IEEE ASSP Magazine. – 1986. – Vol. 3. – No. 1. – P. 4–16. DOI: 10.1109/MASSP.1986.1165342
39. Денисенко, А.А. Исследование методов сокращения размерности в задаче распознавания образов / А.А. Денисенко // Международный журнал прикладных и фундаментальных исследований. – 2020. – № 12. – С. 65–69. DOI: 10.17513/mjpfi.13161
40. Basna, H. Journal of Soft Computing and Data Mining A Review of Principal Component Analysis Algorithm for Dimensionality Reduction / H. Basna, A. Adnan // Applied Soft Computing. – 2021. – No. 2. – P. 20–30. DOI: 10.30880/jscdm.2021.02.01.003
41. Бова, В.В. Методы и алгоритмы кластеризации текстовых данных (обзор) / В.В. Бова, Ю.А. Кравченко, С.И. Родзин // Известия ЮФУ. Технические науки. – 2022. – № 4 (228). – С. 122–143. DOI: 10.18522/2311-3103-2022-4-122-143
42. Орлов, В.И. Применение критерия силуэта в алгоритме автоматической группировки электрорадиоизделий космического применения / В.И. Орлов, Л.А. Казаковцев, И.С. Масич // Вестник Сибирского государственного аэрокосмического университета им. академика М.Ф. Решетнева. – 2016. – Т. 17, № 4. – С. 883–890.
43. Umargono, E. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula / E. Umargono, J. Suseno, S.K. Gunawan // The 2nd International Seminar on Science and Technology. – 2019. DOI: 10.2991/assehr.k.201010.019
44. Hartigan, J.A. A k-means clustering algorithm / J.A. Hartigan, M.A. Wong // Journal of the Royal Statistical Society. Series C (Applied Statistics). – 1979. – Vol. 28, iss. 1. – P. 100–108. DOI: 10.2307/2346830
45. (2023). Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta / I. Firman Ashari, E. Nugroho, R. Baraku, I. Yanda, R. Liwardana // Journal of Applied Informatics and Computing. – 2023. – No. 7. – P. 89–97. DOI: 10.30871/jaic.v7i1.4947

Funding. The study was carried out under the Russian Science Foundation grant No. 22-17-00111.
Conflict of interest. The authors declare no conflict of interest.
The authors' contribution is equivalent.