Petroleum Science and Engineering
2023; 7(2): 35-42
http://www.sciencepublishinggroup.com/j/pse
doi: 10.11648/j.pse.20230702.12
ISSN: 2640-4486 (Print); ISSN: 2640-4516 (Online)

Research on the Application of Bayesian Machine Learning in Reservoir Prediction

Zhiguo Fu[1, *], Kang Chen[2], Juan Liao[3], Long Long[2], Da Peng[2]

[1]School of Geoscience and Technology, Southwest Petroleum University, Chengdu, China

[2]Exploration and Development Research Institute, PetroChina Southwest Oil and Gas Field Company, Chengdu, China

[3]Southwest Institute of Exploration of Geophysics, BGP Inc., CNPC, Chengdu, China

Email address:

fu_zhiguo@aliyun.com (Zhiguo Fu), 202299010021@swpu.edu.cn (Zhiguo Fu)

[*]Corresponding author

To cite this article:

Zhiguo Fu, Kang Chen, Juan Liao, Long Long, Da Peng. (2023). Research on the Application of Bayesian Machine Learning in Reservoir Prediction. *Petroleum Science and Engineering*, *7*(2), 35-42. https://doi.org/10.11648/j.pse.20230702.12

Received: October 28, 2023; **Accepted**: November 21, 2023; **Published**: November 29, 2023

Abstract: As exploration and development of various types of oil and natural gas reservoirs underground continue to evolve, the more complex geological environment for seismic exploration in oil and gas reservoirs become focus increasingly. When the target reservoir is vertically interbedded and nested with various lithological strata, and exhibits poor lateral continuity, it becomes increasingly difficult for humans to distinguish oil and gas reservoirs from such complex backgrounds. It's also challenging to quantitatively assess and determine the accuracy of discrimination and achieve optimal reservoir identification results. In response to this issue, the Bayesian machine learning algorithm is introduced for automated target discrimination, enabling efficient differentiation between dolomite reservoirs, mudstones, and other lithological intercalations. The core of applying the Bayesian classifier is to establish a distribution model for target parameters, which is usually assumed to be a known distribution type such as Gaussian or Cauchy distribution. However, in petroleum seismic exploration, the distribution of oil and gas reservoir parameters is highly irregular and significantly different from these established distribution types, limiting the application of the Bayesian classification method. Therefore, we propose using a radial basis function neural network to estimate the prior distribution probability density of oil and gas reservoir parameters. This approach does not assume the prior distribution to be a certain predetermined model but instead builds the prior distribution model based on the numerical distribution characteristics of the target parameters themselves, enhancing the practicality of the Bayesian classification method. This method replaces manual reservoir identification processes, achieving high-precision, quantitative, and automated discrimination of reservoirs. Applied to actual seismic exploration data in oil fields for gas layer prediction, the discrimination results match with industrial gas wells, demonstrating the feasibility and effectiveness of the method.

Keywords: Machine Learning, Bayes, Classification, Radial Basis Network, Reservoir Prediction

1. Introduction

Using seismic data to obtain geological parameters for underground oil and gas reservoir prediction is a necessary production process in the petroleum exploration industry. After identifying sensitive parameters that reflect the researched wells, constructing feature vectors that quantitatively characterize the objectives to be classified, and selecting and determining class boundaries in the feature space, the goal of identifying reservoir targets and predicting oil and gas can be achieved. However, the majority of software currently used in exploration and production only provides manual classification functions for data processing, such as delineating a linear boundary on the intersection plot of category sample points or delineating an elliptical or polygonal boundary for each category sample point [1-4]. These methods are subjective and significantly influenced by individual understanding, resulting in suboptimal accuracy when manually classifying categories and low work efficiency. Developing advanced and automated methods for classifying geological parameter sample points can efficiently and effectively complete reservoir prediction work, offering significant value in field applications.

Machine learning technologies have evolved from computer pattern recognition methods since the 1970s. [5]. Among them, the Bayesian machine learning classification algorithm has gradually shown good performance in adapting to the number of sample categories and sample capacity. The algorithm has stable classification results, high accuracy, and simple implementation steps, and provides classification results based on automatically calculated data category probabilities without requiring human intervention, theoretically representing the optimal classifier. Bayesian classification has been widely promoted in the field of information processing [6-8], and research has been conducted on the core prior distribution estimation methods [9]. Greenard et al. [10] applied the fast Gaussian transform technique to accelerate the basic kernel density estimation algorithm. Gray et al. [11] proposed a dual-tree recursion algorithm that uses a divide-and-conquer method to implement the basic kernel density estimation algorithm, greatly reducing the computational complexity. The sparse solutions proposed by Hall et al. [12] and Weston et al. [13] reduce the amount of computation by constructing a distribution estimate with a few representative sample points. This is suitable for situations where overall accuracy requirements do not need to be particularly high. Rui Ting et al. [14] described the most relatively stable sample points using a Gaussian model and applied basic kernel density estimation to the remaining small sample points, solving the efficiency problem of classification on dynamic backgrounds. Dong Min et al. [15] solved the kernel density estimation by solving the camera response inverse function through the nonlinear distribution of image edges in image authenticity identification applications. Qiao et al. [16] used the basic kernel density estimation algorithm and histogram for background segmentation and implemented target detection in reverse to meet the real-time requirements of the monitoring system. The above methods mainly improve application efficiency.

There have been only a few applications [17, 18] of Bayesian machine learning classification algorithms in the field of oil and gas exploration, specifically in reservoir prediction. The main challenge lies in estimating the prior distribution of the target categories, which is limited by assumptions about the type of prior distribution, typically assumed to be Gaussian. Here, a radial basis function neural network is used to fit the histogram of the prior distribution of the target categories, thereby obtaining a continuous prior distribution estimate without assuming the distribution type. This approach demonstrates good adaptability to the data and has shown positive results when applied to practical reservoir prediction tasks. Through this research, it is also hoped to contribute to promoting and applying artificial intelligence technology in the petroleum industry.

2. Principle of the Method

2.1. Bayesian Machine Learning Classification Algorithm

The to-be-classified data points are represented with their feature vectors as

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_m \end{pmatrix}^{\mathrm{T}}, \tag{1}$$

where m is the dimension, and x_i, $i = 1, 2, \cdots, m$ is each of component. The probability $P(c_i \mid X)$ of each data point X belonging to each class c_i, $i = 1, 2, \cdots, K$, where K is the number of classes. The Bayesian probability formula is shown in equation (2)

$$P(c_i \mid X) = \frac{P(X \mid c_i) P(c_i)}{P(X)}, \tag{2}$$

$$P(X) = \sum_{i=1}^{K} P(X \mid c_i) P(c_i), \tag{3}$$

where $P(X \mid c_i)$ is the conditional probability of data point X belonging to class c_i, which is the prior distribution; $P(c_i)$ is the probability of class c_i. $P(X)$ is the probability of the data point being X. By calculating equation (2), the probability of the data point belonging to each class is obtained, and the data point is assigned to the class with the highest probability. To calculate $P(c_i \mid X)$, we need to calculate $P(X \mid c_i)$, $P(c_i)$, and $P(X)$ separately. In general classification problems, the number of classes is usually finite and $P(c_i)$ can be estimated from the discrete distribution obtained through sampling. $P(X \mid c_i)$ is typically estimated using a non-parametric kernel density estimation method [19]. Conventional non-parametric kernel density estimation methods have a fixed number of kernels, equal to the number of data points, and noisy data points are also treated as kernel points. We use the radial basis function neural network algorithm to estimate $P(X \mid c_i)$, which allows for adjusting the number of kernels, has some noise resistance, and can obtain better probability density estimation results.

2.2. Radial Basis Function Neural Network

As shown in Figure 1, the Radial Basis Function Neural Network maps the input vector x to a new vector y. For example, the first radial basis neuron $(\vec{c}_1, b_1)$ corresponds to the output of each input vector as $p_{11}, p_{12}, \ldots, p_{1N}$.

For the j-th input x_j, the output of the first radial basis neuron is

$$p_{1j} = \varphi\left(\left\|\vec{x}_j - \vec{c}_1\right\| b_1\right), \tag{4}$$

where $\vec{c}_1$ is the center vector of the 1st radial basis function, and b_1 is its width parameter. Both are collectively referred to as the weights of the radial basis neuron. $\|\square\|$ represents the Euclidean distance between two vectors; the radial basis function can be a Gaussian function.

$$\varphi(t) = \mathrm{e}^{-t^2}. \tag{5}$$

The layer that maps the vector x to the vector p is called the radial basis layer. After the mapping in the radial basis layer, a linear neuron is added as the second layer of the network. It combines the outputs of the radial basis layer linearly to obtain the final outputs $y_1, y_2, \ldots, y_N$. of the entire network. For example, for the j-th input, the final output of the entire network is

$$y_j = \sum_{i=1}^{S} a_i p_{ij} + b_2, \tag{6}$$

where the weights $\vec{a}$, b_2, of the linear layer can be obtained using the conventional least squares method (LSM). For the weight b_1 of the radial basis layer, an appropriate value needs to be determined through experimentation specific to the problem. An important issue in training the radial basis network is determining the number of neurons S in the radial basis layer and the values of the weights $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_S$. Chen et al. [20] proposed an Orthogonal Least Squares method (OLS) that effectively addresses this problem, known as the Radial Basis Network Orthogonal Least Squares Training Algorithm. This algorithm dynamically adjusts the number of kernel functions while controlling the fitting accuracy.

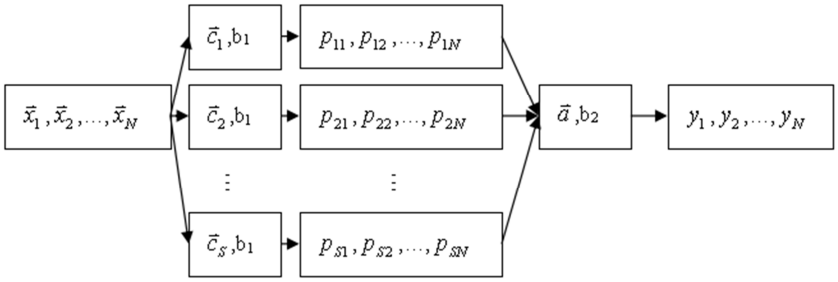

Figure 1. *Radial basis network structure diagram.*

2.3. Model Sample Classification Experiment

2.3.1. Model Data

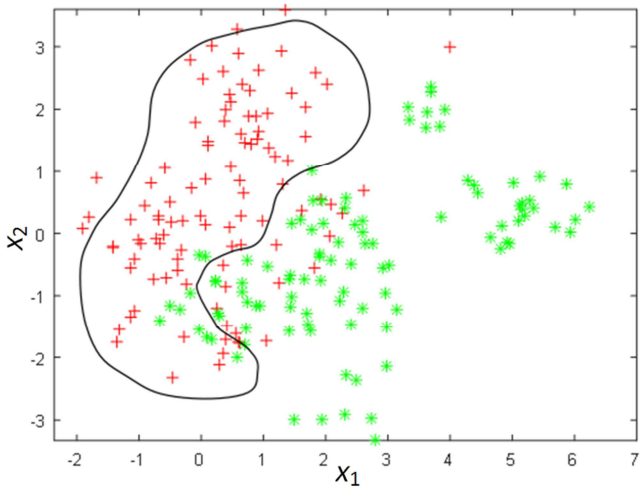

Figure 2. *Classification experiment of samples for model.*

The model sample points are a set of two-dimensional data points X=(x_1, x_2), divided into two categories as shown in Figure 2. Category c_1 is represented by symbols of red "+", and category c_2 green "*". They were generated by adding 200 random points from two-dimensional Gaussian distributions with a standard deviation of 0.2 and randomly distributed center positions. The two categories overlap with each other to represent the common situation of overlapping sample categories in real-world problems.

2.3.2. Distribution Estimation Based on Radial Basis Networks

The prior distribution histograms of the categories can be obtained through frequency statistics. Figure 3a represents the prior distribution histogram of category c_1, and Figure 3b the histogram of category c_2. Radial basis networks are used to fit these two discrete prior distribution histograms, resulting in an estimation of the continuous prior distribution $P(X \mid c_i)$, as shown in Figure 4a, 4b. This estimation is used to calculate the category probabilities using the Bayesian formula.

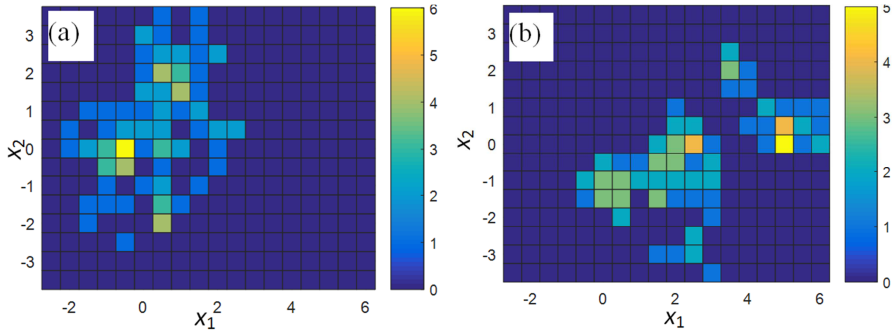

Figure 3. *Statistic of prior distribution by means of histogram.*

2.3.3. Classification

Using the Bayesian probability formula, the probability of each sample point $X=(x_1, x_2)$ belonging to each category can be calculated separately. Here, taking the identification of the "+" sign as an example, the sample points located in the region where $P(X \mid c_1) > P(X \mid c_2)$ are the target category, and the discriminant boundary is determined accordingly, as shown by the black line in Figure 2. The algorithm ensures the lowest possible misclassification rate and maintains good stability in case of sporadic noise in the discrete sample points.

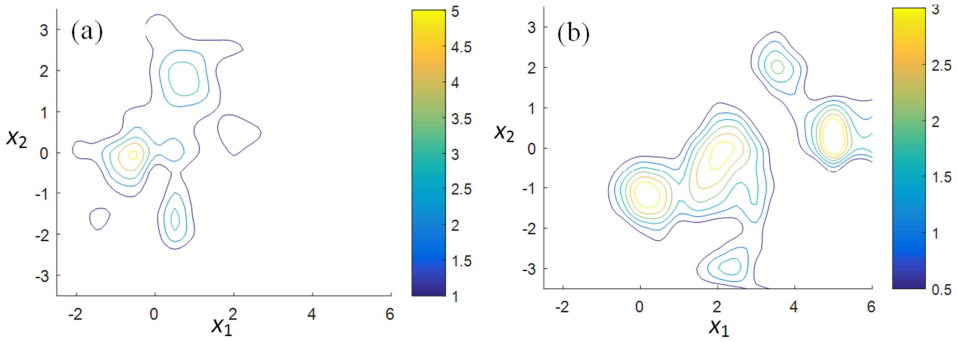

Figure 4. *Statistic of prior distribution by means of histogram Estimation of prior PDF (possibility distribution function) for (a) class* c_1 *and (b) class* c_2.

3. Field Application

3.1. Overview of the Target Stratum

The target stratum for study is the Longwangmiao Formation in the Cambrian System of the Sichuan Basin, which is rich in natural gas. The Longwangmiao Formation averages 76m to 100m thick. It is primarily composed of granular dolomite and fine-crystalline dolomite, with an increased occurrence of limestone in the lower part. Gypsum and salt rocks are commonly interbedded in the middle part, while a small amount of sandstone and mudstone can be found in the upper part. The overlying stratum is the Gaotai formation, mainly comprising clastic sedimentary rocks. It consists of purple-red heterogeneous sandstone and mudstone interbedded with dolomite, with a thickness ranging from 50m to 100m. The underlying Conglomerate Member of the Cambrian Series is a set of clastic sedimentary rocks, with purple-red shale interbeds in the lower part and carbonate rocks interbedded with shale at the top. Its thickness ranges from 130m to 200m (Figure 5) [21, 22].

Period	Series	Formation	Lithologic Symbol	Description	Sedimentary subfacies	Sedimentary Facies
Cambrian	Middle Series	Gaotai Formation		The upper part of this stratigraphic unit consists of dark gray and deep red shale interbedded with mud-limestone and dolomitic limestone. The middle-upper part is composed of light gray sandstone-grainstone. The middle part consists of gray mud-limestone and dolomitic limestone. The lower part is made up of dark gray silty sandstone.	Plate	Mixed deposit limited platform
					Abutment beach	
					Confined lagoon	
					Abutment beach	
	Lower Series	Longwangmiao Formation		The upper part of this stratigraphic unit consists of dark gray sandstone-grainstone and sandy limestone interbedded with argillaceous crystalline limestone. The lower part is composed of dark gray argillaceous crystalline limestone and mudstone-silty shale.	Abutment beach	
					Confined lagoon	
		Changlangpu Formation		This stratigraphic unit consists of dark gray siltstone, argillaceous siltstone, white-weathering quartzose sandstone interbedded with calcareous siltstone, limestone, gray limestone and light gray oolitic peloidal limestone.	foreshore-shoreface	Terrigenous clastic shoreline

Figure 5. *Stratigraphic column.*

The natural gas reservoirs developed in the Longwangmiao Formation are primarily characterized by intergranular and intercrystallite dissolution pores. Figure 6 demonstrates that the reservoir type is based on porosity. The core samples exhibit an uneven distribution of dissolution cavities, resulting in strong heterogeneity. The more developed the dissolution cavities in the reservoir, the better the reservoir's storage performance. Core analysis reveals that the porosity of the reservoir is mainly distributed between 4.0% and 6.0%. The maximum porosity is 11.28%, the minimum 0.32%, and the average 2.75%. Logging data from the study area indicate significant variations in reservoir thickness, ranging from 1m to 30m. The pink bands in Figure 6 represent reservoir intervals with porosity logs exceeding 2%, with segments measuring 22m and 5m in length [23].

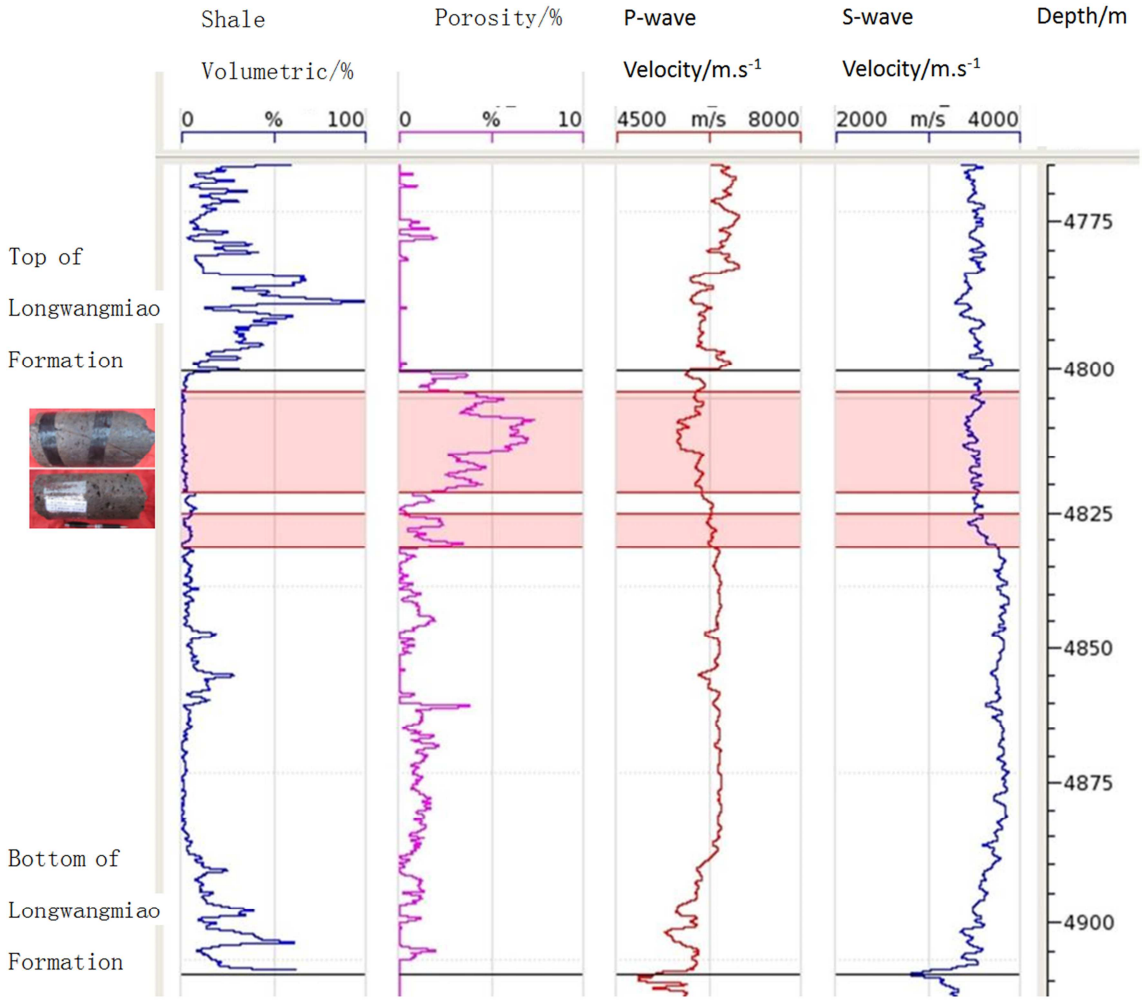

Figure 6. *Samples of rock and logs of reservoir in MX23.*

3.2. Reservoir Prediction Measures

Analysis of the basic data of the target stratum indicates that the challenges in predicting the gas reservoirs in the Longwangmiao Formation are as follows: 1) Complex lithology with strong heterogeneity; 2) Large differences in production rates from individual well tests, ranging from dry wells to high-yield industrial gas wells with millions of cubic meters of production. Additionally, there are difficulties in hydrocarbon detection. To address these challenges in reservoir prediction, the following measures have been taken in data processing. Firstly, using both P-wave and S-wave seismic data as a basis, precise calibration and comparative interpretation of multi-wave well-seismic data were conducted to determine the seismic stratigraphy of the Longwangmiao Formation. Then, combined with the analysis of multi-wave seismic responses of the reservoir, optimized multi-wave elastic parameters were selected, and a detailed rock physics analysis model was established. Finally, the method proposed in this study was used to accurately identify and predict gas-bearing reservoir samples in the Longwangmiao Formation.

3.3. Reservoir Prediction Based on Bayesian Classification

Figure 7 is a collection of sample points of characteristic vectors composed of P-wave impedance and S-wave impedance, which were calculated from the logging data of eight wells in the Longwangmiao Formation within the work

area. The green points represent samples of argillaceous dolomite in the Longwangmiao Formation, the blue points represent samples of dense dolomite, and the red points the samples of gas-bearing reservoirs. It can be observed that the characteristic vectors composed of P-wave impedance and S-wave impedance have good discriminative ability for distinguishing gas-bearing reservoirs and other lithologies. Using the radial basis function neural network, prior distribution density estimation for three categories was established. The contour lines of the prior distributions are plotted in Figure 7 with red, blue, and green representing the prior distribution estimates for gas reservoir samples, dense dolomite samples, and argillaceous dolomite samples, respectively. By applying the Bayesian probability classification method, the probability of each sample point belonging to each category can be calculated. Finally, for a batch of samples, the sample point with the highest probability of being a gas-bearing reservoir can be determined as a reservoir type.

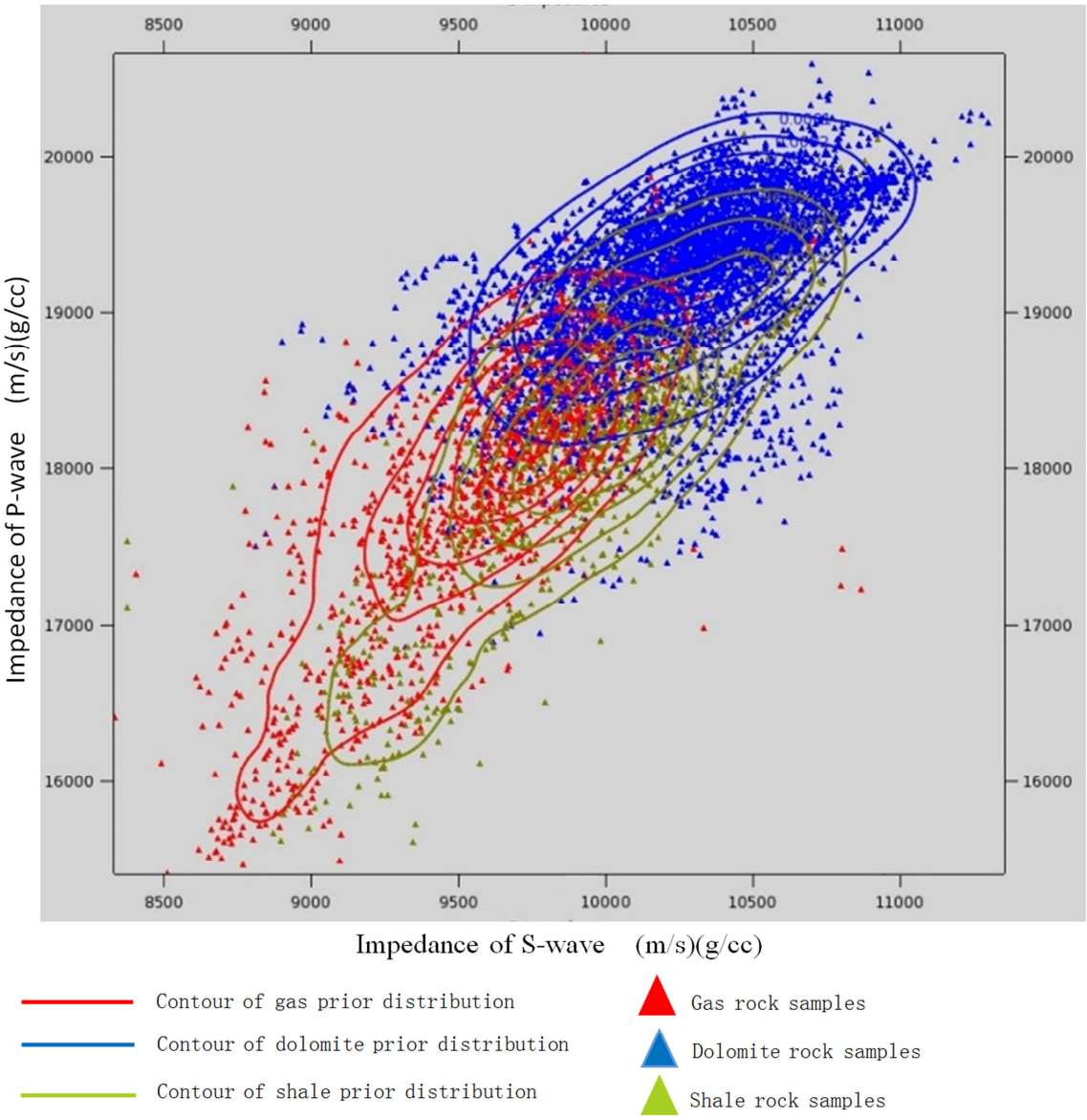

Figure 7. *Prior distributions for 3 lithological types.*

For the entire work area, a three-dimensional data volume of characteristic vectors composed of P-wave impedance and S-wave impedance obtained through multi-wave joint inversion forms a three-dimensional data volume. By applying the gas probability determination method above, a spatial data volume of gas probability for the target formation is obtained. Figure 8 shows two profiles extracted along the survey line at the locations of two industrial gas-producing wells from the gas probability data volume. Warm colors represent high gas probability, while cool colors represent low gas probability. The reservoir in Well MX23 is mainly at the top of the Longwangmiao Formation, with gas production of 110.8 x 10^4 m^3/d and a predicted gas probability exceeding 60%. On the other hand, the gas reservoir is not developed in Well MX208 with a low gas probability. This indicates that the Bayesian learning method effectively identifies gas-bearing reservoirs

automatically.

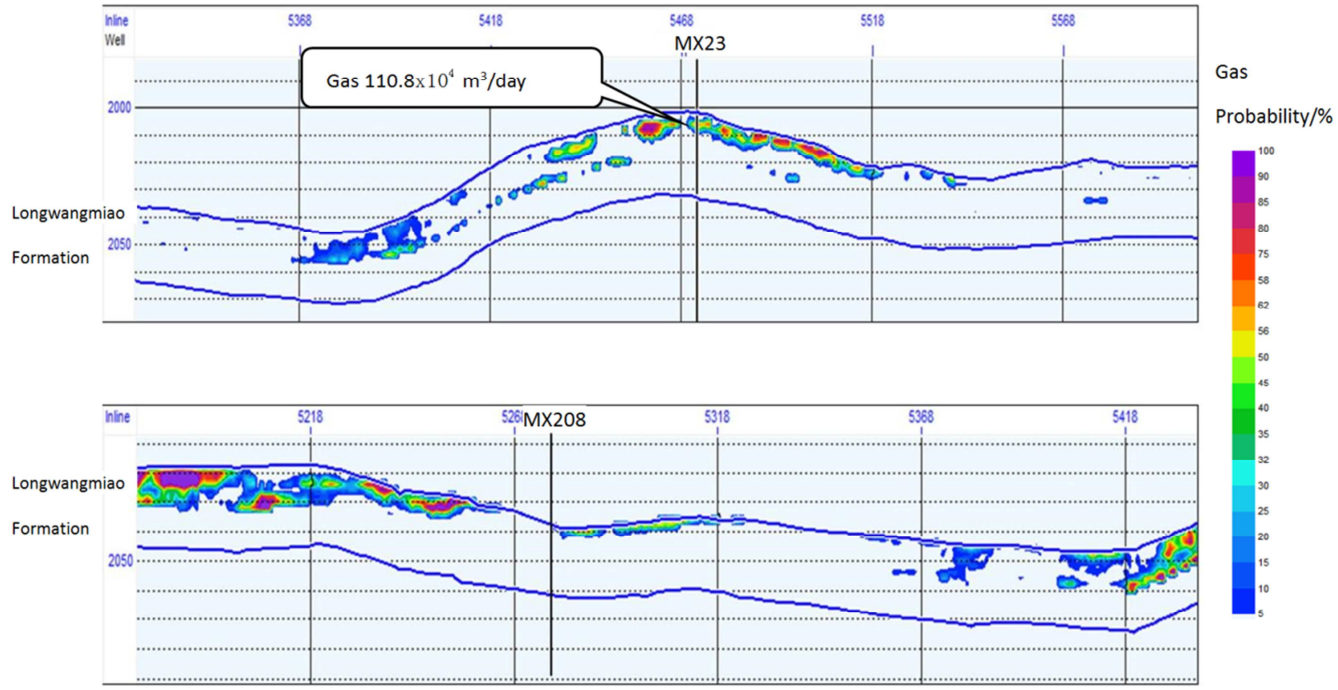

Figure 8. *Probability section along the line at well MX23 and MX208 respectively.*

4. Conclusion

This study has developed the Bayesian machine learning algorithm and gained the following insights:

1) Bayesian machine learning classification demonstrates good adaptability, as it can simultaneously recognize multiple categories and effectively adapt to the needs of reservoir prediction in geologically complex formations.
2) The use of radial basis function neural networks to fit prior distributions of categories eliminates the need to assume distribution types, making it highly practical for handling geological parameters.
3) The application of Bayesian machine learning techniques enables automated identification of reservoirs, thus improving production efficiency.

This method is a strong suit for multi-target recognition. In the next steps, the proposed method can be further extended to predict oil and gas reservoirs under more complex geological conditions. However, for weak signals, such as thin interbeds and tight reservoirs, which have minimal contrast with the background, recognizing these types of targets poses a challenge to the accuracy of the Bayesian method.

Acknowledgments

I would like to express my deep appreciation to the editor assistant, Cindy Black, for continuous encouragement leading to the successful completion of our paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Hartagung R T and Rosid M S. (2022). Implementation of the Poisson Impedance Inversion to Improve Hydrocarbon Reservoir Characterisation in the Poseidon Field, Browse Basin, Australia [J]. *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, 12 (2): 102–114.

[2] Suleman Mauritz Sihotang, Ida Herawati1. (2021). Simultaneous Seismic Inversion for Reservoir Characterization at Poseidon Field, Browse Basin, Australia [J]. *Journal of Materials Science, Geophysics, Instrumentation and Theoretical Physics*, 4 (1): 63-73.

[3] P. S. Nwiyor, E. D. Uko, I. Tamunobereton-ari, et al. (2022). Prediction of Elastic Properties using Crossplot Analysis and Seismic Inversion for Enhanced Quantitative Interpretation of Zeta Field, Niger, Delta, Nigeria [J]. *IARJSET*, 9 (1): 155-163.

[4] Ashraf, Umar, Hucai Zhang, Aqsa Anees, et al. (2020). Controls on Reservoir Heterogeneity of a Shallow-Marine Reservoir in Sawan Gas Field, SE Pakistan: Implications for Reservoir Quality Prediction Using Acoustic Impedance Inversion [J]. *Water* 12, no. 11: 2972.

[5] Tom M. Mitchell. (1997). Machine learning [M]. NYC: McGraw-Hill.

[6] Tou J. T., Gonzalez, R. C. (1974). Pattern recognition principles [M], Massachusetts: Addison-Wesley Publishing Company.

[7] Duda R., Hart, P., Stork, D. (2001). Pattern classification [M]. New York: John Wiley & Sons INC.

[8] Bian Z Q, Zhang X G. (1999). Pattern recognition (2nd Edition) [M]. Beijing: Tsinghua University Press. (In Chinese)

[9] Yin X F, Hao Z F, Yang X W. (2007). Review of fast kernel density estimation algorithms [J]. *Computer Engineering and Applications*, 43 (31): 1-4. (In Chinese)

[10] Greengard L, Strain J. (1991). The fast gauss transform [J]. SIAM J Sci Comput, (2): 79-94.

[11] Gray A G, Moore A W. (2000). N-body problems in statistical learning [J]. *Advances in Neural Information Processing Systems*, 13 (4): 521-527.

[12] Hall P, Wand M P. (1996). On the accuracy of binned kernel density estimators [J]. *Journal of Multivariate Analysis*, 56 (2): 165- 187.

[13] Weston J, Gammerman A, Stitson M O, et al. (1999). Support vector density estimation [M]. Cambridge: MIT Press.

[14] Rui T, Zhou Y, Ma g Y, et al. (2011). Target detection using kernel density estimation and Gaussian model cascade mechanism [J]. *Computer Engineering and Applications,* 47 (18): 1-3. (In Chinese)

[15] Dong M, Zhu H, Xing N, et al. (2012). Calculation method of inverse camera response function of single image based on kernel estimation [J]. *Computer Engineering and Applications*, 48 (10): 171-174. (In Chinese)

[16] Qiao J F, Zhu H, Shi J, et al. (2012). Fast kernel density estimation method for background modeling [J]. *Computer Engineering and Applications*, 48 (5): 192-193. (In Chinese)

[17] Avseth P., Mukerji, T., Mavko G. (2005). Quantitative seismic interpretation [M]. New York: Cambirdge University Press, 278-294.

[18] Bertrand, C., Tonellot, T., Fournier, E. (2012). Seismic facies analysis applied to p and s impedances from pre-stack inversion [C]. *SEG Expanded Abstract, 72'd Annual Meeting, Salt Lake City, Tulsa, Oklahoma, USA: SEG 2012*: 217-220.

[19] P. M. Doyen. (2007). Seismic reservoir characterization: an earth modelling perspective [M]. Netherlands: EAGE.

[20] Chen S C, P. M. Grant, C. F. N. Cowan. (1992). Orthogonal least squares algorithm for radial basis function networks. Radar and signal processing [J]. *IEE Proceedings F*, 139 (349): 302-309.

[21] Xie W R, Yang W, Li X Z, et al. (2018). The origin and influence of the grain beach reservoirs of Cambrian Longwangmiao formation in central Sichuan area, Sichuan basin [J]. *Natural Gas Geoscience*, 29 (12): 1715-1726. (In Chinese)

[22] Zhang C, Yang C C, Liu Y C. (2017). Controlling factors of fluid distribution in the lower Cambrian Longwangmiao formation, Moxi area, Sichuan basin [J]. *Geology and Exploration,* 53 (3): 0599-0608. (In Chinese)

[23] Wang Y P, Yang X F, Wang X Z. (2019). Reservoir property and genesis of powder crystal dolomite in the Longwangmiao formation, Moxi area in central Sichuan basin [J]. *Geological Science and Technology Information*, 38 (01): 197-205. (In Chinese)

Biography

Fu Zhiguo (1977-), male, Doctor, senior engineer, is engaged in research on seismic data interpretation methods, Oil and gas reservoir pattern recognition, and Multi-wave seismic exploration.