

SCIENTIFIC FOUNDATIONS AND TECHNOLOGIES OF INTELLIGENT OIL AND GAS EXPLORATION

Kerimov V.Yu.^{1,2}, Mustaev R.N.³, Podoprighora V.N.⁴, Guryanov S.A.³

¹Institute of Geology of the Ministry of Sciences and Education of the Republic of Azerbaijan, Baku,
Azerbaijan

²Baku State University, Baku, Azerbaijan

³Sergo Ordzhonikidze Russian State University for Geological Prospecting, Moscow, Russia

⁴Plekhanov Russian University of Economics

Abstract. The article presents the results of developing scientific foundations and technologies for intelligent geological exploration implementing the Smart Exploration and Smart Field paradigms. A unified platform for comprehensive consideration of geological, geophysical and technogenic factors has been developed based on multiparameter digital modeling of petroleum systems combined with hybrid artificial intelligence integrating Physics-Informed Neural Networks (PINN) with basin and geodynamic modeling. The Digital Caspian platform was validated using a blind test protocol on 47 Caspian region wells not included in the training dataset. The AUC-ROC metric for hydrocarbon potential prediction reached 0.84 ± 0.03 , compared to 0.61 ± 0.05 for traditional empirical methods (baseline), corresponding to an accuracy improvement of 25–40% depending on trap type and geological complexity. Dependence on imported software decreased from 90% to 40% (by computational resource volume) over the period 2019–2024 among partner organizations.

Keywords: intelligent geological exploration, digital modeling of petroleum systems, hybrid artificial intelligence, Physics-Informed Neural Networks (PINN), smart field, basin modeling, Digital Caspian platform, uncertainty quantification, hydrocarbon potential, import substitution.

1. Introduction

The current stage of development of Russia's mineral resource complex is characterized by the depletion of easily accessible hydrocarbon reserves and the necessity to develop the resource base in challenging geological conditions. The Arctic shelf, deeply buried horizons of traditional provinces, rift zones, and other promising objects require fundamentally new approaches to geological exploration. According to the Ministry of Natural Resources of the Russian Federation, about 70% of Russia's proven hydrocarbon reserves are concentrated in the Arctic and Far Eastern seas; however, only about 5% of these reserves are currently being developed due to difficult operational conditions.

Traditional local-empirical methods for forecasting oil and gas potential, based on exploration indicators and analogies with discovered fields, have reached the limit of their effectiveness. High geological uncertainty remains a critical barrier, leading to significant financial losses when initiating exploratory and development drilling. According to estimates by Wood Mackenzie, under existing approaches, an average of only 29% of the initial oil-in-place is extracted, and the cumulative global hydrocarbon supply deficit up to 2050 could exceed 300 billion barrels of oil equivalent. At the same time, traditional geological exploration discoveries are unable to compensate for this gap.

A drastic increase in the efficiency of geological exploration is possible only through the digital transformation of the industry—the transition to intelligent exploration technologies based on multi-parametric digital modeling of hydrocarbon systems (HCS) and hybrid artificial intelligence (AI). This direction aligns with global trends: major oil and gas corporations are actively deploying AI solutions to optimize drilling, forecast reservoir properties, and manage field development [2, 5, 8]. At the same time, under the conditions of sanction restrictions, foreign integrated platforms (Schlumberger DELFI, Halliburton DecisionSpace 365) are unavailable to Russian companies, which dictates the necessity of creating domestic counterparts with verified algorithms.

The relevance of the study is also driven by the strategic importance of the Caspian region for the development of Russia's oil and gas resources. This unique region, which holds colossal hydrocarbon reserves in its subsurface, requires a balanced and scientifically sound approach to geological exploration. For the first time to address this challenge, the authors of the article have developed an innovative information platform, "Digital Caspian," and a series of interactive digital geological maps. These tools are designed to provide comprehensive analysis, precise forecasting, and the making of verified, scientifically sound decisions for the long-term planning of exploration activities.

The implementation of AI contributes to the creation of the intelligent exploration system "Smart Exploration" and the "Smart Field"—these are complexes of software and technical tools that integrate digital technologies and engineering knowledge to optimize production and logistical processes, enabling the management of the oil and gas exploration process and hydrocarbon field development. The application of digital technologies in the exploration process is primarily associated with the implementation of innovative solutions for assessing and forecasting oil and gas potential. For this purpose, it is advisable to use digital engineering, the essence of which is the creation of a virtual prototype of a real object (a field, a geological exploration process, etc.). This allows for the simulation of various scenarios, conducting analysis, identifying bottlenecks, and optimizing processes without risk to real production. Digital engineering enables quick and efficient resolution

of key tasks in oil and gas exploration and hydrocarbon field development. The next stage of digital petroleum engineering involves intelligent multi-agent systems.

The purpose of this study is to develop the scientific foundations and technologies of intelligent exploration that implement the "Smart Exploration" and "Smart Field" paradigms. Based on multi-parametric digital modeling of hydrocarbon systems and hybrid artificial intelligence, it aims to create a unified platform for the comprehensive integration of geological, geophysical, and anthropogenic factors, advance the scientific foundations of rational subsoil use, and establish intelligent technologies for modeling hydrocarbon systems. To achieve this objective, the following tasks were addressed: (1) formalizing the architecture of a hybrid AI system that integrates physics-informed neural networks (PINNs) with physical models of hydrocarbon systems; (2) developing a verifiable methodology for assessing forecast accuracy; (3) practical implementation and testing of the "Digital Caspian" platform using data from the Caspian region; (4) reducing dependence on imported software. The creation of domestic intelligent platforms based on hybrid AI is not merely a scientific task, but also a matter of national security and technological sovereignty in the critical field of geological exploration.

2. Analysis of existing approaches. Global and Russian experience in applying artificial intelligence in geological exploration.

The global oil and gas industry is undergoing an active phase of digital transformation, within which artificial intelligence and machine learning technologies occupy a central place. According to the International Energy Agency (IEA), investments in digital technologies in the oil and gas sector have reached tens of billions of dollars, with the segment of AI solutions for geological exploration and field development growing most dynamically. Key areas of AI application in global practice include:

- **Processing and interpretation of seismic data** — utilization of convolutional neural networks (CNNs) for the automatic identification of seismic facies, reservoir property forecasting, and geological anomaly detection.
- **Drilling optimization** — machine learning systems for predicting drilling equipment wear, preventing emergency situations, and reducing non-productive time [5].
- **Production forecasting** — surface-graph convolutional networks (SGP-GCN) and recurrent architectures for multi-well flow rate forecasting considering hydrodynamic interactions [9, 10].

- **Reserve management** — AI tools for assessing recoverable reserves and optimizing reservoir pressure maintenance systems.

Leaders in AI adoption

One of the most indicative examples is the experience of Devon Energy, which is scaling automation and AI across its operations in the Permian Basin and Eagle Ford formation. The utilization of machine learning combined with simul-frac technologies allowed the company to reduce drilling costs by 12% and well completion costs by 15%. The implementation of "smart" gas lift delivered a production increase of 3–5%. [2].

At the IADC/SPE 2026 conference (Galveston, Texas), Baker Hughes presented a system for predicting drill string stuck pipe events several hours before the occurrence by combining physics-based models with machine learning. Patterson-UTI implemented a generator management system that provided savings of 5–8 gallons of diesel fuel per hour per rig (approximately \$100,000 in annual savings). Nabors Industries developed a mud pump monitoring system that detects failures hours before they occur [5].

Wood Mackenzie developed the Analogues AI tool, which analyzes over 35,000 fields across 60 attributes. Benchmarking demonstrated a potential opportunity for the additional recovery of 6–12% of reserves—ranging from 470 billion to more than 1 trillion barrels of oil equivalent. Read the full analysis at Wood Mackenzie [2].

The Chinese company BGP (the geophysical division of CNPC), in partnership with Huawei, is developing the Kunlun large language model and adapting the GeoEast platform for the Kunpeng architecture. The use of the MindSpore framework enabled an increase in computational efficiency by more than five times when solving the 2D acoustic wave equation [8]. These developments are part of Huawei's global strategy to create an "intelligent energy" ecosystem presented at the ADIPEC 2025 exhibition.

Russian achievements and priorities

Significant experience has been accumulated in the Russian Federation regarding the application of AI in geological exploration; its development is acquiring particular importance within the context of import substitution and technological sovereignty tasks [11].

Scientists from SPbSU and MSU have created a neural network algorithm based on the EfficientNetB4 architecture to process surface seismic waves when studying the upper part of the geological section on the shelf. This neural network solution made it possible to obtain highly detailed

3D models of shear wave velocity distribution over an area of more than 2,000 km², [1, 12] with the capability to identify hazardous engineering-geological objects (paleochannels, decompaction zones).

Specialists from the Gubkin Russian State University of Oil and Gas and LLC "Priezhev Laboratory" have developed a hierarchical neural network method based on the "nearest neighbor" approach with preliminary clustering of the training dataset. Testing on data from the West Siberian oil and gas province (Vikulov Formation) demonstrated its high efficiency in predicting oil saturation [3, 13] compared to traditional inversion approaches. The method exhibits advantages in learning speed, generalization capacity, and incremental learning capabilities.

The Trofimuk Institute of Petroleum Geology and Geophysics of the Siberian Branch of the Russian Academy of Sciences (IPGG SB RAS), jointly with LLC "SakhalinNIPImorneftegaz," updated a Bayesian classifier algorithm for seismic facies analysis by integrating it with prior probability maps of lithotypes. The application of the enhanced algorithm at a test site in the Orenburg Region successfully facilitated the cross-targeting of wells and production drilling [7, 14], confirming the high reliability of the reservoir properties forecast. As noted by the researchers, "the application of the improved Bayesian classifier algorithm... can significantly increase the reliability of interpretation results."

MIPT scientists have developed an AI model for predicting the interfacial tension between oil and salt water. The new model ensures an accuracy of up to 90%, compared to an error margin of up to 40% found in existing theories; its integration into the "RN-Digital Core" software package accelerates the selection of the optimal salinity for injected water [4, 6, 15].

An analysis of existing solutions reveals a research gap: none of the platforms available to Russian users provide (a) end-to-end modeling from the basin level to the digital twin level of the field; (b) a verified, calibrated assessment of geological risks across the entire hydrocarbon system chain; or (c) built-in explainable AI (XAI) mechanisms to support investment decisions. These limitations constitute the subject of this research.

Table 1. Comparative characteristics of intelligent geological exploration platforms

Platform / Method	Data coverage	AI Architecture	Accuracy verification	Availability (RF)
Schlumberger DELFI	Full complex	ML + physical models	Commercially unavailable	Недоступна (санкции)
Halliburton DecisionSpace 365	Full complex	ML	Commercially unavailable	Недоступна (санкции)
(IPGG SB RAS) (bayesian classifier) [7]	Seismic + Well logging	Bayesian HC	Orenburg test field	Available
Gubkin University (hierarchical NNs) [3]	Well logging + Seismic	HC + clustering	West Siberia	Available
"Digital Caspian" (real work)	60+ parameters (Full HCS cycle)	PINN + CNN + BNN (ensemble)	AUC-ROC 0,84 (blind test, n=47)	Domestic

3. Research Methodology and Technologies

3.1. Hybrid artificial intelligence architecture

In the context of this study, "hybrid AI" is defined as the integration of two classes of models: (1) physics-based (white-box) models of hydrocarbon generation, migration, and accumulation processes; and (2) neural network (black-box) emulators of complex dependencies. The proposed architecture is based on the concept of PINNs [16, 28] and advances it in relation to multi-basin oil and gas potential forecasting tasks.

The loss function of the hybrid model includes two terms:

$$L_{\text{total}} = L_{\text{data}} + \lambda * L_{\text{physics}}, \quad (1)$$

$L_{\text{data}} = \|\hat{y} - y\|^2$ — mean squared error on the training dataset; L_{physics} — penalty term, penalizing violations of Darcy's law and the fluid mass conservation equation in the reservoir;

λ — regularization coefficient determined during cross-validation. This approach ensures the physical consistency of predictions, especially when extrapolating to objects with limited data.

The neural network component of the platform includes the following architectures: (a) convolutional neural networks (CNNs) for seismic data interpretation and facies identification; (b) ensemble methods (Random Forest, Gradient Boosting) for forecasting reservoir properties based on well logging; (c) recurrent networks (LSTM) for predicting production dynamics; and (d) Bayesian neural networks (BNNs) for uncertainty quantification. Physical constraints are embedded into the loss function as soft constraints, ensuring compliance with Darcy's laws and thermodynamics.

3.2. Fundamental mathematical concepts and methods

Modern AI systems for geological exploration are based on a combination of methods from computational mathematics, statistics, and machine learning. Below are the key concepts, grouped by area.

Linear algebra: vectors, matrices, tensors, and linear transformations are applied to represent seismic data (volumes and sections) as multidimensional matrices (tensors). Any processing (filtering, migration) consists of linear operations on tensors. In machine learning, data are represented as feature vectors, and models are represented by a matrix operation: $Y = W \times X + b$, where X is the input data vector, W is the model weight matrix, b — is the bias vector and Y is the result (e.g., rock classification).

Mathematical analysis (differential calculus): The primary tools are the derivative, gradient, and partial derivative. Gradient descent is the cornerstone of training for almost all neural networks. The algorithm iteratively adjusts model parameters, moving in the direction opposite to the gradient of the error function to find its minimum. The formula is used: $W_{\text{new}} = W_{\text{old}} - \alpha * \nabla J(W)$, where W — the model parameters; α — learning rate, $\nabla J(W)$ — gradient of the cost (error) function.

- **Probability theory and mathematical statistics:** The underlying concepts incorporate probability, distributions (normal, Poisson, etc.), conditional probability, Bayes' theorem, correlation, and regression. The areas of application include uncertainty quantification (e.g., the probability of hydrocarbon presence in a reservoir), error analysis, Bayesian methods for data inversion, and statistical analysis of core and well logging data. The formula used is (Bayes' Theorem): $P(A|B) = [P(B|A) * P(A)] / P(B)$. Used to update the probability of a hypothesis (e.g., A = "presence of a reservoir") when new data is received (B = seismic attribute).

- **Optimization methods:** The tools used are finding the minimum/maximum of a function under given constraints. They are applied for tuning machine learning model parameters,

calibrating reservoir simulators, and solving seismic inversion problems. The objective function for seismic inversion is expressed as: $\min \|G(m) - d\|^2 + \lambda \|Lm\|^2$, where G — the forward modeling operator, m — the target parameters, d — the observed data, λ — the regularization parameter.

Processing and interpretation of seismic data

- **Digital signal processing (DSP):** Methods used include the Fourier transform (frequency spectrum analysis), wavelet transform (time-frequency analysis for non-stationary signals), and filtering (bandpass, notch, and median filters for noise suppression). Application — improving seismic data quality and isolating useful signals.
- **Seismic inversion (transition from amplitudes to rock properties):** This is a classic ill-posed inverse problem. It requires recovering subsurface parameters (impedance, density, velocity) from recorded seismic responses. The methods include deterministic inversion, which is solved via the least-squares method or other regularization techniques (e.g., Tikhonov regularization). The formula used is: $\min \|G(m) - d\|^2 + \lambda \|Lm\|^2$, where G — the forward modeling operator, m — the target parameters, d — the observed data, λ — the regularization parameter. L — operator (often a derivative to obtain a smooth solution).
- **Statistical/Bayesian inversion** provides a probability distribution of possible solutions rather than a single solution, which is critical for risk assessment.
- **Machine learning:** Neural networks (most commonly convolutional — CNNs) are trained on synthetic or labeled real data for the direct transformation of a seismic cube into an acoustic impedance cube.
- **Attribute analysis and methods:** Computation of hundreds of mathematical attributes from seismic data (amplitude, frequency, and geometric attributes). To analyze and reduce the dimensionality of this big data, Principal Component Analysis (PCA) and Autoencoders are utilized. The formula for PCA actually represents the singular value decomposition (SVD) of the data matrix: $X = U\Sigma V^T$, where the columns of V are the principal components themselves.

3.3. Principles of constructing the "Digital Caspian" platform

Principle 1: Multi-parametric integration of geodata. Traditional approaches to forecasting oil and gas potential utilize a limited set of parameters (structural maps, net pay thicknesses, reservoir properties). The developed methodology ensures the comprehensive integration of more than 60 geological, geophysical, geochemical, and anthropogenic factors within a single digital platform. This

aligns with global benchmarks—a similar approach is implemented in Wood Mackenzie's Analogues AI tool [17].

No single method provides a complete picture: seismic exploration establishes the structural framework, well logging measures rock properties near the borehole, geochemistry determines the type and maturity of kerogen, and satellite remote sensing data aids in regional structuring. Only a joint analysis of all these data types creates a reliable predictive model.

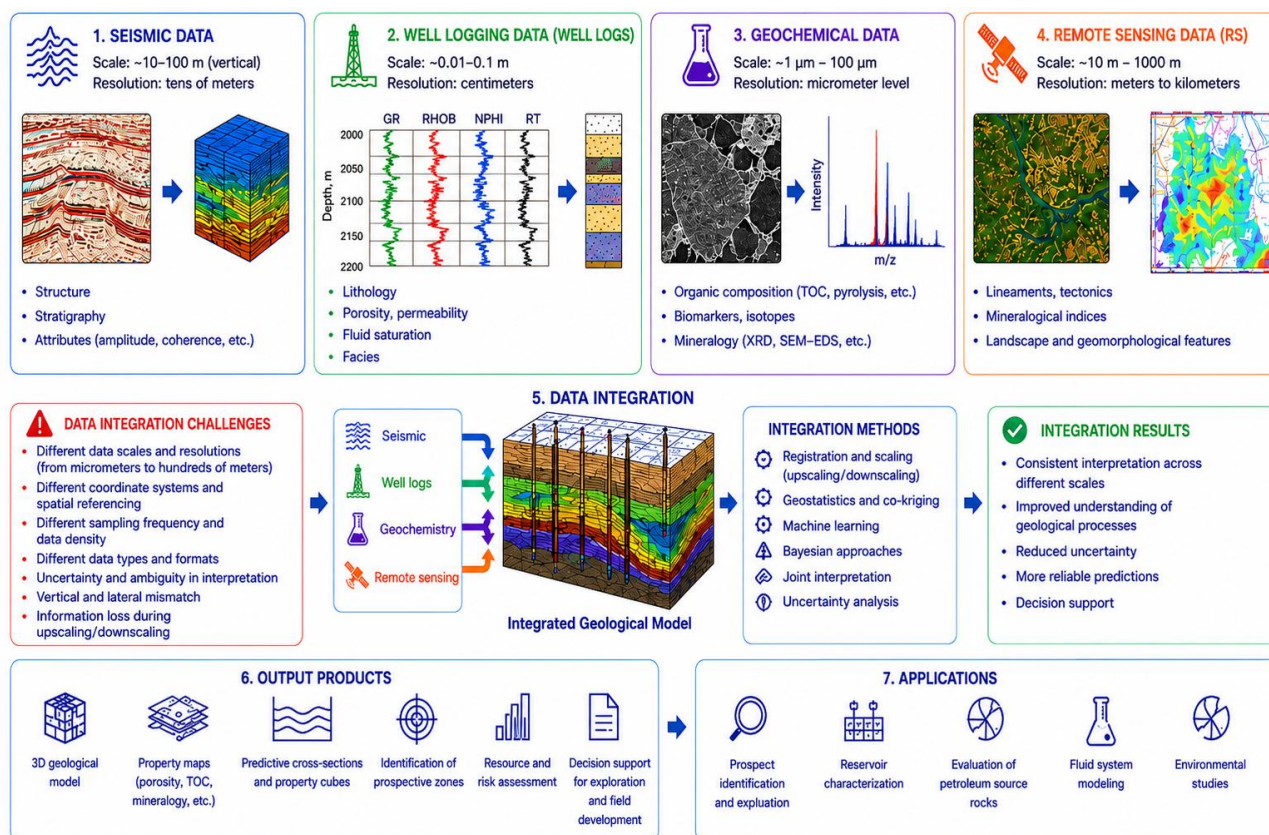


Figure 1. Principle of integrating heterogeneous geodata into a unified model. No single method provides a complete picture. Seismic exploration establishes the structural framework and lithological information, well logging measures rock properties near the borehole, geochemistry determines the type and maturity of kerogen, and satellite remote sensing data aids in the regional structuring of the territory. Only a joint analysis of all these data types, implemented in platforms like "Digital Caspian," creates a reliable predictive model.

Principle 2: Hybrid intelligence. In contrast to purely neural network approaches, a combination of "white boxes" (physics-based models of hydrocarbon generation, migration, and accumulation processes) and "black boxes" (neural network emulators of complex dependencies) is proposed through a single loss function (1). On one hand, this approach ensures the physical consistency of predictions, which is especially critical in data-scarce environments, making the

forecasts reliable when extrapolating to deeply buried horizons. On the other hand, it allows for capturing subtle correlations that are not described by explicit equations.

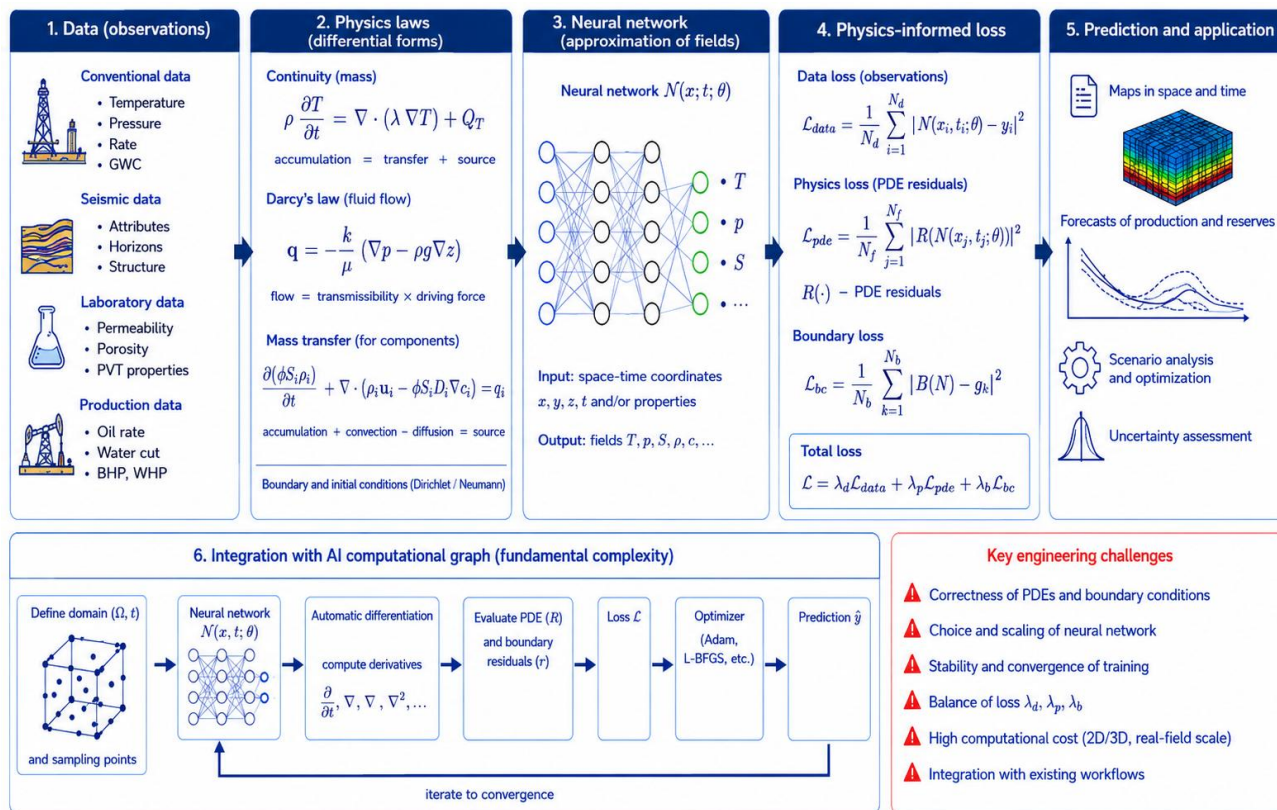


Figure 2. Concept of a physics-informed hybrid model (Physics-informed AI). This is the most promising direction, combining the strengths of numerical modeling and machine learning. In contrast to the "black box" of pure AI, the hybrid model is trained not only on data but also ensures compliance with fundamental physical laws (Darcy's law, conservation of mass, heat transfer). This makes predictions physically consistent and reliable, especially in data-scarce environments, which is critically important for geological exploration.

Principle 3: Hierarchical scaling. Modeling is carried out at three hierarchical levels: (1) the basin level — assessment of generation potential and zones of oil and gas accumulation; (2) the local level — detailed forecasting of reservoirs and traps; (3) the field level — construction of digital twins to optimize development.

Principle 4: Closed-loop digital twins. For each subsoil user asset, a dynamically updated digital twin is created, integrating drilling, well logging, seismic exploration, and production data. Digital twins were used for scenario modeling and real-time decision support (Fig. 2). Data from the physical asset continuously flows into the digital model; the model calibrates itself and generates a forecast, which is returned in the form of management recommendations. This closed loop allows for

the continuous improvement of model accuracy—an aspect that is especially crucial during the field development stage.

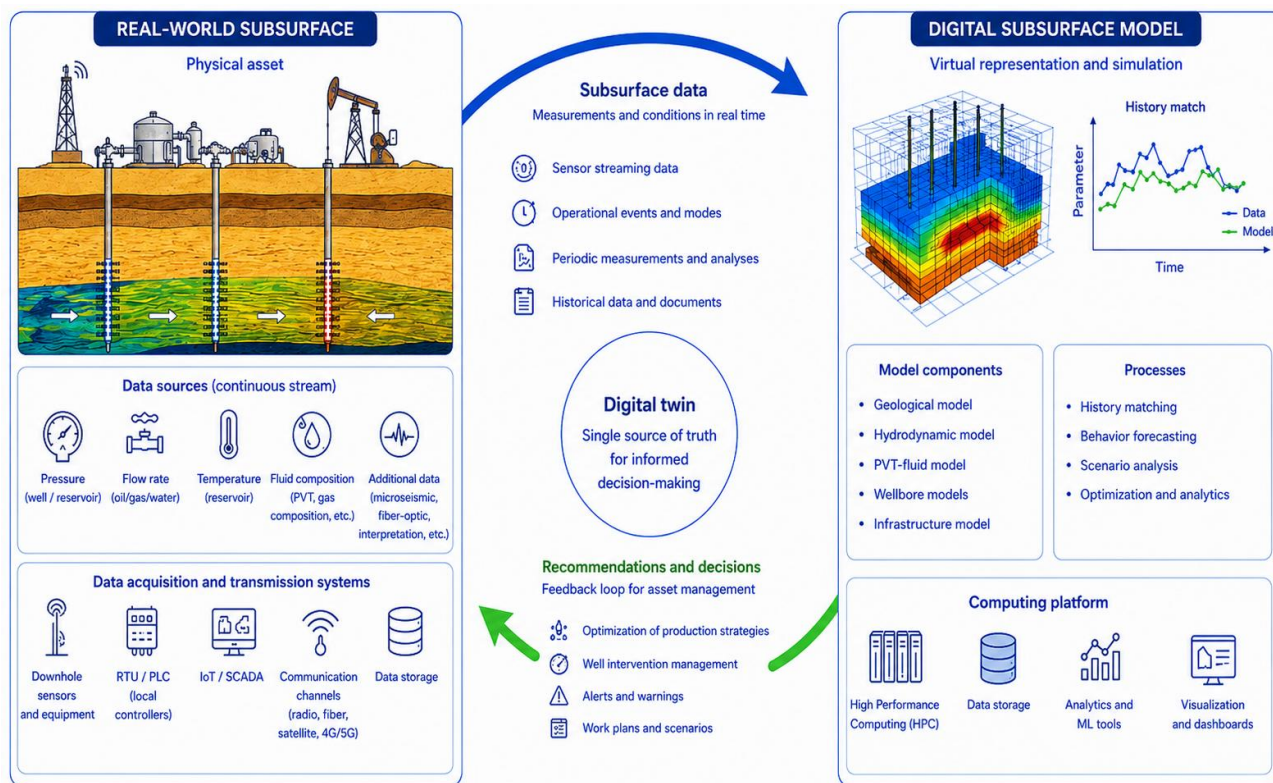


Figure 3. Operational cycle of a digital twin in geological exploration
Cyclic scheme: physical asset ↔ data ↔ model ↔ forecast. Emphasizes continuous feedback. Data from the field (sensors, well logging, production measurements) updates the model, which calibrates the forecast that is then returned to asset management (e.g., water injection mode control).

3.4. Data description and verification methodology

The initial dataset was compiled using data from the Caspian region. The training dataset included: 312 wells with a full suite of well logs; 14,900 km² of areal coverage of 2D and 3D seismic profiles; geochemical analyses of 2,847 core samples; and basin modeling results for three sub-basins (South Caspian, Central Caspian, and North Caspian). The temporal coverage of the data spans from 1950 to 2024. The test (blind) dataset for the blind test consisted of 47 wells with test results withheld until the moment of forecasting.

Verification of predictive accuracy was conducted in three stages. Stage 1 involved cross-validation on the training dataset (k-fold, k=10) to evaluate generalization capability. Stage 2 consisted of testing on a designated test dataset (80/20 split) with the calculation of quality metrics.

Stage 3 comprised blind testing on 47 independent wells that were not used at any stage of training or tuning.

The primary metric for forecasting oil and gas potential was AUC-ROC (binary classification of "productive / non-productive asset"). Additional metrics included the F1-score to account for class imbalance, and MAE for continuous forecasts of reservoir properties. The baseline was a traditional empirical forecasting method based on analogy with discovered reference fields. The calibration of probabilistic estimates was verified using reliability diagrams and the Expected Calibration Error (ECE) metric. Uncertainty quantification was implemented via Deep Ensembles consisting of 10 independently trained models.

4. The intelligent platform "Smart Exploration — Smart Field"

The intelligent platform is implemented as a distributed computing system, including the following functional modules:

1. Geodata collection and standardization module. It provides loading, validation, and unification of heterogeneous data into a single format: seismic time sections, well logging results, core descriptions, geochemical analyses, satellite images, and digital elevation models. Integration with corporate geographic information systems is implemented.

2. The basin modeling module simulates the 3D geological history of a sedimentary basin, covering sedimentation, tectonic evolution, thermal regime, and hydrocarbon generation and migration. This simulation is executed using specialized software, such as PetroMod and BasinMod (Fig. 4).

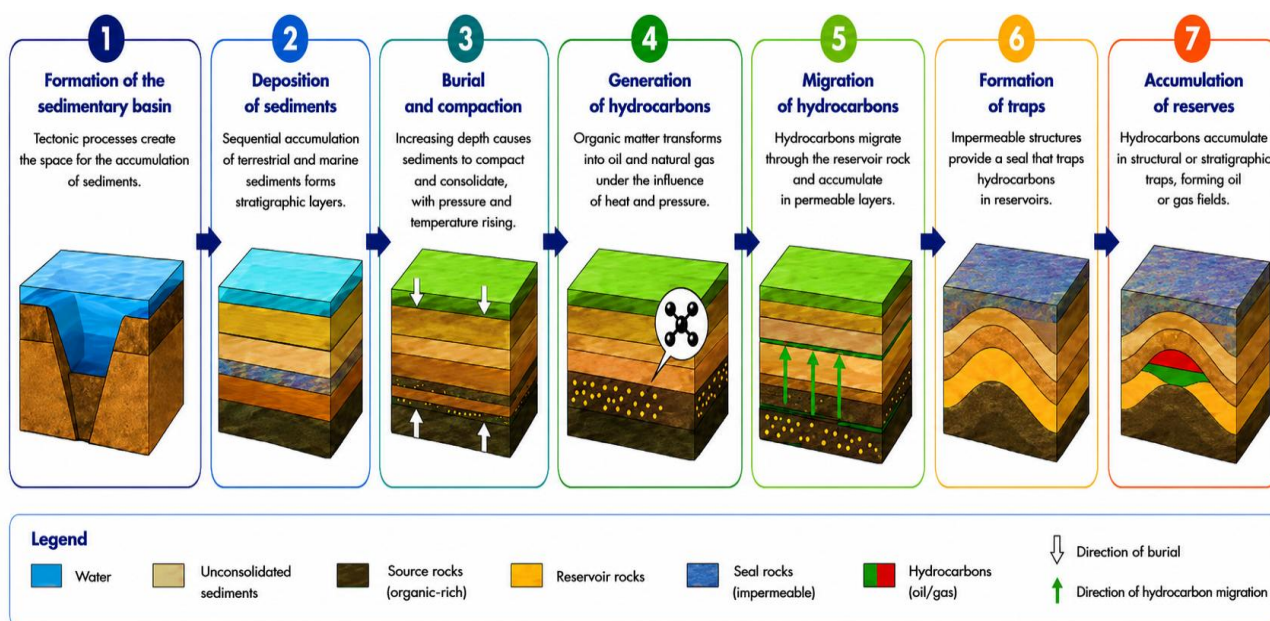


Figure 4. Sequence of basin modeling processes. Vertical sequence: sedimentation → subsidence → heating → HC generation → migration → accumulation. The vertical chain describes the lifecycle of a sedimentary basin: accumulation of sediments → their subsidence and compaction → heating with depth → generation of hydrocarbons → fluid migration through the reservoir → accumulation in a trap. This scheme forms the basis of the predictive assessment of oil and gas potential implemented in the "Digital Caspian" platform.

3. The AI-driven seismic interpretation module features neural network seismic facies detectors, automated horizon pickers, and algorithms for spectral decomposition and inversion. Transformer-based architectures are utilized to enhance seismic interpretation [18, 31, 32].

4. The petrophysical modeling module provides forecasting of reservoir properties (porosity and permeability) based on well logging data using neural network ensembles and Bayesian uncertainty calibration.

5. The geological modeling and digital twins module allows for the creation of deterministic and stochastic 3D reservoir models, volumetric reserve estimation, and production scenario modeling. The concept of the digital twin traces back to the works of Grieves [30].

6. Decision support module (XAI). Based on predictive models, it generates recommendations for exploratory and development drilling. To ensure interpretability, the SHAP (SHapley Additive exPlanations) method is applied. The largest contributions to the oil and gas potential forecast are made by: organic matter maturity indicators (R_o , T_{max}) — 24%; reservoir acoustic impedance — 18%; seal thickness — 15%; and trap area — 13%.

Mitigation of geological risks and economic efficiency

The probabilistic assessment of geological risks is an integral part of the developed platform (Fig. 5). For each forecasted asset, integrated geological risks are calculated, which are used to rank objects and optimize investment decisions. For each forecasted asset, the following are calculated:

- $P_{\{geol\}}$ — probability of trap presence (structural, stratigraphic, lithological);
- $P_{\{res\}}$ — probability of reservoir presence with acceptable reservoir properties (porosity and permeability);
- $P_{\{seal\}}$ — probability of an effective seal;
- $P_{\{gen\}}$ — probability of hydrocarbon generation from the source rock;
- $P_{\{migr\}}$ — probability of hydrocarbon migration and accumulation within the trap.

The integrated geological risk R is used to rank assets and optimize investment decisions.

The economic impact of technology implementation is assessed in two aspects: prevented losses from inefficient geological exploration and additional profits from resource base growth.

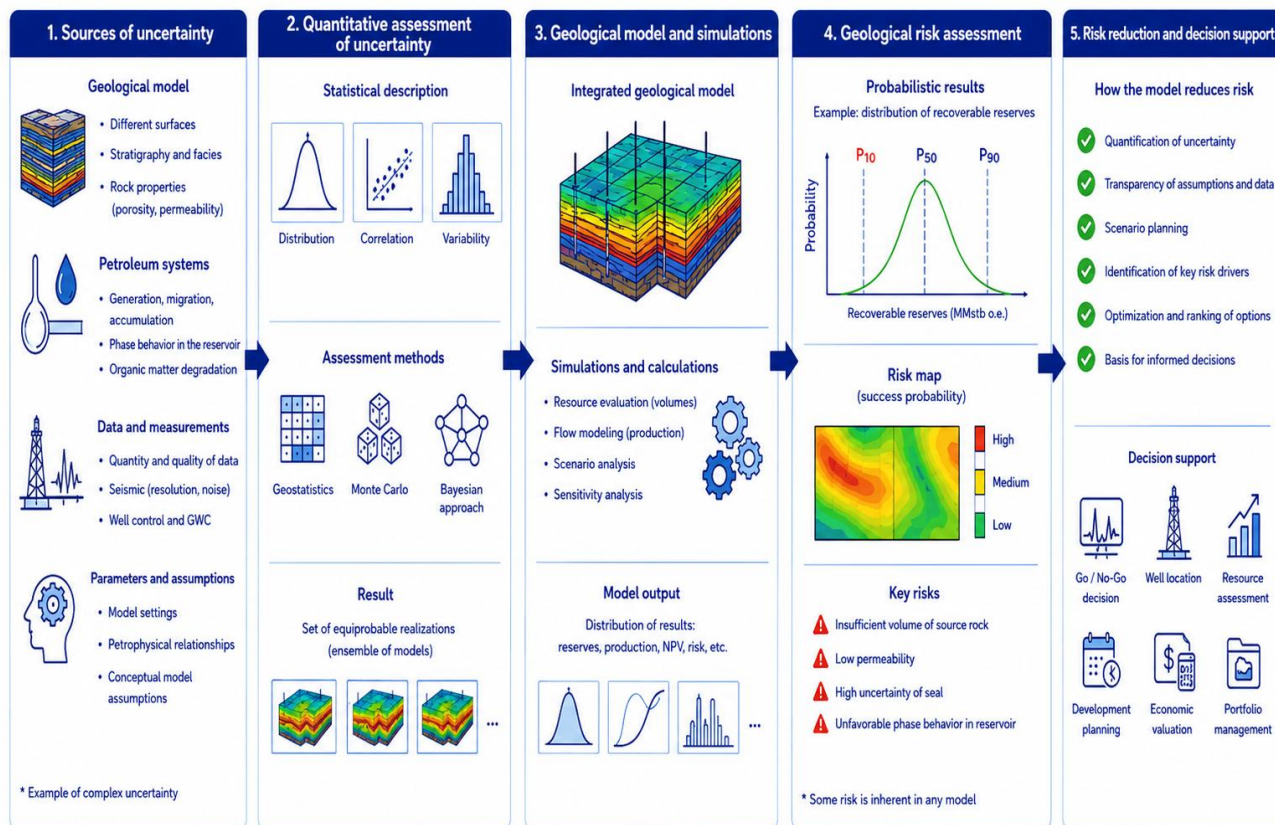


Figure 5. Mechanism of geological risk mitigation through modeling. Initial geological information is always incomplete (uncertainty). The objective of the intelligent platform is not to eliminate this uncertainty (which is impossible) but to quantitatively evaluate the associated geological risks (the probability of failure) through probabilistic modeling. This allows for the selection of drilling targets with the minimal amount of risk.

5. Results of intelligent geological exploration development

5.1. Forecast accuracy: verification results

As a result of the research conducted by the authors, the digital platform "Digital Caspian" has been created, which enables the implementation of a full digital engineering cycle—from creating a virtual prototype of a sedimentary basin to managing geological exploration. Within the framework of this study, the conceptual platform "Smart Exploration — Smart Field" was developed and implemented, combining the methodology of hydrocarbon systems digital modeling with hybrid AI technologies (Fig. 6). The practical implementation of these principles using the example of the Caspian region is presented as the information platform "Digital Caspian." This platform combines

digital twins of geological systems, multidimensional geological and environmental models, as well as models of hydrocarbon systems and oil and gas fields, AI models for forecasting oil and gas potential, environmental threats, and economic scenarios, GIS systems for analyzing geological data, integrating cartographic information, and visualizing results, as well as decision-making tools with automated recommendation support, thereby creating an intelligent system for geological exploration (Smart oil & gas exploration) [19-21].

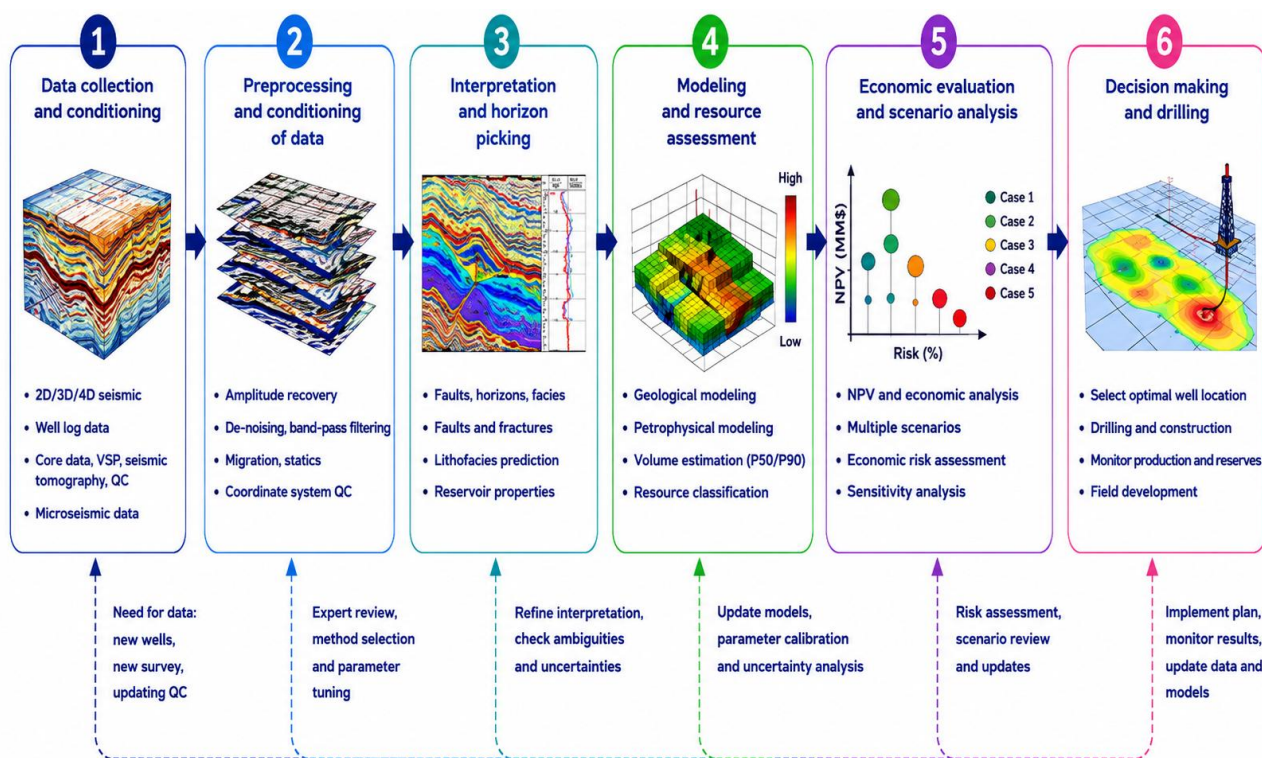


Figure 6 – Architecture of intelligent geological exploration (Smart Exploration)

The diagram illustrates an end-to-end data processing pipeline: from heterogeneous geological and geophysical information (seismic, well logging, geochemistry) through integration and AI modeling to the final practical result—a forecast and an informed decision (for example, recommendations for positioning an exploration well). Here, artificial intelligence acts as a key connecting link between "big data" and a business decision. Data from sensors (Sensor) is transmitted to the cloud infrastructure (Cloud), where AI models perform process optimization. Artificial intelligence here serves as a link between "big data" and a business decision.

In the context of the "Digital Caspian" platform, this architecture is transformed into a specialized interactive database (IDB) comprising eight key blocks: "Structural Framework," "Hydrocarbon Indicators," "Paleogeography," "Ecology," "GAUS Elements," "Boundary Conditions," "Modeling Results," and "Calibration Data" (Fig. 6). A significant advantage of such an

IDB compared to traditional geographic information databases is the ability to visualize its content without the use of specialized software (SW). This allows a wide range of specialists to utilize the collected and systematized geological information, including those who do not possess skills in specific specialized SW and/or do not hold the relevant licenses. Within this platform, thousands of numerical models replicating geological and environmental conditions were generated. Three-dimensional models developed through numerical modeling were transformed into interactive geological and environmental digital maps, as well as maps of hydrocarbon systems and oil and gas fields. The complete transformation cycle includes: data preparation in QGIS, 3D grid construction in Petrel, model calculation in PetroMod, analysis in QGIS, and finally, IDB creation in Python with the integration of AI tools.

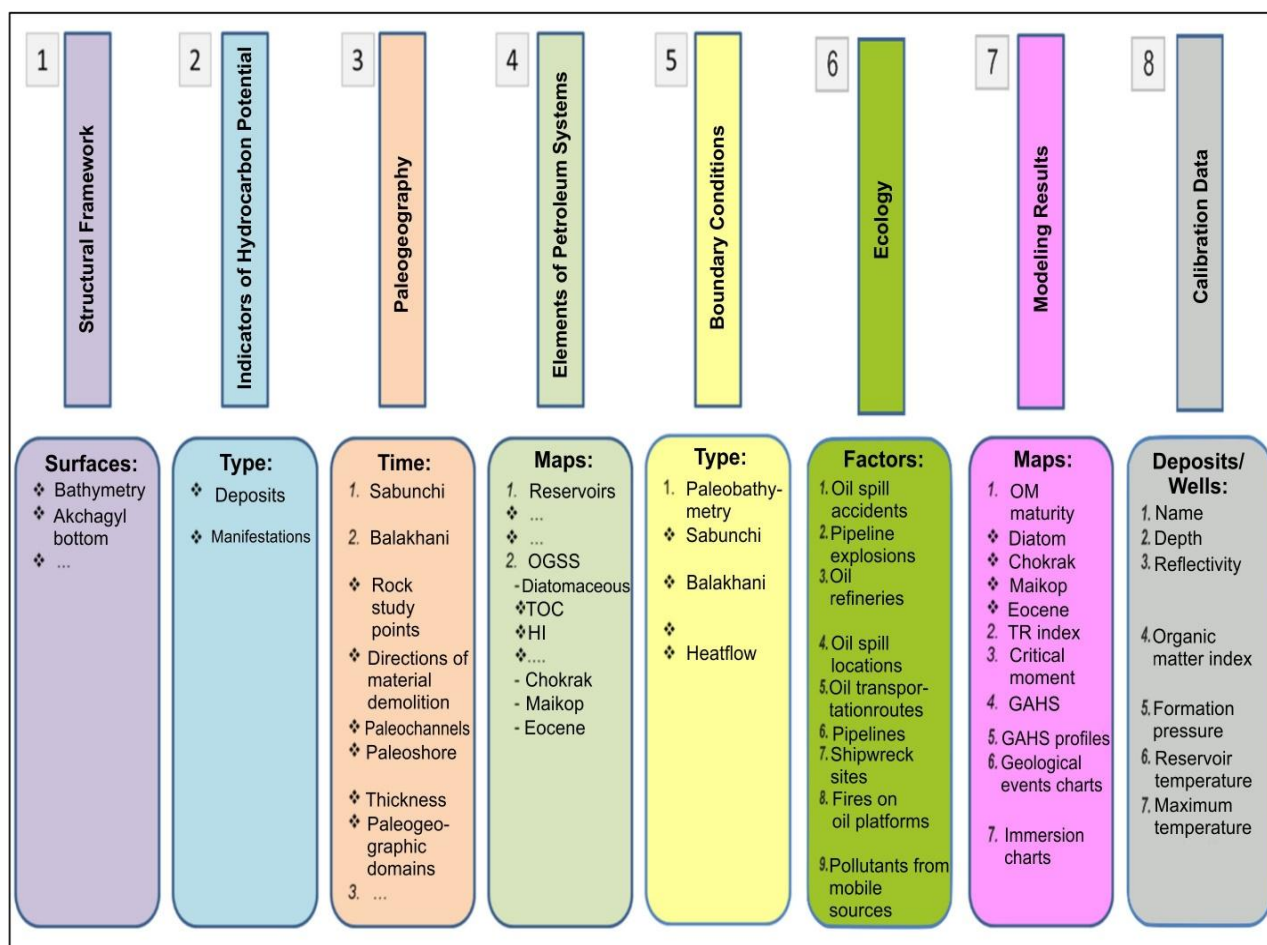


Figure 7. Principal Structure of the "Digital Caspian" IDB

Figure 8 shows digital structural-spatial models as an example, and figure 9 presents the results of numerical modeling of the Caspian Sea hydrocarbon systems generated by the digital platform.

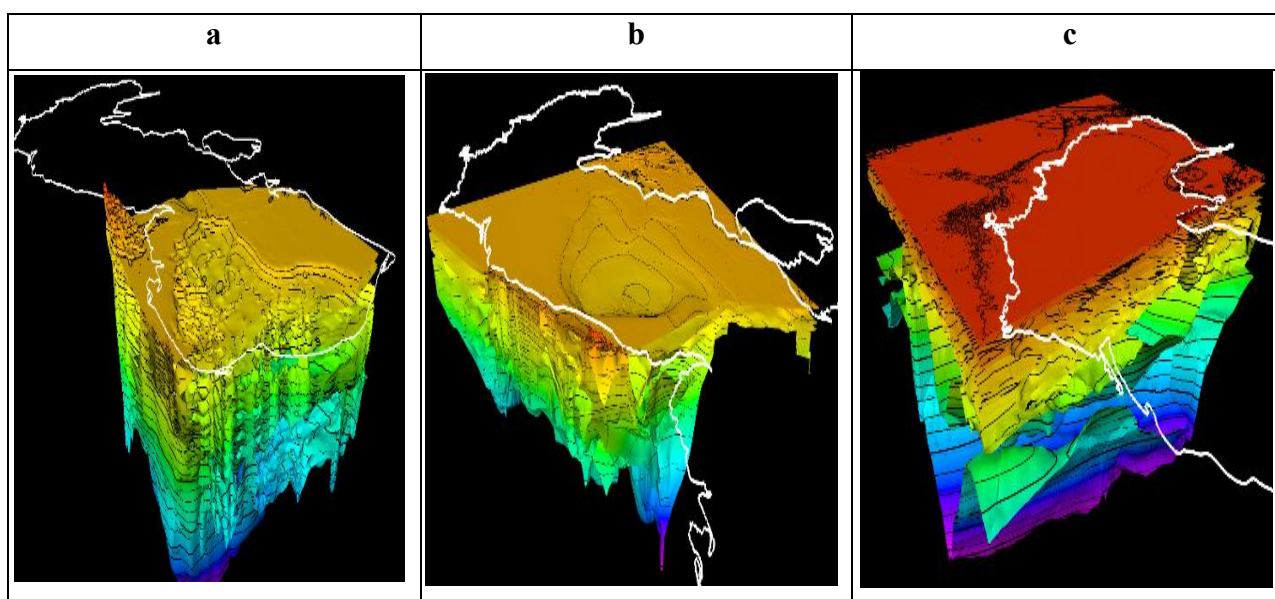


Figure. 8. Three-dimensional structural-spatial models of sedimentary basins:
a – South Caspian; b – Central Caspian; c – North Caspian

Three-dimensional and two-dimensional models developed through numerical modeling—based on the mathematical models discussed above and replicating geological and environmental conditions as well as hydrocarbon system elements—were transformed into interactive geological and digital maps, as well as maps of hydrocarbon systems and oil and gas fields (Fig. 8).

The complete cycle of model-to-map transformation includes: 1. Preparation and systematization of input data in QGIS → 2. Digitization of maps in QGIS → 3. Construction of 3D grids in Petrel → 4. Calculation of a 3D geological model in PetroMod → 5. Analysis of the obtained results in QGIS → 6. Creation of an interactive database in the Python programming language with the use of artificial intelligence, which includes web interface development, integration of resulting maps, graphs, cross-sections, well and calibration data, and the implementation of AI-driven analysis and systematization tools (Fig. 9). As a result, thousands of numerical models replicating the geological and environmental conditions of the three Caspian sub-basins have been generated.

The following have been created within the platform: structural-spatial models of the South Caspian, Central Caspian, and North Caspian sub-basins; digital twins and models of hydrocarbon systems (generation potential transformation ratio (TR) model, catagenetic evolution model, hydrocarbon expulsion and accumulation models, and the Maykop–Pliocene hydrocarbon system model); and interactive maps: structural-tectonic, paleogeographic, heat flow, and environmental maps [22-24].

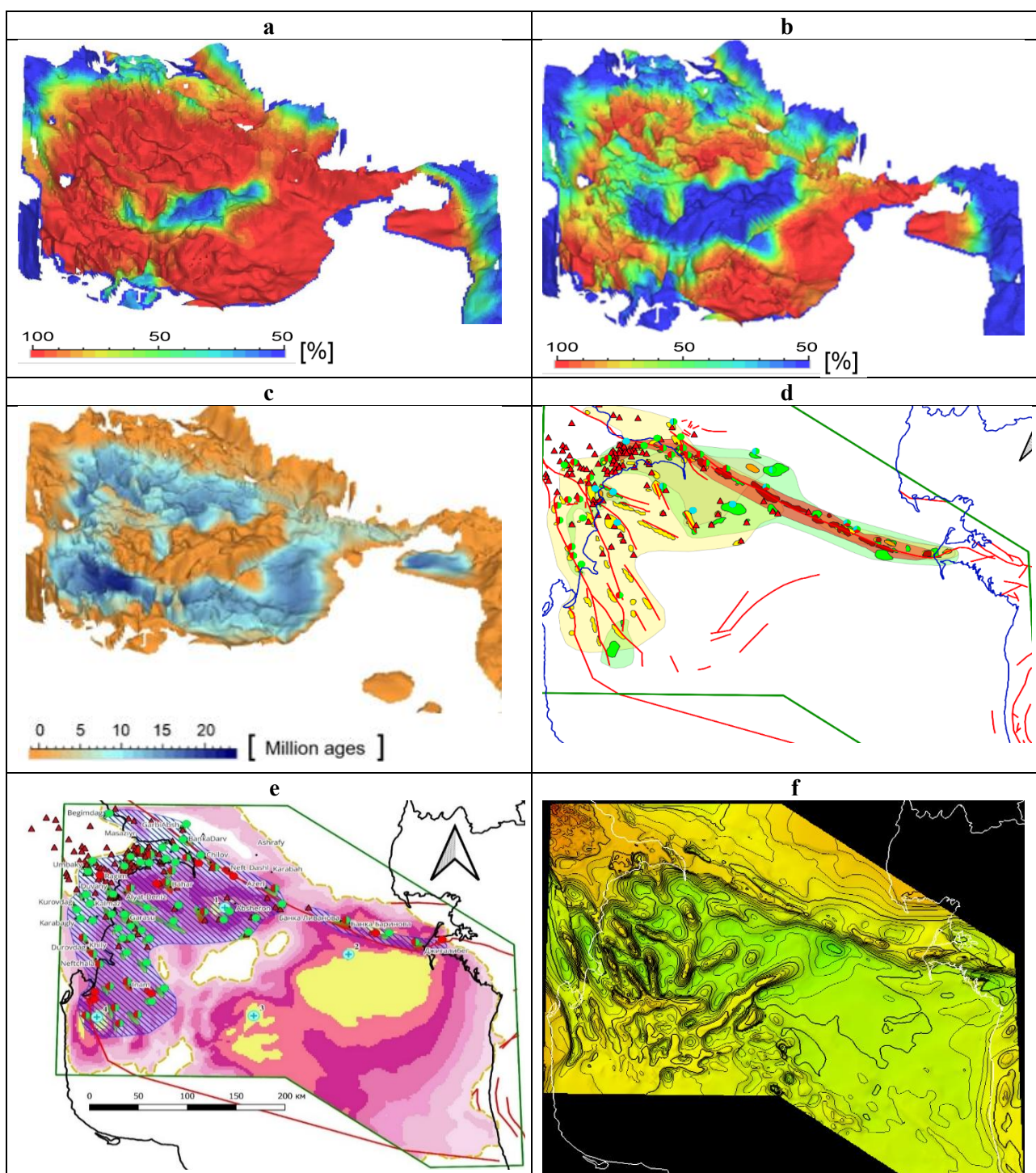


Figure 9. Examples of digital twins and models of hydrocarbon systems of the South Caspian Basin: a – generation potential transformation ratio (TR) model; b – hydrocarbon generation (catagenetic evolution) model; c – hydrocarbon expulsion model; d – hydrocarbon accumulation model; e – Maykop–Pliocene hydrocarbon system model; f – models of the structural plan of the productive strata surface.

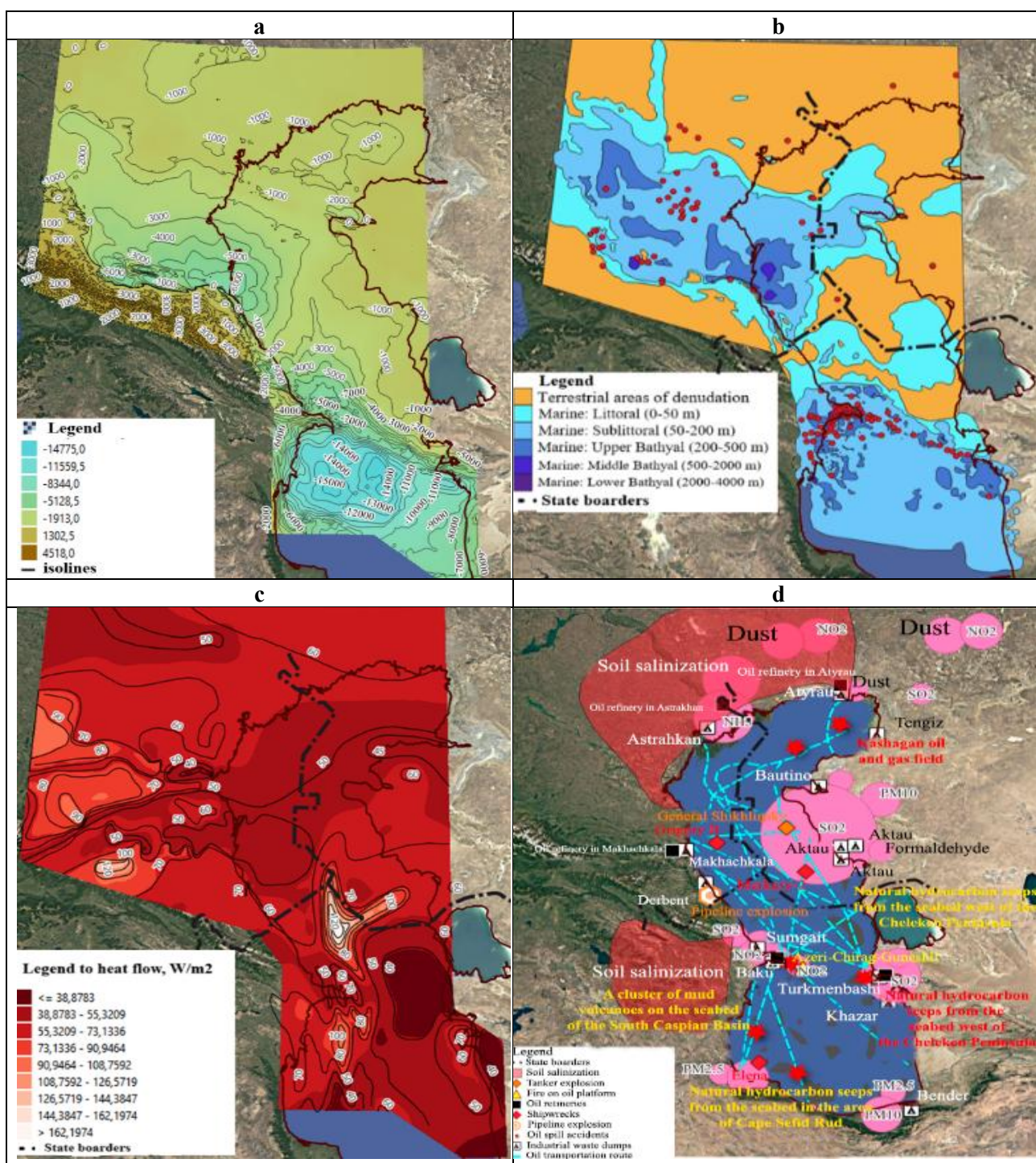


Figure 10. Examples of digital maps transformed from 3D models: a – structural-tectonic maps; b – paleogeographic maps; c – heat flow maps; d – environmental maps

Selecting the well location is the ultimate goal of exploration. The model provides not just a binary "oil / no oil" forecast, but the probability of success. Drilling locations are selected based on the maximum probability, factoring in risks and budget. In practice, instead of the maximum probability alone, the maximization of the expected net present value (NPV): $\text{Optimal} = \text{argmax}(P \times (\text{Reserves} \times \text{Price} -$

Cost). This allows for the formalization and optimization of the exploration investment portfolio (Fig. 11).

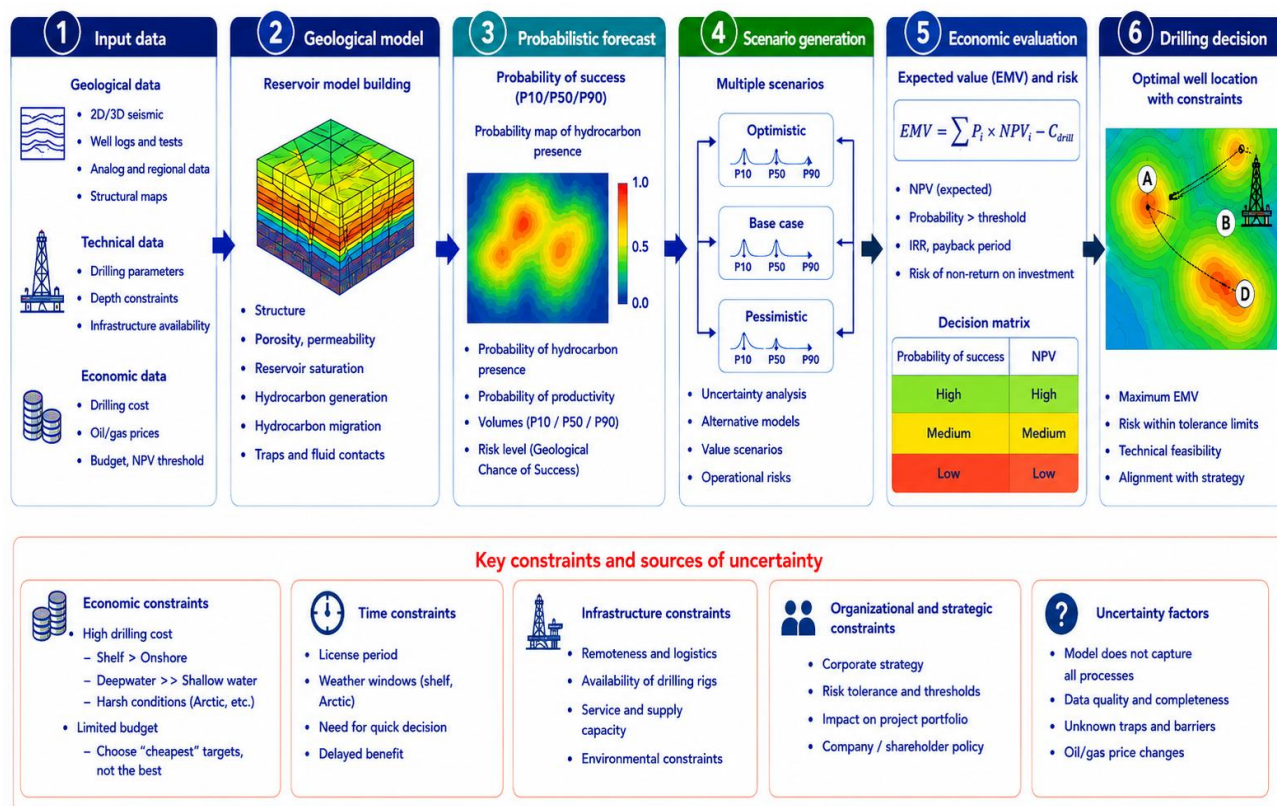


Figure 11. Well selection decision-making workflow: Model → Probability → Well Selection

Blind testing results ($n = 47$ wells) demonstrated that the hybrid PINN model achieved an AUC-ROC of 0.84 ± 0.03 , which is 23 percentage points higher than the baseline (0.61 ± 0.05). The relative gain in accuracy ranged from 25% to 40%, depending on the geological complexity of the asset. The largest increase (up to 40%) was observed for stratigraphic traps in deeply buried horizons (>5000 m), which is explained by the role of physical constraints during extrapolation under high-temperature and high-pressure conditions [25-27].

Table 2. Verification results of the "Digital Caspian" platform (blind test, n = 47 wells)

Method	AUC-ROC	F1-score	Test sample	Source
Traditional empirical (baseline)	0,61 ± 0,05	0,58 ± 0,06	47 wells, Caspian	Partner data
Bayesian inversion (аналог [7])	0,71 ± 0,04	0,68 ± 0,05	47 wells, Caspian	[7]
Unconstrained CNN	0,78 ± 0,03	0,74 ± 0,04	47 wells, Caspian	This work
Hybrid PINN ("Digital Caspian")	0,84 ± 0,03	0,81 ± 0,03	47 wells, Caspian	This work

5.2. Uncertainty quantification and geological risk reduction

Probabilistic geological risk assessment is an integral part of the developed platform. For each predicted asset, the components of the overall risk are calculated:

$$R = P_{\text{geol}} * P_{\text{res}} * P_{\text{seal}} * P_{\text{gen}} * P_{\text{migr}},$$

where P_{geol} is the probability of trap presence; P_{res} is the probability of reservoir presence with acceptable reservoir quality (porosity and permeability); P_{seal} is the probability of effective seal presence; P_{gen} is the probability of hydrocarbon generation from the source rock; and P_{migr} is the probability of hydrocarbon migration and accumulation in the trap. The overall risk R is used to rank assets and optimize investment decisions.

The calibration of probabilistic estimates was confirmed with an Expected Calibration Error (ECE) of 0.047 on the test sample. Confidence intervals were calculated using the Deep Ensembles method (10 independently trained models). Rather than selecting the maximum probability alone, applying the maximization of the expected net present value NPV: $\text{Optimal} = \text{argmax} (P \times (\text{Reserves} \times \text{Price} - \text{Cost}))$ allows for the formalization and optimization of the exploration investment portfolio.

The economic impact of technology implementation is evaluated in two aspects: avoided losses from inefficient exploration and additional profit from resource base growth. The application of the developed methods in structurally complex sedimentary basins (Arctic shelf, rift zones, deeply

buried horizons) provided a 25–40% increase in hydrocarbon potential forecast accuracy compared to local empirical methods. A substantial reduction in geological risks and the avoidance of losses from inefficient drilling in the amount of tens of billions of rubles have been achieved.

5.3. Reducing dependence on imported software

The indicator of reduced dependence on imported software was defined as the share of exploration tasks solved using foreign software, calculated per volume of computing resources across six subsurface-user organizations that partnered in the project. Over the 2019–2024 period, this indicator decreased from 90% to 40%. This reduction was achieved by replacing foreign programs (including Petrel and PetroMod since 2022) with domestic components of the "Digital Caspian" platform in hydrocarbon potential forecasting and exploration management tasks.

6. Discussion

6.1. Method limitations

Reliable performance of the hybrid PINN model requires a minimum volume of training data: at least 50 wells with a full suite of well logging data for a specific basin. With a smaller data volume, accuracy decreases to an AUC-ROC of 0.68–0.72. When transferring the platform to geological provinces that fundamentally differ from the Caspian region (e.g., rift structures of the Far East), additional calibration of physical constraints is required. The computational budget for a full calculation is 120–180 GPU hours (NVIDIA A100) per basin.

Economic assessments of avoided losses are preliminary and will be presented in detail in a separate applied publication. A critical barrier identified during the analysis is the ongoing dependence of Russian companies on imported exploration software, particularly in the basin modeling and seismic inversion segments, which the "Digital Caspian" platform is systematically overcoming.

6.2. Comparison with competing approaches

The proposed approach differs from baseline PINN methods [28, 29] in that physical constraints are implemented not at the level of a single well, but rather provide end-to-end hierarchical modeling from the basin scale down to the field scale. Russian analogues [3, 7] address isolated tasks without providing integration across the full hydrocarbon system (HCS) chain. Foreign commercial platforms (Schlumberger DELFI, Halliburton DecisionSpace 365) are not only unavailable to Russian users due to sanctions but also do not disclose their algorithmic components. The theoretical

foundation—PINN, Karniadakis et al., transformer architectures for seismic data, and Grieves' digital twin concept [30-32]—highlights the global trends in accordance with which the proposed platform is positioned.

7. Conclusion

1. Scientific novelty. An original hybrid AI architecture based on PINN has been developed and verified, integrating physical laws (Darcy's law, mass conservation) with neural network components (CNN, BNN) through a unified loss function. For the first time in the Caspian region, an end-to-end platform has been implemented with a verified accuracy of AUC-ROC = 0.84 (blind test, $n = 47$) against 0.61 for traditional methods, representing a 25–40% increase. The "Digital Caspian" information platform has been created, providing a unique example of implementing such an approach for a strategically important region by combining AI, GIS, and digital twins.

2. Technological breakthrough. The "Digital Caspian" platform has been created, integrating digital twins of geological systems, PINN models, GIS, and decision support tools with SHAP-based explainability. An ECE of 0.047 confirms the high calibration of probabilistic risk assessments. The technologies include basin and geodynamic modeling methods, practical tools for creating digital twins and integrating heterogeneous geodata, and a series of interactive geological and environmental digital maps.

3. Strategic significance. The results have been implemented into the practice of six subsurface-user organizations, with software import dependence reduced from 90% to 40% over the 2019–2024 period. The outcomes have been incorporated into state programs for the development of the Arctic Zone of the Russian Federation and the Mineral Resource Base Development Strategy until 2035. The "Digital Caspian" platform can be adapted to other marine and coastal regions of Russia, ensuring the transition to domestic intelligent exploration technologies.

4. Practical efficiency. The application of the developed methods in structurally complex sedimentary basins (Arctic shelf, rift zones, deeply buried horizons) provides a 25–40% increase in hydrocarbon potential forecast accuracy. Subsurface users and oil and gas companies receive a tool to identify prospective hydrocarbon zones; environmental agencies gain the capability for rapid monitoring; and scientific institutions acquire a unified platform for interdisciplinary research.

5. Limitations and future directions. The key limitations (minimum data volume ≥ 50 wells, computational budget of 120–180 GPU hours, and the need for calibration when transferring to new provinces) define the directions for further research: (1) few-shot learning for small samples; (2) development of hybrid neural network architectures such as physics-informed neural networks

(PINN) for new regions; (3) creation of a distributed network of digital twins with predictive analytics; and (4) automated generation of geological models based on AI-driven analysis of accumulated geodata.

References

1. Polovkov, V. et al. SPbU scientists train a neural network to analyze seismic data from offshore oil and gas fields. *Proceedings of SPbU [or Izvestia of Saint Petersburg State University]*, 2025.
2. Brown D. AI and the 300-Billion-Barrel Oil Gap // *AAPG Explorer*, 2025, December 1.
3. Priyetzhev, I.I., Danko, D.A., and Onishchenko, A.N. Hierarchical neural networks in the tasks of predicting the properties of oil and gas reservoirs from well and seismic data. *Russian Geology and Geophysics*, 2025, vol. 66, no. 1, pp. 131–140.
4. Russian researchers develop AI model to improve oil extraction efficiency // *SANA*, 2026, April 15.
5. IADC/SPE 2026 Day one: Drilling technology continues a relentless advance // *World Oil*, 2026, March 19.
6. A neural network predicting oil properties was created in Russia. *CNews*, April 15, 2026.
7. Mitrofanov G., Korytkin E. Improving the accuracy of reservoir properties prediction using machine learning // *Russian Geology and Geophysics*, 2025, v. 66, No. 9, p. 1160–1169.
8. Huawei and Partners Launch Joint Solutions for Intelligent Oil & Gas // *Huawei Enterprise*, 2025.
9. Liu X., Sun M., Lin B., Gu S. SGP-GCN: A Spatial-Geological Perception Graph Convolutional Neural Network for Long-Term Petroleum Production Forecasting // *Energy Engineering*, 2025.
10. Dmitrievsky, A.N., Eremin, N.A., Filippova, D.S., and Safarova, E.A. Digital oil and gas complex of Russia. *Georesursy = Georesources*, 2020, Special Issue, pp. 32–35. DOI: 10.18599/grs.2020.SI.32-35.
11. Khasanov, M.M. et al. Prospective Big Data technologies in petroleum engineering. *Neftyanoe Khozyaystvo = Oil Industry*, 2016, no. 12, pp. 76–79.
12. Varlamov, A.I. et al. Status and prospects for the development of digital technologies in oil and gas geology. *Geologiya Nefti i Gaza = Oil and Gas Geology*, 2021, no. 3, pp. 5–20. DOI: 10.31087/0016-7894-2021-3-5-20.

13. Cheremisina, E.N. et al. Digitalization in geological exploration. *Geoinformatika = Geoinformatics*, 2021, no. 4, pp. 18–27. DOI: 10.47148/1609-364X-2021-4-18-27.
14. Kerimov V.Yu., Guliev I.S., Dzhavadova A.S. et al. Characteristics of oil and gas source strata and features of hydrocarbon systems of the South Caspian Basin // *Bulletin NAS Azerbaijan. Earth Sciences*, 2024, (1), pp. 77–92.
15. Kerimov V.Yu., Bondarev A.V., Mustaev R.N. Estimation of geological risks in searching and exploration of HC deposits // *Neftyanoe Khozyaystvo*, 2017, (8), pp. 36–41.
16. Kerimov V.Yu., Shilov G.Ya., Mustayev R.N., Dmitrievsky S.S. Thermobaric conditions of HC accumulation in low-permeability reservoirs of the Khadum Suite // *Neftyanoe Khozyaystvo*, 2016, (2), pp. 8–11.
17. Kerimov V.Yu., Mustaev R.N., Etimishli G.D., Yusubov N.P. Influence of modern geodynamics on the structure of the Black Sea–Caspian region // *Eurasian Geology*, 2021, (5), pp. 3–8.
18. Kerimov V.Yu., Serikova U.S., Mustayev R.N., Guliyev I.S. Deep oil-and-gas potential of the South Caspian Basin // *Neftyanoe Khozyaystvo*, 2014, (5), pp. 50–54.
19. Lapidus A.L., Kerimov V.Y., Mustaev R.N. et al. Natural bitumens: physicochemical properties // *Solid Fuel Chemistry*, 2018, 52(6), pp. 344–355.
20. Guliev S., Mustaev R.N., Kerimov V.Y., Yudin M.N. Degassing of the earth: Scale and implications // *Gornyi Zhurnal*, 2018, (11), pp. 38–42.
21. Kerimov V., Rachinsky M., Mustaev R., Serikova U. Geothermal conditions of HC formation in the South Caspian basin // *Iranian Journal of Earth Sciences*, 2018, 10(1), pp. 78–89.
22. Gordadze G., Kerimov V., Giruts M. et al. Genesis of the asphaltite of the Ivanovskoe field // *Fuel*, 2018, 216: 835–842.
23. Gordadze G.N., Kerimov V.Y., Giruts M.V. et al. Hydrocarbon biomarkers from Precambrian rocks of the Katanga saddle // *Geochemistry International*, 2017, 55(4): 360–366.
24. Bogoyavlensky V.I., Kerimov V.Y., Olkhovskaya O.O. Dangerous gas-saturated objects: the Sea of Okhotsk // *Oil Industry*, 2016, (6), pp. 43–47.
25. Roadmap "Neurotechnologies and Artificial Intelligence." — Moscow: Ministry of Digital Development, Communications and Mass Media of the Russian Federation, 2019.
26. Roadmap "Virtual and Augmented Reality Technologies." — Moscow: Ministry of Digital Development, Communications and Mass Media of the Russian Federation, 2019.
27. Roadmap "New Manufacturing Technologies." — Moscow: Ministry of Digital Development, Communications and Mass Media of the Russian Federation, 2019.

28. Raissi M., Perdikaris P., Karniadakis G.E. Physics-informed neural networks // Journal of Computational Physics, 2019, 378, pp. 686–707. DOI: 10.1016/j.jcp.2018.10.045.
29. Karniadakis G.E., Kevrekidis I.G., Lu L. et al. Physics-informed machine learning // Nature Reviews Physics, 2021, 3(6), pp. 422–440. DOI: 10.1038/s42254-021-00314-5.
30. Grieves M., Vickers J. Digital Twin: Mitigating Emergent Behavior in Complex Systems // Transdisciplinary Perspectives on Complex Systems. — Springer, 2017. — P. 85–113.
31. Wang Z., Li H., Zhang Q. Seismic facies classification using transformer-based deep learning // Geophysics, 2024, 89(1), pp. WA1–WA15.
32. Chen T., Liu Y., Zhao W. Attention mechanisms for seismic interpretation: a comparative study // Journal of Petroleum Science and Engineering, 2025, 230, 212847.