

REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

**THE MODELS OF SYNTHESIZED SEISMIC IMPACT
IN THE SHAMKIR WATER RESERVOIR AREA
BASED ON SCENARIO EARTHQUAKES DATASET**

Speciality: 2507.01 – “Geophysics, methods of
geophysical exploration of minerals”

Field of science: Earth sciences

Applicant: **Tural Hikmat Babayev**

Baku – 2026

The work was performed at the Department of Seismology and Seismic Hazard Assessment of the Institute of Geology and Geophysics (currently, the Institute of Geology) of the Ministry of Science and Education of the Republic of Azerbaijan.

Scientific supervisor: Doctor of Earth Sciences, Professor
Gulam Rustam Babayev

Official opponents: Doctor of Physical and Mathematical Sciences, Associate Professor
Adalat Badal Hasanov

Doctor of Philosophy (Ph.D) on Geology and Mineralogy, Associate Professor
Hafiz Ismail Shakarov

Doctor of Philosophy (Ph.D) on Geology and Mineralogy, Associate Professor
Sabina Eldar Kazimova

Dissertation council FD 1.01 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Institute of Geology of the Ministry of Science and Education of the Republic of Azerbaijan.

Chairman of the Dissertation council:

Doctor of Geological and Mineralogical Sciences, Academician, Professor
Akper Akper Feyzullayev

Scientific secretary of the Dissertation council:

Doctor of Philosophy (Ph.D) in Technical Sciences, Associate Professor
Dilgusha Ramzey Mirzayeva

Chairman of the scientific seminar:

Doctor of Geological and Mineralogical Sciences, Professor
Tofiq Rashid Ahmadov

INTRODUCTION

Relevance and processing degree of the topic. The development of the science of seismology has come a long way in recent years. As one of its branches, the seismic hazard assessment has also become a promising field of science with the application of improved methods.

The relative motions of tectonic plates within the Earth's lithospheric layer generate seismic activity by triggering earthquakes. It has been established that Azerbaijan is characterized by several seismically active zones, including the Gusar–Shabran region, the southeastern slope of the Greater Caucasus, the Kur Depression, the northeastern slope of the Lesser Caucasus and the Talysh Mountains. The existence of infrastructure within these seismically active regions underscores the necessity of comprehensive seismic hazard assessment studies. From this perspective, the investigation of regional seismicity and the evaluation of seismic hazard in areas hosting hydrotechnical facilities in Azerbaijan represent an issue of significant contemporary relevance. In this context, seismic hazard assessment based on modern datasets and analytical capabilities in the Shamkir water reservoir area, situated within the Kur intermountain mega-depression between the Greater and Lesser Caucasus mountain-fold systems, is of particular importance.

In addition, seismic hazard studies in the area are important, since the rivers passing through the study area are used for irrigation and other economic activities and the active operation of large reservoirs and hydroelectric power plants creates a possibility of technogenic earthquakes. Studies have been carried out by scientists such as G.R.Babayev, F.A.Kadirov, G.J.Yetirmishli, L.Telesca, P.Talwani, O.Mikhailov, K.Malik and S.Liu in areas where hydrotechnical facilities are located. A number of studies indicate that continuous tremors caused by water-level changes are observed in the Shamkir reservoir area. This demonstrates that the seismic hazard assessment in the area is an urgent and relevant issue today.

Research object and subject. The Shamkir water reservoir and its surroundings, located in the Shamkir region of the Republic of Azerbaijan, were selected as the object of the research. The study area covers the administrative regions of Shamkir, Ganja, Samukh and

Tovuz and has a total area of 2608.674 (Two thousand six hundred eight point six seven four) km². The area includes strategically important facilities such as the Shamkir reservoir, Shamkir Hydroelectric Power Station (HPS), Yenikend reservoir, Yenikend HPS and Ganja International Airport (IA).

The subject of the research is the proposal, presentation and application of an improved modern methodology based on new approaches for seismic hazard assessment in the case of the Shamkir reservoir and its surroundings.

Research aim and objectives. The aim of the dissertation is to compile a modern seismic hazard map in the Shamkir reservoir area based on deterministic assessment of seismic hazard. For this purpose, **the following objectives** are sequentially performed within the framework of the research:

1. Conducting geological-geophysical analyses and studying the site effect in order to take into account the impact of complex geological conditions in predicting seismic hazard in the research area. For this purpose, compiling a distribution map of the average shear wave velocity at a depth of 30 meters (V_{S30}) and performing computer algorithms and using contemporary data to conduct the site classification;

2. Conducting an analysis of recent seismicity in the research area, the use of digital seismic stations in comprehensive implementation of seismic monitoring. Performing and using computer algorithms for the organization and application of a methodology in order to obtain accelerograms of earthquakes in the recent past and compiling a modern digital database;

3. Analyzing the seismic regime in the study area, determining the maximum magnitude, potential earthquake foci based on bibliographic information, as well as statistical research of seismicity catalogs, the proposal, presentation and application of a methodology of generating artificial accelerograms by integrating earthquake and site parameters into modern software and constructing a synthetic database;

4. Proposing, presenting and applying a methodology based on the use of machine learning algorithms in the development of local ground motion models reflecting real conditions by training real and

synthetic databases designed to replace traditional ground motion prediction equations and comprehensively predict peak ground acceleration (PGA) at the scenario earthquake focus;

5. Application of deterministic seismic hazard assessment method to the study area, prediction of the PGA by determining “the worst-case scenario”. Compilation of the attenuation graph and spatial-distribution map of the predicted PGA;

6. To conduct seismic zoning in the study area, a seismic hazard map was compiled in intensity units on the MSK-64 scale using PGA-intensity relation.

Research methods. The deterministic method is applied in seismic hazard assessment using artificial intelligence techniques. The seismic hazard assessment for the study area is carried out in the following sequence:

1. Identification of potential earthquake foci (active tectonic faults) based on geological-geophysical data in the study area;

2. Analysis of the seismic regime in the study area and determination of the maximum magnitude;

3. Selection of the most active earthquake source (Kur active tectonic fault) as a scenario earthquake source and defining the scenario earthquake and magnitude for “the worst-case scenario”;

4. Locating the scenario earthquake at every 10 km along the 170 km section of the Kur active tectonic fault and determining the focal depths based on catalog analyses;

5. Segmentation of the study area: the study area is divided into a grid of square cells of size 10 km * 10 km, the coordinates of the nodal points of the cells are noted. Performing computer algorithms to determine the site characteristics of the segments using the V_{S30} distribution map and site classes;

6. Measurement and calculation of epicentral and hypocentral distances from the scenario earthquake to the grid nodes;

7. Predicting PGA (in gals) with models developed based on scenario magnitude, hypocentral distance, V_{S30} and site class data.

The key points of the defense. The main key points are put forward for the defense:

1. Ground motion modeling in the study area, based on the

integration of modern databases with machine learning (ML) algorithms, taking into account earthquake parameters and site characteristics;

2. Spatial prediction of the peak ground acceleration (PGA) for a scenario earthquake based on ground motion models and compilation of a seismic hazard map for the area.

Scientific novelty of the research. As a result of the research conducted, **the following novelties** were obtained:

1. By performing computer algorithms and integrating contemporary data, a V_{S30} distribution map has been compiled for the area. V_{S30} characterization allows for the site classification in the area, the study of site effect and the consideration of the impact of complex geological conditions in seismic hazard prediction;

2. For the first time, methodologies for compiling and processing modern digital databases using the Python programming language have been proposed for the purpose of ground motion modeling in the study area;

3. By applying a method based on the integration of modern software, including the maximum magnitude set determined based on seismicity analyses in the area, as well as other earthquake and site parameters, artificial accelerograms were generated and a synthetic database was compiled for the study area;

4. Based on the integration of real and synthetic databases compiled in the research area, ML algorithms were used for the first time and local ground motion models (GMM) were developed. Comparative analyses of ML-based models were conducted and the most efficient models were selected and used in seismic hazard assessment;

5. The concept of “the worst-case scenario” was applied within the deterministic approach in the research area. Using the developed local GMMs, PGA was predicted for the specified scenario and finally the ground motion attenuation graph and PGA spatial-distribution map were compiled (in gals);

6. The predicted PGA values were correlated to intensity units on the MSK-64 scale and a seismic hazard map (in intensity units) for the study area was presented.

Theoretical and practical significance of the research. The research is dedicated to seismic hazard assessment in the Shamkir reservoir area, one of the largest reservoirs in Azerbaijan. Theoretically, given that seismic hazard assessment studies are highly relevant at the global scale, the dissertation aims to contribute to the application of modern approaches in Azerbaijani science by integrating methods such as ML algorithms, the development of GMMs and the generation of synthetic datasets used in contemporary scientific practice.

From a practical perspective, the dissertation contributes to ensuring safe conditions by determining the level of earthquake hazard that strategically important facilities may face in the future, as well as the seismic impact to which they may be exposed during earthquakes in scenario-based analyses.

Thus, the proposal, presentation, approval and application of an integrated complex methodology enhanced by modern approaches for seismic hazard assessment and seismic zoning in the Shamkir reservoir area demonstrate the methodological-applied nature and the scientific-practical significance of the research.

Approbation and application. Part of the research conducted in the dissertation was presented at the General Assembly of the European Geosciences Union held in Vienna, Austria (2024), the 15th General Assembly of the Asian Seismological Commission held in Antalya, Türkiye (2024) and the 16th General Assembly held in Tashkent, Uzbekistan (2026) and a number of other international scientific conferences. A total of 13 scientific publications have been published on the topic of the dissertation.

Name of the organization where the dissertation was performed. The dissertation was performed at Geology and Geophysics Institute (now Geology Institute) of the Ministry of Science and Education.

Total volume of the dissertation with characters, indicating the volume of the structural sections of the dissertation separately. The dissertation is presented in a total volume of 216666 characters, including an introduction (10173 characters), 5 chapters (Chapter I – 24051 characters, Chapter II – 28689 characters, Chapter III – 31497

characters, Chapter IV – 53510 characters, Chapter V – 63739 characters), conclusion (3083 characters), appendices (5 tables) and a list of abbreviations and symbols (1924 characters), 49 Figures, 25 Graphs, 27 Tables (22, excluding tables in the Appendices) and 196 references.

The personal contribution of the author. In the dissertation, the author characterized the geological–geophysical and tectonic–geodynamic conditions of the study area. He demonstrated and applied methodologies for obtaining, processing and analyzing the international seismicity catalog and analyzed the seismicity characteristics of tectonic faults. For the first time, he developed and used GMMs for the study area through the integration of ML algorithms, applied modern software for the compilation and processing of digital databases and the use of synthetic database generation methodologies, as well as implemented algorithms using the Python programming language. Based on these results, he assessed the seismic hazard in the Shamkir reservoir area and produced a modern seismic hazard map for the study area.

Acknowledgements. I would like to express my deep gratitude to my scientific supervisor, Professor G.R. Babayev, for his continuous support in completing my dissertation, for his significant contribution to my scientific development and for his valuable advice; to Academician F.A.Kadirov for his periodic and dedicated attention to my dissertation and his attentive guidance; and to Academician Ak.A.Alizadeh for the opportunities and conditions created for carrying out the dissertation research and for supporting and acknowledging my achievements.

I would like to express my heartfelt gratitude to all those close to me, especially my family members and the department staff I worked with, for their moral and material support throughout the completion of my dissertation.

GENERAL CHARACTERIZATION OF WORK

CHAPTER I. GEOPHYSICAL, GEOLOGICAL, TECTONIC AND GEOMORPHOLOGICAL CHARACTERISTICS OF THE STUDY AREA

This chapter is devoted to providing detailed and comprehensive information about the study area. A comprehensive analysis of the geological-geophysical conditions and tectonic-geodynamic characteristics of the area during seismic hazard assessment constitutes an essential initial stage of the study.

Accordingly, this chapter presents information about the study area, characterizes the site composition based on the shear-wave velocity at a depth of 30 m and emphasizes the relevance of the study by presenting strategically important facilities located in the area.

The research presented in the dissertation was conducted in the area covering the Shamkir reservoir, located in the Shamkir region of the Republic of Azerbaijan and the surrounding territories. The study area includes the administrative regions of Shamkir, Ganja, Samukh and Tovuz. It covers an area of 2608.674 (Two thousand six hundred eight point six seven four) km². Figure 1 presents the study area.



Figure 1. Description of the study area¹

¹ Image of Azerbaijan: [Electronic resource] / Google LLC, Google Earth Software.
– Mountain View, California, U.S., 2024. URL: <https://earth.google.com/web/>.

The study area is characterized by weakly and poorly lithified Pliocene and Quaternary sediments and volcanic formations. In the southern part, Upper Pliocene Agchagil sediments are predominant. These sediments consist of clastic conglomerates, sandstones, silty claystones, clayey siltstones and volcanic ash [1].

The United States Geological Survey (USGS)² provides global V_{s30} data using methods based on topographic and relief indicators, as well as referring to the information based on various combinations of lithological, geological and geomorphological indicators. By processing these data using the Python programming language, a V_{s30} map of the study area was compiled (Figure 2).

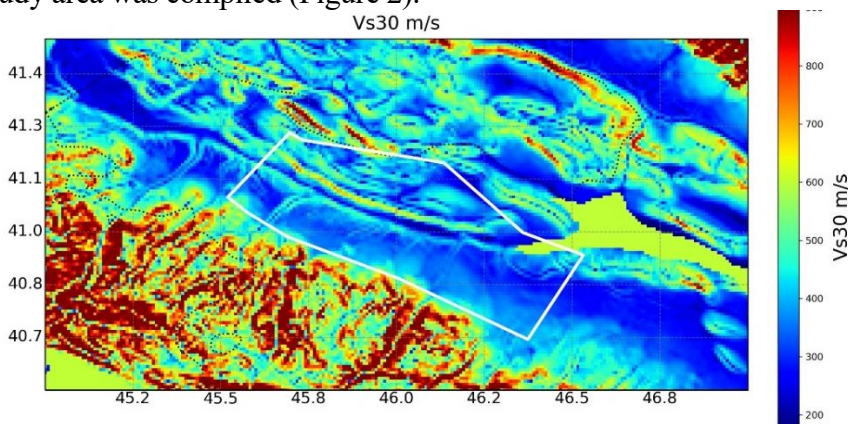


**Figure 2. V_{s30} map of the study
(Using and processing USGS² data, compiled by: T.H.Babayev)**

The study area includes several strategically important facilities: the Shamkir and Yenikend reservoirs and HPSs, as well as Ganja IA. Thus, against the background of active tectonic and geodynamic movements, the potential seismotectonic activity and seismogenic characteristics of the longitudinal Kur deep fault extending through the area and representing a potential source of large-magnitude earthquakes should be taken into account in seismic hazard assessment.

² Global Vs30 Data: [Electronic resource] / USGS: United States Geological Survey. – Reston, Virginia, United States, 2020. URL: https://apps.usgs.gov/shakemap_geodata/vs30/global_vs30.grd.

For this purpose, Chapter I is devoted to a detailed investigation of the study area from geophysical, geological and tectonic perspectives, as well as to the compilation of a generalized deep fault map based on various sources (Figure 3).

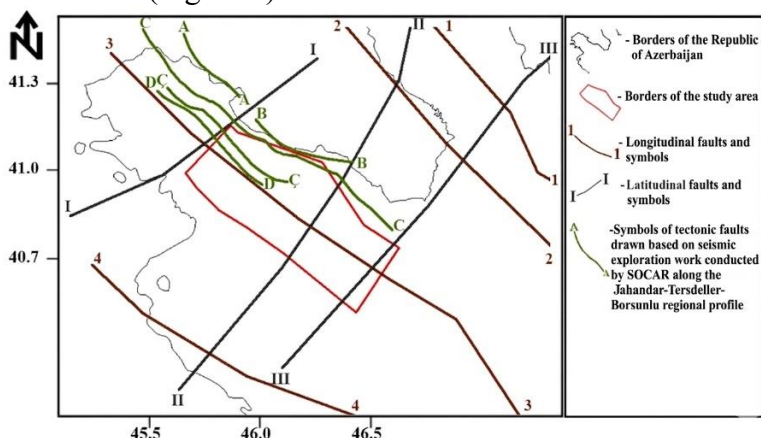


Figure 3. Generalized tectonic depth fault map for the study area, completed by the author of the dissertation based on the works of G.J.Yetirmishli et al.³, T.Babayev et al. [9] and the 2011 SOCAR project results⁴ (Based on data, compiled by: T.H.Babayev)

The map presented in Figure 3 is used in the subsequent chapters for seismic hazard assessment in the study area.

CHAPTER II. HISTORICAL SEISMIC EVENTS IN THE STUDY AREA AND THE CHRONOLOGICAL DEVELOPMENT OF SEISMIC HAZARD STUDIES

This chapter presents information on the history of seismicity research, an overview of historically devastating earthquakes, the

³ Yetirmishli, G.J. About the seismic risk of Sabayil district of Baku / G.J.Yetirmishli, T.Y.Mammadli, R.B.Muradov [et al.] // Seismoprognois Observations in the Territory of Azerbaijan, – 2017. 14, 1, – p. 12-21.

⁴ Aliyeva, I.P., Yunusova, S.A. Tectonic scheme of Neogene structures presented in the report of seismic exploration studies conducted in 2010 along the Jahandar–Tersdeller–Borsunlu regional and 2D profiles in the Kur–Gabiirri interfluve and Ganja OGR: [Electronic resource] / Exploration Geophysics Department of SOCAR Geophysical Research Institute. – Baku, 2011. – CD-ROM (in Azerbaijani).

development of seismic hazard assessment methodologies and the integration of Artificial Intelligence as a modern approach.

There are several strong earthquakes recorded in the history of Azerbaijan. For instance, the devastating earthquake of magnitude 6.7 at a depth of 12 km in 427 AD is considered one of the strongest earthquakes in the country's history. In 1139, an earthquake of magnitude 6.8 at a depth of 15 km occurred in the Ganja (Goygol) area, leading to the formation of Goygol, one of the most notable natural landmarks of Azerbaijan.

Improved approaches such as probabilistic and deterministic methods are used in seismic hazard assessment.

The initial concept of the probabilistic seismic hazard assessment (PSHA) method was proposed by C.A.Cornell⁵ in the 1960s. The main purpose of this method is to estimate the annual probability of exceeding ground motion parameters within a given time interval. PSHA is used for seismic hazard assessment over large areas such as cities or regions of countries.

The deterministic seismic hazard assessment (DSHA) method was developed in the 20th century. In this method, the most hazardous scenario for the study area is analyzed by calculating peak ground motion parameters. This approach is preferred for structures of special strategic importance, such as reservoirs, power plants and tunnels, as well as for relatively small study areas. Using this method, the author conducted studies in the Shamakhi–Ismayilli seismic zone [3, 4], Shamkir reservoir area [5] and the Caspian Sea region [7].

In the first quarter of the 21st century, technological development worldwide has been reflected in science, as in every field. The rapid advancement of digitalization is also observed in our country. In seismology, digitalization and the use of modern technologies have become widespread both globally and locally.

A recent development in global science is the application of Artificial Intelligence (AI) and Machine Learning (ML). AI has shown significant potential in seismic signal identification and has opened new research opportunities in recent years.

⁵ Cornell, C.A. Engineering seismic risk analysis // Bulletin of the Seismological Society of America, – 1968. 58 (5), – p. 1583-1606.

ML is applied in earthquake seismology for the development and analysis of seismic catalogs, focal mechanism analysis and ground motion prediction, among other tasks.

Although ML and AI techniques play an important role in methodological advancements, the complete exclusion of physical modeling from traditional approaches may lead to undesirable outcomes. To avoid such issues, hybrid approaches integrating ML methods with classical seismological techniques are proposed.

CHAPTER III. DIGITAL MODELING OF EARTHQUAKE EVENTS AND ANALYSIS OF DISTRIBUTION DYNAMICS USING MODERN TECHNOLOGIES

This chapter provides a detailed analysis of the seismicity of the study area from the past century to the present based on earthquake catalogs. The catalogs of the International Seismological Centre and the Republican Seismic Survey Center are analyzed. The author then presents the methods for constructing and processing the database for the most recent period.

F.A.Kadirov, G.J.Yetirmishli, A.H.Kadirov⁶ compiled a maximum magnitude map of Azerbaijan based on the catalog for 1902-2010. The author adapted the map to the study area applying georeferencing (Figure 4).

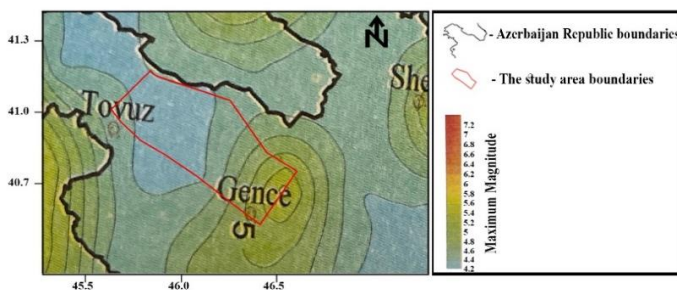


Figure 4. Map of probable maximum magnitude calculated based on the 1902-2010 catalog (Compiled by: F.A.Kadirov, G.J.Yetirmishli, A.H.Kadirov⁶, processed and adapted to the study area: T.H.Babayev)

⁶ Kadirov, F.A., Yetirmishli, G.J., Kadirov, A.H. Assessment of the maximum magnitude and intensity of earthquakes in the territory of Azerbaijan // – Baku: Proceedings of the Institute of Geology, – 2008. №36, – p. 81-87 (in Russian).

The study area is characterized by maximum magnitudes ranging from 4.2 to 5.8 (Figure 4). Seismicity in the area is associated with both active tectonic movements and the major Shamkir and Mingachevir reservoirs [2]. Based on catalog analysis around the Shamkir reservoir, the study area was divided into three groups according to seismicity characteristics and classified using the b-value of the Gutenberg-Richter (G-R) law. The northern part of the area is characterized by seismicity associated with the Shamkir and Mingachevir reservoirs [2, 6]. A decrease in b-value from north to south across the depression-mountain transition was observed, indicating the presence of compressional-tensional deformation [2].

To characterize seismicity in the area, a catalog of 132 earthquakes with local magnitudes between 1.3 and 4.8, recorded from 24/03/1974 to 29/12/2022, was considered.

Although anthropogenic factors have recently been identified as a contributing cause of seismicity in the area, the occurrence of natural seismicity associated with active faults crossing the study area remains evident. To clarify this, a depth distribution map of earthquake hypocenters along generalized tectonic faults is presented in Figure 5.

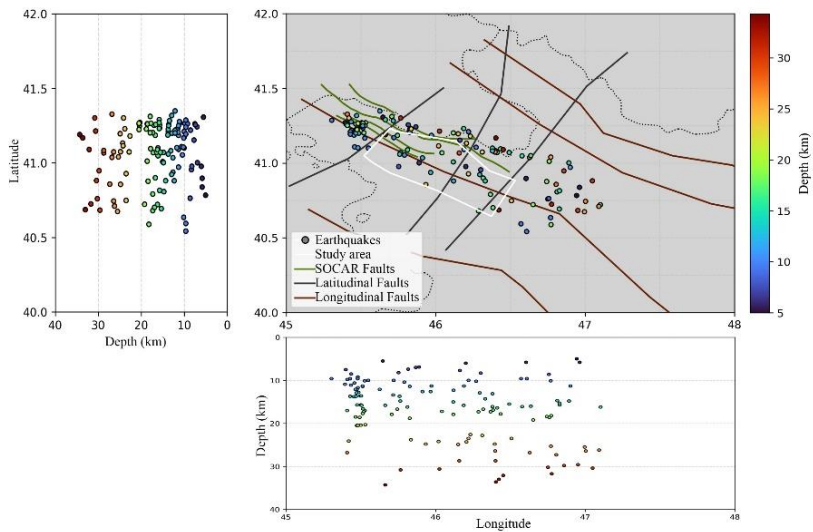


Figure 5. Epicentral distribution of earthquakes along faults and depth variations (Compiled by: T.H.Babayev)

The G-R law is applied to the catalog using a clustering approach [10]. The a- and b-values of the G–R law are calculated by dividing the catalog into sub-clusters (Figure 6). For this purpose, the magnitude of completeness, maximum magnitude and magnitude bin are determined [2]. These steps are repeated for each cluster (Graph 1). Finally, spatial distribution maps of the a- and b-values are compiled for the study area (Figures 7a, b).

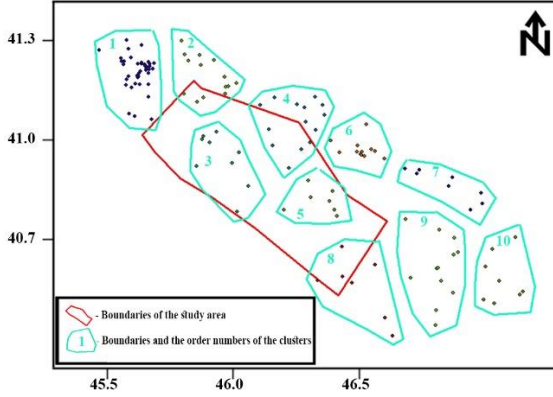
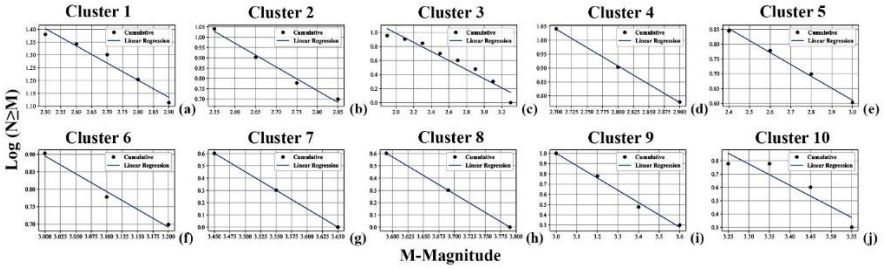


**Figure 6. Epicentral distribution of the catalog by clusters
(Compiled by: T.H.Babayev)**



**Graph 1. Illustration of the G-R law for clusters and calculation of
b-value by linear regression (Compiled by: T.H.Babayev)**

Maximum magnitude uncertainty is calculated for each cluster [10]. It is defined as the difference between the observed maximum magnitude and the expected maximum magnitude in each cluster. The expected maximum magnitude is expressed as the ratio of the a-value to the b-value.

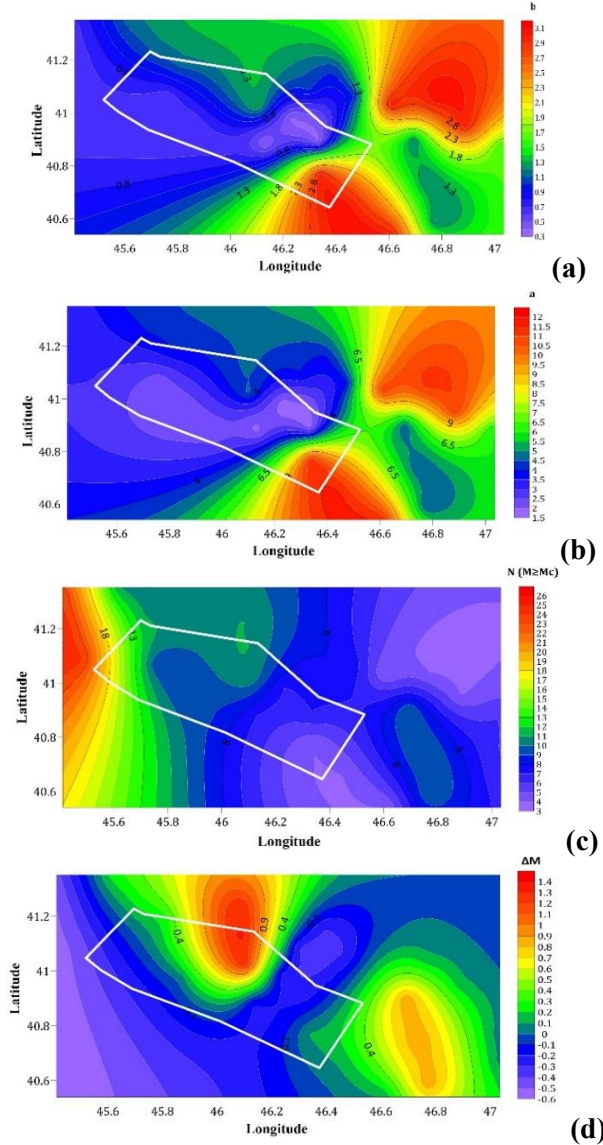


Figure 7. Distribution maps of seismicity indicators calculated for the catalog over the study area. (a) b-value. (b) a-value. (c) N: number of earthquakes with local magnitude greater than or equal to the magnitude of completeness. (d) Maximum magnitude uncertainty (Compiled by: T.H.Babayev)

Thus, the closeness of the difference between the expected and observed maximum magnitudes to zero depends on the accurate estimation of the a- and b-values. A negative difference indicates that the expected maximum magnitude is greater than the observed maximum magnitude, while a positive difference indicates the opposite [10].

The spatial distribution maps of the calculated seismicity indicators are shown in Figure 7.

Figure 7a shows that the study area is mainly characterized by low b-values. The a-value exhibit a similar spatial trend to the b-value (Figures 7a, b). This indicates that small-magnitude earthquakes occur relatively frequently, whereas large-magnitude earthquakes are rare. The decrease in the number of earthquakes satisfying the $M \geq M_c$ condition with increasing a- and b-values indicates a lower occurrence of large-magnitude events (Figure 7c). The increase in b-value from southwest to northeast can be interpreted as a result of compression associated with active tectonic movements in the indicated direction. Maximum magnitude uncertainty in the study area is generally close to zero (Figure 7d), indicating satisfactory accuracy of the calculated a- and b-values.

Thus, this chapter provides an overview of the seismicity of the study area, demonstrates the methodologies for obtaining, processing and analyzing the seismicity catalog and analyzes the seismicity characteristics of tectonic faults.

CHAPTER IV. THEORETICAL BASIS FOR THE APPLICATION OF GROUND MOTION MODELS IN DETERMINISTIC SEISMIC HAZARD ASSESSMENT AND APPLICATION OF MACHINE LEARNING ALGORITHMS

This chapter contains the theoretical basis and methodological sequence for the application of the deterministic approach, testing of ground motion models against a database and their development and application using machine learning algorithms.

The DSHA method evaluates the possible “worst-case scenario” considering the geological and tectonic characteristics of the region. By the “worst-case scenario,” a realistic, justified and geologically and tectonically expected scenario earthquake with the greatest impact in the study area is meant.

Acceleration is a parameter that describes the change in velocity of a body with respect to time, regardless of direction. PGA is defined as the maximum absolute value of ground motion acceleration recorded in a given area during the propagation of earthquake waves. PGA values are expressed in units of g (gravitational acceleration of the Earth) or cm/s^2 . 1 gal equals 1 cm/s^2 and 1 g equals 981 gals.

Using real data, a ground motion model (GMM) or ground motion prediction equation (GMPE) can be developed for any region. The larger the amount of real data, the higher the accuracy of the developed models or equations. However, the limited availability of recorded data reduces the reliability of prediction accuracy.

For instance, the applicability of a GMPE developed for the Caucasus region is tested for the study area. First, a database consisting of real PGA values is compiled.

To construct the database, a catalog covering the period 2022–2024 is compiled using RSSC data.

The catalog consists of 18 earthquakes with local magnitudes ranging from 3.0 to 4.9 (Figure 8). A database consisting of 77 records is compiled based on station recordings of these earthquakes.

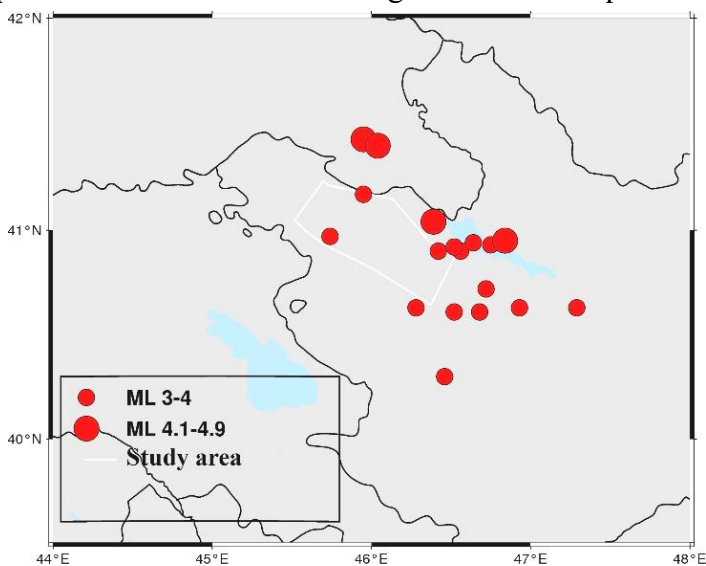


**Figure 8. Magnitude distribution of earthquakes
(Compiled by: T.H.Babayev)**

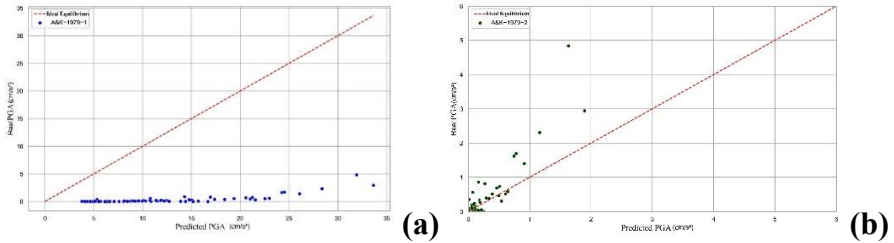
Next, the GMPE is tested against the database. GMPE (eq-s 1 & 2) presented by F.F.Aptikayev and Y.F.Kopnichyev⁷ is widely used in seismic hazard assessment studies conducted in various regions of Azerbaijan. The equation is conventionally denotes as A&K–1979.

$$\log PGA_{A\&K-1979-1} = 0.28M - 0.8 \log R + 1.70, \quad (1)$$

$$\log PGA_{A\&K-1979-2} = 0.8M - 2.3 \log R + 0.80, \quad (2)$$

where, M is the magnitude of the scenario earthquake, R is the distance from its hypocenter to the study area. $\log PGA_{A\&K-1979-1}$ is the logarithm to the base ten of the predicted PGA by the equation for cases greater than or equal to 160 cm/s^2 and $\log PGA_{A\&K-1979-2}$ is the logarithm to the base ten for cases less than or equal to 160 cm/s^2 .

The real PGA values are compared with those predicted by A&K–1979–1 and A&K–1979–2 (Graph 2).

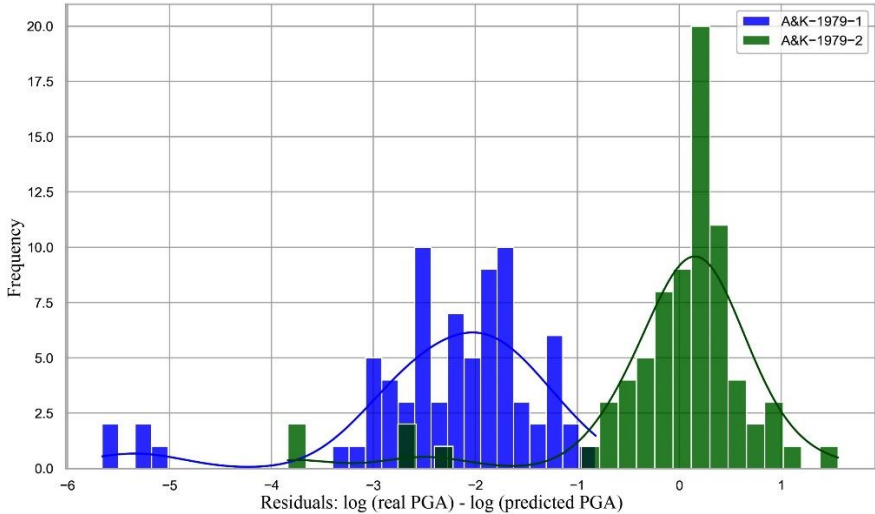


Graph 2. Comparison of PGA predicted by the A&K–1979 equations with the acceleration obtained from seismic records of earthquakes that occurred in and near the study area. (a) A&K–1979–1. (b) A&K–1979–2 (Compiled by: T.H.Babayev)

The red dashed line shown in Graph 2a and Graph 2b represents the “ideal equilibrium” line, corresponding to the condition where the PGA values obtained from seismic records are equal to those predicted by the equation. Thus, for a model or equation to be considered reliable or ideal, the data points are expected to lie on or close to the “ideal

⁷ Aptikayev, F.F., Kopnichev, Y.F. Consideration of the earthquake source mechanism in the prediction of strong ground motion parameters // – Doklady of the USSR Academy of Sciences, – 1979. Vol. 247, № 4, – p. 822-825 (in Russian).

equilibrium” line. The deviation of the points (PGA values) from this line indicates a mismatch between predictions and observed data. The residuals shown in Graph 3 raise doubts regarding the applicability of the equations to the study area.



Graph 3. Residuals. Blue for equation A&K-1979-1, green for equation A&K-1979-2 (Compiled by: T.H.Babayev)

Residuals are defined as the logarithmic difference between the observed and predicted PGA values. Positive residuals indicate that the predicted values are lower than the observed values, whereas negative residuals indicate that the predictions are overestimated. Thus, Graphs 2 and 3 show that the A&K-1979-1 equation overestimates the real values, while the A&K-1979-2 equation underestimates them [9, 13]. For this reason, a methodology for constructing a local GMM using ML algorithms is presented and applied to the Shamkir reservoir area.

In the GMMs constructed in the dissertation using four algorithms (ANN, XGBoost, RF and SVM), the input features include local magnitude (M_L), V_{S30} (m/s) and hypocentral distance from the station (km), while PGA (gal) is used as the target variable. When the input features are provided, the models predict the target variable, namely PGA (gal).

Initially, a GMM is constructed for the study area using the ANN algorithm [11]. ANN is a computational model that mimics the information processing mechanism of the human brain. It consists of interconnected layers of neurons, where each neuron processes input data and transmits it to the next layer. The ANN model is designed with a multi-layer perceptron structure, including an input layer, two hidden layers with 64 and 32 neurons, respectively and an output layer.

Next, modeling is performed using the XGBoost algorithm. XGBoost constructs an additive model in a sequential manner and minimizes the residual errors of previous iterations. The XGBoost model is configured with a learning rate of 0.05, 500 estimators (trees) and a maximum tree depth of 4.

Then, a GMM is constructed using the RF regression algorithm. The RF algorithm operates by constructing decision trees in parallel. The model consists of 500 decision trees.

The final model is developed using the SVM algorithm. The Support Vector Regressor function is applied. The model hyperparameters are set to $C = 100$ and $\varepsilon = 0.1$.

A comparative analysis of the constructed GMMs is performed based on statistical results for the entire dataset (Table 1).

Table 1

**General statistical indicators used for evaluating
the developed models (Compiled by: T.H.Babayev)**

	R^2	MAE	RMSE	MSE	Bias	Deviation (sigma)
ANN	0.3270	0.5761	0.9146	0.8365	0.0892	0.9103
XGBoost	0.7816	0.3075	0.5210	0.2714	-0.0027	0.5210
RF	0.8050	0.2858	0.4923	0.2423	-0.0433	0.4904
SVM	0.6255	0.3797	0.6823	0.4655	0.1187	0.6719

R^2 – Coefficient of Determination, MAE – Mean Absolute Error, RMSE – Root Mean Square Error, MSE – Mean Squared Error.

Ensemble-based algorithms (XGBoost and RF) provide the highest accuracy and the lowest error. Although SVM does not show as strong a training fit as the ensemble models, it achieves the highest R^2 value (0.5005) among the four models on the test set. ANN shows the weakest performance across all evaluation metrics, which is attributed to its high sensitivity to the limited number of input features.

According to the statistical indicators, the RF model is considered to provide the best predictive performance for the study (Table 1).

The database used for training the developed GMMs is limited in size in terms of the number of observations [11]. Thus, 80% of the data was used for model training, while the remaining 20% was used to test predictive performance. However, the limited size and insufficient representativeness of both the training and testing datasets make the applicability of the developed models uncertain. The more diverse data an ML algorithm is exposed to, the greater the variability it can learn, thereby improving its robustness and predictive capability.

As a solution, it is proposed to construct a synthetic database and develop GMMs using ML algorithms based on this synthetic dataset [9].

CHAPTER V. DETERMINISTIC SEISMIC HAZARD ASSESSMENT IN THE STUDY AREA. COMPILATION OF A SET OF SYNTHETIC ACCELEROGRAMS AND ITS INTEGRATION INTO MACHINE LEARNING ALGORITHMS

This chapter presents the application of the DSHA method to the study area. First, a scenario earthquake is selected, followed by the calculation of the distance to the scenario source, segmentation of the study area, determination of site characteristics for each segment and prediction based on the scenario earthquake. A synthetic accelerogram database is established using maximum magnitudes determined in the study area (Figure 4) and ML-based GMMs are developed [12]. The most efficient GMM is used to predict PGA for seismic hazard assessment and a seismic hazard map is compiled.

Due to its high seismic activity, the approximately 170 km section of the Kur fault within and near the study area is considered as the scenario earthquake source. To assess “the worst-case scenario”, the magnitude of the scenario earthquake is set to 6.0. For this scenario, conditional epicenters are defined at 10 km intervals along the considered 170 km section of the Kur fault (Figure 9).

In the next step, nodal points of square segments with a width and length of 10 km are defined and a total of 47 nodal points are established across the study area (Figure 9).

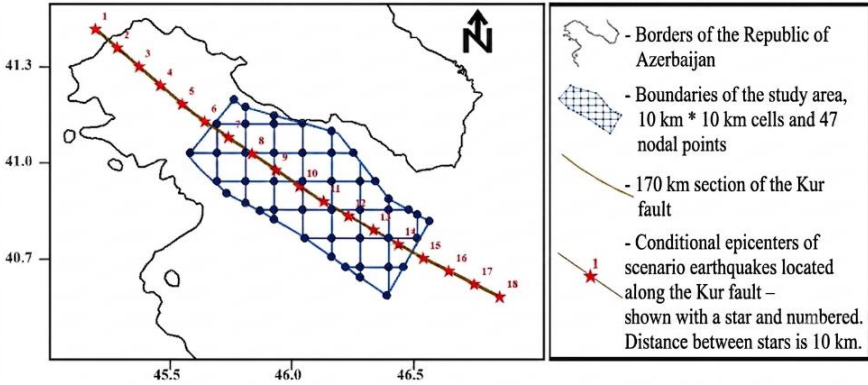


Figure 9. Representation of the conditional epicenters determined for the scenario earthquake and segmentation of the study area
(Compiled by: T.H.Babayev)

Site properties are characterized by determining V_{S30} values and site classes at 47 nodes. The site classification map of the study area is shown in Figure 10.

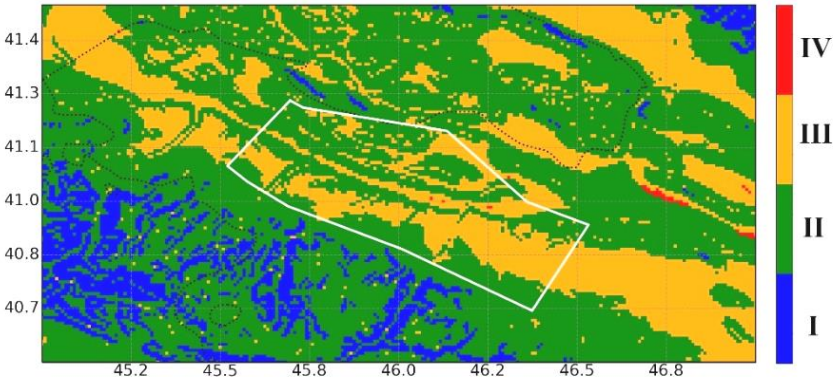


Figure 10. Site classification map of the study area. Site classes I, II, III and IV are shown in the map with corresponding colors
(Compiled by: T.H.Babayev)

The study area is characterized by site classes II and III (with amplification factors of 1.0 and 1.2, respectively) (Figure 10). Due to the presence of the Kur River and its tributaries, the geological composition of the area is dominated by river terraces and alluvial, alluvial–diluvial and alluvial–proluvial deposits composed of clays, sands and gravels. The presence of a moist geological environment results in low shear-wave propagation velocity (Figure 2), indicating a strong resonance effect during ground motion.

In cases where real data do not satisfy quantitative and qualitative requirements, the generation of synthetic data is a widely used approach. In this study, synthetic accelerograms are generated using the Seismoartif⁸ software. The generation method proposed by B.Halldorsson and A.S.Papageorgiou⁹ is applied. Using this method, a synthetic database consisting of 3013 artificial accelerograms is created [12].

Thus, in the next step, the constructed synthetic database is used to train more complex forms of ANN, XGBoost, RF and SVM algorithms and new GMMs are developed (Table 2).

Table 2
Models showing the best results based on statistical indicators
evaluated through integration of the synthetic database
(Compiled by: T.H.Babayev)

	R^2	MAE (gal)	RMSE (gal)	MSE (gal ²)	Bias	Deviation (sigma, gal)
Training	RF	RF	RF	RF	ANN	RF
Testing	ANN	SVM	ANN	ANN	ANN	XGBoost
Model	RF	RF	RF	RF	ANN	RF

R^2 – Coefficient of Determination, MAE – Mean Absolute Error, RMSE – Root Mean Square Error, MSE – Mean Squared Error.

⁸ SeismoArtif v2024 Release-1 Build-2 (64-bit): [Electronic resource] / Seismosoft ltd. – Pavia, Italy, 2024. URL: <https://seismosoft.com/products/seismoartif/>.

⁹ Halldorsson, B., Papageorgiou, A.S. Calibration of the specific barrier model to earthquake of different tectonic regions // Bulletin of the Seismological Society of America, – 2005. 95 (4), – p. 1276-1300.

Based on the comparison of the compiled GMMs, the ANN and RF models demonstrate the most advanced results among the four models (Table 2). Thus, combining predictions from both models in the application of the DSHA method in the Shamkir reservoir area is considered a more robust and comprehensive approach. The magnitude of the scenario earthquake ($M = 6$), calculated hypocentral distances, V_{S30} and site class data for each node are used as input features in both models (ANN and RF). Based on the PGA predictions from both models, a general PGA distribution map is compiled for the study area (Figure 11).

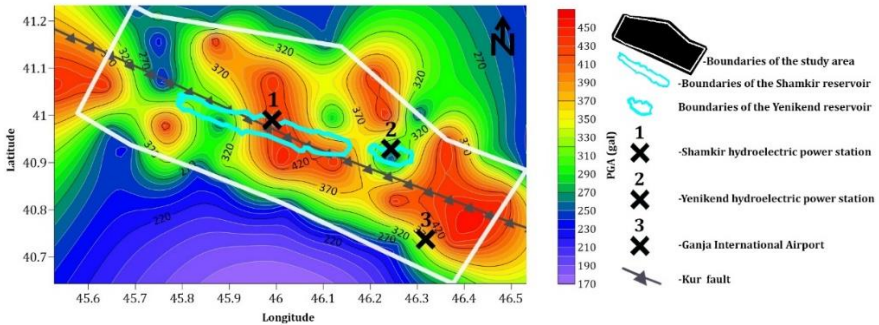


Figure 11. Predicted PGA distribution map for “the worst-case scenario” ($M = 6$) defined for the study area (Compiled by: T.H.Babayev)

Strategic infrastructure facilities in the study area are located within a heterogeneous seismic environment. The Shamkir and Yenikend reservoirs and HPSs, as well as Ganja IA, are directly influenced by the Kur active fault zone and exhibit hazard levels in the range of 260–450 gals (Figure 11). Areas are observed where seismic wave energy shows a nonlinear increase due to both the source mechanism and the resonance properties of local lithological layers.

The Shamkir reservoir, which is the main object of the study, as well as the Shamkir HPS, exhibit the highest predicted PGA value (450 gal) for the entire area.

Based on the predicted PGA values, a seismic hazard map in intensity units is compiled for the study area (Figure 12).

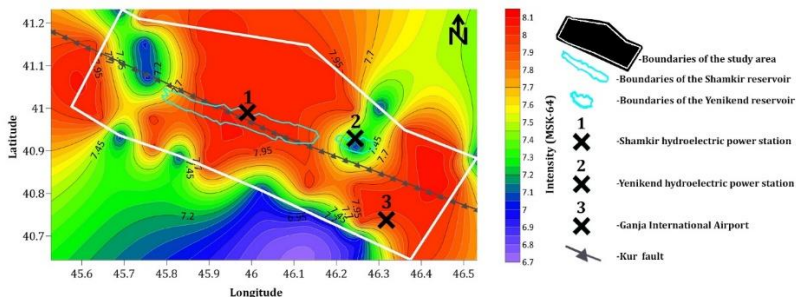


Figure 12. Seismic hazard map in intensity units, prepared based on the predicted PGA values for “the worst-case scenario” ($M = 6$) defined for the study area (Compiled by: T.H.Babayev)

The study area is characterized by intensity levels of VII–VIII (Figure 12). The seismic impact of the active Kur fault extending along the study area is clearly expressed in intensity terms. Thus, the map indicates intensity VIII in the vicinity of the fault and intensity VII with increasing distance from it. This also reflects the law of ground motion attenuation, whereby seismic vibrations decrease with increasing distance from the source. The Shamkir and Yenikend reservoirs and HPSs, as well as Ganja IA, are characterized by intensity VIII. The area is therefore subject to strong seismic impact. The seismic hazard potential in the area includes small-scale landslides, various types of damage to buildings, partial collapse of old structures and the formation of large ground cracks. In addition, due to the widespread presence of low-density loose sands typical of fluvial environments, moist and water-saturated medium-density fine-grained and silty sands, moist semi-solid clays, loams and relatively porous loams, the formation of large and hazardous ground fissures is characteristic and the resonance effect during earthquakes is significant.

The author’s characterization of the study area in several studies based on catalog and tectonic fault analysis [2], as well as the application of other scenario variations of the DSHA method [5], again yielding intensity values of VII and VIII, supports the present results and the newly applied approach. Thus, the new and modern methodology, tools and data used in the dissertation demonstrate that they produce reliable results and are self-consistent. In addition, within the

framework of the relativity approach, the distance of the selected scenario earthquake from the study area is also important. Thus, scenario earthquakes defined at excessively small distances and with large magnitudes may lead to overestimated PGA and intensity values [8]. However, the results of the current study show that the selected scenario earthquake is consistent with realistic conditions.

Therefore, the dissertation presents the results of seismic hazard assessment for the Shamkir reservoir area. The approaches used in the study, based on international practice and the proposed methodology are accepted, presented and applied by the author as new and modern methods considered necessary for application, providing comprehensive results and complying with accepted standards. Thus, the results obtained in the dissertation are of scientific and practical importance for seismic hazard and risk assessments in the study area, which includes strategically important facilities.

CONCLUSION

The results of the research conducted on the dissertation are presented as follows:

1. The site effect in the area was assessed and it was determined that the average shear wave velocity at a depth of 30 m (V_{S30}) varies in the range of 200–600 m/s. The area is characterized mainly by low V_{S30} indicators. The lowest velocities are observed along the Kur fault in the Shamkir reservoir basin. The coverage of the area with fluvial sediments and the presence of an active tectonic fault demonstrate a high seismic hazard potential. In the low-velocity zones (LVZ), loose alluvial sediments and the abundance of groundwater enhance the resonance effect, reducing the velocity of shear waves and increasing their amplitude [12];
2. A database of 18 earthquakes ($M_L \geq 3$) that occurred in the study area in 2022-2024, consisting of a total of 77 station records, was compiled. The database includes the date, location, magnitude, depth, PGA (gal) of the earthquakes, stations' ID, V_{S30} value, epicentral and hypocentral distances. The earthquakes range in local magnitude from 3.0 to 4.9, hypocentral distance from 32 to 284 km

and PGA reaches 5 gal [9, 11];

3. 3013 artificial accelerograms were generated for site classes II and III with a magnitude of 4.2–5.8, a hypocentral distance of 1–140 km and a PGA of 165–733 gals using the SeismoArtif software. The accelerograms were validated with a maximum mean error of 10% and a coefficient of variation of 15% (with only one exception, 17%) [12];
4. Four ML algorithms, ANN, XGBoost, RF and SVM, were applied to develop the GMMs. The coefficients of determination of the models built with real data were 0.3270, 0.7816, 0.8050 and 0.6255, respectively (approximately 33–81% explanatory and generalizing ability). The models built with synthetic data showed higher generalization and prediction accuracy with coefficients of determination of 0.8436, 0.8472, 0.8506 and 0.8387, respectively [12];
5. The final results are presented as a ground motion attenuation graph, PGA spatial distribution map and seismic hazard map, obtained from hybrid predictions of ANN- and RF-based GMMs developed by training synthetic data. The attenuation graph demonstrates the dependence of PGA values in the range of 230–450 gals on the hypocentral distance. The obtained graph reflects that the model takes into account not only the geometric propagation factor, but also complex geological conditions;
6. According to the spatial-distribution map of PGA, the seismic impact of the Shamkir and Yenikend reservoirs and HPSs, Ganja IA is predicted to be between 260–450 gals, being directly within the influence zone of the Kur active fault zone. Shamkir and Yenikend reservoirs are characterized by 270–450 gals and 260–330 gals, respectively and Shamkir and Yenikend HPSs are characterized by 450 gal and 270 gal, respectively. Ganja IA is characterized by 360 gal. The highest PGA value (450 gal) in the study area is predicted at Shamkir HPS;
7. According to the presented seismic hazard map, the study area is generally characterized by an intensity of VII–VIII (MSK-64). The seismic hazard map reflects an intensity of VIII around the fault and an intensity of VII getting away from it. Shamkir and Yenikend reservoirs and HPSs, Ganja IA are characterized by an intensity of

VIII. Thus, dangerous seismic effects such as damage, partial collapse, small-scale landslides and the formation of large cracks on the ground are predicted in a number of parts of these infrastructures. The results are confirmed by comparison with a number of other studies [1, 2, 5].

The main content of the dissertation are reflected in the following scientific works:

1. Babayev, T.H., Babayev, G.R. Tectonic, Geological, Geomorphological and Seismic Analysis of Shamkir Region and Surroundings Based on Bibliographical Review // – Baku: PAHTEI-Proceedings of Azerbaijan High Technical Educational Institutions, – 2023. Vol 33 (04), Issue 10, – p. 73-81 (in Azerbaijani).
2. Babayev, T.H. Seismic analysis of the Shamkir reservoir area through insights into the dynamics of the earthquake characteristics / T.H.Babayev, Y.N.Aliyev, I.B.Muradi [et al.] ANAS Transactions, Earth Sciences, – Baku: – 2023. № 2, – p. 51-63.
3. Babayev, G.R., Babayev, T.H., Aliyev, Y.N., Muradi, I.B. Numerical earthquake ground motion modeling for southern slope of Greater Caucasus // VII International Conference "Seismology and Engineering Seismology" dedicated to the 100th anniversary of the birth of the Nationwide Leader H.Aliyev, – Baku: Materials of Conference, – 7-9 June, – 2023, – p. 264-268.
4. Babayev, G., Babayev, T., Telesca, L. Deterministic ground motion modeling with target earthquakes and site effects in eastern Azerbaijan // Arabian Journal of Geosciences, – 2024. 17 (61), – p. 1-12.
5. Babayev, T.H. Recent scenario-based ground motions and intensity modeling of Shamkir water reservoir area // PAHTEI-Proceedings of Azerbaijan Higher Technical Educational Institutions, – 2024. 26 (3-4), – p. 6-14.
6. Babayev, T., Babayev, G. Spatial distribution of b-value on the Mingachevir water reservoir area for seismicity analysis applying the microzonation method of Babayev G. et. al., 2020 // EGU General Assembly 2024, – Vienna, – 14-19 April, – 2024, – 1 p.
7. Babayev, T., Muradi, I., Babayev, G. Offshore ground motion estimation for the Caspian Sea based on the encompassing scenario

- epicenters // Caspian Basin Conference, – Baku: Book of Abstracts, – 3-6 June, – 2024, – p. 50-51.
8. Babayev, T., Babayev, G. Far and remote scenario earthquakes in seismic hazard assessment // 15th General Assembly of the Asian Seismological Commission, – Belek, Antalya: Book of Abstracts, – November 3-7, – 2024, – p. 43-49.
 9. Babayev, T. Development of ann-based data-driven ground motion model for Azerbaijan using temporal earthquake records of 2022–2024 / T.Babayev, G.Babayev, S.Irawan [et al.] // *Frontiers in Earth Science*, – 2025. 13, – p. 1-18.
 10. Babayev, T.H. Aliyev, Y.N., Babayev, G.R. Clustering analysis of 100-year international seismic catalogue for Azerbaijan // *Geology and Geophysics of Russian South*, – 2025. 15 (1), – p. 105-118.
 11. Babayev, T.H. Development of a regional ground model for the Shamkir reservoir area using artificial intelligence algorithms // Republican scientific-practical conference of young scientists, specialists and students dedicated to the “Year of Constitution and Sovereignty” and the 80th anniversary of ANAS, titled “Interdisciplinary Research in Earth Sciences: Achievements and Perspectives”, – Baku, – 18-19 December, – 2025, – p. 11-13 (in Azerbaijani).
 12. Babayev, T. Machine learning-based ground motion simulation for seismic hazard assessment of critical water infrastructure in Azerbaijan (Case study: Major Shamkir water reservoir) / Tural Babayev, Gulam Babayev, Emil Bayramov [et al.] // *PLOS One*, – 2026. 21 (4), – p. 1-26.
 13. Babayev, T., Babayev, G., Bayramov, E. ANN-Based Ground Motion Model for Azerbaijan Using 2022–2024 Earthquake Records // 16th General Assembly of Asian Seismological Commission, – Tashkent, Uzbekistan, – April 25-30, – 2026, – p. 287.

The defense will be held on 26 October 2026 at 14³⁰ at the meeting of the Dissertation council FD 1.01 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Institute of Geology of the Ministry of Science and Education of the Republic of Azerbaijan.

Address: Az1073, Azerbaijan, Baku, H.Javid ave., 119.

Dissertation is accessible at the library of the Institute of Geology of the Ministry of Science and Education Republic of Azerbaijan.

Electronic version of the abstract is available on the official website of the Institute of Geology of the Ministry of Science and Education Republic of Azerbaijan.

Abstract was sent to the required addresses on 25 September 2026.

Signed for print: 23.09.2026

Paper format: 60x84^{1/16}

Volume: 37434

Number of hard copies: 20