Development of Avalanche Prediction Algorithms Based on a Set of Parameters

Natalya Denissova [1], Olga Petrova [2], Evgeny Fedkin [1], Gulzhan Daumova [2,*], and Evgeny Sergazinov [1]

[1] Department of Information Technology, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan

[2] School of Geosciences, D. Serikbayev East Kazakhstan Technical University, Ust-Kamenogorsk, Kazakhstan

Email: NDenisova@edu.ektu.kz (N.D.); OPetrova@edu.ektu.kz (O.P.); evgeny.fedkin@edu.ektu.kz (E.F.); GDaumova@edu.ektu.kz (G.D.); serg@edu.ektu.kz (E.S.)

*Corresponding author

***Abstract*—This article presents the results of a study conducted to develop an avalanche predictive model based on a set of climatic data. The research area includes the territory of East Kazakhstan, where a sharply continental climate prevails with hot summers and cold and snowy winters. With climate change, despite the low altitudes in this mountainous area, the problem of avalanche safety is acute in the region. To compile a avalanche predictive model, meteorological data from regional weather stations for 23 years (2001–2024) and meteorological observations in avalanche-prone areas for 19 years (2005–2024) were analyzed. This information was compared with the recorded data on spontaneous avalanches over the past 11 years (2013–2024). A database was created to carry out the research. The meteorological data is analyzed using mathematical statistics methods with the construction of probable trends of regional climatic changes. MATLAB data analysis has shown a significant relationship between sudden warming, increased wind speed, and precipitation that precedes avalanches. The analysis showed the need to take these parameters into account when developing a forecast model, as the likelihood of dangerous weather events will increase every year. The avalanche prediction was performed using regression analysis (logistic regression). The Loginom Community statistical software package is used for this purpose. The quality of the constructed predictive model was assessed. In the future, it will be used to predict spontaneous avalanche based on observations of meteorological data in avalanche-prone areas of the East Kazakhstan region.**

***Keywords*—databases, logical database schema, regression analysis of data, logistic regression, probability of event, model training, climate change, avalanches, monitoring system**

I. Introduction

Global climate change has various regional manifestations. At the same time, meteorological factors have significant variability in their spatial and temporal distribution. Their influence is reflected in the frequency and severity of various dangerous natural disasters. This includes an increase in avalanche activity in the regions [1–3]. Therefore, the possibility of forecasting them is an important task of modern science. The construction of predictive avalanche hazard models should consider many factors to improve the quality of the forecast.

The literature review shows that to identify patterns and predict avalanche hazards, special attention is paid to statistical methods, among which regression analysis occupies an important place. But at the same time, very little attention is paid in general to forecasting natural hazards using the logistic regression method. Therefore, this study will allow us to introduce new information based on the study of real data for the development of this area.

II. Literature Review

The influence of climate change on the formation and characteristics of avalanches is manifested in an increase in the number of heavy snowfalls, an increase in air temperature and wind speed. These changes lead to an increased risk of avalanches [4, 5]. Climate change affects the area and timing of snowfall [6, 7]. This is reflected in noticeable regional differences [8–10].

Global warming is changing the nature of snowfall in many parts of the world. Periods of snowfall and snowmelt are also changing along with rising global temperatures. There is a correlation between snowfall and air temperature, which is associated with an increased moisture content in the atmosphere. This has significantly affected the frequency, timing, scale, and types of avalanches in various mountainous regions of the planet [11–14]. The above-mentioned scientific studies note that changes in snow characteristics depend on meteorological factors, atmospheric circulation, and soil conditions, which, in turn, affect avalanche activity. Many years of research experience in predicting and identifying the causes of avalanches show that the avalanche formation process is influenced not only by the above factors, but also by their complex combination.

Among the various approaches to predicting avalanche hazards, special attention is paid to statistical methods that allow us to identify patterns between meteorological conditions and the probability of avalanches.

Manuscript received September 6, 2024; revised October 28, 2024; accepted March 11, 2025; published June 26, 2025.

Statistical methods for predicting avalanche hazards can be divided into several categories:

1) Qualitative methods: based on empirical observations and expert assessments. Forecasters analyze current meteorological conditions, comparing them with historical avalanche data, and based on this, draw conclusions about the possible danger.
2) Quantitative (analytical) methods: mathematical models are used to quantify avalanche hazards. These models consider various meteorological parameters, such as precipitation intensity, air temperature, wind speed, and others.
3) Climatic and meteorological methods: based on long-term climatic data and statistical relationships between meteorological conditions and the frequency of avalanches in a particular region.

In international practice, various statistical methods are used to assess avalanche hazards, among which logistic regression occupies a significant place.

Logistic regression is widely used to analyze and predict natural risks. For example, Sujatha and Sridhar [15], Chowdhury *et al.* [16] have demonstrated its use for mapping the likelihood of landslides based on data on terrain, soil properties, and climatic conditions. The results obtained confirmed the high accuracy of the model and became the basis for developing recommendations to reduce the risks of destruction.

Logistic regression is one of the most effective tools for predicting the probability of avalanches. It allows you to model the relationship between a binary variable (the presence or absence of an avalanche) and a set of independent meteorological variables.

In the of Gauthier *et al.* [17], weather conditions were analyzed using logistic regression to calculate the probability of avalanches. The developed linear regression model made it possible to classify situations according to five levels of avalanche danger. The obtained models demonstrated high efficiency in predicting days with increased avalanche activity.

Nosrati *et al.* [18] analyzed parameters such as height, slope, vegetation, and climatic conditions that can affect avalanches. Since avalanche-prone zones occupy a much smaller area compared to the general territory, logistic regression of rare events was used to identify the main factors of avalanche formation. The analysis showed that the curvature of the relief, height, rock outcrops, slope and exposure are key factors of snow accumulation.

Chourot and Martin [19] compared logistic regression and a new method based on thresholds for snowfall intensity and duration to analyze weather conditions conducive to avalanche formation. The results of the cross-validation showed that the linear regression model has greater accuracy and reliability because it considers more meteorological factors.

Thus, the use of logistic regression to predict avalanches based on meteorological data is a reasonable and effective approach. Logistic regression allows you to model probabilistic events and establish relationships between independent variables and binary outcomes. The use of such models helps to increase the accuracy of forecasts and, consequently, improve precautions and safety measures in avalanche-prone regions.

The key aspects of the research conducted on the development of a predictive model are shown in Fig. 1.

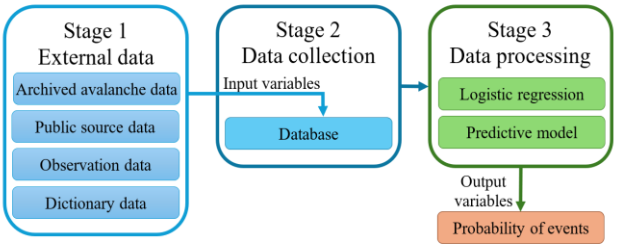


Fig. 1. Key aspects of the study.

III. Methods

At the first stage of building an avalanche predictive model, historical and archival data on meteorological parameters and data on snow cover in avalanche-prone areas of Eastern Kazakhstan were collected in a single database. Statistical data processing has shown that regional manifestations of the global climate change process are expressed in an increase in avalanche factors. There is a gradual increase not only in the average annual amount, but also in the amount of precipitation in winter. At the same time, the amount of precipitation falling in one day increases. A feature of the climate of Eastern Kazakhstan was the revealed redistribution of precipitation between summer and winter, towards an increase in precipitation in winter.

Data collection in the created database for East Kazakhstan was carried out from various sources: dictionary data; archival data; data from public sources; observation data. Further, the analysis of these data allows for operational monitoring of avalanche hazards and serves as a source for predicting avalanche hazards in the observed avalanche-prone areas.

The prediction of spontaneous avalanches was performed using regression analysis as a tool that allows us to determine the relationship between a set of input factors and a dependent variable. To compile a predictive model, it was necessary to determine the probability of an event occurring—an avalanche at an avalanche-prone area based on a combination of weather conditions.

To calculate the logistic regression of predicting spontaneous avalanches, as part of the study, we identified a set of input variables and used the Loginom Community statistical software package.

Based on the scenario and the developed database, logistic regression training was performed. The evaluation of the quality of the constructed model was performed by calculating the AUC ROC coefficient and the quality of data recognition.

The conducted research will help to create an algorithm for an automated avalanche hazard monitoring system for the East Kazakhstan region. The resulting regression analysis is the basis of this algorithm.

IV. Result and Discussion

In the ongoing research, the authors are developing an avalanche hazard monitoring system for Eastern Kazakhstan. The area under study is shown in Figs. 2 and 3.

The mountainous regions of East Kazakhstan have developed a tourist infrastructure and an extensive network of roads. Many people live and work in avalanche risk areas. Extreme weather conditions due to climate change are becoming more frequent: heavy precipitation, sudden warming in winter, and strong winds. This leads to spontaneous avalanches. In this regard, the development of an avalanche hazard monitoring system is very important for the region. It will be a software and hardware complex for analyzing the avalanche situation and transmitting appropriate warnings. A set of monitoring system sensors has already been developed. For its further work, it is necessary to develop software algorithms that will process various data (historical and current) meteorological parameters and information about avalanche-prone areas. The research conducted by the authors is devoted to the creation of an avalanche predictive model, which is the basis of this algorithm.

The authors studied data on 497 avalanche-prone areas in the region, of which 325 threaten objects and people's lives. Information about them is given in Table I. The source of this data is digitized information from observation logs on snow measuring routes in the East Kazakhstan region and emergency information from the State Institution Kazselezashchita on avalanches. This data is included in the developed database.

Fig. 2. Location of the research area on the world map.

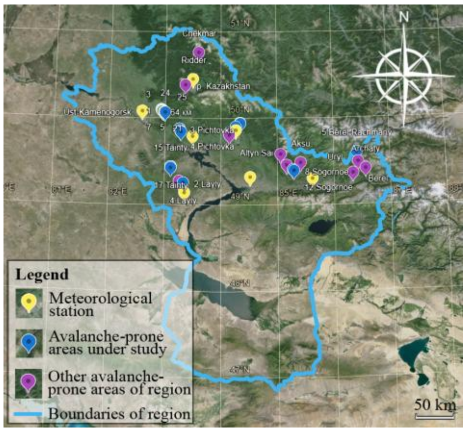


Fig. 3. Location of avalanche areas in East Kazakhstan.

The analysis of climatic data was carried out for 7 regional weather stations of the National Hydro-meteorological Service of Kazakhstan Kazhydromet. It is located near the avalanche-prone areas under study. Information about this data is shown in Table II.

TABLE I. Information about Avalanche-prone Areas

Data type	Data subtype	Max	Min	Average	Data acquisition period, year
Snow height (cm)	On snow measuring route	148	5	77	2005–2023
Avalanches	Number of avalanches	142	17	65	2013–2024
	Volume of snow in avalanches, m^3	593.680	67.675	220.500	2013–2024

TABLE II. Data from Meteorological Stations

Data type	Data subtype	Max	Min	Average	Data acquisition period, year
Air temperature (°C)	Average per day per year	4.4	1.3	3.2	2001–2023
	Average per day in winter	−5.2	−12.1	−7.8	2001–2023
The amount of precipitation, (mm)	Average per day per year	1.7	0.9	1.3	2001–2023
	Average per day in winter	1.7	0.7	1.1	2001–2023
	The amount for the year	621	319	485	2001–2023
	The amount for the winter	302	131	204	2001–2023
Relative humidity of the air (%)	Average per day per year	72	63	68	2001–2023
	Average per day in winter	76	70	73	2001–2023
Lack of air saturation	Average per day per year	6.05	3.75	4.71	2001–2023
	Average per day in winter	1.64	1.06	1.43	2001–2023
Snow height (cm)	Average per day per season	37	14	25	2001-2023
Wind speed (m/s)	Average per day per year	2.4	1.8	2.2	2001–2023
	Average per day in winter	2.4	1.6	2.1	2001–2023

The greatest attention in the study was paid to avalanche-prone areas, where the most vulnerable infrastructure facilities are located, and the largest number of spontaneous avalanches occurred. An example of such a risky situation is shown on Fig. 4. Here, in the Zubovsk village, at the foot of the mountain, there are residential buildings and a school in an avalanche-prone area. Preventive avalanche descent is not possible here due to the proximity of buildings. Therefore, the installation of an avalanche hazard monitoring system is most important for this site.

Spontaneous avalanches are possible from slopes with a steepness of more than 20°, especially if the slope does not have vegetation of shrubs and trees. The steepness of the slopes of East Kazakhstan is shown in Fig. 5.

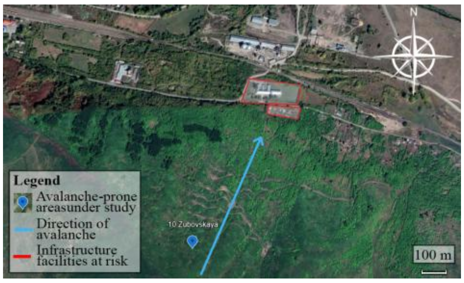


Fig. 4. Location of infrastructure facilities near an avalanche-prone area.

Information from all data sources has been collected into a single digital database, based on which an algorithm for analyzing these parameters will be developed specifically for the East Kazakhstan region. Meteorological data are not the only important components of it. During the study of avalanche-prone areas of the East Kazakhstan region, the following information objects were identified: area, avalanche-prone area, avalanche-trapping collection, meteorological data, morphological type, type of slope exposure, type of vegetation, avalanche-trapping vegetation, degree of avalanche danger, device, observation parameter, observation data, preventive descends, spontaneous avalanche.

A database based on the MySQL database management system was created to store the collected data. Google Earth Pro was used to form the cartographic basis of the project.

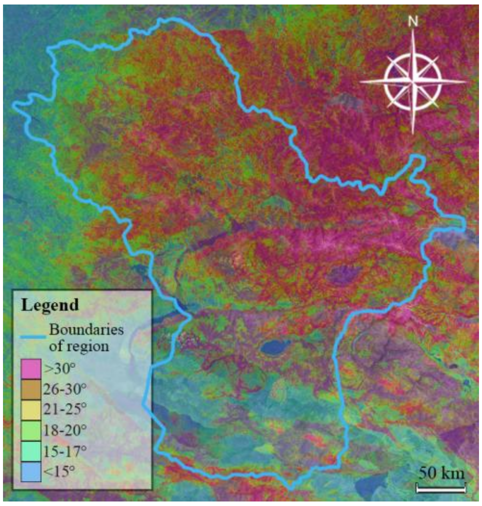


Fig. 5. Terrain steepness of the East Kazakhstan region.

To understand the interaction between previously identified information objects in the database, Fig. 6 shows the logical schema of the database. The information objects selected for the study are shown, as well as the relationships between them and their attributive composition.

Fig. 6 shows the tables in the database, their attribute composition, the primary keys of the tables and the relationships between the tables.

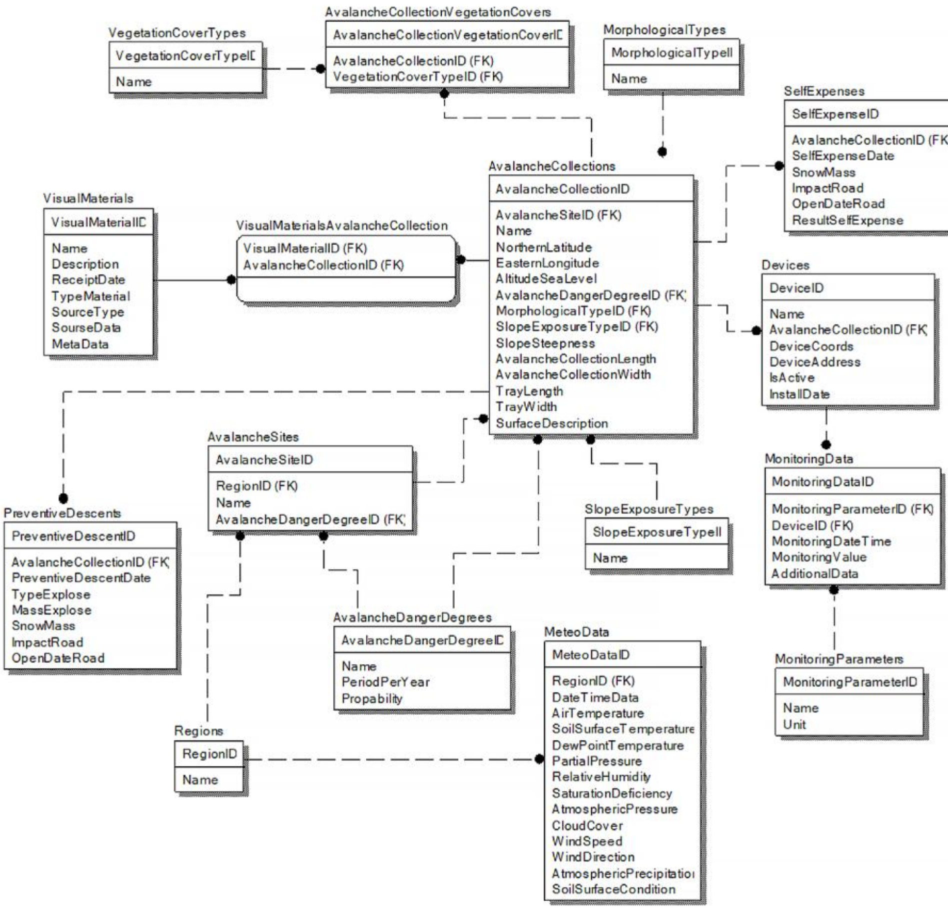


Fig. 6. The logical schema of the database.

The data in the database created is collected from various data sources. For the specified database, data sources can be roughly divided into the following categories: Dictionary data—reference data of various types, avalanche-prone areas records—information on avalanche sites observed in the monitoring and forecasting system (operational logs of preventive descents and spontaneous avalanches of the State Institution Kazselezashchita); data from public sources—data from public sources (databases of the National Hydrometeorological Service of Kazakhstan Kazhydromet, etc.) that are required for predicting avalanches; observation data—data from various metering devices (stationary devices, UAVs, etc.).The distribution of information objects of the database being developed by data sources is shown in Fig. 7 for clarity.

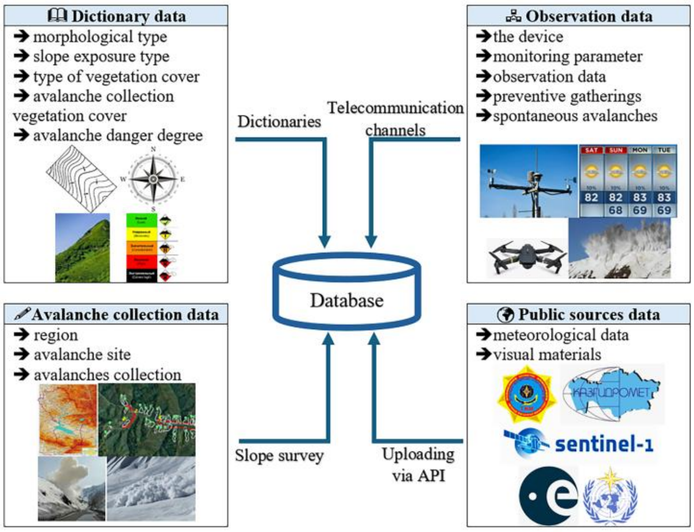


Fig. 7. Distribution of database information objects by data sources.

The unified database created within the framework of this study for collecting data on avalanche-prone areas allows for operational monitoring of avalanche hazard and also serves as a data source for predicting avalanche danger in the observed avalanche-prone areas.

In general, a software package has been designed and developed to collect and process data on snow cover in avalanche-prone areas, which includes the following components:

- A mobile application that is used to collect data on an avalanche-prone area (temperature, weather conditions, snow cover). This application was developed for the Android operating system, which allows you to install it on many mobile devices that run on this operating system.
- A MySQL database in which the information collected in the mobile application is entered.
- API interface for interaction between the database and the mobile application. Interaction with API data via REST-requests (Representational State Transfer). To protect the transfer of data and credentials, the specified API interface uses the SSL/TLS protocol (Secure Sockets Layer / Transport Layer Security).

The scheme of operation of the developed software package is shown in Fig. 8.

The climate data was processed using Excel software, and data trends were compiled using polynomial dependencies of the 2nd degree. The data analysis showed the changes that have occurred with the regional climate over a long period of time. Temperature trends are consistent with global changes and indicate a gradual increase in air temperature in the area (Fig. 9).

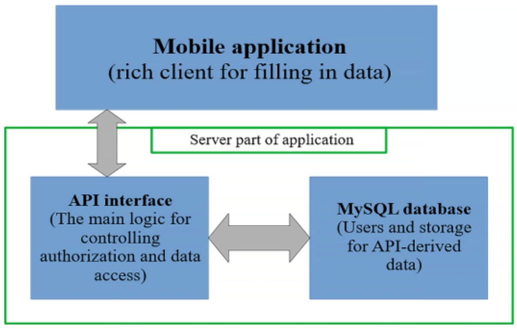


Fig. 8. Operation diagram of the mobile application.

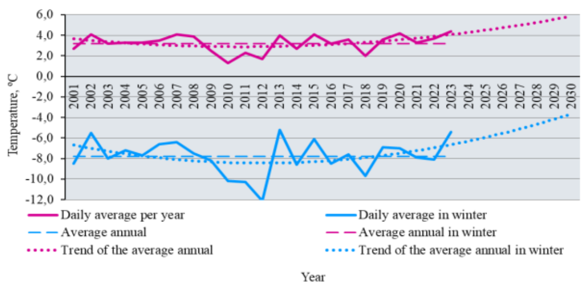


Fig. 9. Average statistical air temperature data for the study area and forecast for their changes.

Fig. 10 shows data on the amount of precipitation in this area. As can be seen from the graph, there is a gradual increase in precipitation, both the annual average and an increase in precipitation in winter. Both indicators have consistently exceeded the average values for the study period since 2009. This period is shown in the figure with a red bracket at the top.

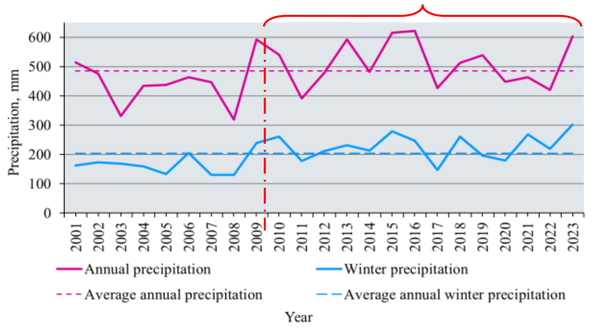


Fig. 10. Average annual precipitation and average winter precipitation for the study area.

Snowy winters are common in East Kazakhstan, but previously the greatest amount of precipitation fell during the warmer months. An analysis of the data showed that there is a steady trend of increasing precipitation in winter (Fig. 11). This can be explained by the fact that the air temperature is rising, and the summer is becoming more arid. The air temperature also rises in winter, which contributes to greater saturation of the air with water vapor and more precipitation. This also entails a change in the conditions for the formation and descent of avalanches in the direction of increasing avalanche risks.

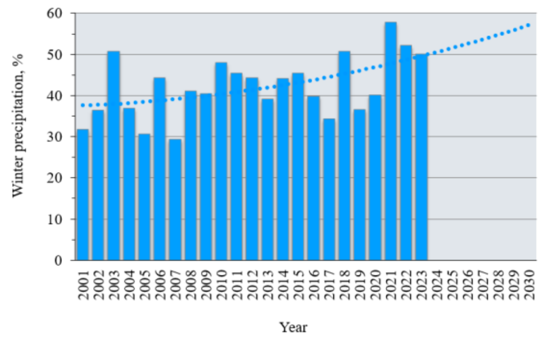


Fig. 11. Winter precipitation as a percentage of average annual precipitation for the study area.

Data research also showed that there is a tendency to increase the amount of precipitation per day in winter. Fig. 12 shows the average value of this indicator, and the trend line constructed according to a polynomial dependence of the 2nd degree. This is done only for the winter period, as this is what is important in assessing avalanche risks, as well as in order not to overload the chart with data. Comparing the graphs in Figs. 10 and 12, we can see that since 2009, the average amount of precipitation per day in winter has increasingly exceeded the average value of this indicator. According to the constructed model, it is expected that in the future the amount of precipitation that fell at the same time in one day in winter will also increase. This means that the number of days with heavy precipitation increases, which becomes one of the decisive factors leading to an inc rease in avalanche hazard.

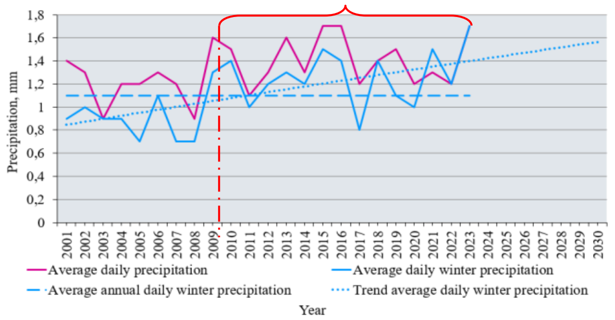


Fig. 12. Average daily precipitation for the study area.

As the analysis has shown, the microclimate of the territory of East Kazakhstan is changing, and these trends will continue in the coming years. The climate will become more humid and warmer, which will inevitably affect the avalanche situation in the region. All these conditions will affect the formation of snow cover. The average snow cover height for each year is shown in Fig. 13. There is already a tendency to exceed the annual average.

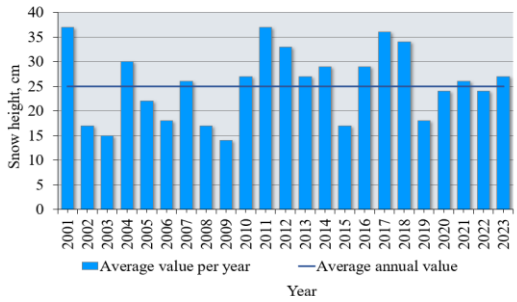


Fig. 13. Snow depth and trend for the study area.

In Kazakhstan, the average annual wind speed and its extremes are decreasing, as observed by the National Meteorological Service of Kazakhstan Kazhydromet. However, there is a tendency for them to increase in the study area. This is shown in the graph in Fig. 14. The region tends to increase the average wind speed both throughout the year and in winter. Such a wind regime of the territory will contribute to more active wind transport of snow on the slopes. Wind transport of snow along the slope is one of the factors contributing to avalanches. The wind in the mountains leads to the formation of canopies and snow deposits on the slopes. Subsequently, their collapse leads to avalanches. An increase in wind speed will lead to increased local changes in snow cover on the slopes and, consequently, to an increase in avalanche hazards in the region.

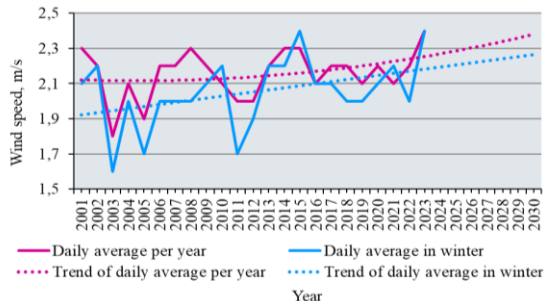


Fig. 14. Wind speed and trends in its changes for the study area.

All climatic changes contribute to an increase in avalanche activity in the region. This is confirmed by the avalanche observation data shown in Figs. 15 and 16. The winters of 2020–2021 and 2021–2022 were particularly difficult. The forecast for the coming winter, in accordance with the trend lines, also shows an increase in the avalanche danger level. There is an obvious connection between the general trends of increasing air temperature, precipitation and wind speed in winter and an increase in the number of avalanches and their volumes in recent years in Eastern Kazakhstan.

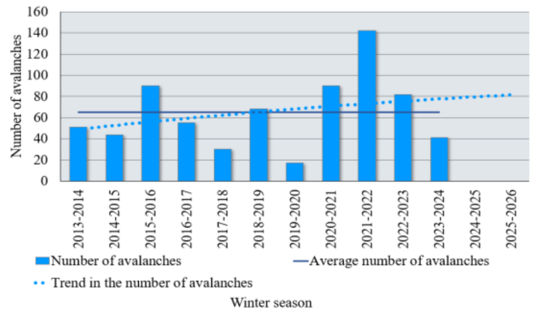


Fig. 15. The number of avalanches in East Kazakhstan region for various winter seasons.

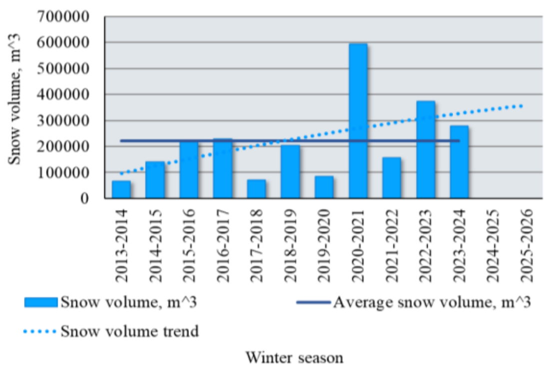


Fig. 16. Snow volume of avalanches in East Kazakhstan region for various winter seasons.

Having considered the change in climatic data for East Kazakhstan, we see that the main factors of avalanche risks for this territory are changing. The analysis revealed that there are obvious links between an increase in air temperature, an increase in precipitation and wind speed in winter and an increase in the number of avalanches and their volumes in recent years in the territory of East Kazakhstan:

- Climate change: the climate is becoming more humid and warmer.
- Increased precipitation: There is a steady increase in both average annual and winter precipitation, which leads to increased avalanche risks.
- Snow cover changes: average snow cover heights are increasing, which is also associated with an increase in avalanche activity.
- Wind and avalanches: there is an increase in the average wind speed in the region, which contributes to the active wind transport of snow on the slopes. This increases the risk of the formation of canopies and snow deposits, which can lead to avalanches.
- Increased avalanche risk: all climatic changes, including increased precipitation, temperature, and wind speed, contribute to an increase in avalanche activity in the region, and forecasts for future winters also predict an increase.

Meteorological parameters (air temperature, wind speed, and precipitation) affect spontaneous avalanches. Using modeling, we analyzed 5 avalanche-prone areas (Pikhtovka, Prohodnaya, Tainty, Sogornoye-Barlyk, Bogatyrskaya mine), where spontaneous avalanches occurred at different times. For the study, meteorological parameters were taken 3 days before the incident, on the day of the avalanche, and 3 days after the incident.

Table III shows data on these avalanche-prone areas. The date of the avalanche is shown in bold. Information about avalanches is shown in Table IV. Statistical analysis methods of the MATLAB program were used to evaluate the relationships between these variables.

TABLE III. METEOROLOGICAL DATA ON SPONTANEOUS AVALANCHES

Avalanche-prone areas	Date	Wind speed (x) (m/s)	Precipitation (y) (mm)	Air temperature (z) (°C)
	22.12.2015	2.4	0.2	−14.2
	23.12.2015	5.0	1.4	−8.5
	24.12.2015	3.3	2.2	−2.3
Pikhtovka	**25.12.2015**	**2.6**	**4.4**	**−0.2**
	26.12.2015	3.0	0.3	1.3
	27.12.2015	4.5	0	0.9
	28.12.2015	4.3	0	−0.1
	06.01.2019	0	-	−25.7
	07.01.2019	6.3	-	−8.7
	08.01.2019	4.5	-	−8.0
Prohodnaya	**09.01.2019**	**5.8**	**1.1**	**−0.4**
	10.01.2019	1.8	8.4	−3.2
	11.01.2019	0	1.8	−13.4
	12.01.2019	0	1.5	−13.1
	20.01.2020	1.1	1.7	−8.7
	21.01.2020	2.1	5.9	−8.2
	22.01.2020	3.4	6.7	−6.6
Tainty	**23.01.2020**	**3.0**	**4.9**	**−6.2**
	24.01.2020	1.6	-	−7.5
	25.01.2020	2.5	7.4	−4.0
	26.01.2020	3.0	6.7	−4.6

Sogornoye-Barlyk	22.11.2021	8.4	-	0.5
	23.11.2021	2.5	3.4	−4.4
	24.11.2021	1.4	10.9	−1.9
	25.11.2021	**4.5**	**-**	**1.9**
	26.11.2021	3.5	3.8	−4.0
	27.11.2021	3.3	-	−17.8
	28.11.2021	6.1	-	−11.4
Bogatyrevskaya mine	04.01.2023	3.0	0.4	−4.9
	05.01.2023	2.0	19.7	−3.9
	06.01.2023	0.3	16.6	−3.2
	07.01.2023	**0.5**	**4.6**	**−1.6**
	08.01.2023	1.6	3.0	−5.0
	09.01.2023	1.0	2.3	−1.1
	10.01.2023	0.8	14.1	−6.9

TABLE IV. INFORMATION ABOUT AVALANCHES

№	Avalanche-prone area	Specifications	Consequence
1	Pikhtovka	The reason: heavy snowfall, increased wind and poor visibility. The avalanche point was above the road. Avalanche volume 500 m^3.	There are no casualties or damage.
2	Prohodnaya	The reason is an increase in air temperature, heavy precipitation and increased wind. The avalanche point was above the road. The avalanche volume is 4300 m^3.	The snow completely blocked the road. There are no casualties or damage.
3	Tainty	The reason: heavy snowfall, increased wind and poor visibility. The avalanche point was above the road. The avalanche volume is 1100 m^3.	There are no casualties or damage.
4	Sogornoye-Barlyk	The reason: increased air temperatures and precipitation. The avalanche point was above the road. The avalanche volume is 250 m^3.	The snow completely blocked the road. There are no casualties or damage.
5	Bogatyrevskaya mine	The reason: increased air temperatures and precipitation. The avalanche point was above the road. The avalanche volume is 3500 m^3.	There are no casualties or damage.

The results are shown in Figs. 17–21, and the dependency models are presented in Eqs. (1)–(5). Variable 1 (x)—wind speed, variable 2 (y)—precipitation, variable 3 (z)—air temperature (Table V).

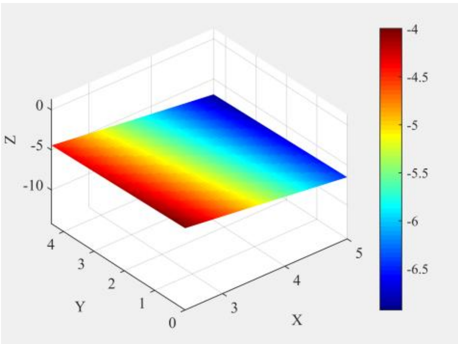


Fig. 17. Graph of dependence of variables for Pikhtovka.

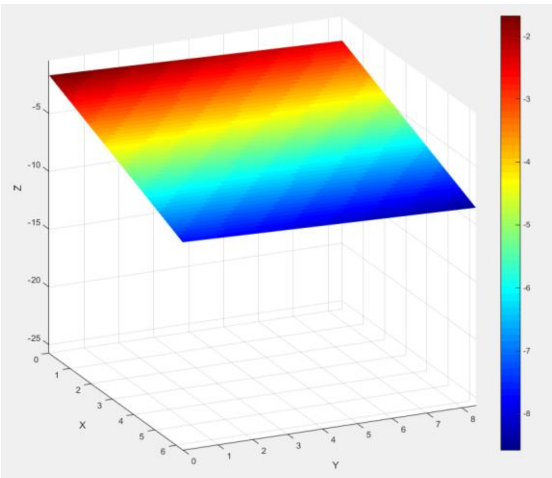


Fig. 18. Graph of dependence of variables for Prokhodnaya.

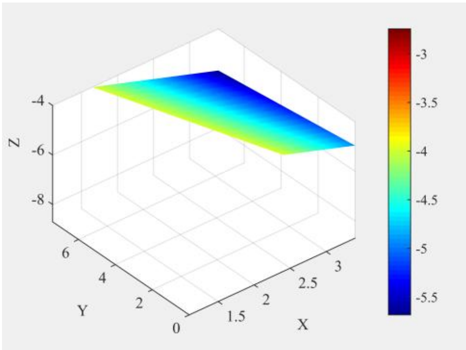


Fig. 19. Graph of dependence of variables for Tainty.

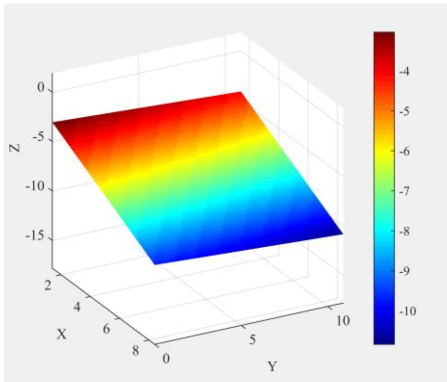


Fig. 20. Graph of dependence of variables for Sogornoye-Barlyk.

Thus, meteorological conditions—a sharp warming, an increase in wind speed and heavy precipitation, contributing to snow melting and its transfer, are the main factors influencing the spontaneous avalanche, regardless of the year and month. These three factors are interrelated, as warming in winter is usually accompanied by increased wind and brings precipitation to the region.

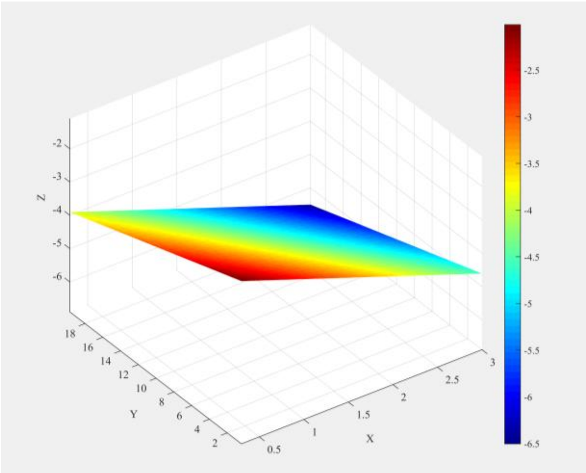


Fig. 21. Graph of dependence of variables for Bogatyrevskaya mine.

TABLE V. DEPENDENCY MODELS

№	Avalanche-prone area	Model	
1	Pikhtovka	$z = -8.687 + 0.9804x + 1.027y$	(1)
2	Prohodnaya	$z = -14.6872 + 2.2983x + 0.8736y$	(2)
3	Tainty	$z = -10.3211 + 0.2617x + 0.5909y$	(3)
4	Sogornoye-Barlyk	$z = -6.2887 + 0.2519x + 0.370y$	(4)
5	Bogatyrevskaya mine	$z = -1.6846 - 0.9627x - 0.098y$	(5)

As mentioned earlier, as part of our research, we collected historical meteorological data at avalanche collections in the East Kazakhstan region. Based on the collected data, it is possible to predict spontaneous avalanches using regression analysis.

Regression analysis is one of the main statistical tools that allow you to determine the relationship between a set of input factors and a dependent variable. When constructing a regression, coefficients are determined for each input variable, which determines the degree of influence of each input factor on the value of the output variable.

When predicting spontaneous avalanches, we need to determine the probability of an avalanche event at an avalanche collection based on weather conditions, which in this case will act as input variables. Logistic regression is the most suitable tool for solving this problem.

Logistic regression is used when there is an output variable obeying the binomial distribution law. Since the output value is binary, the specified regression type calculates the probability of assigning the regression value to one of the two possible values of the regression output variable.

To calculate the logistic regression of predicting spontaneous avalanches in the framework of the study, we determined the following set of input variables, shown in Table VI.

The data listed in Table VI for calculating logistic regression was obtained from a database on avalanche weather conditions developed as part of this study. In our case, the recorded spontaneous avalanches on a certain date on a given avalanche collection will be the output variable. To extract data from the database, we created views (Supplementary).

TABLE VI. INPUT VARIABLES FOR CALCULATING LOGISTIC REGRESSION

№	Variable	Explanation
1	avg_temperatures_day_1	average daytime temperature on the day of the avalanche
2	avg_temperatures_day_2	average daytime temperature the day before the avalanche
3	avg_temperatures_day_3	average daytime temperature 2 days before the avalanche
4	avg_temperatures_decade_1	average daytime temperature on the day of the avalanche
5	avg_temperatures_decade_2	average daytime temperature the day before the avalanche
6	avg_temperatures_decade_3	average daytime temperature 2 days before the avalanche;
7	rainfalls_value_1	the amount of precipitation on the day of the avalanche
8	rainfalls_value_2	the amount of precipitation the day before the avalanche
9	rainfalls_value_3	the amount of precipitation 2 days before the avalanche
10	snows7_average_1	average snow cover at 7:00 a.m. on the day of the avalanche
11	snows7_average_2	average snow cover at 7:00 a.m. the day before the avalanche
12	snows7_average_3	average snow cover at 7:00 a.m. 2 days before the avalanche
13	snows7_maximum_1	the maximum snow cover is at 7:00 a.m. on the day of the avalanche
14	snows7_maximum_2	the maximum snow cover is at 7:00 a.m. the day before the avalanche
15	snows7_maximum_3	the maximum snow cover is at 7:00 a.m. 2 days before the avalanche
16	snows19_average_1	average snow cover at 19:00 on the day of the avalanche
17	snows19_average_2	average snow cover at 19:00 the day before the avalanche
18	snows19_average_3	average snow cover at 19:00 2 days before the avalanche
19	snows19_maximum_1	the maximum amount of snow cover is at 19:00 on the day of the avalanche
20	snows19_maximum_2	the maximum amount of snow cover is at 19:00 the day before the avalanche
21	snows19_maximum_3	the maximum snow cover is at 19:00 2 days before the avalanche
22	temperatures7_value_1	temperature at 7:00 on the day of the avalanche
23	temperatures7_value_2	temperature at 7:00 the day before the avalanche
24	temperatures7_value_3	temperature at 7:00 2 days before the avalanche
25	temperatures19_value_1	temperature at 19:00 on the day of the avalanche
26	temperatures19_value_2	temperature at 19:00 the day before the avalanche
27	temperatures19_value_3	temperature at 19:00 2 days before the avalanche

We used the Loginom Community statistical software package to perform logistic regression calculations. In this software package, a scenario was created for calculating logistic regression.

The created package connects to the developed database, extracts data, and uses it to train logistic regression.

For a general understanding, Fig. 22 shows a diagram of the interaction of the system components.

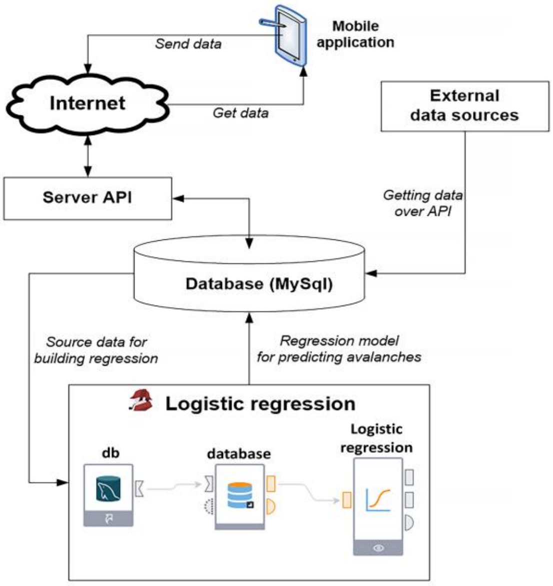
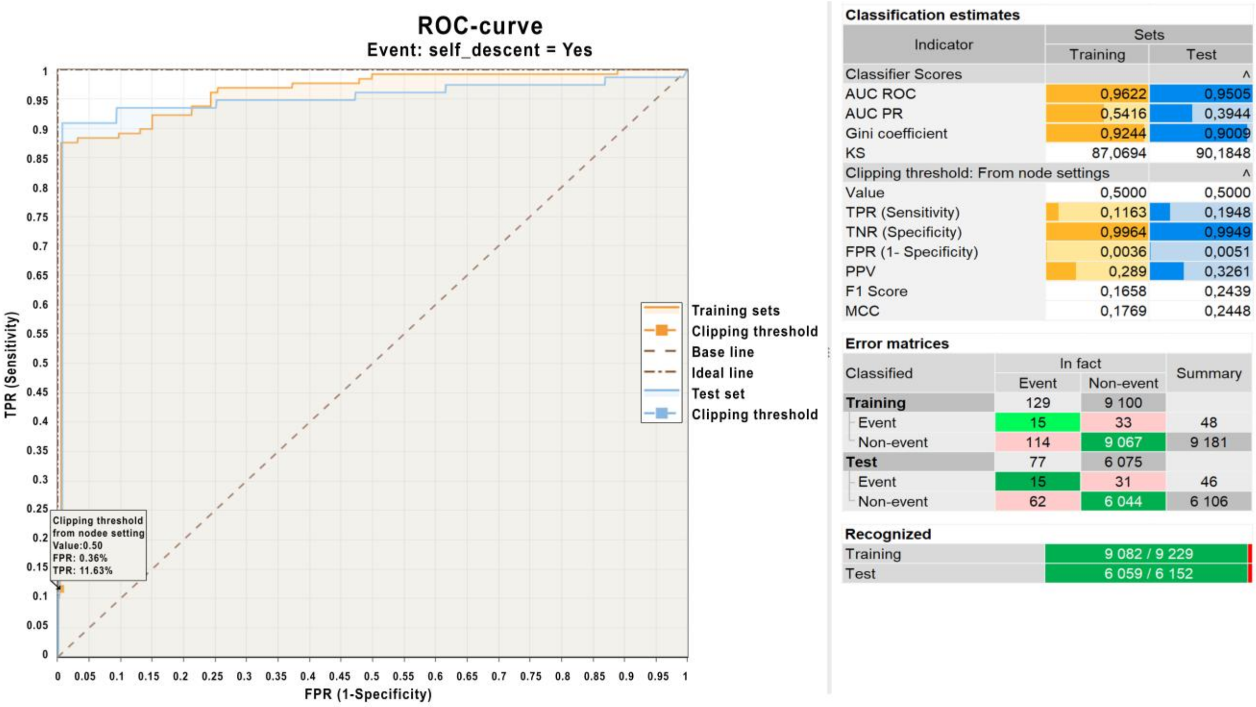


Fig. 22. Component interaction diagram.

The avalanche hazard analysis and forecasting system based on logistic regression using the Loginom software platform is included in the created information system and operates using the MySQL database and the mobile application described above.

Also in this statistical package, we evaluated the quality of the constructed logistic regression model. Fig. 23 shows a graph and data on the estimation of the constructed logistic regression. As can be seen from the data shown in Fig. 23, the constructed regression model for predicting avalanches is of a high quality. This is evidenced by the following indicators:

- A high value of the AUC ROC coefficient, which has a value of 0.9622 for the training set and a value of 0.9505 for the test set, since for this indicator a value of 0.9 or more indicates a high quality of forecasting.
- High quality of value recognition—over 98% for the training set (9082 out of 9229) and the test set (6059 out of 6152).



Classification estimates

Indicator	Sets: Training	Test
Classifier Scores		
AUC ROC	0,9622	0,9505
AUC PR	0,5416	0,3944
Gini coefficient	0,9244	0,9009
KS	87,0694	90,1848
Clipping threshold: From node settings		
Value	0,5000	0,5000
TPR (Sensitivity)	0,1163	0,1948
TNR (Specificity)	0,9964	0,9949
FPR (1- Specificity)	0,0036	0,0051
PPV	0,289	0,3261
F1 Score	0,1658	0,2439
MCC	0,1769	0,2448

Error matrices

Classified	In fact: Event	Non-event	Summary
Training	129	9 100	
Event	15	33	48
Non-event	114	9 067	9 181
Test	77	6 075	
Event	15	31	46
Non-event	62	6 044	6 106

Recognized

Training	9 082 / 9 229
Test	6 059 / 6 152

Fig. 23. Quality assessment of the regression model.

After training the model, we obtained a regression coefficient, the value of which is given in Table VII.

TABLE VII. THE VALUE OF LOGISTIC REGRESSION COEFFICIENTS

№	Names of input fields	Regression coefficient
1	The constant	−2.6596
2	avg_temperatures_day_1	0.0397
3	avg_temperatures_day_2	0.3580
4	avg_temperatures_day_3	0.0271
5	avg_temperatures_decade_1	0.1441
6	avg_temperatures_decade_2	0.0778
7	avg_temperatures_decade_3	0.0879
8	rainfalls_value_1	−0.7986
9	rainfalls_value_2	−0.9923
10	rainfalls_value_3	−0.8184
11	snows7_average_1	−0.0794
12	snows7_average_2	−0.0289
13	snows7_average_3	0.0144
14	snows7_maximum_1	0.0256
15	snows7_maximum_2	0.0852
16	snows7_maximum_3	0.0041
17	snows19_average_1	0.0254
18	snows19_average_2	0.0247
19	snows19_average_3	0.0240
20	snows19_maximum_1	0.0169
21	snows19_maximum_2	−0.0595
22	snows19_maximum_3	−0.0789
23	temperatures7_value_1	0.1553
24	temperatures7_value_2	0.0886
25	temperatures7_value_3	0.1139
26	temperatures19_value_1	−0.0106
27	temperatures19_value_2	−0.1273
28	temperatures19_value_3	−0.0346

The logistic regression we constructed in this part of the study can later be used to predict spontaneous avalanches based on observations of meteorological data in avalanche-prone areas of the East Kazakhstan region.

V. Conclusion

The study analyzed weather data on 497 avalanche-prone areas of the East Kazakhstan region. An analysis of climate change has shown that: the average annual temperature and precipitation in the region are increasing, especially in winter, which increases the risk of avalanches; wind speed is also increasing, which contributes to the formation of snow deposition; there is a relationship between climate change and an increase of number of avalanches.

In addition to the weather factors presented in the study, an avalanche can be triggered by humans and other conditions. Avalanches have already claimed many lives around the world this winter [20]. And all the cases were provoked by the victims themselves. But at the same time, the avalanche risk in these cases has already been estimated above 3 (according to the North American avalanche danger scale) due to weather conditions. Meteorological factors are the primary factors that increase avalanche risk.

The main factors of avalanches are a sharp increase in temperature, increased wind and heavy precipitation. Real cases of spontaneous avalanches have been analyzed for several sites and probabilistic models have been constructed.

As part of the study, historical weather data was used to predict spontaneous avalanches using regression analysis. Logistic regression, suitable for binary outcomes, allows you to predict the probability of an avalanche based on weather conditions.

A set of input variables is used to build the model. The recorded spontaneous avalanches act as the output variable. Forecasting was carried out using the Loginom Community statistical package.

The evaluation of the model showed a high quality of forecasting. Recognition accuracy has exceeded 98%. The obtained regression coefficients can be used to predict avalanches based on meteorological data in the East Kazakhstan region.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Natalya Denissova: Conceptualization, Methodology, Investigation, Formal analysis, Writing—original draft, Writing—review & editing. Olga Petrova: Conceptualization, Methodology, Investigation, Formal analysis, Writing—original draft, Writing—review & editing. Evgeny Fedkin: Methodology, Investigation, Data curation, Writing—original draft, Resources. Gulzhan Daumova: Conceptualization, Methodology, Investigation, Formal analysis, Writing—original draft, Writing—review & editing. Evgeny Sergazinov: Investigation, Data curation, Resources. All authors had approved the final version.

Funding

The article presents the results of scientific research obtained during the implementation of a scientific and technical program BR21882022 on the topic: "Research of avalanche activity in the East Kazakhstan region for development of monitoring systems and scientific substantiation of their placement" within the framework of program-targeted financing.

References

[1] E. Hawkins, D. Frame, L. Harrington, M. Joshi, A. King, M. Rojas, and R. Sutton, "Observed emergence of the climate change signal: From the familiar to the unknown," *Geophysical Research Letters*, vol. 47, e2019GL086259, 2020. doi: 10.1029/2019GL086259

[2] H. Wen, X. Wu, X. Shu, D. Wang, S. Zhao, G. Zhou, and X. Li, "Spatial heterogeneity and temporal tendency of channeled snow avalanche activity retrieved from Landsat images in the maritime snow climate of the Parlung Tsangpo catchment, southeastern Tibet," *Cold Regions Science and Technology*, vol. 223, 104206, 2024. doi: 10.1016/j.coldregions.2024.104206

[3] N. Eckert, C. Corona, and F. Giacona, "Climate change impacts on snow avalanche activity and related risks," *Nature Reviews Earth & Environment*, vol. 5, pp. 369–389, 2024. doi: 10.1038/s43017-024-00540-2

[4] Q. Li, T. Yang, H. Zhou, and L. Li, "Patterns in snow depth maximum and snow cover days during 1961–2015 period in the Tianshan Mountains, Central Asia," *Atmospheric Research*, vol. 228, pp. 14–22, 2019. doi: 10.1016/j.atmosres.2019.05.004

[5] J. Hao, X. Zhang, P. Cui, L. Li, Y. Wang, G. Zhang, and C. Li, "Impacts of climate change on snow avalanche activity along a transportation corridor in the Tianshan mountains," *International Journal of Disaster Risk Science*, vol. 14, pp. 510–522, 2023. doi: 10.1007/s13753-023-00475-0

[6] F. Giacona, N. Eckert, C. Corona, R. Mainieri, S. Morin, M. Stoffel, B. Martin, and M. Naaim, "Upslope migration of snow avalanches in a warming climate," in *Proc. the National Academy of Sciences*, 2021. doi: 10.1073/pnas.2107306118

[7] L. R. Mudryk, P. J. Kushner, C. Derksen, and C. Thackeray, "Snow cover response to temperature in observational and climate model ensembles," *Geophysical Research Letters*, vol. 44, pp. 919–926, 2017. doi: 10.1002/2016GL071789

[8] X. Zhong, T. Zhang, S. Kang, and J. Wang, "Snow depth trends from CMIP6 models conflict with observational evidence," *Journal of Climate*, vol. 35, pp. 1293–1307, 2022. doi: 10.1175/JCLI-D-21-0177.1

[9] J. I. López-Moreno, N. Callow, and H. McGowan, "Marginal snowpacks: The basis for a global definition and existing research needs," *Earth Science Reviews*, vol. 252, 104751, 2024. doi: 10.1016/j.earscirev.2024.104751

[10] Z. Tang, G. Deng, G. Hu, H. Zhang, H. Pan, and G. Sang, "Satellite observed spatiotemporal variability of snow cover and snow phenology over High Mountain Asia from 2002 to 2021," *Hydrology*, vol. 613, 128438, 2022. doi: 10.1016/j.jhydrol.2022.128438

[11] K. E. Kunkel, D. A. Robinson, S. Champion, X. Yin, T. Estilow, and R. M. Frankson, "Trends and extremes in Northern hemisphere snow characteristics," *Current Climate Change Reports*, vol. 2, pp. 65–73, 2016. doi: 10.1007/s40641-016-0036-8

[12] L. Quante, S. N. Willner, R. Middelanis, and A. Levermann, "Regions of intensification of extreme snowfall under future warming," *Scientific Reports*, vol. 11, 16621, 2021. doi: 10.1038/s41598-021-95979-4

[13] J.-S. Hao, Y. Wang, and L.-H. Li, "Snowpack variations and their hazardous effects under climate warming in the central Tianshan Mountains," *Advances in Climate Change Research*, vol. 15, no. 3, pp. 442–451, 2024. doi: 10.1016/j.accre.2024.06.001

[14] K. Laute and A. A. Beylich, "Potential effects of climate change on future snow avalanche activity in western Norway deduced from

meteorological data," *Geografiska Annaler: Series A, Physical Geography*, vol. 100, no. 2, pp. 163–184, 2018. doi: 10.1080/04353676.2018.1425622

[15] E. R. Sujatha and V. Sridhar, "Landslide susceptibility analysis: A logistic regression model case study in Coonoor, India," *Hydrology*, vol. 8, no. 1, 41, 2021. doi: 10.3390/hydrology8010041

[16] M. S. Chowdhury, M. N. Rahman, M. S. Sheikh, S. M. A. Sayeid, K. H. Mahmud, and B. Hafsa, "GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh," *Heliyon*, vol. 10, no. 1, e23424, 2024. doi: 10.1016/j.heliyon.2023.e23424

[17] F. Gauthier, D. Germain, and B. Hétu, "Logistic models as a forecasting tool for snow avalanches in a cold maritime climate: Northern Gaspésie, Québec, Canada," *Natural Hazards*, vol. 89, pp. 201–232, 2017. doi: 10.1007/s11069-017-2959-3

[18] K. Nosrati, S. Kiashemshaki, and M. M. Hoseinzadeh, "Prediction of snow avalanche-susceptibility in Meigun-Shemshak road using rare events logistic regression," *Journal of Geography and Environmental Hazards*, vol. 5, no. 1, pp. 55–68, 2016. doi: 10.22067/geo.v5i1.47116

[19] A. Chourot and J. P. Martin, "Comparison of logistic regressions and snowfall intensity–duration threshold as forecasting tools for direct-action snow avalanches in the presidential range, New Hampshire, USA," *Natural Hazards*, vol. 93, pp. 1649–1656, 2018. doi: 10.1007/s11069-018-3361-5

[20] Bridger-Teton Avalanche Center. Recent Accidents. [Online]. Available: https://bridgertetonavalanchecenter.org/recent-accidents/