

Geomatics for Environmental Monitoring, Analysis and Forecast

Valery G. Gitis and Alexander B. Derendyaev(✉)

Institute for Information, Transmission Problems RAS (Kharkevich Institute), Moscow, Russia
gitis@iitp.ru, wintsa@gmail.com

Abstract. We suggest geomatic technology for monitoring natural processes, which is implemented on three web GIS platforms: (1) http://distcomp.ru/geo/arctic/—monitoring the hydroecological situation in the Arctic, (2) http://distcomp.ru/geo/2/, http://distcomp.ru/geo/3/—analysis of seismic fields and (3) http://distcomp.ru/geo/prognosis/—automatic prediction of earthquakes. Platforms combine two levels of geodata analysis. The first level supports automatic data processing and simple analysis tools that are suitable for any Internet user. The second level is designed for detailed data analysis performed by a specialist. Thus, users of the platform have the opportunity to receive preliminary information about the processes in the environment and conduct research.

Keywords: Web-based GIS platform · Arctic · Seismic fields · Automatic forecasting of earthquakes · Machine learning · Method of the minimum area of alarm

1 Introduction

Monitoring of natural and man-made processes is used for regular observation, assessment, and prediction of the state of the environment. The solution to these problems is connected with the analysis of large volumes of heterogeneous spatial and spatio-temporal data. The characteristics of such data require the use of specialized methods and technologies that are implemented in web GIS.

The main set of GIS operations for monitoring and analyzing natural processes were considered in [1]. Currently, a web GIS is usually used for visual analysis of meteorological fields. For example, the https://www.windy.com and https://www.ventusky.com/ systems provide up-to-date and predictive information for a large number of meteorological processes, as well as relevant information about the infrastructure. Both GIS have high-quality interactive 4D visualization of temporal processes with animation elements. The presence of animated cartography and interactive graphics, the possibility of joint visualization of several processes, user-friendly interface contribute to the understanding of the meteorological situation.

The work was supported by Russian Foundation for Basic Research, project No. 17-07- 00494.

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11622, pp. 200–215, 2019.
https://doi.org/10.1007/978-3-030-24305-0_16

When developing technology, we proceeded from the fact that monitoring data in many cases should be used to predict hazardous phenomena. Therefore, it is desirable that the monitoring system supports not only convenient data visualization but also analytical functions. Information obtained from the system can be divided into two groups. The first group includes information that can be obtained from initial observations automatically using predefined data processing methods. The second group includes information for which an extraordinary complex analysis of data is required. We believe that any Internet users, including employees of federal and regional services and administrations, as well as research staff, can be interested in the first group of information. The second group of information is requested by problem domain specialists performing scientific research. The first group of users requires automatic loading, processing, visual representation of results and a simplified interface. The second group of users requires a set of tools needed for joint data analysis, as well as an ability to download and integrate additional data from remote servers and local network. In systems for seismic hazard analysis and earthquake prediction, it is advisable to combine both levels of data analysis. The first level should support monitoring of seismic fields and the forecast of earthquakes, providing automatic loading and processing of data and providing the operator with simple analysis tools in combination with a visual representation of the results. The second level is designed for detailed analysis of hypotheses, which an expert can formulate at the first level.

The idea of two-level analysis of geographic data is universal. It has proven itself in the ESRI solution [2], in which editing and administration of maps are performed using ArcGIS Online, and in-depth data analysis is supported using ArcGIS Desktop [3].

The monitoring platforms implemented in this technology combine automatic data analysis, accessible to any Internet user, with detailed analysis performed by a specialist. Thus, platforms are convenient for all users. They receive preliminary information about the processes in the environment and have the opportunity to conduct research. Platforms consist of two web GIS. The first GIS automatically loads and processes the source data, provides the user with analytical tools with intuitive operations and a simplified interface, starts calculations on a remote server and prepares a GIS project for the second GIS. The second GIS is a multifunctional system focused on the analysis of spatio-temporal processes, which is performed by a specialist.

In Sects. 2, 3 and 4 we present three geomatic technologies. They are implemented on the following network platforms: (1) http://distcomp.ru/geo/arctic/ is the platform for monitoring of hydroecological processes in the Arctic, (2) http://distcomp.ru/geo/2/ and http://distcomp.ru/geo/3/ are the platforms for analysis of seismic fields and (3) http://distcomp.ru/geo/prognosis/ is the platform for automatic prediction of earthquakes.

2 The Technology of Monitoring and Analysis of the Hydroecological Situation in the Arctic

Natural and anthropogenic processes in the Arctic substantially affect the nature of the entire Earth [4, 5], and high sensitivity of the Arctic to the increase in the average temperature here plays a unique role. Warming in the Arctic leads to ice cover

reduction, causes reflection of solar rays, and, therefore heating of the whole Earth. Discharge of cold fresh water as a result of ice melting reduces seawater salinity in the near-Arctic regions, which causes changes in oceanic currents. Melting of permafrost and reduction of the ice cover can lead to discharged large methane volumes to the Earth's atmosphere, therefore promoting global warming [6].

The Arctic is notably rich in hydrocarbon, mineral and biological resources, which makes it a region of growing economic activity for their exploration. Hydrocarbons are mined on the Arctic shelf, rare earth minerals are drilled on the coasts, and the marine environment is widely used for fishing and transit. Economic activity, storage of industrial waste and various emergencies can significantly affect the ecology of the Arctic. It is clear that the anthropogenic impact on natural processes in the Arctic will only increase in the future. Note that the peculiarity of the Arctic environment is its long-term recovery after adverse anthropogenic influences [7]. The above factors show why the Arctic is today one of the most important objects for monitoring natural and anthropogenic processes.

Arctic monitoring technology is implemented on a network platform (http://distcomp.ru/geo/arctic), which consists of two web GIS [8, 9]. The first GIS, GIS GeoMonitor, is built in a client-server architecture with a thin client. It supports automatic loading and processing of data, a visual representation of spatio-temporal processes, marine vehicles and earthquake epicenters, measurement of hydrometeorological fields and point data attributes, and also forms a GIS project and starts modeling hydrocarbon spills on a remote server.

GIS GeoTime 3 is realized in a thick Java-based client-server architecture (http://www.geo.iitp.ru/GT3/). It is a big multi-functional system focused on the analysis of spatio-temporal environmental processes and earthquake forecasting research. It is launched from the GIS GeoMonitor page with the data that is used for analysis. Tools of GeoTime 3 makes it possible to supplement the GIS project with tile maps and 2D, 3D, 4D point, vector and grid data from remote and local servers. GIS Geo-Time 3 supports the joint processing of all these types of data. The leading interactive operations supporting visual analysis are the following: presentation of data in the form of maps, cuts, plots, diagrams, and tables; animation visualization of one or more spatio-temporal processes; measurement of distances, areas and reading the attributes; evaluation of statistics for points and polygons.

The platform provides users with access to the analysis of three types of spatial and spatio-temporal data: re-analysis of historical data, operational observations, and forecast. The information resources of the platform include: (1) Reanalysis data of ERA-Interim for 1979–2015 (ECMWF); (2) Data of operative mid-term (0–120 h) forecast of the GFS and NCEP/NOAA meteorological fields (Roshydromet); (3) Data on movements of vessels (the North Sea Route Administration); (4) Data of operative forecast of vessels icing rate (Pacific Marine Environmental Research Laboratory); (5) Operative ice cover maps of the Arctic Ocean (the Arctic and Antarctic Research Institute); (6) Forecast data of hydrocarbon spill modeling (State Oceanographic Institute); (7) Earthquake catalogs (International Seismological Centre); (8) Google Maps tiled maps.

On the start page of the platform, there is access to three applications. The first application incorporates the climatic fields of the Arctic Zone for the last 37 years.

For this purpose, it employs the ERA-Interim. The fields of climatic parameters are calculated from the following hydrometeorological characteristics: (1) Temperature of the underlying surface, K; (2) Air temperature at a height of 2 m, K; (3) Near-surface pressure, hPa; (4) Integrity of ice cover (ratio between the ice-covered area and the total area of the cell of the gridded field); (5) Divergence of the field of the wind speed at a height of 10 m; (6) Rotation of the field of the wind speed at a height of 10 m. The last two characteristics are calculated from meridional and zonal components of the wind speed (in m/s). The fields of the mentioned parameters are presented in the interval of latitudes from 55° to 85° N, on the grid of 0.75° × 0.75° × 30 days.

The data refer to the following three time intervals: T_1 = 01.01.1979–01.01.1998, T_2 = 01.01.2006–01.01.2015 and T_3 = 01.01.1979–01.01.2015. For each month and for each of the above hydrometeorological characteristics, GIS GeoMonitor presents field maps for the following four climatic indicators: (1) Average monthly values in intervals T_1 and T_2; (2) Mean square deviations of values in intervalsT_1 and T_2; (3) Values of significance of changes in the mean values of characteristics in intervals T_1 and T_2 (based on the Student statistics); (4) Trends of characteristics in the interval T_3.

As an example, let us consider the analysis of the indicator estimating the significance of changes in the temperature regime of the underlying surface at arctic latitudes. Figure 1 demonstrates the GIS GeoMonitor screenshots with the maps of significance of changes in average monthly temperatures of the underlying surface for odd months (January, March, May, July, September, and November) of intervals T_1 (1979–1998) and T_2 (2006–2015), i.e., the Student statistics that equals the difference (normalized to the mean square deviation) between mean values in intervals T_2 and T_1. It can be seen in the maps that temperatures in "winter" months (January and November) of 2006–2015 considerably increased (light gray tones) in comparison with the period of 1979–1998, whereas temperatures in "summer" months (May and July) decreased (dark gray tones). Additionally, it is seen in the maps that temperature changes are spatially inhomogeneous.

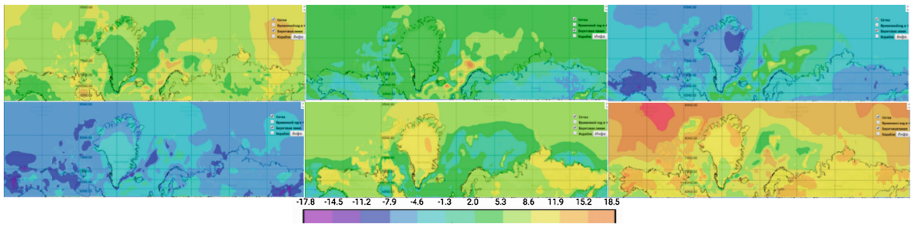


Fig. 1. The maps of the significance of changes in the average monthly temperature between time intervals 2006–2015 and 1979–1998, from left to right and from top to bottom: January, March, May, July, September, November.

Figure 2 shows the screenshot from GIS GeoTime 3 with the ice cohesion map in January as a percentage, averaged over 2006–2016 and a graph of average monthly ice cohesion for the point 70.385° N, 54.489° E.

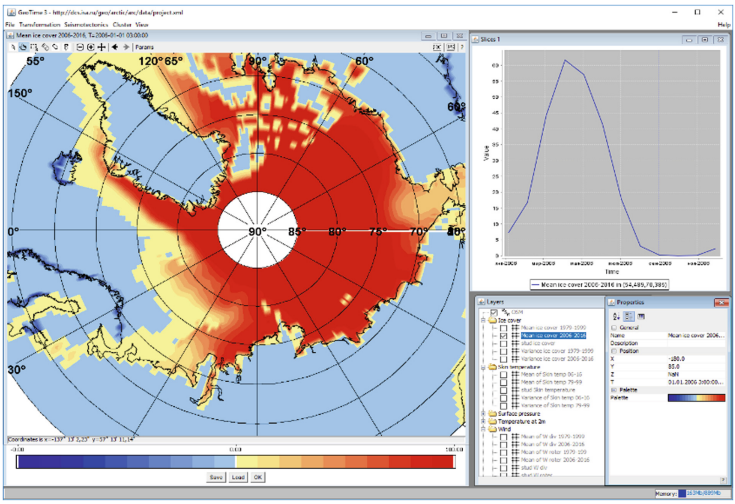


Fig. 2. Ice cohesion in January as a percentage averaged over 2006–2016. The right is a graph of average monthly ice cohesion for the point 70.385° N, 54.489° E.

The second and third applications launch the GIS GeoMonitor with modules to analyze the operational and forecast situation in the Barents and White seas, as well as in the Bering Sea region. Analysis of operative data and forecast of the environmental situation is of significant interest to all participants of sea-based activities. Due to the intensified anthropogenic load in the region, the operative data module contains, apart hydrometeorological characteristics, such data as operative locations and movement directions of vessels, forecasts of their icing rates, and forecast of the evolution of possible emergencies related to hydrocarbon spills.

The data of monitoring of the operative situation in the White and Barents seas are automatically updated. They consist of the observed values of parameters and forecast. According to these data, the GeoMonitor GIS contains the set of the following spatio-temporal fields: (1) Modulus of the wind speed at a height of 10 m from underlying surface, m/s; (2) Intensity of precipitation, mm/h; (3) Atmospheric pressure at the underlying surface, hPa; (4) Temperature of the underlying surface, °C; (5) Air temperature at a height of 2 m, °C; (6) Specific air humidity (water mass in 1 kg of air), g/kg; (7) Icing rate or ice augmentation, mm/h (light icing is 0.1–6.858 mm/h, moderate icing is 6.858–20.32 mm/h, and heavy icing is more than 20.32 mm/h); (8) the ice cohesion.

The interactive user interface of the GIS GeoMonitor allows reading of the attributive data: these data for vessels are the name, type, movement direction, and speed; for earthquake epicenters, these data are the occurrence time, magnitude, and hypocentral depth. An example of the presentation of operational data in the Bering Sea region is shown in Fig. 3.

An authorized user of the GIS GeoMonitor can also perform analysis of oil product distribution after a spill. Analysis of the distribution of hydrocarbon spills proceeds as follows. First, in the GIS GeoMonitor dialog box, the modeling parameters are set by

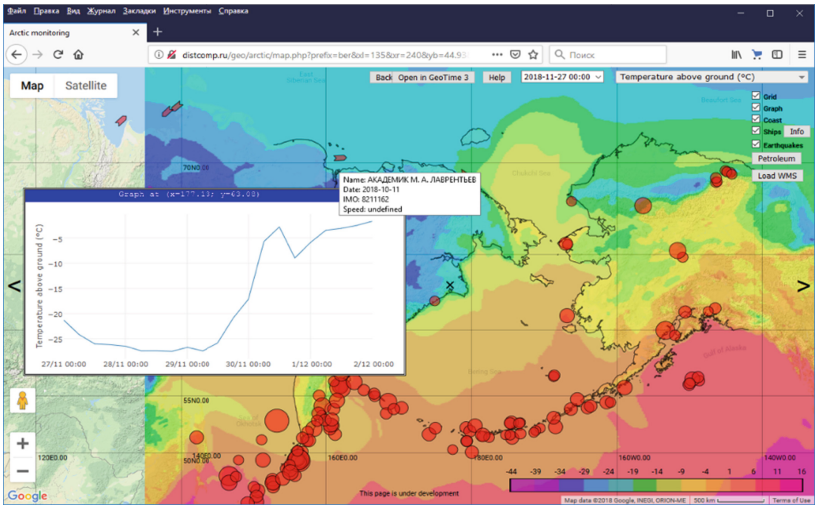


Fig. 3. GeoMonitor working window: a slice of the spatio-temporal air temperature field at the height of 2 m dated 11.27.2008 00:00 and the temperature forecast graph at the point marked with a cross. Earthquake epicenters marked with red circles. (Color figure online)

the operator: the execution time of the prediction of the distribution of oil or oil products, visualization step, type of oil product, time of the accident, coordinates of the spill point, intensity and duration of oil discharge. Further, according to these data, the remote server calculates the coordinates, shape, and thickness of the film for each modeling stage, taking into account the coastline, ice conditions, meteorological forecast and other data on the state of the environment [10]. According to the report on the completion of the next iteration, the GIS GeoMonitor loads the spill spot with differentiation across the thickness of the oil film. After that, the user can display the spatio-temporal oil spill model for each simulation stage on an interactive map.

Predicted positions and shapes of the oil spill at 2, 4, 14, 16, 18, and 19 h after the spill are shown in Fig. 4. The spill point is marked with a cross.

3 The Platform for Monitoring of Seismic Activity Fields

Seismic activity characterizes the intensity of the process of accumulation and release of seismic energy. Long-term seismic activity maps show the spatial distribution of seismic process intensity. Spatial and temporal variations of seismic activity indicate the processes of seismic quietness and activation. It is known that anomalous manifestations of these processes may precede strong earthquakes [11]. Seismic quietness is associated with the process of accumulation of seismic energy in the preparation of an earthquake [12]. The increase in seismic activity reflects changes in the geological environment, which may relate to both the preparation of a strong earthquake and the appearance of post-seismic deformations [13].

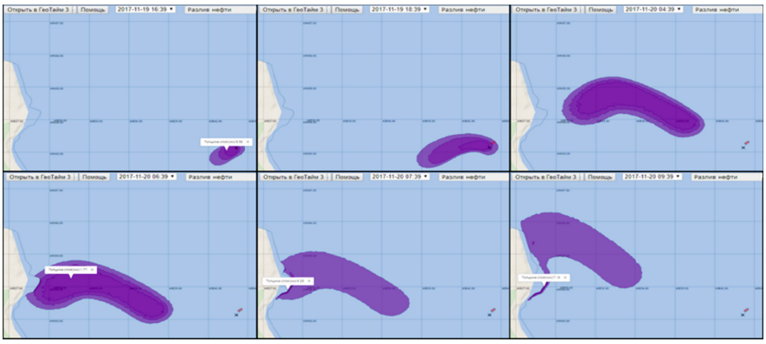


Fig. 4. Predicted positions and shapes of the oil spill at 2, 4, 14, 16, 18, and 19 h after the spill. The spill point is marked with a cross.

The GIS platform represents a seismic process with the help of spatial and spatio-temporal grid-based fields. It consists of two Web GISs. GIS SeismoMap provides automatic loading, processing, and analysis of seismic data, and GIS GeoTime 3 supports a comprehensive analysis of the seismic process.

GIS SeismoMap is realized in a thin client-server architecture. It calculates seismic grid fields, represents the results in the geographical context of the Google Maps API, and it is intended for both professionals and a wide range of Internet users. GIS SeismoMap daily downloads regional catalogs of earthquakes for 24 seismically active regions. GIS calculates two types of fields: spatial (2D) fields characterizing the background regime of the seismic process, and spatio-temporal (3D) fields characterizing the behavior of the seismic process on the test time interval. The background seismic regime is represented by 2D grid fields, the values of which are averaged over the interval of 540 weeks. These include the field of averages of b-value, the field of seismic activity, the field of root-mean-square deviations of seismic activity, the field of quantiles of seismic activity distribution. The test seismic regime is represented by the following 3D grid fields: the field of test seismic activity with a time step of 7 days the values of which are estimated over the intervals 100 day, the field of variation (anomaly) of the test seismic activity, the values of which are equal to the normalized difference of the test and background activity (the field of the seismic activity change points), the field, the values of which are equal to those quantile distribution of background seismicity, those correspond to the values of the test activity, and the field of the ratios of the current values of seismic activity to the background ones. The GIS SeismoMap user interface supports the visualization of 2D and 3D seismic fields and earthquake epicenters, reading of earthquake attributes and plotting of time series graphs in interactively selectable points. The fields of the density of epicenters, seismic activity, and b-value were calculated using the method of local kernel regression [14]. Student's t-statistic is used to assess the field of seismic activity anomalies. It is defined for each grid node as the ratio of the difference of average values of the current (196 days) and background (3650 days) intervals to the standard deviation of this difference. Positive t-statistics values correspond to higher values on the test interval.

Screenshots of three time slices of quantiles of the background seismic activity corresponding to the values of test seismic activity are shown in Fig. 5. The earthquake epicenters are in the interval of 60 days from the date of the slice. The first two slices show a significant decrease in seismic activity in the depicted circle. The third slice shows in this circle an earthquake with a magnitude of $M = 6.5$, the depth of the epicenter $H = 35$ km, which occurred on 05.09.2018.

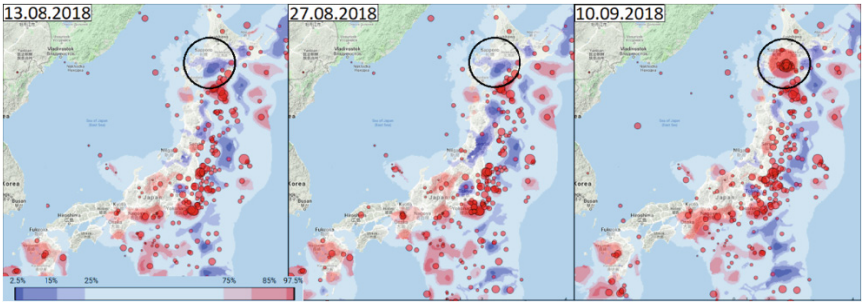


Fig. 5. Time slices of quantiles of the background seismic activity corresponding to the values of test seismic activity. The earthquake epicenters are in the interval of 60 days from the date of the slice.

4 Earthquake Forecasting

4.1 Platform

Earthquakes occur when the stress in rocks of the earth's crust exceeds the ultimate strength. This causes movements within the earth's crust. This phenomenon accompanies the sudden release of energy in the Earth's lithosphere that creates seismic waves, which causes great social and economic damage. Spatio-temporal forecast consists of estimating the location, time and magnitude of the future earthquake [15, 16]. To find the prediction rule, machine learning methods are used, for example, the Bayesian approach assuming that the prediction features are independent [17], the well-known M8 algorithm [18], artificial neural networks [19]. The complexity of the problem is determined mainly by two factors: insufficient knowledge on the mathematical models of earthquake preparation processes, as well as incompleteness of observations about the seismotectonic processes. At the same time, the existing knowledge and available data allow for the use of modern information technologies in order to promote systematic analysis of seismic processes and forecasting the earthquakes [20, 21]. The seismic and geodynamic monitoring data necessary for this purpose are published in real time on a number of remote servers.

The earthquake forecasting platform (http://distcomp.ru/geo/prognosis) is constructed using the same technology as the above-mentioned systems. It consists of two Web GI Systems: GeoPrognosis and GeoTime 3. GIS GeoPrognosis is built in a client-server architecture with a thin client. It operates regularly with constant step Δt. At the moment of forecasting t the system performs the following actions: loads new input

data, converts the input data to grid-based fields, calculates the feature fields, calculate the forecast field, which is a function of the feature fields, calculates the alarm area, tests the forecast outcome for the interval $(t - \Delta t, t)$, computes the forecast statistics, and forms a GIS project for launching the GIS GeoTime 3. Thus, for any forecast point in time t, the work of the platform can be divided into two intervals: the interval for feature field processing and training from start time to t, and the forecast interval from t to $t + \Delta t$. At each step, the training time of the forecast field increases by Δt and the alarm area is corrected.

4.2 Method of the Minimum Area of Alarm

The minimum alarm zone method was proposed in [22]. Here we briefly repeat the main aspects of the method.

Let the spatial and spatio-temporal properties of the seismic process are represented by grid-based fields of features. The feature fields $\mathbf{F}_i$, $i = 1, \ldots, I$, are interpolated to a unified grid with a step $\Delta x \times \Delta y \times \Delta t$. The values of these fields at grid node n correspond to the vectors of the I-dimensional feature space $\mathbf{f}^{(n)} = \{f_i^{(n)}\}$.

A forecast field $\mathbf{\Phi}$ is a function of the feature fields. It is trained using retrospective data that contains a sample set of target earthquakes $q = 1, \ldots, Q$ with the magnitudes $M \geq M^*$ and a set of grid feature fields $\mathbf{F}_i$, $i = 1, \ldots, I$, which describe spatial (quasi-stationary) and spatio-temporal (dynamic) properties of the seismotectonic process.

The method of the minimum area of alarm uses the following data model.

1. Strong earthquakes are preceded by anomalous (improbable) values of the feature fields.
2. The feature fields are constructed in such a way that for anomalies, the values of some of them are close to maximum or minimum. To simplify the explanation in this section, we assume that the anomalies refer only to the *largest* values of the feature fields.
3. If the $\mathbf{f}^{(q)}$ is an anomaly vector, preceding the target event q, then any vector $\mathbf{f}$ with the components $f_i \geq f_i^{(q)}$ for all $i = 1, \ldots, I$ can also precede a similar target event (*monotonicity condition*).

The vectors for which $\mathbf{f} \geq \mathbf{f}^{(q)}$ componentwise we will call *the base vectors of the feature space*. The nodes of the grid of the forecast field with the values $\varphi \geq \varphi^{(q)}$ we will call *the base nodes of the forecast field*.

From the assumptions it follows that the earthquake forecast can be carried out using the simplest threshold decision rule. If the value of the forecast field $\varphi^{(n)} \geq \theta$, then spatio-temporal *alarm cylinders* are created at all base nodes of the forecast field with the values $\varphi \geq \varphi^{(n)}$. The alarm cylinder of the grid node n with the coordinates $(x^{(n)}, y^{(n)}, t^{(n)})$ has center of base in the node $(x^{(n)}, y^{(n)}, t^{(n)})$, the base radius R and the element $[(x^{(n)}, y^{(n)}, t^{(n)}), (x^{(n)}, y^{(n)}, t^{(n)} + \mathrm{T})]$. From this it follows that for a given value of the threshold θ an earthquake with the epicenter coordinates (x^*, y^*, t^*) will be detected if and only if the cylinder with the center of the base (x^*, y^*, t^*), radius R and the element $[(x^*, y^*, t^* - \mathrm{T}), (x^*, y^*, t^*)]$, contains at least one grid node with the value $\varphi^{(n)} \geq \theta$. This cylinder will be called a *precursor cylinder*.

The alarm field detects an earthquake if its epicenter falls within an area consisting of a combination of alarm cylinders (*alarm area*). The quality of the forecast field at threshold θ is determined by two values: (1) the ratio of a number of correctly detected events $Q^*(\theta)$, to all Q target events: $U(\theta) = Q^*(\theta)/Q$ (*probability of detection*) and (2) the ratio of a number of grid nodes $L^*(\theta)$ of the alarm area, to a number of grid nodes L of all area under study: $V(\theta) = L^*(\theta)/L$ (*volume of alarm*).

The basis of our machine learning method is nonparametric new one-class classification algorithm. It has two peculiar properties compared to standard ones [23, 24]. The first relates to the data model. The model postulates two properties of anomalous objects: (1) anomalous objects are unlikely, and some of their properties take values close to the maximum (or minimum) among the sample, and (2) the vectors of the space of features, which are componentwise larger (or smaller) of the vector corresponding to the anomalous object, can also be anomalous objects. This model allows one to build a classification rule from a set of anomalous objects. In this case, normal objects are considered statistically through the probability of detecting anomalous objects by a random forecast. The second difference is that the algorithm allows constructing a forecast function that optimizes the probability of detecting anomalous objects in the training sample if the probability of a random forecast is not more than predetermined.

For training, we have a sample of occurred target earthquakes and a set of feature fields.

At the first step, the algorithm should move from a sample set of target earthquakes to a sample set of target earthquake precursors. A *precursor* of the earthquake q is the vector $\mathbf{f}^{(q)}$ of a feature space which has the minimum volume of alarm $v^{(q)} = L^{(q)}/L$ among all vectors corresponding to the grid nodes of the cylinder of precursor of the event q, where L is the number of all grid nodes of the analyzed area, $L^{(q)}$ is the number of nodes in the grid of the alarm areas generated by the base points of the vector $\mathbf{f}^{(q)}$.

There are three the most important versions of the algorithm.

The first version of the algorithm is to construct the forecast field so that as the threshold θ decreases, the training earthquakes are detected in sequence in which the corresponding alarm volumes increase $v^{(Q)} \le v^{(Q-1)} \le \ldots \le v^{(2)} \le \mathrm{v}^{(1)}$ (this version is selected for testing). The version consists of the following steps.

To create a training sample $\{\mathbf{f}^{(q)}, v^{(q)}\}$, where $\mathbf{f}^{(q)}$ is the precursor of the earthquake q, $v^{(q)}$ is the corresponding alarm volume.

To order the precursors $\mathbf{f}^{(q)}$, $q = 1, \ldots, Q$, by increasing the alarm volume: $v^{(Q)} \le v^{(Q-1)} \le \ldots \le v^{(q)} \le \ldots \le v^{(2)} \le \mathrm{v}^{(1)}$.

To assign to the nodes of the grid of the forecast field $\mathbf{\Phi}$ a value of 0.

To replace the value of 0 by Q at the nodes of the grid of the forecast field, for which the monotonicity condition, $f_i^{(n)} \ge f_i^{(Q)}$ for all $i = 1,\ldots, I$, is satisfied in the feature space; to replace the value of 0 by $Q - 1$ at the nodes of the grid of the forecast field, for which the monotonicity condition, $f_i^{(n)} \ge f_i^{(Q-1)}$ for all $i = 1, \ldots, I$ is satisfied in feature space, and then, successively, in the same way, to replace the values 0 by $Q + 1 - q$.

Obviously, the choice of the order of the earthquake precursors at the 2nd step of the algorithm determines the dependence $U(V)$ obtained from the forecast field. The 2nd version of the algorithm makes it possible to optimize the forecast field so that when

the next target earthquake is detected, the alarm volume increases by a minimum value. To do this, one should arrange the precursors so that when changing from event detection $q + 1$ to event q, the increase in alarm volume is minimal. Here, at each transition from the previously selected event $q + 1$ to q, a small search through the remaining q events is required. The 3rd version of the algorithm allows one to optimize the forecast field so that it detects the maximum number of target earthquakes with a total alarm volume of less than or equal to the predetermined value. In this case, you need to perform a full search of the selected number of events.

4.3 Testing

The method of minimum area of alarm was tested on the platform of automatic earthquake forecast (http://distcomp.ru/geo/prognosis). The system tests the data with a constant step Δt. On each step (at time t) the raster fields of features are computed, the alarm area is trained based on data before the time t, and the system tests for time since t till $t + \Delta t$ if the alarm area covers an epicenter of the target earthquake. Then at time $t + \Delta t$ the training time is increased by Δt, the alarm zone is updated and the test is repeated.

Testing of the forecast method should provide an opportunity to compare different methods of solving the problem on the same indicators of the forecast quality. In this method, we use two quality indicators: the probability of detecting the target events from the test interval $U = Q^*/Q$ and the volume of alarm $V = L^*/L$. The number of target events Q is determined by a set of test samples, the number of target events detected Q^* is determined by the results of the forecast, the analysis area and its size L is selected at the beginning of the test, the size of the alarm zone L^* is determined by the training data.

In the following test experiments, the area of analysis was chosen such that in any circle with radius R = 100 km for the period 1984–1993 at least 300 earthquake epicenters with a magnitude of at least minimum magnitude of completeness in earthquake catalog were recorded. This condition allows one to select a seismically active area for analysis but does not ensure its seismic homogeneity. Therefore, the indicator of the volume of alarm obtained during testing should be considered only in the context of the selected area of analysis. At the same time, the choice of the field of analysis according to a formal rule makes it possible to compare the results of the forecast obtained using various methods and according to different data.

One way to assess the quality of a forecast is to compare a regular forecast obtained by the algorithm under analysis (*regular forecast*) with a random one. We will assume that the forecast is random if the values of the forecast field are selected from a segment in accordance with a uniform distribution. Obviously, for this probabilistic model, the alarm volume V_r is equal to the probability of a random prediction U_r. It follows from this that comparing the probability of a regular forecast U with the probability of a random forecast U_r for the same alarm volumes $V = V_r$ is equivalent to comparing U with the corresponding alarm volume V. If, at the same time, a sample of target events were cleaned of aftershocks and foreshocks, then by proposing the independence of target events and using the binomial distribution model for them, we could build a confidence interval for estimating U.

In the number of articles, the results of a regular forecast are compared with the results of a forecast by a stationary field. In papers [16, 24, 25], the regular forecast is compared with the forecast by the 2D field of seismic activity (or earthquake epicenter density). The result of the comparison makes it possible to evaluate the efficiency of a regular forecast in relation to the forecast by the field F, which is based only on the spatial heterogeneity of the seismic process. Comparison of results can be done in two ways. In one method the probabilities of regular prediction of Q target earthquakes are compared with the results of predicting the same earthquakes by a stationary field F (for example, F is a 2D field of the earthquake epicenter density). Another method uses the Gutenberg-Richter model [26].

Testing was performed for two regions: the Mediterranean and California. The Mediterranean region: 10°–30° E, 34°–47° N. Input data: earthquakes for the period from 27.05.1983 till 14.02.2018 with magnitudes $M \geq 2.7$ and depths of hypocenters $H \leq 160$ km from the International Seismological Centre catalog [27]. Target earthquakes: magnitudes $M \geq 6.0$ and hypocenter depths $H \leq 60$ km. California region: 126°–114° W, 32°–43° N. Input data: earthquakes for the period of 01.01.1983–15.02.2018 with magnitudes $M \geq 2.0$ and depths of hypocenters $H \leq 160$ km from the NEIC USGS catalog [28]. For the forecast, the target earthquakes with magnitudes $M \geq 5.7$ have been selected.

The following 6 fields of features were analyzed for forecasting:

- $\mathbf{F_1}$ is 3D field of the density of all considering earthquakes in the region.
- $\mathbf{F_2}$ is 3D field of mean magnitudes among all considering earthquakes in the region.

The estimation of 3D fields of $\mathbf{F_1}$ and $\mathbf{F_2}$ is performed with the method of local kernel regression. The kernel function for the n-th earthquake has the form $K_n = [\mathrm{ch}^2(r_n/R)^2\ \mathrm{ch}^2(t_n/T)]^{-1}$, where $r_n < R\varepsilon$, $t_n < T\varepsilon$ are the distance and time interval between the n-th epicenter of the earthquake and the node of the 3D grid of the field, $\varepsilon = 2$, $R = 50$ km, $T = 100$ days for $\mathbf{F_1}$ and $R = 100$ km, $T = 730$ days for $\mathbf{F_2}$.

- $\mathbf{F_3}$ is 3D field of negative temporal anomalies of the density of earthquakes.
- $\mathbf{F_4}$ is 3D field of positive temporal anomalies of the density of earthquakes.
- $\mathbf{F_5}$ is 3D field of positive temporal anomalies of mean earthquake magnitude.

To estimate the field of $\mathbf{F_3}$, $\mathbf{F_4}$, $\mathbf{F_5}$, the Student's t-statistic was used, which is defined for each grid node as the ratio of the difference of average values of the current (196 days) and background (3650 days) intervals to the standard deviation of this difference. Positive t-statistics values correspond to higher values on the test interval.

- $\mathbf{F_6}$ is 2D field of the density of earthquake epicenters: kernel smoothing in the interval 1988-2008 the parameter $R = 50$ km.

The grid fields for the Mediterranean were calculated in a grid step $\Delta x \times \Delta y \times \Delta t = 0.2° \times 0.13° \times 49$ days. The forecast field was trained from 1998 until the next step of the forecast after 2008. The radius of the alarm cylinder is $R = 20$ km and the element is $T = 50$ days. Testing is performed in the interval of 2008–2019. There are 11 target earthquakes in the analysis area. We used the method of stepwise selection to find the most informative fields of features. The algorithm selected the $\mathbf{F}_3$ and $\mathbf{F}_6$ fields to construct the alarm field.

We compare the earthquake prediction probabilities obtained using different fields of features in Table 1: U_1 is the forecast probability using the earthquake density field 2D ($\mathbf{F}_6$), U_2 is the probability using 3D field of negative earthquake density anomalies ($\mathbf{F}_3$), $\mu(V)$ is the probability of forecast by 2D field of earthquake epicenters density, obtained using the Gutenberg-Richter model, and U_3 is the probability using $\mathbf{F}_3$ and $\mathbf{F}_6$ fields. We can see that the highest probability of a successful forecast occurs when the fields $\mathbf{F}_3$and $\mathbf{F}_6$ are used together. When $V = 0.2$ ($U_r = 0.2$), the ratios for the prediction probability obtained with $\mathbf{F}_3$ and $\mathbf{F}_6$ fields to the prediction probabilities obtained with 2D earthquake density field ($\mathbf{F}_6$) and for the field calculated using the Gutenberg-Richter ratio, are equal respectively $U_3(0.2)/U_1(0.2) = 0.91/0.64 = 1.49$ and $U_3(0.2)/\boldsymbol{\mu}(0.2) = 0.91/0.41 = 2.2$.

Table 1. Comparison of the probabilities of earthquake forecast for the Mediterranean.

V_{learn}: alarm volumes for learning interval	0.01		0.05		0.1		0.15		0.20	
Test indicators	V_{test}	U_{test}	V_{test}	U_{test}	V_{test}	U_{test}	V_{test}	U_{test}	V_{test}	U_{test}
Field F_6	0.00	0.00	0.03	0.09	0.10	0.09	0.15	0.45	0.20	0.64
Field F_3	0.00	0.00	0.02	0.09	0.20	0.36	0.29	0.45	0.35	0.64
$\mu(V)$: probability for the field F_6 obtained by the model	–	0.00	–	0.1	–	0.26	–	0.35	–	0.41
Fields F_3 and F_6	0.03	0.00	0.08	0.36	0.15	0.45	0.24	0.55	0.32	0.91

Table 1 shows the values of two types of alarm volumes: V_{learn} is the alarm volume received in accordance with the training data, and V_{test} is the alarm volume corresponding to the alarm volume V_{lean}, but observed on the test data. You can see that when testing in almost all cases, except for testing the 2D-field $\mathbf{F}_6$, the volumes of V_{test} are greater than V_{lean}. This is explained by the fact that the number of recorded earthquakes in a region changes over time (Fig. 6). The number of earthquakes is influenced by the development of a seismic network and natural changes in the seismic process. Figure 6 shows that the number of earthquakes increases significantly in the test interval. An increase in the density of earthquake epicenters leads to an increase in the field values of the $\mathbf{F}_3$ function, which ultimately leads to an increase in the volume of anxiety during testing.

The grid fields for California were calculated in a grid step $\Delta x \times \Delta y \times \Delta t = 0.125° \times 0.11° \times 49$ days. The radius of the alarm cylinder is $R = 14$ km and the element is $T = 100$ days. Testing of the earthquake forecast was performed for the interval 2009–2018. There were 9 target earthquakes. The algorithm selected tree fields of features for the construction of the alarm field: $\mathbf{F}_4$, $\mathbf{F}_5$, and $\mathbf{F}_6$.

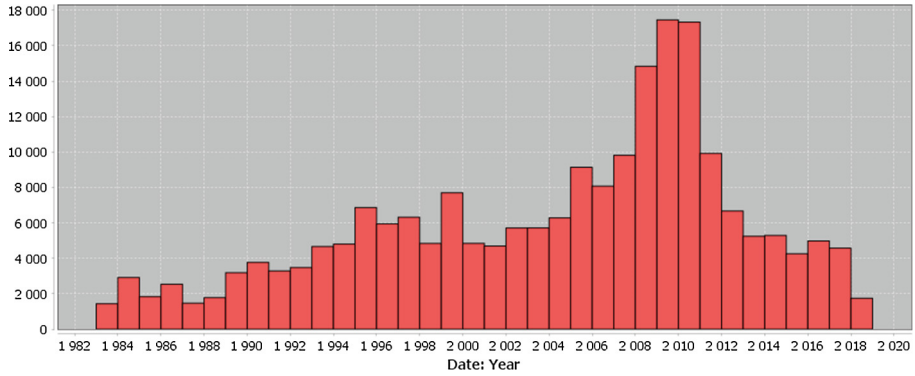


Fig. 6. Histogram of the number of earthquakes with a magnitude greater than 2.7 and a depth of hypocenters less than 160 km.

Table 2 shows the probabilities of earthquake forecast for California.

Table 2. Probabilities of earthquake forecast for California.

Volume of alarm V_{learn}	Volume of alarm V_{test}	Number of correct forecasts	Forecast probability U
0.01	0.01	1	0.11
0.05	0.06	4	0.44
0.1	0.13	4	0.44
0.15	0.13	4	0.44
0.2	0.25	8	0.89

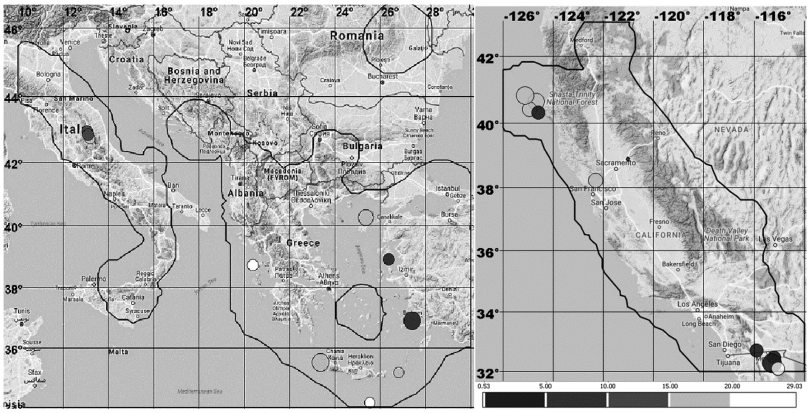


Fig. 7. Areas of analysis and tested target epicenters of earthquakes in 2009–2018: (a) the Mediterranean, (b) California. Shades of grey indicate the minimum volume of alarm with which the epicenter was forecasted. The darkness of grey decreases in accordance with the volume of the alarm: 0.05, 0.1, 0.15, 0.2. White color indicates that an earthquake is not forecasted with an alarm volume of less than 0.2.

Figures 7(a) and (b) show the test results for both regions. They depicted polygons selected as the area of analysis, and circles are the target epicenters of earthquakes in 2009–2018 with $M \geq 6.0$ for Mediterranean and $M \geq 5.7$ for California.

5 Conclusion

We suggested technology designed for systematic analyzing hazardous phenomena in the natural environment and three its implementations: the platform for monitoring of natural and anthropogenic processes in the Arctic, the platform for the analysis of seismic fields and the platform for the automatic forecast of earthquakes. The technology combines methods of automatic processing of monitoring data of primary observations, fairly simple methods of visual data analysis and analytical tools for the comprehensive study of spatio-temporal processes and prediction of hazardous phenomena. The technology has two advantages: (1) simplicity of configuration for a new problem area and (2) a combination of interactive easy analysis supported by intuitive operations and a simplified user interface with a detailed, comprehensive analysis of spatio-temporal processes intended for specialists.

Earthquake prediction platform uses the method of the minimum area of alarm. Test results showed satisfactory forecast quality. In addition, after the launch of the automatic earthquake prediction platform in February 2018, a Greek earthquake of 25.10.2018 with a magnitude of 6.6–6.8 was predicted. The epicenter of the earthquake got into the alarm zone 26.09.2018. At the same time, these results are preliminary. Indeed, testing was conducted only for three regions and for the simplest seismological characteristics. Therefore, we believe that the results show only the effectiveness of the method and the operability of the forecast platform.

References

1. Tsou, M.H.: Integrating Web-based GIS and image processing tools for environmental monitoring and natural resource management. J. Geogr. Syst. **6**(2), 155–174 (2004)
2. Dangermond, J.: Geodesign and GIS–designing our futures. In: Peer Reviewed Proceedings of Digital Landscape Architecture, Anhalt University of Applied Science, Germany (2010)
3. Harder, C.: The ArcGIS Book: 10 Big Ideas about Applying Geography to Your World. Esri Press, Redlands (2015)
4. Kattsov, V.M., Porfir'ev, B.N.: Climatic changes in the Arctic: consequences for the environment and the economy. Arct. Ecol. Econ. **2**(6), 66–78 (2012)
5. Overland, J.E., Wang, M., Walsh, J.E., Stroeve, J.C.: Future Arctic climate changes: adaptation and mitigation time scales. Earth's Future **2**(2), 68–74 (2014)
6. Anisimov, O.A.: Potential feedback of thawing permafrost to the global climate system through methane emission. Environ. Res. Lett. **2**(4) (2007)
7. Zelenina, L.I., Fed'kushova, S.I.: Prediction and consequences of climate fluctuation of the Arctic region. Arktikai Sever **5**, 1–5 (2012). (In Russian)
8. Gitis, V., Derendyaev, A., Weinstock, A.: Web-based GIS technologies for monitoring and analysis of spatio-temporal processes. Int. J. Web Inf. Syst. **12**(1), 102–124 (2016)

9. Gitis, V.G., et al.: Technology of monitoring and analysis of the hydrometeorological situation in the Arctic. J. Commun. Technol. Electron. **63**(6), 691–705 (2018)
10. Zatzepa, S.N., et al.: Modelling of oil spills in sea for planning on guaranteeing of ecological safety by realization of oil and gas plans. Part 1: methodology. ProblemyArktikiiAntarktiki **4** (106), 27–39 (2015)
11. Huang, Q., Sobolev, G.A., Nagao, T.: Characteristics of the seismic quiescence and activation patterns before the M = 7.2 Kobe earthquake, January 17,1995. Tectonophysics **337**(1–2), 99–116 (2001)
12. Kagan, Y.Y., Jackson, D.D.: New seismic gap hypothesis: five years after. J. Geophys. Res. Solid Earth **100**(B3), 3943–3959 (1995)
13. Chen, C.C., et al.: The 1999 Chi-Chi, Taiwan, earthquake as a typical example of seismic activation and quiescence. Geophys. Res. Lett. **32**(22) (2005). https://doi.org/10.1029/2005GL023991
14. Gitis, V.G., Derendyaev, A.B., Pirogov, S.A., Spokoiny, V.G., Yurkov, E.F.: Earthquake prediction using the fields estimated by an adaptive algorithm. In: Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics, Article No. 30 (2017)
15. Keilis-Borok, V., Soloviev, A.A. (eds.): Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. Springer, Heidelberg (2013)
16. Kossobokov, V., Shebalin, P.: Earthquake prediction. In: Keilis-Borok, V.I., Soloviev, A.A. (eds.) Nonlinear Dynamics of the Lithosphere and Earthquake Prediction. SSSYN, pp. 141–207. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-662-05298-3_4
17. Zavyalov, A.D.: Extended Forecast of Earthquakes: Fundamentals, Methodology, Realization. Nauka, Moscow (2006). (in Russian)
18. Kossobokov, V.G.: User manual for M8. In: Healy, J.H., Keilis-Borok, V.I., Lee, W.H.K. (eds.) Algorithms for Earthquake Statistics and Prediction, vol. 6, pp. 167–222 (1997)
19. Bhatia, A., Pasari, S., Mehta, A.: Earthquake forecasting using artificial neural networks. In: International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences (2018)
20. Rhoades, D.A.: Application of the EEPAS model to forecasting earthquakes of moderate magnitude in southern California. Seismol. Res. Lett. **78**(1), 110–115 (2007)
21. Bishop, C.M.: Machine Learning and Pattern Recognition. Information Science and Statistics. Springer, Heidelberg (2006)
22. Gitis, V.G., Derendyaev, A.B.: Earthquake prediction learning using the least alarm method. J. Commun. Technol. Electron. **63**(6), 680–690 (2018)
23. Khan, S.S., Madden, M.G.: A survey of recent trends in one class classification. In: Coyle, L., Freyne, J. (eds.) AICS 2009. LNCS (LNAI), vol. 6206, pp. 188–197. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17080-5_21
24. Molchan, G.M.: Earthquake prediction as a decision-making problem. Pure. appl. Geophys. **149**(1), 233–247 (1997)
25. Kossobokov, V.G., Romashkova, L.L., Keilis-Borok, V.I., Healy, J.H.: Testing earthquake prediction algorithms: statistically significant advance prediction of the largest earthquakes in the Circum-Pacific, 1992–1997. Phys. Earth Planet. Inter. **111**(3–4), 187–196 (1999)
26. Gutenberg, B., Richter, C.: Frequency of earthquakes in California. Bull. Seismol. Soc. Am. **34**(4), 185–188 (1944)
27. International Seismological Centre: Internatl. Seismol. Cent., Thatcham, United Kingdom (2015). http://www.isc.ac.uk
28. Masse, R.P., Needham, R.E.: NEIC-The National Earthquake Information Center. Earthquakes Volcanoes (USGS) **21**(1), 4–44 (1989)