



HAL
open science

Probabilistic Estimates of Hypocenters from the Data of Kamchatka Seismic Network Stations

D. V Droznin, S. Ya. Ya Droznina, S. L Senyukov, D. V Chebrov, Nikolai M. Shapiro, P. N Shebalin

► **To cite this version:**

D. V Droznin, S. Ya. Ya Droznina, S. L Senyukov, D. V Chebrov, Nikolai M. Shapiro, et al.. Probabilistic Estimates of Hypocenters from the Data of Kamchatka Seismic Network Stations. *Fizika zemli / Izvestiya Physics of the Solid Earth*, 2019, 55 (4), pp.677-687. 10.1134/S1069351319040049 . hal-03088558

HAL Id: hal-03088558

<https://hal.science/hal-03088558>

Submitted on 15 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Probabilistic Estimates of Hypocenters from the Data of Kamchatka Seismic Network Stations

D. V. Droznin^a, S. Ya. Droznina^a, S. L. Senyukov^a, D. V. Chebrov^a,
N. M. Shapiro^{b, c}, and P. N. Shebalin^{d, *}

^a*Kamchatka Branch of the Geophysical Survey of the Russian Academy of Sciences,
Petropavlovsk-Kamchatskii, 683006 Russia*

^b*Institut de Physique du Globe de Paris, Paris, France*

^c*Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, 123242 Russia*

^d*Institute of Earthquake Prediction Theory and Mathematical Geophysics,
Russian Academy of Sciences, Moscow, 117997 Russia*

**e-mail: shebalin@mitp.ru*

Received September 10, 2018; revised November 28, 2018; accepted December 3, 2018

Abstract—A new approach is proposed for determining earthquake hypocenters aimed at a more comprehensive characterization of its uncertainty and ambiguity. Application of the new approach to study the seismic focal subduction zones and volcanic seismicity is discussed by the example of the data of the Kamchatka Branch of the Geophysical Survey of the Russian Academy of Sciences.

Keywords: hypocenter, one-dimensional traveltime curve, posterior distribution of hypocenter estimates, residual, distribution model of residuals, correlation of residuals, systematic errors of coordinate determination

DOI: 10.1134/S1069351319040049

INTRODUCTION

Earthquake location (determination of hypocentral coordinates and origin time) is one of the key problems in seismology. The earthquake catalogs obtained as a result of this procedure are the initial data for solving many problems of tectonics and geodynamics, as well as for evaluation of seismic regimes and seismic hazard assessment. The methods for analyzing the seismicity are constantly developing and improving. Here, one of the main lines of development is to refine the resolution of the analysis in time and space. When it comes to the application of modern analysis methods to the real data, the accuracy of the results is frequently limited not by the methods themselves but by the incompleteness of the catalogs and by the errors of hypocenter location. Therefore, for the adequate analysis of seismicity, it is vital to reliably characterize the quality of the used catalogs overall and to take into account the quality of each individual solution.

This work is aimed at the development of a method that allows reliable evaluation of the determination quality of hypocentral coordinates and origin time. The uncertainty in the hypocenter location can significantly increase for the earthquakes recorded by only a few stations or in the case when

most stations are located far from the earthquake's epicenter and do not provide reasonable azimuthal coverage of the epicenter. The first situation is typical of the weak volcanic earthquakes and for moderate tectonic earthquakes on the periphery of the responsibility zone of the Kamchatka seismic network (Chebrov et al., 2018). The second situation is often observed with surface earthquakes in the subduction zone. Thus, the problem of accurately estimating the quality of the hypocenter's location is pertinent for most earthquakes detected in the active seismic and volcanic regions of the Far East of the Russian Federation.

Practically all the methods for hypocenter location are based on the minimization of the residuals or differences between the calculated and actual arrival times of seismic waves at the recording stations. As the objective function, the sum of squared residuals is used. The nonlinear dependence of traveltimes on the distance between the epicenter and the station makes this optimization problem fundamentally nonlinear and, hence, the situations are possible when nonlinearity leads to the substantial ambiguity of the solution up to the existence of more than one minimum of the objective function. Efficient methods for earthquake location appeared in the 1960s through the application

of computers. As discussed in (Husen and Hardebeck, 2010), at the first steps, because of the limited computing power, the optimization problem was solved by iterative methods. The HYPO71 program (Lee and Lahr, 1975) or the method that has long been used in Kamchatka (Gusev, 1979) are well known examples of this approach.

The iterative methods have the advantage that they are computationally fast; however, in the case of a strong nonlinearity of the problem, they yield distorted estimates of the solution quality. Sometimes, as a measure of the solution's quality, some authors misleadingly use the root mean square residual of the arrival times of the waves assuming that this quantity should be as small as possible for a good solution. The falsity of this approach is evident from a simple consideration that the close-to-zero value of the residual is obtained with a very small number of stations, whereas the quality of the solution should increase with a large number of stations. Therefore, a more correct approach is to use the so-called error ellipses; however, their calculation based on iterative methods is frequently challenging.

Therefore, recently it has become increasingly popular to use statistical methods for earthquake location (Sambridge and Kennet, 1986; Lomax et al., 2000; Minson and Lee, 2014). These methods are not limited to searching for the extremum of the objective function but evaluate its distribution in the entire space of the possible hypocenter locations and origin times. Thus, the domain of reliable solutions is identified. For this, a combination of the existing models of traveltimes of the waves and the corresponding errors in the quality of the a priori information are frequently used for calculating the distributions of the posterior probability of the locations of specific hypocenters based on the specific data on P - and S -wave arrival times at specific stations using Bayes' theorem. Thus, instead of a single point of the most probable location of the hypocenter, this approach yields clouds of points for each of which the probability is calculated to be the locus of the real hypocenter. In this work, we demonstrate the use of Bayes' theorem for locating the earthquakes in the Kamchatka subduction zone and in the region of active Kamchatka volcanoes based on the data of the Kamchatka Branch of the Geophysical Survey of the Russian Academy of Sciences (KB GS RAS). After describing the basic theory underlying the method and the practical implementation of the algorithm, we discuss how the non-optimal station layout may lead to ambiguity in the solutions for different types of earthquakes.

INITIAL DATA

The information base for our study is the results of processing the earthquakes from 2010 to 2018 based on

the data from the system of detailed seismological observations of the KB GS RAS (Chebrov et al., 2013). The information resources <http://www.emsd.ru/ts/all.php> (Chebrov et al., 2010) and Seismological Data Information System (SDIS) (<http://sdis.emsd.ru/main.php>) (Chebrova et al., 2015) have been created at the KB GS RAS for storing and accessing the information on the earthquakes. In this work, we used the arrival times of the P - and S -waves and the earthquake catalog for the study period compiled from the calculations by the DIMAS program (Droznin and Droznina, 2010), which includes the hypocenter determinations for more than 150 000 tectonic and volcanic events. The theoretical traveltimes of seismic waves were calculated from the P - and S -wave regional traveltime curve from the software package (Mel'nikov, 1990) which is used in the laboratory of seismic and volcanic activity research of the KB GS RAS to determine the parameters of the tectonic earthquakes and, due to the absence of the refined local velocity models, volcanic earthquakes. The earthquakes of the North and Avacha groups of volcanoes for which local traveltime curves are used (Senyukov, 2006) are the exception.

CALCULATION METHOD

The main purpose of this work is to represent the calculated hypocenters of Kamchatka's tectonic and volcanic earthquakes in the form of a three-dimensional probability distribution of the hypocenter locations instead of the traditional form. As a probability distribution, we use the posterior four-dimensional probability distribution of the hypocenter location and origin time calculated by the Bayes formula based on the a priori constraints on the probable solution, the actual arrival time data (in our case, for P - and S -waves), the physical model of seismic velocities (traveltime curve), and the distribution model of the residuals between the actual and calculated traveltime (Tarantola and Valette, 1982). In the general form, this solution (Minson, Lee, 2014) is

$$p(\varphi, \lambda, h, t | \mathbf{d}) = F p(\mathbf{d} | \varphi, \lambda, h, t) p(\varphi, \lambda, h, t), \quad (1)$$

where \mathbf{d} is the data vector of all the recorded arrival times of seismic P - and S -waves at all stations; $p(\varphi, \lambda, h, t | \mathbf{d})$ is the sought posterior probability of the estimates of the hypocenter's location and origin time; $p(\varphi, \lambda, h, t)$ is the a priori distribution of the probable solution; $p(\mathbf{d} | \varphi, \lambda, h, t)$ is the probability of realization of the arrival times with a given hypocenter location and origin time and with a certain error model, which is referred to as the likelihood; and F is the normalization factor equal to the inverse quantity to the integral of $p(\mathbf{d} | \varphi, \lambda, h, t) p(\varphi, \lambda, h, t)$ over all possible (φ, λ, h, t)

values. We note that the Bayes theorem allows the likelihood function of the data for the specific hypocenter and origin time to be converted into the probability of the hypocenter location and origin time under the specific data on the arrival times.

Formula (1) gives the four-dimensional distribution of the solution including the origin time. The probability distribution of the hypocentral coordinates alone or the origin time alone can be obtained by integrating over all the possible origin times or over all the possible locations of the hypocenter, respectively. In Bayes' theory, these distributions are referred to as marginal distributions. The obtained posterior distributions of the origin times typically have a small scatter compared to the times between the separate events; therefore, in this work we focus on the posterior distributions of the coordinates of the hypocenters.

The likelihood function $p(\mathbf{d}|\varphi, \lambda, h, t)$ is typically assumed (Tarantola and Valette, 1982; Minson and Lee, 2014) in the form of a multivariate normal distribution:

$$p(\mathbf{d}|\varphi, \lambda, h, t) = \frac{1}{(2\pi)^{N_d} \sqrt{C_\chi}} e^{-\frac{1}{2}(\mathbf{d}-\hat{\mathbf{d}})^T C_\chi^{-1} (\mathbf{d}-\hat{\mathbf{d}})}, \quad (2)$$

where N_d is the dimension of data vector \mathbf{d} (number of detected arrivals); $\hat{\mathbf{d}}$ is the vector of the calculated arrival times depending on parameters (θ, φ, h, t) ; and, respectively, $\mathbf{d} - \hat{\mathbf{d}}$ is the vector of residuals, C_χ is the covariance matrix of the residuals, and T denotes transposition of a vector. If we assume the independence of the residuals and the total error value, the covariance matrix is the diagonal with the entries equal to the variance of the residuals. The solution corresponding to the maximum of likelihood (2) is in this case equivalent to the least-square solution of the minimization of the total mean square residual of wave arrivals (Tarantola and Valette, 1982). For locating the hypocenters, different modifications of the probabilistic approach also use the likelihood in form (2). At the same time, the problem of determining the covariance matrix C_χ is solved in different ways. Tarantola and Valette (1982) postulate the correlation between the arrival times at different stations which depends on the inter-station distance; however, the theoretical error in determining the wave traveltime for different stations is assumed to be identical and independent of hypocentral distance. This error is assumed to be 0.2 s. In (Minson and Lee, 2014), which is devoted to hypocenter determinations from historical data with the use of the more advanced traveltime curve, it is assumed that the traveltime calculation errors are negligible compared

to determination accuracy of arrival times. Lomax et al. (2000) pointed out the complexity of the realistic determination of the covariance matrix. At the same time, it is evident that the size of the error ellipse of the hypocenter's determination drastically depends at least on the error values (the diagonal elements of the covariance matrix C_χ). Therefore, for obtaining an adequate idea of the hypocenter's determination error, it is required to have realistic determinations of the covariance matrix C_χ .

In our work, we assume the dependence of the diagonal matrix elements on the distance between the hypocenter and the corresponding station. We also assume that the off-diagonal elements scale with the values of the corresponding diagonal terms; however, also the dependence on the distance to the hypocenter and/or the dependence between the stations of the corresponding correlation coefficients is allowed. Constructing the covariance matrix is complicated by the fact that for some stations, the arrival times of the S - or P -wave can be absent. Below we present the general form of the residual vector and the covariance matrix C_χ under the described assumptions. The lower index corresponds to a certain station and the upper index indicates the P - or S -wave. Initially, there are N^{PS} pairs of matrix rows and columns and the components of the residual vector corresponding to the stations for which there are the arrivals of both the P - and S -waves; then, there are N^P elements corresponding to the stations for which there are only the arrivals of P -waves; and finally, there are N^S elements corresponding to the stations for which there are only where only the arrivals of S -waves. The total number of wave arrivals is $2N^{PS} + N^P + N^S$ and the total number of stations is $N^{PS} + N^P + N^S$. The residual vector is composed of the differences between the actual and calculated wave arrival times. The calculated time is determined as $t + \delta t_p$ for P -wave or $t + \delta t_s$ for the S -wave where the traveltimes δt_p and δt_s of P - and S -waves from point (φ, λ, h) to the station are calculated from the time–distance (traveltime) curve.

$$\mathbf{d} - \hat{\mathbf{d}} = \left\{ t_p^1 - \delta t_p^1, t_s^1 - \delta t_s^1, t_p^2 - \delta t_p^2, t_s^2 - \delta t_s^2, \dots, \right. \\ \left. t_p^{N^{SP}} - \delta t_p^{N^{SP}}, t_s^{N^{SP}} - \delta t_s^{N^{SP}}, t_p^{N^{SP}+1} - \delta t_p^{N^{SP}+1}, \dots, \right. \\ \left. t_p^{N^{SP}+N^P} - \delta t_p^{N^{SP}+N^P}, t_s^{N^{SP}+N^P+1} - \delta t_s^{N^{SP}+N^P+1}, \dots, \right. \\ \left. t_s^{N^{SP}+N^P+N^S} - \delta t_s^{N^{SP}+N^P+N^S} \right\}, \quad (3)$$

$$\begin{aligned}
 & C_\chi \\
 & \left[\begin{array}{cccc}
 \sigma_p^2(\Delta_1) & \sigma_p(\Delta_1)\sigma_s(\Delta_1)\rho_{PS}(\Delta_1) & \sigma_p(\Delta_1)\sigma_p(\Delta_2)r_{PP}(\delta_{1,2}, \Delta_{1,2}) & \dots \\
 \sigma_p(\Delta_1)\rho_{PS}(\Delta_1) & \sigma_s^2(\Delta_1) & \sigma_p(\Delta_2)\sigma_s(\Delta_1)r_{SP}(\delta_{1,2}, \Delta_{1,2}) & \dots \\
 \sigma_p(\Delta_2)r_{PP}(\delta_{1,2}, \Delta_{1,2}) & \sigma_p(\Delta_1)\sigma_s(\Delta_1)r_{PS}(\delta_{1,2}, \Delta_{1,2}) & \sigma_p^2(\Delta_2) & \dots \\
 \sigma_p(\Delta_1)\sigma_s(\Delta_2)r_{PS}(\delta_{1,2}, \Delta_{1,2}) & \sigma_s(\Delta_1)\sigma_s(\Delta_2)r_{SS}(\delta_{1,2}, \Delta_{1,2}) & \sigma_s^2(\Delta_2) & \dots \\
 \dots & \dots & \dots & \dots \\
 \sigma_p(\Delta_1)\sigma_s(\Delta_{N^{PS}})r_{PS}(\delta_{1,N^{PS}}, \Delta_{1,N^{PS}}) & \sigma_s(\Delta_1)\sigma_s(\Delta_{N^{PS}})r_{SS}(\delta_{1,N^{PS}}, \Delta_{1,N^{PS}}) & \dots & \dots \\
 \sigma_p(\Delta_1)\sigma_p(\Delta_{N^{PS}+1})r_{PP}(\delta_{1,N^{PS}+1}, \Delta_{1,N^{PS}+1}) & \sigma_p(\Delta_1)\sigma_p(\Delta_{N^{PS}+1})r_{PP}(\delta_{1,N^{PS}+1}, \Delta_{1,N^{PS}+1}) & \dots & \dots \\
 \dots & \dots & \dots & \dots \\
 \sigma_p(\Delta_1)\sigma_p(\Delta_{N^{PS}+N^P})r_{PP}(\delta_{1,N^{PS}+N^P}, \Delta_{1,N^{PS}+N^P}) & \sigma_p(\Delta_2)\sigma_p(\Delta_{N^{PS}+N^P})r_{PP}(\delta_{2,N^{PS}+N^P}, \Delta_{2,N^{PS}+N^P}) & \dots & \dots \\
 \sigma_s(\Delta_1)\sigma_s(\Delta_{N^{PS}+N^P+1})r_{SS}(\delta_{1,N^{PS}+N^P+1}, \Delta_{1,N^{PS}+N^P+1}) & \sigma_s(\Delta_2)\sigma_s(\Delta_{N^{PS}+N^P+1})r_{SS}(\delta_{2,N^{PS}+N^P+1}, \Delta_{2,N^{PS}+N^P+1}) & \dots & \dots \\
 \dots & \dots & \dots & \dots \\
 \sigma_s(\Delta_1)\sigma_s(\Delta_{N^{PS}+N^P+N^S})r_{SS}(\delta_{1,N^{PS}+N^P+N^S}, \Delta_{1,N^{PS}+N^P+N^S}) & \sigma_s(\Delta_1)\sigma_s(\Delta_{N^{PS}+N^P+N^S})r_{SS}(\delta_{1,N^{PS}+N^P+N^S}, \Delta_{1,N^{PS}+N^P+N^S}) & \dots & \dots
 \end{array} \right] \quad (4)
 \end{aligned}$$

We assume here that the values of the diagonal matrix terms only depend on distance Δ_i from point (φ, λ, h) to the i th station. For computational convenience, as a measure of distance, here we use the calculated P -wave traveltime δt_i . ρ_{SP} — is the coefficient of correlation between the residuals of different types of waves. Here, we also assume for a while that this coefficient can also depend on distance. The covariance terms of the matrix that correspond to each of two different stations i and j are in scale with the square root of the corresponding diagonal terms. The respective coefficients r_{PP} , r_{SS} , and r_{PS} , which have the sense of correlation coefficients, are assumed to depend on both the distance between these stations $\delta_{i,j}$ and on the mean distance $\Delta_{i,j} = 1/2(\Delta_i + \Delta_j)$. The analysis of the dependence of parameters σ_p, σ_s , and ρ_{PS} on the hypocentral distance and the dependence of the parameters r_{PP}, r_{SS} , and r_{PS} on the mean hypocentral distance and on inter-station distance, as well as the construction of the model of residuals, are addressed in the next section.

In the case when it is possible to neglect the terms of the covariance matrix that correspond to the different stations and to only consider the correlation between the residuals of the P - and S - waves at the same station, matrix C_χ in Eq. (2) takes on the piecewise-diagonal form which allows us to obtain the explicit expression for the inverse matrix, and Eq. (1) can then be represented in the linear form:

$$\begin{aligned}
 p(\varphi, \lambda, h, t) = & F_2 \prod_{i=1}^{N^{PS}} \frac{1}{2\pi\sigma_p\sigma_s\sqrt{(1-\rho_{PS}^2)}} \\
 \times e^{-\frac{1}{1-\rho_{PS}^2} \left\{ \frac{\delta p_i^2}{2\sigma_p^2} - \rho_{PS} \frac{\delta p_i \delta s_i + \delta s_i^2}{\sigma_p\sigma_s + 2\sigma_s^2} \right\}} & \prod_{j=1}^{N^P} \frac{1}{\sqrt{2\pi}\sigma_p} e^{-\frac{\delta p_j^2}{2\sigma_p^2}} \prod_{k=1}^{N^S} \frac{1}{\sqrt{2\pi}\sigma_s} e^{-\frac{\delta s_k^2}{2\sigma_s^2}}. \quad (5)
 \end{aligned}$$

Here, similarly to (1), F_2 is the normalization coefficient. Quantities σ_p, σ_s , and ρ_{PS} can depend on the mutual position of a given station and the point with coordinates (φ, λ, h, t) .

When using the Bayes theorem, as the a priori distribution, it is most common to consider the so-called non-informative prior distribution—the uniform distribution of a parameter within the given limits. We only know the limits within which the values of the parameter should lie but do not know which of its values are more probable. In the hypocenter location problem, the limits for the position of a hypocenter are typically known. In this work, the problem is solved iteratively in two stages because the computer’s random access memory is only sufficient for a limited domain size under a given resolution of the problem’s solution. In other words, the solutions for large domains can only be found on a spatiotemporal grid with a large spacing. We use here the idea of the approach that is referred to in the English-language literature as the *importance sampling method* (Lomax et al., 2000).

For speeding up the computations, at each iteration step, we use a $41 \times 41 \times 41 \times 41$ numerical grid. At the first iteration step of the first stage, the boundaries of the uniform prior distribution are specified in the form of a parallelepiped whose horizontal projection somewhat amply describes the boundaries of responsibility of the KB GS RAS (Levina, 2008; Chebrov, 2018) and whose depths span the interval from -5 to 700 km. The prior time limits for the interval of the values are specified as the minimum and maximum possible for a given point with the theoretical traveltime taken into account. At each subsequent step of the first stage, the region is halved with a center at the point of the maximum $p(\varphi, \lambda, h, t)$ until a given

accuracy (the step value) is achieved. At the second stage, the region is gradually enlarged by shifting the grid with the preservation of the step size in all the parameters, and the posterior probability $p(\varphi, \lambda, h, t)$ is calculated. The process is stopped when the probability of a hypocenter's location within a current grid domain exceeds 99%. This procedure guarantees that the integral of the non-normalized posterior probability within the domain is at least 100 times as large as the similar integral beyond the domain if it were possible to carry out calculations with the same resolution within the initial limits.

A convenient representation of the results of calculating the posterior distribution of hypocenter location is in the form of the lines of a probability level which contour the spatial domains where the probability density integral at the points ordered decrementally in probability density is equal to the probability level, say, of 90%.

THE DISTRIBUTION MODEL OF RESIDUALS OF P - AND S -WAVES

The residuals' errors and the covariance matrix C_χ overall have two additive components: the theoretical component C_T and the data or real component C_I , $C_\chi = C_T + C_I$ (Tarantola, 1982; Gusev, 1979). The theoretical component is due to the inaccuracy of the traveltime curve and inhomogeneity of the medium. The real component is due to the inaccuracy of the data. In the present-day conditions when the coordinates of the seismic stations are known with a very high degree of accuracy and the timing of seismograms is carried out by synchronization with the reference time signal of GPS satellites, the real component can only be due to the inaccurate picking of the arrival times. In practice, it is very difficult to separate the components; therefore, in this work we opted to study the empirical regularities using extensive factual data. From the Kamchatka catalog, we selected the earthquakes with the energy class $K^F \geq 7$ (the energy class based on the S -wave determined from S.A. Fedotov's nomogram (Fedotov, 1972)) for which the station database has at least five arrivals of P -waves and at least five arrivals of S -waves. We compiled the database of the optimal residuals—the differences between the calculated (with the hypocenter location and origin time according to the catalog data) and real arrival times of seismic waves δ^p and δ^s . The database includes more than 10^6 values of the residuals.

The purpose of studying the residuals in this section is to determine the dependence of parameters σ_p, σ_s , and ρ_{PS} on the hypocentral distance (the distance from the supposed hypocenter with the coordinates (φ, λ, h) to the respective station) and the depen-

dence of parameters r_{PP}, r_{SS} , and r_{PS} on the mean hypocentral distance and on the distance between the corresponding stations.

The diagonal elements of the covariance matrix C_χ may depend on the relative position of the earthquake source and the station (Gusev, 1979). A convenient measure of this relative position in the hypocenter location procedure is in terms of the calculated travel-time δt_p of the P -wave since this quantity is directly used in the calculations. As shown in Fig. 1, the scatter of the residuals increases with the increase of δt_p .

The residuals of the P - and S -waves at the same station cannot be regarded as independent. In the calculations of the hypocenter, this effect may prove very significant, especially in the cases when at some of the stations only one of the P - or S -wave arrivals is detected. In any case, if the correlation of the residuals is neglected, the accuracy of hypocenter determination is artificially overrated. In principle, the degree of correlation between the residuals can vary from station-to-station and depending on the location of the earthquake source; moreover, both the theoretical and real components of the covariance may contribute to this effect. Here we assume that these details are insignificant compared to the averaged model of the correlation between the residuals of P - and S -waves. Figure 2 shows the dependence of the residuals for the S -wave as a function of the residual of the P -wave for the same station and the same earthquake, constructed over the entire database of the residuals described above. The correlation coefficient ρ_{PS} is 0.55.

For studying the dependence of the errors in σ_p and σ_s and in the correlation coefficient ρ_{PS} on δt_p , we ordered the database incrementally in t_g^p and then calculated the root mean square deviations and the sample correlation coefficient, as well as the mean δt_p for a group of the data using the moving windows that include 1000 pairs of P - and S -wave residuals at the same station from the same earthquake. The respective dependences are shown on the loglog scale in Fig. 3.

As seen from the figure, the dependences of the errors in σ_p and σ_s are fairly closely approximated by the straight lines on the loglog scale at $\delta t_p \geq 10$ s. The deviation from these linear dependences at small σ_p and σ_s towards high values can be interpreted as the relative growth of the contribution of the real component in the covariance matrix: it is likely that the picking errors of P - and S -wave arrival times are about 0.3 and 0.5 s, respectively. The coefficient of correlation ρ_{PS} of the P - and S -residuals, as it turned out (Fig. 3c), slightly depends on t_g^p . Therefore, as the final model describing the dependence of errors in σ_p and σ_s on

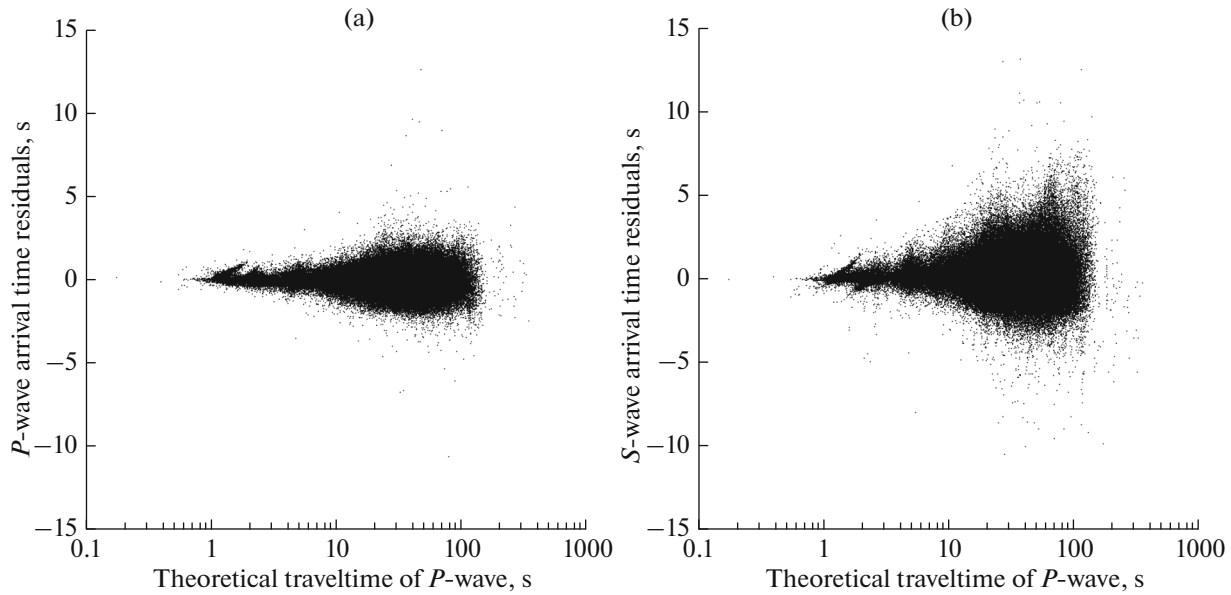


Fig. 1. Dependence of difference of calculated and real arrival times of (a) P - and (b) S -waves on calculated P -wave traveltime δt_p over all stations and all earthquakes with energy class $K^F \geq 7$ of 2010–2018.

δt_p and the model of ρ_{PS} , we accepted the following relations:

$$\begin{aligned} \sigma_p &= \max\{0.3, 0.14\delta t_p^{0.42}\}, \\ \sigma_s &= \max\{0.5, 0.16\delta t_p^{0.53}\}, \\ \rho_{PS} &= 0.55. \end{aligned} \quad (6)$$

In the analysis of the coefficients corresponding to different station pairs, two different situations should

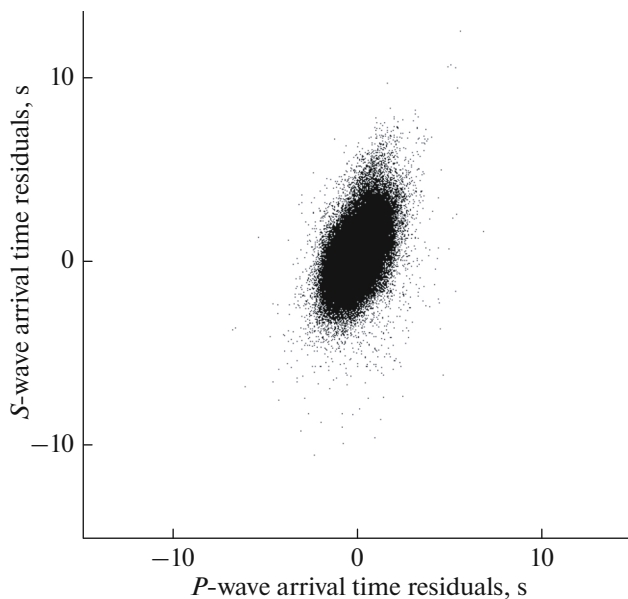


Fig. 2. Difference of calculated and real times of arrival of S -wave versus corresponding difference of P -wave over all stations and all earthquakes with class $K^F \geq 7$, 2010–2018.

be considered. The correlation of the residuals at stations close to each other should be small if the distance from both stations to the hypocenter is small. At large hypocentral distances, in contrast, the correlation of stations very close to each other should be close to unity. However, in this case, the real component of the errors can reduce this correlation because of the relatively significant picking error of the arrival times. In this work, we model the first situation by the condition ($\delta \geq \Delta$), where δ is the distance between a pair of stations and Δ is the mean distance from the hypocenter to each of the two stations. In this case, we assume zero values for coefficients r_{pp} , r_{ss} , and r_{ps} . For the second situation ($\delta < \Delta$), in the moving window in δ with a width of 0.1 degree, we estimate the sample covariance for the residuals normalized to the corresponding errors determined by formula (6). Due to the normalization, coefficients r_{pp} , r_{ss} , and r_{ps} are dimensionless and have the sense of the correlation coefficients. Here, the normalization takes into account the increase of the errors and, accordingly, all the covariance terms with the growth of hypocentral distance. The results are shown in Fig. 4.

Eventually, we accepted the following relations as the model:

$$\begin{aligned} r_{pp}(\delta) = r_{ss}(\delta) &= \begin{cases} 0.55e^{-\delta/0.15}, & \delta < \Delta \\ 0, & \delta \geq \Delta \end{cases}, \\ r_{ps}(\delta) &= \begin{cases} 0.3e^{-\delta/0.15}, & \delta < \Delta \\ 0, & \delta \geq \Delta \end{cases}. \end{aligned} \quad (7)$$

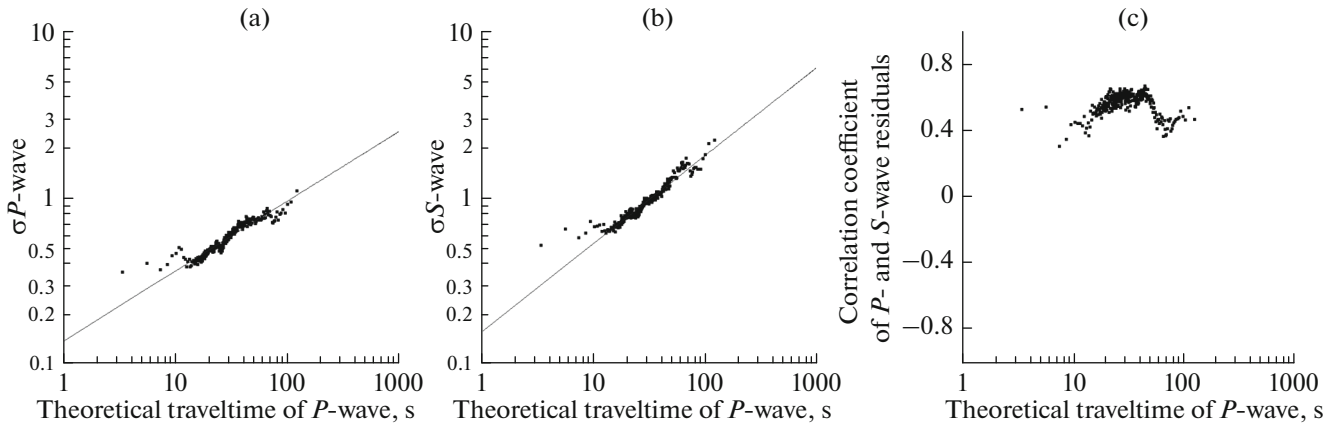


Fig. 3. Dependence of errors in (a) σ_p and (b) σ_s and (c) coefficient of correlation of residuals ρ_{ps} on calculated traveltime δt_p of P -wave propagation from hypocenter to station. Straight lines show linear approximation of dependences: $\sigma_p(\delta t_p) = 0.14\delta t_p^{0.42}$ and $\sigma_s(\delta t_p) = 0.16\delta t_p^{0.53}$.

EXAMPLES OF POSTERIOR DISTRIBUTIONS OF HYPOCENTER LOCATIONS

In this section, we present four examples illustrating the application of the developed technique for constructing the confidence areas of determination of the hypocenter and origin time.

Figure 5 shows the solution for the Okhotsk Sea earthquake of May 24, 2013, $M_w = 8.3$ (Sil'nye..., 2014). Due to the fact that the earthquake was recorded by many stations whose locations provided good azimuthal coverage of the epicenter, we managed to obtain a fairly accurate estimate despite the large depth of the earthquake source. The linear size of the 95% confidence area is about 10 km horizontally and about 30 km along the depth. A similarly accurate solution is also obtained for the shallow Blizhne-Aleutskoe (Near Islands Aleutian or Near Aleutian) earthquake of July 17, 2017, $M_w = 7.8$ (Fig. 6). The linear size of the 95% confidence interval is about 5 km for the epicenter and about 30 km for the depth of the source. We note that these estimates characterize the highest possible accuracy of hypocenter location for the earthquakes in this regions with the existing network of seismic stations.

If an earthquake is recorded by only a few stations and if these stations provide poor azimuthal coverage of the epicenter, any point estimate of the hypocenter may prove to be ambiguous because of the large extent of the area of the optimal solution. Figure 7 shows the confident area of the solution determined by the developed technique for one of the aftershocks of the Ipyrskoe earthquake of March 13, 2013, $M_w = 5.8$ (Sil'nye..., 2014). The linear shape of the area of solution's maximum explains the previously revealed artifact of the linearly shaped cloud of aftershocks from this earthquake (Saltykov and Droznina, 2014). This

configuration of the confidence area of the solution is due to the unfavorable location of the stations around the epicenter. The point solutions for the different aftershocks are randomly spread along this area.

A poor station layout can lead to a solution with two well-expressed maxima. This especially concerns

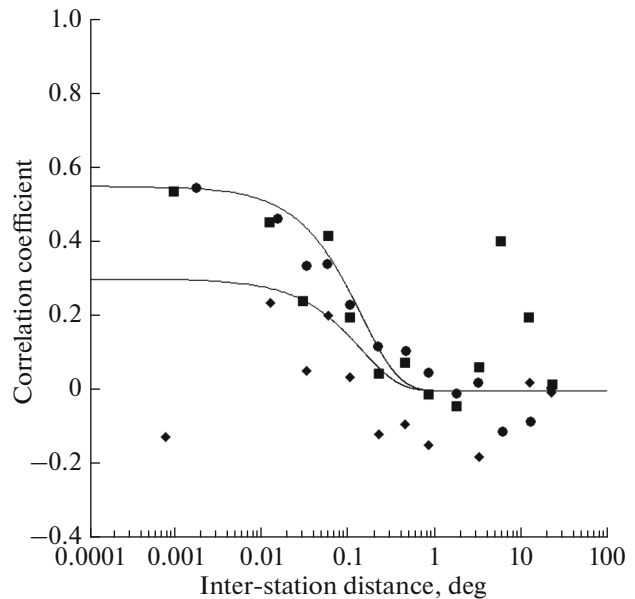


Fig. 4. Dependence of estimates of coefficients r_{pp} , r_{ss} and r_{ps} on inter-station distance δ (degrees) under condition ($\delta < \Delta$). Circles, squares, and diamonds mark estimates of parameters r_{pp} , r_{ss} and r_{ps} , respectively. Solid and dashed lines indicate exponential approximations of dependences $r_{pp}(\delta|\delta < \Delta) = r_{ss}(\delta|\delta < \Delta) = 0.55e^{-\delta/0.15}$ and $r_{ps}(\delta|\delta < \Delta) = 0.3e^{-\delta/0.15}$, respectively.

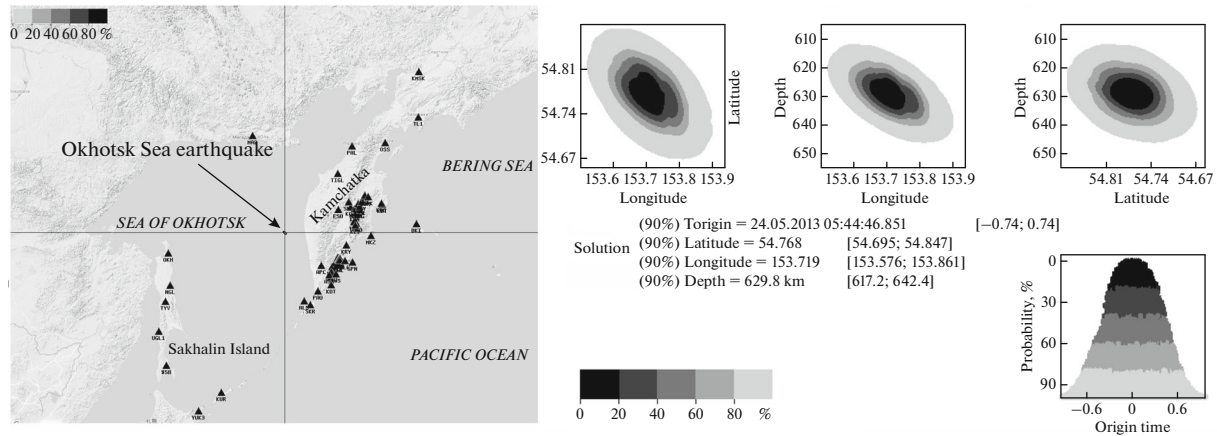


Fig. 5. Posterior distributions of probabilistic estimates of hypocenter location and origin time for Okhotsk Sea earthquake of May 24, 2013, $M_W = 8.3$. In map (left), triangles indicate stations whose data used for solution. Solution area is marked by intersection of lines and by level lines (solution area is very small on map scale). In diagram (right), lines of posterior probability level are shown for three projections as well as level lines of marginal distribution for origin time. Level lines from 0 to 95% are shown in gradations of gray.

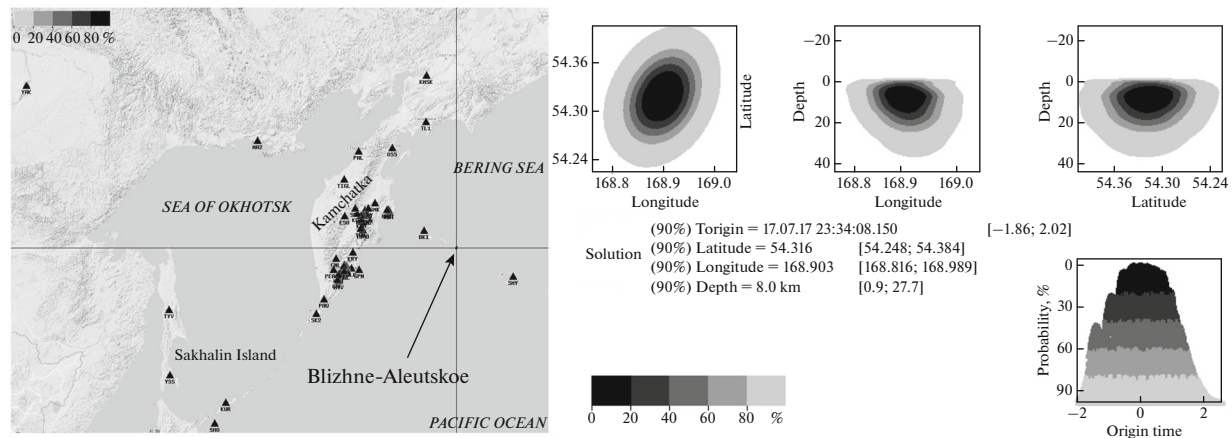


Fig. 6. Posterior distributions of probabilistic estimates of hypocenter location and origin time for Blizhne-Aleutskoe earthquake of July 17, 2017, $M_W = 7.8$. Designations are in Fig. 5.

recording the events on volcanoes. Figure 8 shows an example of a solution with two maxima. The electronic supplement contains also the other examples showing how the lack of seismic stations may sometimes result in obtaining very ambiguous solutions of the hypocenter's location.

DISCUSSION OF RESULTS

The point estimate of the hypocenter (as clearly demonstrated in Fig. 7) may prove to be just one of many almost equally probable candidate solutions. These situations are quite frequent in Kamchatka and the Kuril Islands where the seismic stations are mainly located on one side of the hypocenter. This means that the spatial density of the distribution of hypocenters routinely calculated from their point estimates can drastically differ from the real hypocenters' density. Thus, the estimates of the seismic hazard, the geometry of the Benioff zone, etc., based on the data of stan-

dard earthquake catalogs can be substantially distorted. The posterior probability estimate based on the Bayesian theorem has the advantage that it allows obtaining realistic confidence areas of parameter estimates (Holschneider, 2012). If applied to the analysis of the hypocenters' distribution, this approach, for example, will make it possible to detect systematic distortions of the spatial configuration of the seismically active regions. At the same time, it should be borne in mind that the posterior probability is a conditional probability one of the conditions for which is a specific velocity model. Therefore, this probability cannot be unquestionably considered as the probability of the true position of the hypocenter.

In order to take full advantage of the possibility of detecting systematic errors of point solutions, a new type of electronic catalog of the earthquakes needs to be created in which, in addition to the maximum-likelihood point solution, the hypocenter will also be represented in terms of the spatial distribution of the probability of its

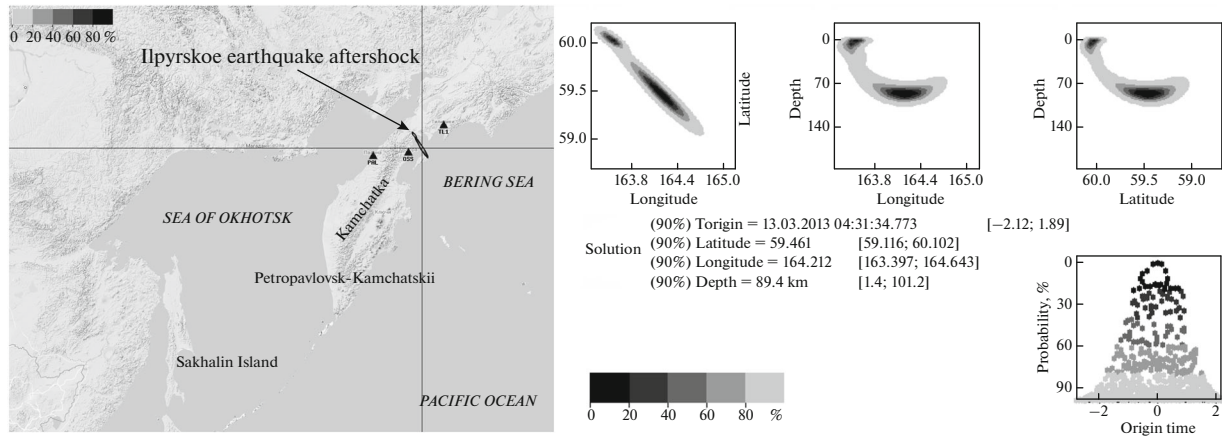


Fig. 7. Posterior distributions of probabilistic estimates of hypocenter location and origin time for aftershock of Ilpyrskoe earthquake of March 13, 2013, $M_w = 5.8$. This aftershock with magnitude $M_L = 4.0$ occurred 1 h 19 min after main shock. Designations are in Fig. 5.

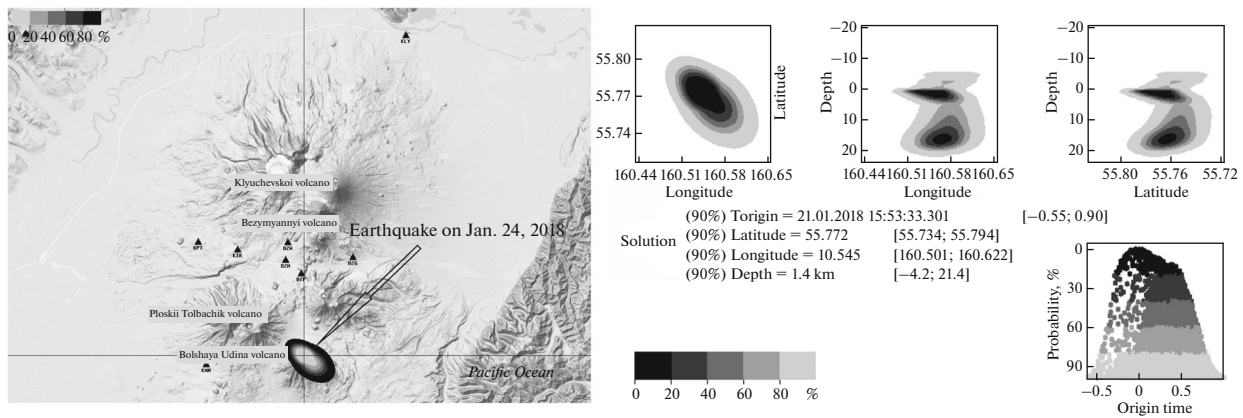


Fig. 8. Posterior distributions of probabilistic estimates of hypocenter location and origin time for seismic event in region of Udina volcano of January 24, 2018, $M_L = 1.5$. Designations in Fig. 5.

coordinates. The solution of this problem is associated with a number of organizational and technical tasks, including the development of the form of representation of the data and organizing access to these data.

The obtained confidence areas of the hypocenter location for the two strongest earthquakes that occurred in the vicinity of Kamchatka during the past ten years (the Okhotsk Sea earthquake with $M_w = 8.2$ in 2013 and the Blichne-Aleutskoe earthquake with $M_w = 7.8$ in 2017) demonstrate the highest possible solution accuracy with the existing configuration of the seismic network. For a more accurate location of hypocenters, it is clearly required to develop the station network, including the deployment of ocean-bottom stations. Developing the network is also necessary for studying volcanic seismicity.

In this work we assume that the distribution of the residuals is normal with a zero mean and not station-specific. At the same time, from the experience of processing the Kamchatka data it follows that the P -wave arrival times are sometimes selected by the staff with a

certain delay; therefore, the real distribution of P -wave residuals can have a positive mean. The amplitudes of various types of noise affecting the variance of the residuals are different for different stations. We hope that the impact of both factors on the obtained distributions of hypocenter location is insignificant. At the same time, exploring these factors can constitute the subject of an independent study, which are obviously important in the problems of hypocentry and constructing the tomographic models.

In this work, we used the regional traveltime curve of P - and S -waves (Mel'nikov, 1990) for the corresponding one-dimensional velocity model of the medium. The same approach is currently used in the routine processing at the KB GS RAS. In the one-dimensional traveltime curve, the traveltimes of the waves depend only on the source depth and epicentral distance but not on the geographic position of the source and the stations. The use of the one-dimensional traveltime curve leads to systematic errors in the theoretical arrival times of the waves. These systematic

errors can be due to the heterogeneities in the Earth's interior. In this case, they contain relevant information which can be used in seismic tomography. Importantly, the one-dimensional character of the travel-time curve has no fundamental importance in the scope of the hypocenter location method suggested in this work. Instead of this hodograph, the theoretical arrival times calculated from three-dimensional velocity models can be used. With this approach, the calculated times δt_p and δt_s in Eq. (3) should not be taken from the travel-time curve but from the table calculated based on the three-dimensional model on the grid for all the possible positions of the sources at all stations. This approach is in principle possible with the modern computing power and memory capacity of the computers. Although this problem falls beyond the scope of this paper, we would like to note that for improving the hypocenter location accuracy, it would probably be reasonable in the nearest future to try the up-to-date three-dimensional tomographic models based on the body seismic waves for the Kamchatka region overall (Gorbatov et al., 1999; Gontovaya et al., 2007; Koulakov et al., 2011a) and their more detailed versions for the volcanic regions (Gontovaya et al., 2004; Koulakov et al., 2011b; 2017). Seismic velocities in the layers closest to the surface can be determined with the use of surface wave tomography (Gordeev et al., 2009; Koulakov et al., 2014; Yanovskaya, 2015; Droznina et al., 2017).

CONCLUSIONS

In this work, for processing the data on the earthquakes that occurred in Kamchatka and its adjacent regions and were recorded by the stations of the KB GS RAS network of detailed seismological observations, we use for the first time the statistic method for determining the hypocenter location and origin time in which the solution is the probability distribution of these parameters.

When implementing this method, we constructed the model of the distribution of the residuals which allows for the correlation between the arrival times of different types of seismic waves. The model takes into account the dependence of the parameters on the distance from the source to the stations and on the interstation distance. The model is based on the huge data accumulated over the previous observation period which allowed us to obtain realistic empirical estimates for the errors of hypocenter location.

A two-level iterative scheme for discretizing the time and space of the earthquakes is suggested for practical calculations. Its use noticeably speeds up the computations, which will allow this method to be used in the future for the routine processing at the KB GS RAS. The method is also applicable for the other regions provided that an adequate model of the residuals is constructed for a given region.

Based on a number of examples, it is shown that with the existing mutual configuration of the Kamchatka regional seismic network and the seismic focal zone, in some cases, substantial systematic distortions of hypocentral coordinates can be observed. These distortions sometimes yield misleading notions of the configuration of the active seismic zones associated, e.g., with the volcanic systems or aftershock regions of large earthquakes. Since the analysis of the spatiotemporal structure of seismicity is frequently used for studying the deep structure of the medium, it is vital to have an instrument for evaluating the contribution of the hypocenter location errors in the subsequent estimates of the location and configuration of the geological and tectonic structures.

We hope that introducing the technique presented in this paper into the routine practice of earthquake processing at the KB GS RAS will supply the earthquake catalogs with information about the real confidence areas of the hypocenters' solutions. In many cases, this will help avoid misleading conclusions in the spatial analysis of seismicity.

FUNDING

The study was conducted with the support of the Russian Federation Ministry of Science and Education (project no. 14.W03.31.0033) in partial fulfillment of a state contract of the KB GS RAS within the research project no. AAAA-A16-116070550057-7 "Conducting continuous seismological, geophysical, and geodynamical monitoring on global, federal, and regional scales, as well as development and implementation of new technologies for processing and systems analysis of large data volumes."

REFERENCES

- Chebrov, V.N., Bakhtiarova, G.M., Droznin, D.V., Dubrovskii, N.I., Kugaenko, Yu.A., Levina, V.I., Pantyukhin, E.A., Senyukov, S.L., and Sergeev, V.A., The Kamchatka branch, Geophysical Survey RAS, information resources on the Internet, in *Problemy kompleksnogo geofizicheskogo monitoringa Dal'nego Vostoka Rossii. Trudy Vtoroi nauchno-tekhnicheskoi konferentsii* (Problems of Integrated Geophysical Monitoring of the Russian Far East, Proc. Second Scientific and Technical Conference), Petropavlovsk-Kamchatskii, October 11–17, 2009, Petropavlovsk-Kamchatskii: GS RAN, 2010, pp. 302–305.
- Chebrov, V.N., Droznin, D.V., Kugaenko, Yu.A., Levina, V.I., Senyukov, S.L., Sergeev, V.A., Shevchenko, Yu.V., and Yashchuk, V.V., The system of detailed seismological observations in Kamchatka in 2011, *J. Volcanol. Seismol.*, 2013, vol. 7, no. 1, pp. 16–36.
- Chebrov, D.V., Droznina, S.Ya., Senyukov, S.L., Shevchenko, Yu.V., and Mityushkina, S.V., Kamchatka and Commander Islands, in *Zemletryaseniya Rossii v 2016 godu* (Earthquakes in Russia in 2016), Obninsk: GS RAN, 2018, pp. 67–73.

- Chebrova, A.Yu., Chebrov, V.N., Matveenko, E.A., Tokarev, A.V., and Chemarev, A.S., The integrated information system of seismological data in the Kamchatka Branch of the Geophysical Survey of the Russian Academy of Sciences as of mid-2015, *Sovremennye metody obrabotki i interpretatsii seismologicheskikh dannyykh. Materialy Desyatoi Mezhdunarodnoi seismologicheskoi shkoly* (Modern methods for processing and interpretation of seismological data, Proc. 10th Int. Seismological School), Azerbaijan, September 14–18, 2015, Obninsk: GS RAS, 2015, pp. 356–360.
- Droznin, D.V. and Droznina, S.Ya., Interactive DIMAS program for processing seismic signals, *Seism. Instrum.*, 2011, vol. 47, no. 3, pp. 215–224.
- Droznina, S.Ya., Shapiro, N.M., Droznin, D.V., Senyukov, S.L., Chebrov, V.N., and Gordeev, E.I., S-wave velocity model for several regions of the Kamchatka Peninsula from the cross correlations of ambient seismic noise, *Izv., Phys. Solid Earth*, 2017, vol. 53, no. 3, pp. 341–352. <https://doi.org/10.1134/S1069351317030028>
- Fedotov, S.A., *Energeticheskaya klassifikatsiya Kurilo-Kamchatskikh zemletryasenii i problema magnitud* (Energy Classification of the Kuril-Kamchatka Earthquakes and the Problem of Magnitudes), Moscow: Nauka, 1972.
- Gontovaya, L.I., Khrenov, A.P., Stepanova, M.Yu., and Senyukov, S.L., The depth model of the lithosphere in the region of the Klyuchevskaya group of volcanoes, Kamchatka, *J. Volcanol. Seismol.*, 2004, no. 3, pp. 3–11.
- Gontovaya, L.I., Gordienko, V.V., Popruzhenko, S.V., and Nizkous, I.V., A depth model for the upper mantle of Kamchatka, *Vestn. KRAUNTS Nauki Zemle*, 2007, vol. 9, no. 1, pp. 90–104.
- Gorbatov, A., Dominguez, J., Suarez, G., Kostoglodov, V., Zhao, D., and Gordeev, E., Tomographic imaging of the P-wave velocity structure beneath the Kamchatka Peninsula, *Geophys. J. Int.*, 1999, vol. 137, no. 2, pp. 269–279.
- Gordeev, E.I., Droznina, S.Ya., and Shapiro, N.M., Crustal and uppermost mantle structure in the Pacific, North American, and Eurasian Plate junction, *Dokl. Earth Sci.*, 2009, vol. 28, no. 7, pp. 1198–1201.
- Gordeev, E.I., Fedotov, S.A., and Chebrov, V.N., Detailed Seismological Investigations in Kamchatka during the 1961–2011 Period: Main Results, *J. Volcanol. Seismol.*, 2013, vol. 7, no. 1, pp. 1–15.
- Gusev, A.A., Computer-based location of close Kamchatka hypocenters, *Vulkanol. Seismol.*, 1979, no. 1, pp. 74–81.
- Holschneider, M., Narteau, C., Shebalin, P., Peng, Z., and Schorlemmer, D., Bayesian analysis of the modified omori law, *J. Geophys. Res.*, 2012, vol. 117, B05317. <https://doi.org/10.1029/2011JB009054>
- Husen, S. and Hardebeck, J.L. Earthquake location accuracy. Community Online Resource for Statistical Seismicity Analysis. Available at <http://www.corssa.org>. 2010. <https://doi.org/10.5078/corssa-55815573>
- Koulakov, I.Yu., Dobretsov, N.L., Bushenkova, N.A., and Yakovlev, A.V., Slab shape in subduction zones beneath the Kurile-Kamchatka and aleutian arcs based on regional tomography results, *Russ. Geol. Geophys.*, 2011a, vol. 52, pp. 650–667.
- Koulakov, I., Gordeev, E.I., Dobretsov, N.L., Vernikovskiy, V.A., Senyukov, S., and Yakovlev, A., Feeding volcanoes of the Kluchevskoy group from the results of local earthquake tomography, *Geophys. Rev. Lett.*, 2011b, vol. 38, p. L09305.
- Koulakov, I., Jaxybulatov, K., Shapiro, N.M., Abkadyrov, I., Deev, E., Yakovlev, A., Kuznetsov, P., Gordeev, E., and Chebrov, V., Asymmetric caldera-related structures in the area of the Avacha group of volcanoes in Kamchatka as revealed by ambient noise tomography and deep seismic sounding, *J. Volcanol. Geotherm. Res.*, 2014, vol. 285, pp. 36–46. <https://doi.org/10.1016/j.jvolgeores.2014.08.012>
- Koulakov, I., Abkadyrov, I., Arifi, N.A.I., Deev, E., Droznina, S., Gordeev, E.I., Yakovlev, A., Khrepy, S.E.I., Kulakov, R.I., Kugaenko, Y., Novgorodova, A., Senyukov, S., Shapiro, N., Stupina, T., and West, M., Three different types of plumbing system beneath the neighboring active volcanoes of Tolbachik, Bezymianny, and Klyuchevskoy in Kamchatka, *J. Geophys. Res. Solid Earth*, 2017, vol. 122, pp. 3852–3874. <https://doi.org/10.1002/2017JB014082>
- Lee, W.H.K. and Lahr, J.C., HYPO71 (revised): A computer program for determining hypocenter, magnitude, and first motion pattern of local earthquakes, *U.S. Geol. Surv. Open-File Report 75-311*, 1975.
- Levina, V.I., Ivanova, E.I., Lander, A.V., and Guseva, E.M., Kamchatka and Commander Islands, in *Zemletryaseniya Severnoi Evrazii v 2002 godu* (North Eurasia Earthquakes in 2002), Obninsk: GS RAN, 2008, pp. 215–225.
- Lomax, A., Virieux, J., Volant, P., and Berge-Thierry, C., Probabilistic earthquake location in 3-D and layered models, in *Advances in Seismic Event Location*, Thurber, C.N. and Rabinowitz, N., Eds., Dordrecht: Kluwer, 2000.
- Mel'nikov, Yu.Yu., Program package for determining the coordinates of hypocenters of Kamchatka earthquakes on PC, *Vulkanol. Seismol.*, 1990, no. 5, pp. 103–112.
- Minson, S.E. and Lee, W.H.K., Bayesian historical earthquake relocation: an example from the 1909, Taipei earthquake, *Geophys. J. Int.*, 2014, vol. 198, pp. 1419–1430. <https://doi.org/10.1093/gji/ggu201>
- Saltykov, V.A. and Droznina, S.Ya., The March 13, 2013 Ilpyrskoye earthquake $M_W = 5.8$ (the northern Kamchatka): abnormal length of focal zone as an artifact, *Vestn. KRAUNTS*, 2014, vol. 24, no. 2, pp. 130–136.
- Sambridge, M. and Kennett, B.L.N., A novel method for hypocenter location, *Geophys. J. R. Astron. Soc.*, 1986, vol. 87, pp. 679–697.
- Senyukov, S.L., Monitoring the activity of Kamchatka volcanoes by remote observation instruments in 2000–2004, *Vulkanol. Seismol.*, 2006, no. 3, pp. 68–78.
- Sil'nye kamchatskie zemletryaseniya 2013 g.* (Kamchatka Strong Earthquakes of 2013), Chebrov, V.N., Ed., Petropavlovsk-Kamchatskii: Novaya kniga, 2014.
- Tarantola, A. and Valette, B., Inverse problems = quest for information, *J. Geophys.*, 1982, vol. 50, pp. 159–170.
- Yanovskaya, T.B., *Poverkhnostno-volnovaya tomografiya v seismologicheskikh issledovaniyakh* (Surface Wave Tomography in Seismological Studies), St. Petersburg: Nauka, 2015.

Translated by M. Nazarenko