Oceanology, Vol. 45, No. 4, 2005, pp. 488–499. Translated from Okeanologiya, Vol. 45, No. 4, 2005, pp. 518–530.

MARINE PHYSICS

On Long-Term Tsunami Forecasting

E. A. Kulikov[1,2], A. B. Rabinovich[1,2], and R. E. Thomson[2]

[1] *Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia*

[2] *Institute of Ocean Sciences, Sidney, British Columbia, Canada*

Received May 17, 2004; in final form, December 23, 2004

Abstract—Data for tsunamigenous earthquakes and tsunami events from the Expert Tsunami Database for the Pacific (ETDB) and from the Tsunami Database of the National Geophysical Data Center (NGDC) are used for long-term tsunami forecasting and for the determination of tsunami run-up distribution functions. The comparative analysis is exemplified using the data for the Pacific coast of South America from 5° to 35° S (adjacent to Peru and northern Chile). The calculated recurrence periods and tsunami wave heights from the corresponding data were compared with each other and with the estimations from other independent sources. A stochastic model with a lognormal alongshore tsunami run-up distribution was found to be quite efficient for the region under study. Based on the ETDB data, we expect tsunami wave heights of 2.7, 5.1, 10.2, and 16.3 m for 10-, 20-, 50-, and 100-year periods, while, from the NGDC data, we obtained 3.0, 5.7, 13.3, and 25.3 m wave heights, respectively. The significant differences in the results arise from the differences in the two datasets.

1. INTRODUCTION

Tsunamis are among the most dangerous natural phenomena. These catastrophic waves, which are usually formed as the result of strong underwater earthquakes, are most frequently observed at the coasts of the Pacific Ocean. The period from 1992 to 2003 was characterized by an unusually high number of destructive tsunamis, more than 20, including the tsunamis on July 12, 1993, in the Sea of Japan; October 4, 1994, near Shikotan Island (Kuril Islands); July 17, 1998, on the coast of Papua New Guinea; and June 23, 2001, on the southern coast of Peru. These tsunamis caused enormous economic damage and killed more than 4000 people.

The problem of long-term tsunami forecasting is especially important for the populated part of the Pacific coasts of Japan, Russia, the United States, Indonesia, and South America. Statistical analysis of the data for historical earthquakes and tsunamis allows us to estimate the degree of danger and risk related to the constructions on the coast and other types of economic and life activities. The calculation of the probability of earthquakes and tsunamis of certain force on a given interval of the coast is usually presented in the form of recurrence graphs. The standard periods of recurrence characterize not only the frequency of rare extreme events (with recurrence of 50, 100, and 200 years) but also the frequency of moderate events with periods of recurrence equal to 5, 10, and 20 years. Actually, the planning of economic activity and any new construction in seismically active and tsunami risk zones of the Pacific coast requires preliminary estimates of the possible floodings related to tsunamis.

The main problem in the analysis of the statistics of tsunamigenous earthquakes and tsunami heights is the completeness and quality of the data. The data coverage of the coast is directly related to its population and the availability of sea level recorders; over significant intervals of the coastline, the data of the flooding are absent. One has to take into account that systematic data gathering about powerful marine events is usually the most important issue in the estimates of extreme flooding. On the other hand, the historical evidence contains the data referring only to the strongest events, whose recurrence is most important for the assessment of extreme floodings. In order to estimate tsunami risks, it is necessary to correctly combine the comparatively complete statistics of tsunamis in 1900–2003 (including weak events) with the fragmentary historical information about catastrophic floodings on the coast in the previous period.

In this study, we discuss the problem of the probability description of the tsunami hazard based on the statistical analysis of historical data and tsunami heights. The requirement of the statistical reliability of the estimates obtained as a result of such an analysis frequently contradicts with the necessity of local forecasts for relatively small regions of the coast. Even in the regions with comparatively dense populations and the presence of observation networks, the number of tsunami records at each specific point is small. However, for selected tsunami hazard zones, it is possible to distinguish segments with a relatively uniform distribution of tsunami recurrence along the coast. In this case, it is possible to form a representative data sample about the tsunami heights for the corresponding part of the coast. However, in this case, the statistical conclusions have a regional character. The calculated values of the probability are related to the entire interval of the coast considered. This means that the event with the given probability would occur somewhere in the study region of the coastline. In order to pass from the probability at some place to the probability of the event at a specific point of the coastline, it is necessary to carry out additional research and calculations. The empirical estimate of Solov'ev [6] is known, in which the maximum tsunami height recorded is related to the mean value of the flood level at the coast. Another effective approach is local tsunami zoning (see, for example, [2, 13]). In this paper, we estimate the extreme tsunami heights with given recurrences for a specific region and discuss a stochastic model for calculating the relation between the maximal and mean heights of tsunami waves on the coast.

In order to develop the method for estimating the tsunami hazard, we chose a region of the coastline in Peru and northern Chile from 5° to 35° S. Several factors determined the choice of this region. First, this zone is characterized by one of the highest levels of seismicity and tsunami hazard. Second, the coastline and shelf in this region are relatively straight and uniform. There are no clearly manifested individual resonant particularities of the topography, which are characteristic, for example, of Japan and the Kuril Islands. Third, the tsunami sources in this region are located approximately uniformly along the coast and the seismicity of the entire region is approximately the same. Owing to the latter feature, we excluded from the consideration the region of southern Chile (south of 35° S), where the seismicity is significantly higher (see, for example, [8, 14, 17]).

The information about the historical tsunamis on the coasts of Peru and northern Chile were taken from known databases: the Tsunami Database of the National Geophysical Data Center for Natural Hazards (NGDC) located in Boulder, Colorado (United States) and the Expert Tsunami Database for the Pacific (ETDB/PAC) developed at the Tsunami Laboratory of the Institute of Computational Mathematics and Mathematical Geophysics of the Siberian Branch of the Russian Academy of Sciences (ICMMG) in Novosibirsk [14].

2. HISTORICAL DATA ABOUT TSUNAMIS AND EARTHQUAKES ON THE COAST OF PERU AND CHILE

The coast of Peru and Chile is characterized by one of the highest levels of seismicity and tsunami hazard. Tectonic interaction of the oceanic plate with the continental platform in the subduction zone is the main source of the high seismicity [9, 16, 19] and, hence, the source of the high level of the tsunami hazard at the coast.

The importance of the estimate of the tsunami recurrence for this region is also due to the fact that, frequently, tsunamis generated near the coast of Peru and Chile cross the ocean and cause catastrophic aftereffects on the opposite coast of the Pacific Ocean. For example, the tsunami in Chile on May 12, 1960, killed 61 people and destroyed about 450 houses on the Hawaii Islands and caused significant damage on the coasts of Japan [7]. It is interesting that the highest tsunami waves registered on the western coast of the Sea of Okhotsk were related to this tsunami: 2.7 m on the coast of Sakhalin and greater than 4 m in the region of Magadan [4]. Tsunamis caused by earthquakes in the region of Peru were repeatedly recorded on the coasts of Japan [15].

The first historical information about the seismicity and tsunamis on the coast of Peru and Chile are related to the 16th century, when Spanish conquistadors settled in this region. Some information about the events of the far past can be obtained from the legends of the people populating Peru and Chile [8]. The scientific study of South American tsunamis started with the paper by Berninghausen [10], in which 49 tsunamis from 1575 to 1960 were listed. Later, other papers were published dedicated to the study of the tsunami hazard in this region [17, 18, 22]. According to the chart of tsunamigenous earthquakes from 1562 to 1960 prepared by Solov'ev and Go [8], the entire zone of the Peru and Chile coast is a region of high tsunami risk. In general, here, we shall use the results of papers [6, 8, 21, 22].

As one might expect, the quality of the historical data about the tsunami manifestation on the coast or strength (magnitude) of the earthquakes is the decisive factor for calculating the statistical parameters of the tsunami risk. Unfortunately, the completeness and reliability of the data strongly depend on time. The known historical information about catastrophic events in the 16th–19th centuries is usually related to the powerful events, when only catastrophic events were documented, that is, tsunamis with heights exceeding 3–5 m. The beginning of the instrumental epoch led to a sharp increase in the number of the events recorded, which is associated with the appearance of sea level gauges, seismographs, and the corresponding services responsible for the observations of seismicity and marine tides.

Another problem, which appears in the analysis of the statistics of tsunamis in South America, is the inequality in the data coverage of the coast. The major part of the information is related to the populated parts of the coast adjacent to the cities of Santiago, Valparaiso, Concepcion, Lima, Callao, and Pisco. Long-term instrumental measurements (records of coastal tide gauges) are available only at selected places (Valparaiso, Antofagasta, and Arica in Chile; Matarani, Callao, and Pisco in Peru). The most important aspect of the study of tsunamigenous earthquakes and tsunamis is the data systematization, reducing them to a common standard, and the quality estimate. It is noteworthy that, while the data about the seismicity are sufficiently presented enough in the form of periodically renewed catalogues and databases on electronic media, the systematization of the information about tsunamis (including tsunamigenous earthquakes) are actually only at the stage of formation. At present, a scientific project is planned concerning the development of the Global His-

Table 1. Information about tsunamis on the coasts of Peru and Chile (5°–35° S) from the NGDC and ETDB databases

Database	Period of observations, years	Number of single tsunami observations	Number of tsunami events	
			total	applicable for analysis
NGDC	1562–2001	318	189	71
ETDB	1513–2001	270	139	92

torical Tsunami Database (GTDB), a global database on tsunamis, which would unite all the materials known from the existing databases on tsunamis in the World Ocean. At present, two databases, the NGDC and ETDB, which are available for public access, are known better than the others. These databases combine in one place the historical data about tsunamigenous earthquakes and tsunami heights on the coasts. In our publication, we do not put forward the objective to analyze the quality and completeness of the data contained in one or another database. The choice of the NGDC and ETDB for the analysis is determined first by the fact that they are widely known and available on the Internet.

The NGDC database contains information about 189 tsunamis, which occurred from 1562 to 2001 in the study region of the Peru and northern Chile coast between 5° and 35° S, while the ETDB database contains data about 139 events from 1513 to 2001. The main characteristics of the information used from the NGDC and ETDB are given in Table 1. We note that, among the data about the tsunami events presented, only about 40% of the events from the NGDC and about 70% of the events from the ETDB appeared applicable for the statistical analysis.

The majority of the rejected data about tsunamis do not contain estimates of the wave height on the coast; therefore, they are not applicable for the current analysis.

The ETDB database contains information about more than 500 earthquakes with $M_s \geq 6.0$ in the zone adjacent to the coast of Peru and Chile from 5° to 35° S, which occurred from 1471 to 2001. Among these, only 30% are tsunamigenous. A chart of the seismicity of the Pacific coast of South America according to the ETDB data is shown in Fig. 1. One can see that the epicenters are located along the coastline following the geometry of the subduction zone. One of the particular features of this region is the immediate proximity of the tsunami sources to the coast.

A graph of the time dependence of the tsunamigenous earthquakes with $M_s \geq 5.5$ is shown in Fig. 2. It is easy to see how the representativeness of the data changes with time. One of the most important requirements of the statistical analysis of the recurrence of random events for the time series of observations is the condition of stationarity. It is clear that this time series does not satisfy this requirement. However, it can be divided into two approximately stationary time series: 1500–1900 and 1901–2001. In the former sample, the data about weak earthquakes (with magnitudes $M_s <$ 7.5) are practically absent. In the latter sample (1901–2001), which is based on the instrumental data, the earthquakes with $M_s \geq 6.0$ are represented well. The seismic process, which is related to the class of Poisson processes [6], can be naturally considered to be stationary. The variations in the quality and quantity of the data are evidently caused by the development of the systems of observations and improving of the instruments. Before 1900, all the values of the magnitudes in the period are based exclusively on indirect data about the strong seismic events. Only in the 20th century did the appearance of seismographic instruments make it possible to widen the range of the events recorded.

It is reasonable to divide all the data about the earthquakes in the study region available into two approximately stationary time series: 1513–1900, in which the seismic events with $M_s > 7.5$ are presented sufficiently completely, and the period 1901–2001, in which the time series of events with magnitudes $M_s > 6.0$ can be considered stationary.

The time series of the tsunami heights recorded on the coast of Peru and northern Chile from 1575 to 2001 is presented in Fig. 3. Similarly to the data about the magnitudes of earthquakes, one can see the same tendency in the distribution of the known tsunamis. The period from 1575 to 1900 is characterized by the absence of events with tsunami heights smaller than 3 m. The statistics of small tsunamis actually starts only in the 20th century. After 1900, the time series of tsunami heights can be considered approximately stationary.

The spatial distribution of the tsunami heights recorded along the coast is shown in Fig. 4. Regardless of the unequal tsunami height data coverage of the coast, one can conclude that the degree of tsunami hazard is approximately the same for all the regions of the coastline. It is clear that this fact is related to the seismic and topographic homogeneity (see Fig. 1).

The supposition about the uniform distribution of the tsunami hazard along the coast of Peru and northern Chile is of course true only for the open parts of the coast with a relatively straight coastline. At the points located in harbors, bays, and river mouths, the effect of local tsunami intensification can be manifested. In these cases, additional analysis is required for obtaining the values of the possible tsunami heights, for example, using numerical modeling.

3. RECURRENCE OF TSUNAMIGENOUS EARTHQUAKES ON THE COAST OF PERU AND NORTHERN CHILE

The probability of a seismic event $M \geq M_0$ is usually presented in the form of the periods of recurrence T or

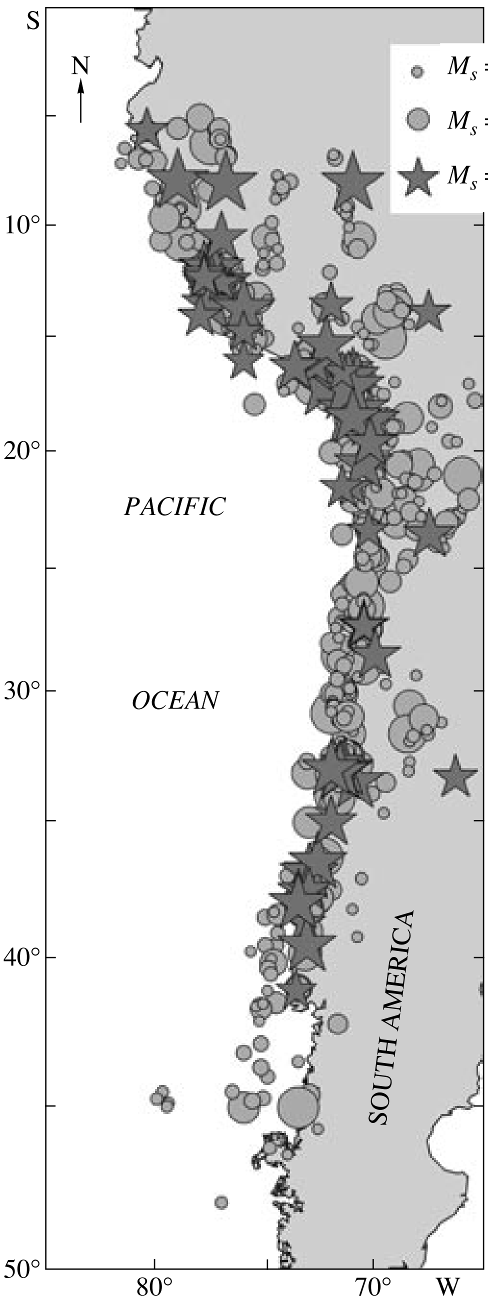

Fig. 1. Chart of earthquakes epicenters with $M_S \geq 6.0$ for the coasts of Peru and Chile plotted on the basis of data from the ETDB data (1471–2001).

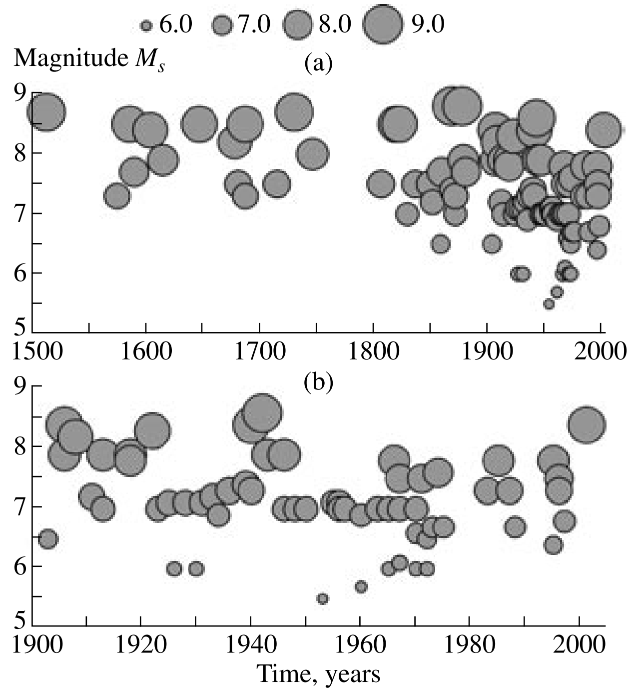

Fig. 2. Time series of the tsunamigenous seismic events recorded on the coasts of Peru and northern Chile (a) from 1513 to 2001 and (b) from 1901 to 2001 on the basis of the ETDB data. The scale of the earthquake magnitudes is shown.

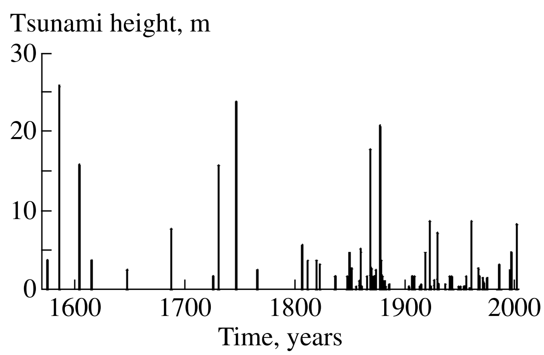

Fig. 3. Time series of the maximal tsunami heights recorded on the coasts of Peru and Chile from 1575 to 2001 on the basis of the ETDB data.

cumulative frequency (1/*T*). According to Gumbel [3], the estimate of the recurrence period *T* is calculated for a sorted sample of *N* random numbers x_i such that $x_1 \leq x_2 \leq \ldots \leq x_N$ using the following relation:

$$T(x_i) = \frac{N+1}{N-i+1}\frac{T_N}{N}, \qquad (1)$$

where *i* is the sequential number of event x_i in the sorted time series, *N* is the number of random numbers, and T_N is the total time of the observations corresponding to the given sample.

The distribution of the recurrence periods of the magnitudes of tsunamigenous earthquakes $M_s \leq 5.5$ on the basis of the ETDB data for the period 1901–2001 is shown in Fig 5a. In addition, the recurrence estimates are plotted and calculated for magnitudes $M_s \leq 8.3$ for the sample from 1513 to 2001. The scale of the graph along the abscissa axis is $\sqrt{1/T}$ (*T* is in years). It is seen

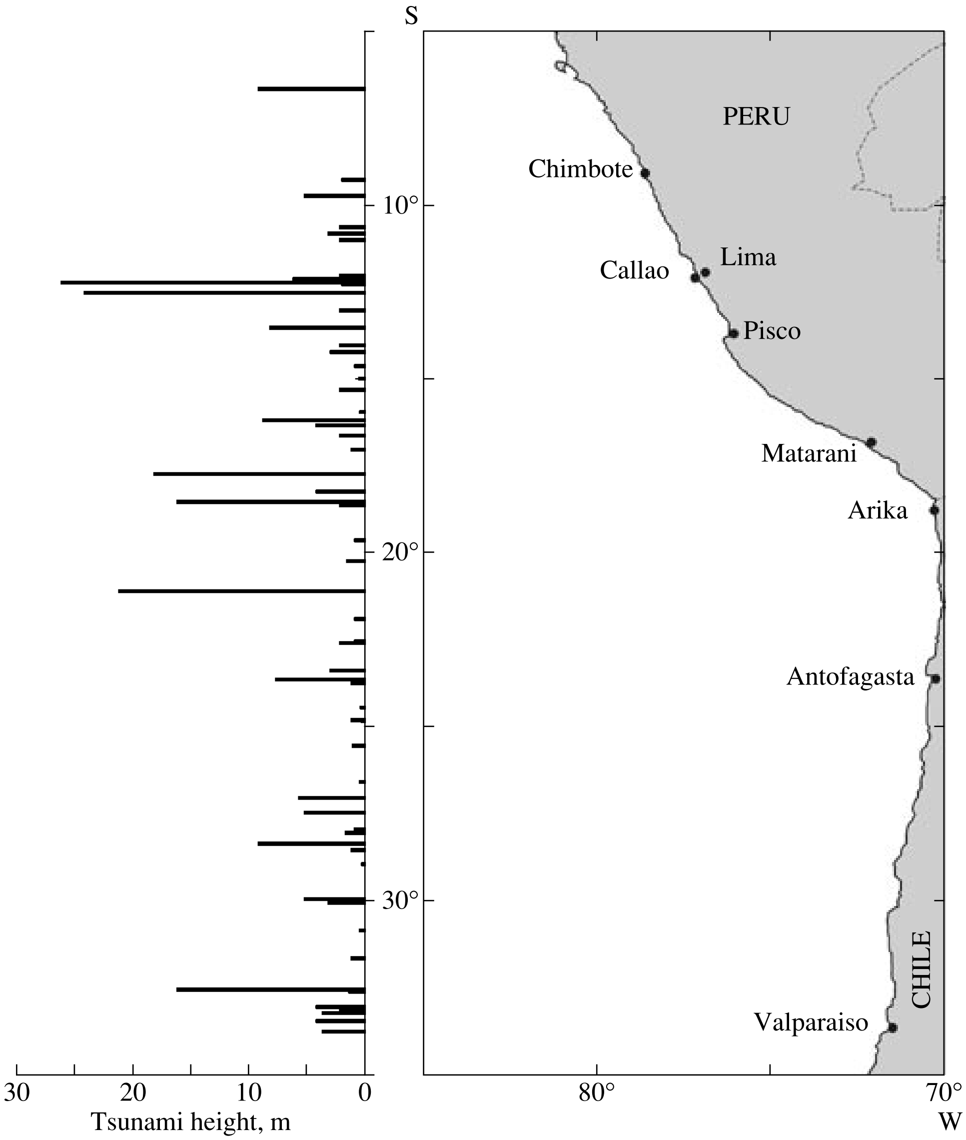

Fig. 4. Distribution of the recorded maximal tsunami heights recorded on the coasts of Peru and northern Chile from 1575 to 2001 on the basis of the ETDB data.

that, for great recurrence periods, the value of the magnitude tends to a limiting value $M_\infty = 9.0$. The same estimates are shown in Fig. 5b on other scale ($\ln T$ along the abscissa axis and $\ln(9.0 - M)$ along the ordinate axis. The asymptote $(9.0 - M)^2 \sim \ln T$ is also shown in this graph.

In order to describe the distribution of the probability of the earthquake magnitudes, Silgado [22] used the models of asymptotic limiting distributions of Gumbel of the first and third types [3]. The probability that the magnitude of the expected event appears smaller than M is written as follows:

$$F(M) = \exp(-\alpha e^{-\beta M}) \tag{2}$$

first kind of the limiting distribution;

$$F(M) = \exp\left(-\left(\frac{M_\infty - M}{M_\infty - \mu}\right)^k\right) \tag{3}$$

third kind of the limiting distribution,

where α, β, μ, k, and M_∞ are the empirical parameters statistically determined from the observations. The recurrence period can be calculated from the following relation:

$$T_e(M) = \frac{\bar{T}}{[1 - F(M)]}, \tag{4}$$

where $\bar{T} = T_N/N$ is the recurrence period of the event with the minimal amplitude presented in the sample.

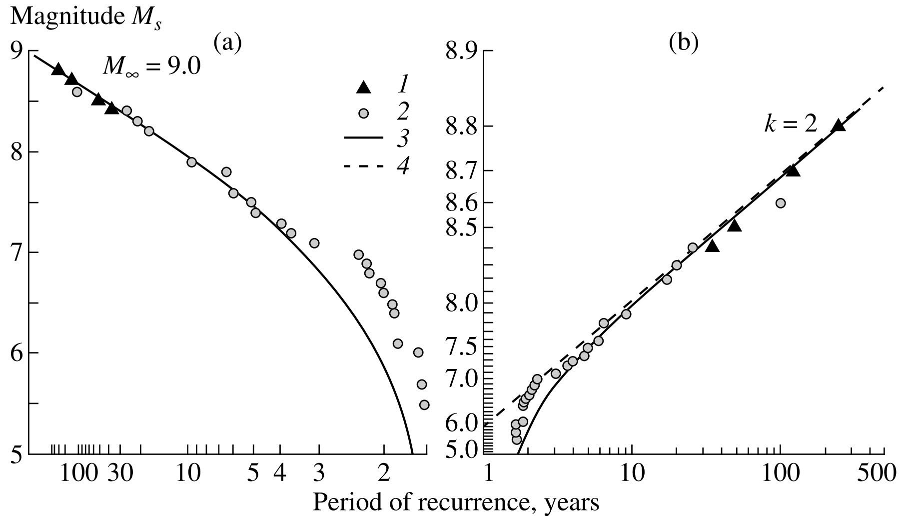

Fig. 5. Graphs of the recurrence of tsunamigenous earthquakes on the coasts of Peru and Chile on the basis of the data (*1*) from 1513 to 2001 and (*2*) from 1900 to 2001. The solid line (*3*) shows the approximation corresponding to the limiting asymptotic distribution of the 3rd type; the dashed line (*4*) shows the asymptote in relation (5). The scale of the mapping on the graph is as follows: (a) $\sqrt{1/T}$ along the abscissa axis; (b) $\ln T$ along the abscissa axis, and ln (9.0 – M) along the ordinate axis.

The solid line in Figs. 5a and 5b is the curve corresponding to relation (4) of distribution (3) with the parameters $M_\infty = 9.0$, $\mu = 6.5$, and $k = 2.0$. Parameters μ and k were selected for the best approximation of the recurrence periods depending on the magnitude of the earthquake along the asymptote in relation (4), which is shown in Fig. 5b with the dashed straight line:

$$\ln(M_\infty - M) = \frac{1}{k}(M_\infty - \mu) - \ln T. \qquad (5)$$

It is important to pay attention to the existence of an absolute limit in the third kind of the Gumbel asymptotic distribution. The existence of the upper limit in relation (3) for the distribution of the magnitudes ($\lim\limits_{T \to \infty} M = M_\infty$) can be explained by two physical causes: the restriction of the possible length of the faults in the subduction zone and the limit of the strength of the Earth's crust in the given region.

The calculated values of the earthquake magnitudes for the main recurrence periods on the coasts of Peru and northern Chile are given in Table 2. The calculation was performed using approximation (3).

The estimates calculated by Silgado [22] using the statistics for 1749–1974 are included into the table for comparison. It is important to note that, for the 50-, 100-, and 200-year recurrence periods, the values of the magnitudes obtained by Silgado are smaller approximately by 0.3–0.5 than those based on the ETDB data. This can be explained by the fact that the statistics used by Silgado in [22] were not sufficiently complete.

4. ANALYSIS OF THE RECURRENCE OF TSUNAMI HEIGHTS ON THE COAST OF PERU AND NORTHERN CHILE

Historical tsunami materials in the NGDC and ETDB databases are presented as two types of information: Events Databases (EDB) and Run-up Databases (RDB). The first sample from the EDB contains the magnitude of the earthquake, the coordinates of the epicenter, and the maximal wave height observed on the coast in the selected region for each tsunami event. The data set from the RDB is grouped by individual events. All the known characteristics of tsunami waves recorded for each tsunami event on the coast in the given region are presented. Unfortunately, for the majority of the events, only the maximal tsunami height is known; therefore, for the statistical analysis of the recurrence of the tsunami heights on the coast, one has to use the EDB data.

The distributions of the values of the recurrence periods calculated on the basis of the ETDB data on the maximal tsunami heights for the periods of 1901–2001 and 1575–2001 are shown in Fig. 6. The graphs are plotted on a scale that is usually referred to as probability paper [3]. On this scale, the values of the recurrence

Table 2. Calculated values of the magnitudes of tsunamigenous earthquakes for the given recurrence periods

Reference	Recurrence period, years						
	2	5	10	20	50	100	200
ETDB (1513–2001)	5.88	7.46	7.96	8.28	8.55	8.68	8.78
Silgado (1749–1974)	–	–	–	–	8.04	8.35	8.47

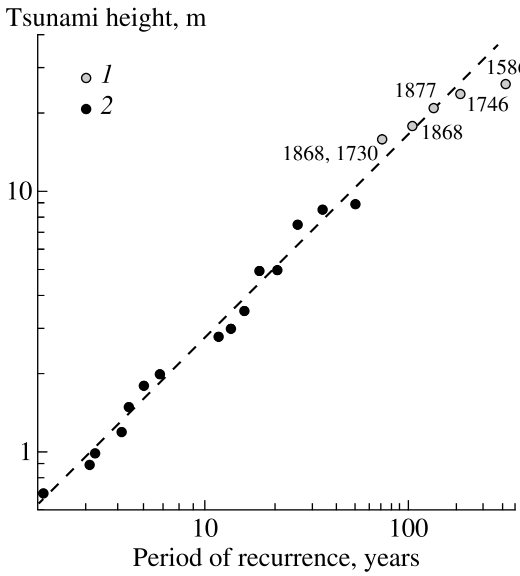

Fig. 6. Graphs of the recurrence of the maximal tsunami heights on the coasts of Peru and northern Chile (*1*) from 1901 to 2001 and (*2*) from 1575 to 1900. The dashed line shows the approximations corresponding to lognormal distributions.

periods corresponding to a lognormal distribution should fall on a straight line.

The lognormal distribution is frequently used to describe the statistics of flooding. The physical basis for application of the model of lognormal distribution for the description of the statistics of tsunami heights is given in Section 6. The corresponding straight line approximating the distribution of tsunami heights relative to their recurrence periods is shown in Fig. 6. It is seen that the observations agree well with the model of the lognormal distribution.

The estimates of the tsunami heights for several recurrence periods corresponding to the lognormal distribution approximation (based on the data from 1575 to 2001 for tsunamis with heights greater than 10 m and based on the data from 1901 to 2001 for tsunamis with heights greater than 0.5 m) are given in Table 3. It is interesting to compare these estimates with those obtained earlier on the basis of the NGDC data [21]. The tsunami heights corresponding to the recurrence periods of 5, 10, and 20 years differ only slightly; meanwhile, for the periods of 50 and 100 years, the estimates of the NGDC are significantly greater. Since, in both cases, the method applied was the same, the difference in the estimates is explained by the differences in the contents of the ETDB and NGDC. A direct comparison indicates that many events contained in one database are absent in the other. However, in the majority of cases, both databases give equal tsunami heights for coinciding events. In this study, we do not put forward the objective to compare the completeness and quality of the data in the ETDB and NGDC databases. However, it is important to note that the differences in the database components leads to significant differences in the estimates of the tsunami heights for the large recurrence periods of 50 and 100 years.

Table 3. Calculated tsunami heights (m) for different recurrence periods

Reference	Recurrence period, years					
	5	10	20	50	100	200
ETDB (1575–2001)	1.3	2.7	5.1	10.2	16.3	24.9
NGDC (1901–2001)	1.3	3.0	5.7	13.3	25.3	–

5. ESTIMATE OF THE EMPIRICAL RELATION BETWEEN THE EARTHQUAKE MAGNITUDE AND TSUNAMI

The study of the relations between the strengths of an earthquake and a tsunami is of special interest in the estimates of the tsunami energy. It is known that the tsunamis of seismic origin are formed as a result of residual displacements of the sea bottom formed along seismic tectonic faults. It is generally accepted that the energy of the generated sea waves can be approximately estimated knowing the energy (magnitude) of the earthquake [6]. It is also agreed that the measure of the tsunami energy is the tsunami magnitude $m = \log_2 h$, or its intensity $i = \log_2 h + 0.5$, where h is the tsunami height at the coast. The direct calculation of the regression between the magnitude of seismic events M_s and the magnitude (m) or intensity (i) of tsunamis for tsunamigenous earthquakes gives approximately this correlation. It is natural that the type of empiric relations obtained have a regional meaning and characterize particular features of the seismically active region.

The application of the empirical relation between the earthquake and tsunami magnitudes in the operative tsunami forecast has a practical importance. Immediately after recording an earthquake close to the coast, the operative tsunami service should estimate in advance the probability of the tsunami generation by this event and its possible height on the coast. If there are no pressure gauges in the open ocean, this estimate is impossible without application of an approximate estimate of wave heights on the basis of the measured magnitude of the earthquake M_s.

Silgado [22] obtained the following relation for the regions of Peru and Chile on the basis of the data available for 1749–1974:

$$\log(h) = 0.79M - 5.70. \qquad (5a)$$

An estimate of the empirical relation between the earthquake magnitudes and the intensities of tsunamis can be obtained using the means of the ETDB software. The corresponding relation for the events with $M_s > 7.5$

for the sample covering 1575–2001 has the following form:

$$i = 1.86M - 13.2. \tag{6}$$

Figure 7 illustrates the efficiency of relations (5) and (6) for tsunami height forecast.

6. STOCHASTIC MODEL FOR THE DISTRIBUTION OF TSUNAMI HEIGHTS ON THE COAST

It is known that the lognormal probability distribution has exact physical grounds in the problems of wave propagation in a randomly inhomogeneous medium [3]. This method is used for describing random fluctuations of the brightness of stars (stellar scintillation) caused by chaotic turbulent fluctuations of the refraction coefficient in the atmosphere.

The probability density of the lognormal distribution $p(x)$ is given by the relation

$$p(x) = \frac{1}{x\delta\sqrt{2\pi}}\exp\left(-\frac{\ln^2(x/\beta)}{2\delta^2}\right), \tag{7}$$

where x is the random value, β is the median of the distribution, and δ is the root-mean-square deviation of $\ln(x/\beta)$. A transition to the normal Gaussian distribution can be written by means of the transformation $y = \frac{\ln(x/\beta)}{\delta}$, where y is distributed normally with a dispersion equal to 1. It is natural that this probability density and the values of the corresponding constants β and δ are related to the selected period of observations. In order to compare the statistics of the data of observations related to different periods of observations, the recurrence periods are evaluated in years, i.e., the estimated probability is related to the conventional period of observations equal to one year.

The physical grounds of the model of the lognormal distribution of tsunami heights on the coast was probably first suggested by Go in 1987 [see 1, 12]. Later, this model was significantly developed in [5]. Random dispersion of tsunami amplitudes over the coast can be described similarly to the method used for estimating the distribution of the brightness of stars in the sky. A wave propagating coastward over an irregular bottom passes regions with different coefficients of amplification (or attenuation), which can be considered as a random sample from a set of ray tubes connecting the region of the source and the coastal regions. A scheme for forming chaotic amplitudes of a tsunami wave in the course of its propagation along the ray tube in the direction to the coast can be presented in the following manner. Let us divide the path of the wave propagation into N segments. The deviations of the bottom profile from the mean value form conventionally random fluctuations of the transmission coefficient (amplification or attenuation) of the wave $k_i \approx 1$. The amplitude of the tsunami on the coast h can be expressed as a product of N coefficients

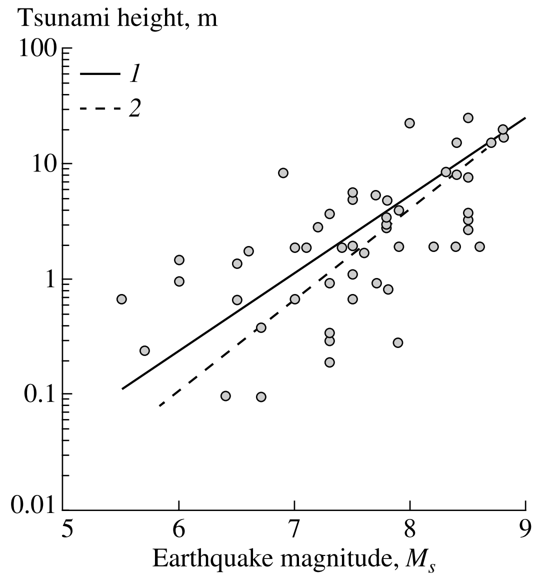

Fig. 7. Relation between the maximal tsunami height on the coasts of Peru and northern Chile and the tsunamigenous earthquake magnitude M_s. (*1*) regression; (*2*) correlation estimate based on the Silgado data [22].

$$h = k_1 k_2 k_3 \ldots k_N h_0, \tag{8}$$

where h_0 is the amplitude of the wave near the source. The multiplicative character of this process of forming chaotic amplitudes leads to a lognormal probability distribution of h.

After finding the logarithm of (8), we obtain:

$$\ln(h) = \ln(k_1) + \ln(k_2) + \ln(k_3) + \ldots + \ln(k_N) + \ln(h_0). \tag{9}$$

The values of $\ln(k_i)$ are small random numbers. According to the limiting theorem, regardless of the form of the distribution of $\ln(k_i)$, the distribution of $\ln(h)$ tends to the normal distribution law (Gaussian distribution) at $n \longrightarrow \infty$, while the distribution of the tsunami height on the coast h in the limit becomes lognormal (7). We note that the measure of dispersion δ is not related to the height of the tsunami in the source h_0. It actually characterizes the measure of the tsunami wave dispersion by random bottom irregularities, and it is constant for all events. The value h_0 characterizes the amplitude (median) of the random values in the distribution so that $h_0 \sim \beta$. Go [1, 12] suggested using the parameter $\beta = e^{\langle \ln(h_j) \rangle}$ (where $\langle\,\rangle$ is the sign of averaging, and h_j is the set of the tsunami heights observed on the coast) to estimate the magnitude of a tsunami instead of the commonly accepted parameter $m = \log_2 h_{\max}$.

Unfortunately, this parameter is difficult to estimate for the majority of the historical events for which the available information is fragmentary. However, the present-day organization of the tsunami service and the results of studies on the coasts for estimating the tsunami run up carried out in the flooding zones during international expeditions [5] make possible the calculation of annual mean values of the tsunami height on the coast with a sufficient statistical reliability. We can give examples of publications [11, 20] for the catastrophic earthquakes in Peru on February 21, 1996, and June 23, 2001, and paper [5], in which the authors consider 11 different events. It is clear that the measure of the dispersion $\delta = \sqrt{\langle \ln^2(h_j/\beta)\rangle}$ is a useful additional characteristic of the tsunami hazard on the coast.

The distribution of tsunami heights on the coast near the tsunami source of February 21, 1996, is shown in Fig. 8a. Despite the small distance to the source, the distribution of the heights is sufficiently random. These measurements are plotted on the graph in Fig. 8b as the probability distribution (frequency) of tsunami run-up heights, which does not exceed a given value. The dashed straight line corresponds to a lognormal distribution with $\delta = 0.44$ and $\beta = 2.42$ m. One can only be surprised by such good agreement between the character of the distribution of the tsunami heights recorded and the theoretical model.

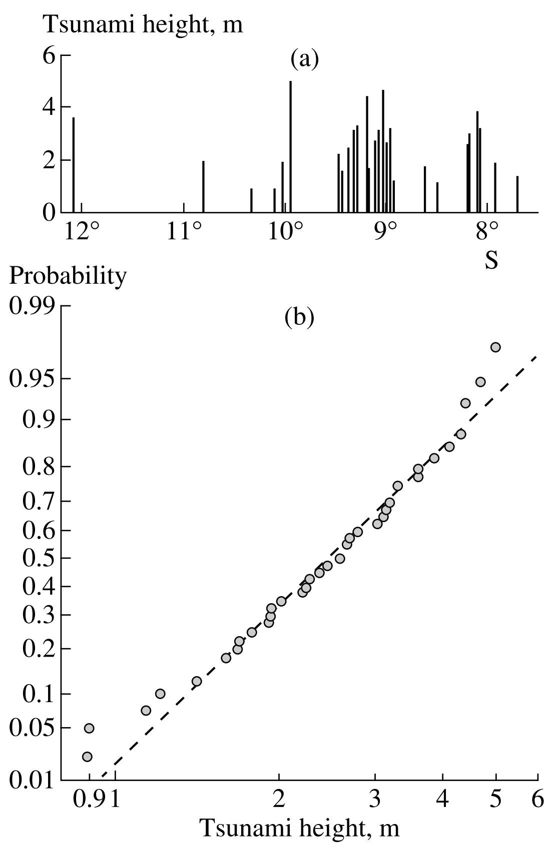

Fig. 8. (a) Distribution of the tsunami heights recorded on November 12, 1996 along the coast of Peru and (b) the distribution of the frequencies (probability) of tsunami heights recorded on the coast. The dashed line corresponds to a lognormal distribution with $\delta = 0.44$ and $\beta = 2.42$ m.

The results of the calculation of the parameters δ and β for three historical tsunamis on the coast of Peru are given in Table 4. Additional estimates of the maximal run-up values are given for all the events in 1901–2001.

It is interesting to compare the values of the dispersion parameter δ for different events. It is clear that the described scheme of forming chaotic tsunami heights on the coast is a rough approximation. The minimal value (0.44) was obtained for the tsunami of February 21, 1996, and the maximal one (1.16) was obtained for the tsunami of October 17, 1966. The mean value obtained for the distribution of the maximal tsunami heights in 1901–2001 is equal to 1.11. The values of β characterize the magnitude of each tsunami, which appeared maximal for the event on June 23, 2001.

The values of the maximal tsunami heights recorded on the coast are usually used in estimating the recurrence of tsunami events. It is possible to prove that the distribution of these values also correspond to the lognormal distribution law. Since the measure of the dispersion of the recorded values of the tsunami heights on the coast is constant for all events, the sample from the set of heights for each event would have the same dispersion. According to Gumbel [3], "If we are interested in longer periods of recurrence, it is enough to consider only the set of maximal values without taking into account all the values exceeding the levels assumed as the basic one." This means that we can interpret the distribution of the maximal heights of the historical tsunamis similarly to the distribution on the coast, i.e., as the lognormal distribution with a dispersion parameter δ, specified properties of the sea bottom in the region, and a parameter of the tsunami risk β, which is determined as the median (geometrical mean) of the distribution of $h_k^{\max}$ of all the maximal tsunami heights recorded on the coast.

Thus, the approximation of the empirical distribution of the recurrence of the tsunami heights by a function corresponding to the lognormal distribution law allows us to not only interpolate and extrapolate the recurrence values but also to estimate the measure of dispersion of tsunami waves by topography separately for each individual region as well as the characteristic tsunami height. It is likely that the introduction of such statistical parameters would be useful for the purposes of tsunami zoning for the comparison of different regions of tsunami hazard.

The statistical estimates obtained are of a regional character. The actual calculated probability means the

Table 4. Results of calculations of the parameters of the lognormal tsunami height distribution on the coast of Peru

Event, date	Magnitude, M_s	Tsunami intensity, I	Number of measurements	Measure of dispersion, δ	Characteristic tsunami height, β, m
Oct. 17, 1966	7.8	1.5	17	1.16	0.40
Feb. 21, 1996	7.3	1.9	39	0.44	2.42
June 23, 2001	8.4	2.6	33	0.73	2.73
1901–2001*	–	–	42	1.11	1.15

* The calculations were performed on the basis of the statistics of the maximal heights for all the tsunami events.

probability of the fact that the event would occur somewhere on the coast of Peru and Chile from 5° to 35° S. In order to pass from the probability at any place to the probability of the event at a specific point of the coastline, it is necessary to use additional models. The authors of [2, 13] suggested an effective method for estimating local tsunami risks. However, in a number of cases, when we consider the characteristics of the coast as a whole, it is reasonable to use the concept of the mean wave height on the coast. The relevancy of using this characteristic is doubtful. Here, we shall consider a stochastic model of the correlation between the maximal heights recorded on the coast with its mean value over the entire time series of observations.

Solov'ev [6] suggested using an empirical relation for the dependence of the mean height of the coast flooding $\bar{h}$ on the maximal run up of a tsunami $h_{\max}$:

$$h_{\max} = 1.1\bar{h} + 0.22\bar{h}^2. \qquad (10)$$

One can construct a similar empirical correlation on the basis of the general properties of the distribution of the tsunami heights on the coast as a sample from the general set of random values. Let us consider the case when we have N recorded values of tsunami heights on the coast for a given event. Assuming that this set of heights is a sample from a single general set, we can use approximation [3] to estimate the maximal value in the sample of random numbers:

$$h_N^{\max} \le \bar{h}_0 + \sigma\sqrt{N-1}, \qquad (11)$$

where σ is the root-mean-square deviation of the distribution of the tsunami heights. It is clear that the value of σ is proportional to the mean height $\bar{h}_0$. Let us also consider that the number of records is approximately proportional to the mean tsunami wave height. This is reasonable because the length of the coast subjected to flooding is linearly related to the height of the wave. In this case, for a sufficiently large N, it is possible to approximate the character of the dependence of $h^{\max}$ on $\bar{h}_0$ as

$$h^{\max} \sim \bar{h}\sqrt{\bar{h}} = \bar{h}^{3/2}. \qquad (12)$$

A graph of the distribution of the dependence of $h^{\max}$ on $\bar{h}_0$ for 17 tsunami events on the Peru and northern Chile coasts from 1901 to 2001 is shown in Fig. 9. The data by Solov'ev [6] for 10 tsunami events in the Pacific Ocean are added to the plot. The straight line $h^{\max} \sim \bar{h}^{3/2}$ is plotted on the graph presented on a logarithmic scale. It is seen that law (12) is followed quite satisfactorily.

It is noteworthy that this stochastic model leads to a contradiction. Actually, relations (11) and (12) mean that, if the coast is studied in more detail, i.e., with a greater number of tsunami measurements on the coast, this would result in greater values of $h^{\max}$. However, this statement is true only within the assumption about the absolute randomness of the distribution of the wave heights: all the measurements are statistically independent and not correlated. In practice, a physical restriction of the correlation scale exists for the variations of the height of flooding on the coast caused by the smoothness of the topography and characteristic wave

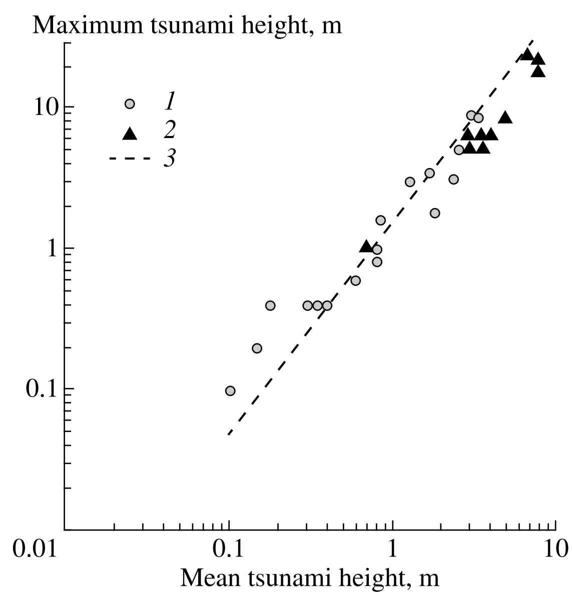

Fig. 9. Relation between the maximal and mean tsunami heights recorded on the coast for different tsunami events. (*1*) The ETDB data for the coasts of Peru and Chile (1901–2001); (*2*) the data from the article by Solov'ev [6]; (*3*) approximation (12).

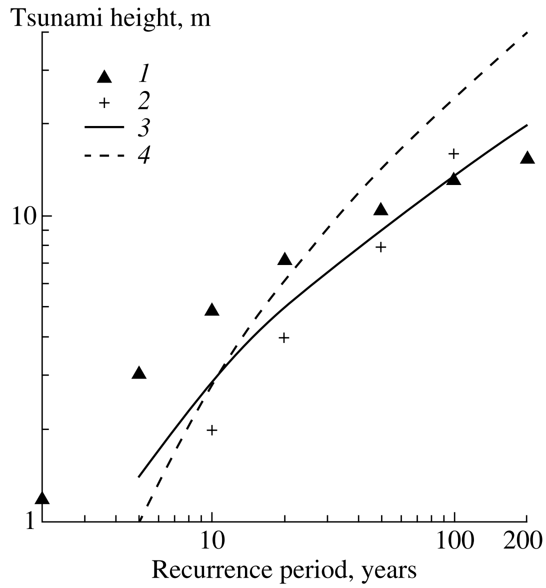

Fig. 10. Summary graph of tsunami height recurrence on the coasts of Peru and Chile. (*1*) Silgado's data [22]; (*2*) Solov'ev's data [6]; (*3*) lognormal distribution calculated on the basis of the ETDB data; (*4*) lognormal distribution calculated on the basis of the NGDC data (1901–2001) [21].

length. In the open part of the coastline, the scale of the tsunami height variability is rarely smaller than 10 km. In the coastal research, only the measurements of extreme tsunami run ups on the coast are usually documented.

The correlation between the maximal and mean tsunami heights observed on the coast was considered in [5]. The authors give an empirical relation $h^{\max}/\bar{h}_0 < 4.5$. It is noteworthy that this relation does not coincide with (11), in which $h^{\max}/\bar{h}_0$ is determined by the character of the tsunami height distribution, by the number of observations, and, generally speaking, is not limited.

This analysis demonstrates that the use of the mean and maximal tsunami heights as statistical characteristics of the tsunami hazard is erroneous. Both values depend on the number of the tsunami observations on the coast. The mean value (as well the maximum one) increases with the increase in the size of the sample for the same random value with a lognormal distribution. As was suggested in [1, 2], an alternative is using the median (geometrical mean) for the tsunami heights recorded on the coast. The mathematical expectation of the value, unlike the mean value, tends to its limit β when the sample size increases.

7. CONCLUSIONS

A comparison of the results of the statistical analysis of the tsunami height data on the coasts of Peru and northern Chile contained in the ETDB and NGDC databases demonstrated significant differences in the estimates of the tsunami heights for long recurrence periods: 50 and 100 years. It was found that these differences are related to the differences in the sets of events presented in these catalogues. Actually, these differences demonstrate the necessity for standardization of the databases and bringing them to conformity.

The recurrence of tsunami heights on the coasts of Peru and Chile is well described by the lognormal distribution. For the periods of recurrence equal to 5, 10, 20, 50, 100, and 200 years, the following estimates of tsunami heights were obtained: 1.3, 2.7, 5.1, 10.2, 16.3, and 24.9 m, respectively.

For the sake of comparison, the estimates obtained from different sources and from the ETDB data are shown together in Fig. 10 (see [6, 21, 22]). The results obtained from the NGDC data in [21] seem to be overestimated for recurrence periods greater than 20 years as compared to other estimates. The obtained estimates are most close to the results published by Solov'ev in 1972 [6].

Unlike the distribution of the probability of earthquake magnitudes, which has a limit value $M_\infty = 9.0$, no limit of this kind was distinguished in the probability distribution of tsunami heights on the coast.

The developed stochastic model of the distribution of tsunami heights on the coast for an individual event relates the maximal tsunami height recorded on the coast to its mean value as $h_{\max} \sim \bar{h}^{3/2}$.

It is planned to use the estimates of the parameters of the lognormal tsunami height distribution in the problems of tsunami zoning as additional characteristics of the tsunami hazard: the geometrical mean tsunami height β (the median of the probability density distribution) and the degree of dispersion (scattering) of tsunami heights on the coast δ.

ACKNOWLEDGMENTS

In conclusion, the authors thank Vyacheslav Gusyakov, head of the Tsunami Laboratory at ICMMG (Novosibirsk), and Garry Rodgers, director of the Pacific Geophysical Center (Sidney, Canada) for useful remarks and consulting. This study was supported by the Russian Foundation for Basic Research, project no. 03-05-64583.

REFERENCES

1. Ch. N. Go, "On the Statistical Assessment of the Height Distribution of Tsunami Waves along the Coast," in *Geodynamics of the Tectonosphere of the Zone of Junction of the Pacific Ocean and Eurasia* (IMGiG, DVO RAN, Yuzhno-Sakhalinsk, 1997), pp. 73–79 [in Russian].
2. Ch. N. Go, V. M. Kaistrenko, and K. V. Simonov, "On the Possibility of Local Long-Term Tsunami Forecast," in *Operative and Long-Term Tsunami Forecasts* (DVNTs Akad. Nauk SSSR, Vladivostok, 1983), pp. 150–162 [in Russian].

3. E. Gumbel, *Statistics of Extremes* (Mir, Moscow, 1965) [in Russian].

4. Kh. S. Kim and A. B. Rabinovich, "Tsunamis on the Northwestern Coasts of the Sea of Okhotsk," in *Natural Catastrophes and Hazards in the Far East Region* (DVO AN SSSR, Vladivostok, 1990), pp. 206–218 [in Russian].

5. E. N. Pelinovsky and I. A. Ryabov, "Distribution Functions of the Tsunami Run-up Heights (From the Data of International Expeditions of 1992–1998)," Okeanologiya **40** (5), 645–652 (2000) [Oceanology **40** (5), 757–766 (2000)].

6. S. L. Solov'ev, "Recurrence of Earthquakes and Tsunamis in the Pacific Ocean," in *Tsunami Waves* (Yuzhno-Sakhalinsk, 1972), pp. 7–47 [in Russian].

7. S. L. Solov'ev and Ch. N. Go, *Tsunami Catalogue of the Western Coasts of the Pacific Ocean* (Nauka, Moscow, 1974) [in Russian].

8. S. L. Solov'ev and Ch. N. Go, *Tsunami Catalogue of the Eastern Coasts of the Pacific Ocean* (Nauka, Moscow, 1975) [in Russian].

9. S. L. Beck and L. J. Ruff, "Great Earthquakes and Subduction along the Peru Trench," Phys. Earth Planet. Inter. **57**, 199–224 (1989).

10. W. H. Berninghausen, "Tsunamis Reported from the West Coast of South America, 1562–1960," Bull. Seismol. Soc. Am. **52** (4), 915–921 (1962).

11. J. Bourgeois, C. Petroff, H. Yeh, *et al.*, "Geologic Setting, Field Survey and Modeling of the Chimbote, Northern Peru, Tsunami of 21 February 1996," Pure Appl. Geophys. **154** (3/4), 513–540 (1999).

12. Ch. N. Go, "The Limitation of Tsunami Height," in *Proc. IUGG/IOC Intern. Tsunami Symp.* (Wakayama, Japan, 1993), pp. 87–93.

13. Ch. N. Go, V. M. Kaistrenko, and K. V. Simonov, "A Two-Parameter Scheme for Tsunami Hazard Zoning," Mar. Geod. **9** (4), 469–476 (1985).

14. V. K. Gusiakov, *ETDB/PAC Expert Tsunami Database for the Pacific, Version 4.9 of December 31, 2003 (CD-ROM)* (Tsunami Laboratory, ICMMG SD RAS, Novosibirsk, 2003).

15. T. Hatori, "Colombia–Peru Tsunamis Observed along the Coast of Japan: Tsunami Magnitude and Source Areas," in *Tsunamis—Their Science and Engineering*, Ed. by K. Iida and T. Iwasaki (Terra Sci. Publ., Tokyo, 1983), pp. 173–183.

16. J. A. Kelleher, "Rupture Zones of Large South American Earthquakes and Some Predictions," J. Geophys. Res. **77** (11), 2087–2103 (1972).

17. P. A. Lockridge, *Tsunamis in Peru–Chile. World Data Center A for Solid Earth Geophysics, Report SE-39* (US Department of Commerce–NOAA, Boulder, CO, 1985).

18. C. Lomnitz, "Major Earthquakes and Tsunamis in Chile during the Period 1535 to 1955," Geolog. Rundsch. **59** (3), 938–960 (1970).

19. E. Norabuena, L. Leffler-Griffin, A. Mao, *et al.*, "Space Geodetic Observations of Nazca–South America Convergence across the Central Andes," Science **279**, 358–362 (1998).

20. E. A. Okal, L. Dengler, S. Araya, *et al.*, "Field Survey of the Camana, Peru Tsunami of June 23, 2001," Seism. Res. Lett. **73** (6), 907–920 (2002).

21. A. B. Rabinovich, E. A. Kulikov, and R. E. Thomson, "Tsunami Risk Estimation for the Coasts Peru and Northern Chile," in *Proc. IUGG/IOC Intern. Tsunami Symp.* (WA, CD, Seattle, 2001), pp. 281–291.

22. E. Silgado, "Recurrence of Tsunamis in the Western Coast of South America," Mar. Geod. **1** (4), 347–354 (1978).