Pure appl. geophys. 164 (2007) 593–603
0033–4553/07/030593–11
DOI 10.1007/s00024-006-0170-7

Pure and Applied Geophysics

Conditional Probability Approach of the Assessment of Tsunami Potential: Application in Three Tsunamigenic Regions of the Pacific Ocean

K. Orfanogiannaki and G. A. Papadopoulos

Abstract—We develop stochastic approaches to determine the potential for tsunami generation from earthquakes by combining two interrelated time series, one for the earthquake events, and another for the tsunami events. Conditional probabilities for the occurrence of tsunamis as a function of time are calculated by assuming that the inter-arrival times of the past events are lognormally distributed and by taking into account the time of occurrence of the last event in the time series. An alternative approach is based on the total probabilitiy theorem. Then, the probability for the tsunami occurrence equals the product of the ratio, r (= tsunami generating earthquakes/total number of earthquakes) by the conditional probability for the occurrence of the next earthquake in the zone. The probabilities obtained by the total probability theorem are bounded upwards by the ratio r and, therefore, they are not comparable with the conditional probabilities. The two methods were successfully tested in three characteristic seismic zones of the Pacific Ocean: South America, Kuril-Kamchatka and Japan. For time intervals of about 20 years and over the probabilities exceed 0.50 in the three zones. It has been found that the results depend on the approach applied. In fact, the conditional probabilities of tsunami occurrence in Japan are slightly higher than in the South America region and in Kuril-Kamchatka they are clearly lower than in South America. Probabilities calculated by the total probability theorem are systematically higher in South America than in Japan while in Kuril-Kamchatka they are significantly lower than in Japan. The stochastic techniques tested in this paper are promising for the tsunami potential assessment in other tsunamigenic regions of the world.

Key words: Tsunami potential, stochastic approaches, Pacific Ocean.

Introduction

The stochastic forecasting of strong tsunami occurrences in tsunamigenic seismic zones is of great importance for the development of reliable tsunami hazard assessment and risk mitigation strategies. However, no standard methodologies have been developed so far to forecast tsunami occurrences. Nevertheless, some statistical and probabilistic methods have been proposed and two main types of approaches can be recognized. The first focuses on the estimation of the recurrence of tsunami

Institute of Geodynamics, National Observatory of Athens, 11810 Athens, Greece.
E-mails: korfanogiannaki@freemail.gr; g.papad@gein.noa.gr

wave generation in particular tsunamigenic sources or zones. A classic effort of this type was presented by SOLOVIEV (1970) for the tsunami recurrence in the Pacific in terms of statistical parameters and probabilities based on relations of magnitude-frequency for earthquakes and tsunamis. The magnitude-frequency statistics for earthquakes combined with the conversion of earthquake occurrence estimates in tsunami probabilities was introduced by TINTI (1993) for the tsunami hazard assessment in Italy. The tsunami intensity-frequency statistics was used by PAPADOPOULOS (2003) to calculate mean recurrence and probabilities of occurrence of tsunamis of different intensity levels in the Corinth Gulf, Greece.

The second type of approach emphasizes the estimation of the recurrence of tsunami wave height in particular coastal sites. LIN and TUNG (1982) combined seismological and hydrodynamic models and analyzed tsunami hazard as the probability of the event that the water elevation at a site exceeds an arbitrary but specified level. Similarly, combining the wave height estimate from a numerical experiment with the probability evaluation of tsunami occurrence in Japan, RIKITAKE and AIDA (1988) evaluated probabilities of a site being hit by a tsunami, of which the wave height exceeds certain levels. According to GO *et al.* (1985), tsunami recurrence for tsunami hazard assessment is described by two parameters: Frequency of occurrence of large tsunamis and coefficient of wave amplification near the shore. The joint probability method was applied by several authors (e.g., SANCHEZ and FARRERAS, 1987; RABINOVICH *et al.*, 1992) for the calculation of expected runup height due to tsunamis and other rapid sea-level changes, like astronomical tides and storm surges.

Very recently, GEIST and PARSONS (2006) developed probabilistic tsunami hazard analysis (PTHA) from the standpoint of integrating computational methods with empirical analysis of past tsunami runup. PTHA is derived from probabilistic seismic hazard analysis, with the main difference being that PTHA must account for far-field sources.

In this paper we introduce the concept of *tsunami potential* which is defined as the probability for the generation of one or more tsunami waves, of a prescribed magnitude or intensity range, in a particular tsunamigenic zone within a particular time window. This definition implies that the approach developed here is of the first type since it does not describe probabilities of tsunami occurrence in coastal sites but probabilities for tsunami generation in tsunami sources. The definition of tsunami potential proposed here is similar to that introduced by NISHENKO (1985) for seismic potential.

Methodology

Postulate that in a seismic zone one or more seismic sources are capable of producing tsunamis and that the tsunami generation in the zone is a point process in time. Subsequently, the tsunami potential of the zone could be approached by considering a probability model for the tsunami generation process in time. As explained in the introductory section, however, such an approach is not new. On the other hand, the calculation of the tsunami probability depends on the data completeness. For very strong events the tsunami catalogues may be complete for the last 100 years or more. However, for strong or moderate tsunamis the catalogues are complete only for the last few decades, which implies that the number of tsunami events available is relatively low, even in the most active seismic zones of the Pacific Ocean, and that the assessment of tsunami potential in probabilistic terms may suffer from reduced reliability. In addition, a hiatus in the high magnitude tsunamis has been observed at least in some regions, like in Sanriku, NE Japan (ADAMS, 1972a,b), which implies that in the existing catalogues the number of high magnitude tsunamis is underestimated due to long repeat time.

To overcome these difficulties of tsunami statistics for the assessment of tsunami potential one may rely on earthquake statistics and calculate the probability that the next earthquake in the zone, whenever it occurs, will be a tsunamigenic one. This implies that the calculation of probabilities is based on the earthquake time series. Tsunami catalogues possibly are more complete than the earthquake catalogues in the last 100 years or so although the total number of events inserted in the complete parts of the catalogues is certainly higher in the earthquake catalogues than in the tsunami catalogues. This is because only a fraction of earthquakes is tsunamigenic.

For the application of probabilistic approaches we implement the total probability theorem, hereafter called TPT for reasons of brevity, combined with a model of conditional probabilities. It is assumed that tsunami generation due to non-seismic sources in the zone, like landslides or earth slumps, is negligible, that is tsunamis are produced only by earthquakes.

Let E, E^c be a partition of the sample space Ω, where E is an earthquake event and E^c is the complementary event of E. Then, according to TPT the probability, $P(T)$, of occurrence of a tsunami event T becomes

$$P(T) = P(E) \cdot P(T|E) + P(E^c) \cdot P(T|E^c), \tag{1}$$

where $P(T/E) = r$ is the ratio of tsunami generating earthquakes over the total number of earthquakes, and $P(E^c)$ is the probability of tsunami occurrence, given that no earthquake has occurred. $P(T/E^c)$, however, equals zero since the assumption has been made that only earthquakes and no other causes generate tsunami waves in the zone. As a consequence the expression (1) is reduced to

$$P(T) = r \cdot P(E), \tag{2}$$

which is the probability of tsunami generation in association with the next earthquake E in the zone; $P(E)$ is the probability for the occurrence of the next earthquake within time interval from t_0 to $t_0 + t$. Probability $P(E)$ becomes simple Poissonian by assuming that the time distribution of earthquakes in the zone is random. The random model implies that the probability for the occurrence of an earthquake is independent of the time of occurrence of the last earthquake. However, other probability models could be adopted as well. We selected to follow the model of conditional probability for the earthquake occurrence as it was introduced by the WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES (1988). This model introduces some memory in the system by means of considering the time elapsed since the last earthquake occurrence in the zone.

The assumption is made that the interarrival time, τ, of successive earthquake events follows a lognormal distribution, that is log τ, is distributed normally. The probability density function of τ is

$$f(\tau) = \frac{1}{\sqrt{2\pi}\sigma\tau} \exp\left[-(\ln\tau - \mu)^2/2\sigma^2\right], \tag{3}$$

where μ = mean of ln τ, σ = standard deviation of ln τ. The probability that the recurrence interval is shorter than τ is

$$F(\tau) = \int_0^{\tau} f(\tau')\, \tilde{d}\tau'. \tag{4}$$

The conditional probability that the next earthquake will occur within the time interval from t_0 to $t_0 + t$ equals to

$$P(E_c) = \frac{F(t_0 + t) - F(t_0)}{1 - F(t_0)}. \tag{5}$$

From equation (5) one may calculate the conditional probability, $P(E_c)$, to observe one earthquake, *Eq1*, in a time interval from t_0 to $t_0 + t$ under the condition that no earthquake occurred after the last earthquake event, *Eq0*, in the seismic zone; where t_0 is a prescribed date after the occurrence of the last event. Then inserting $P(E_c)$ in formula (2) we get the probability, $P(T)$, for the next earthquake to generate a tsunami. In this approach the probability for more than one earthquake event can be calculated. However, for reasons of simplicity one may consider that the calculation is updated after the occurrence of the earthquake *Eq1* by inserting in the data set the time interval from *Eq0* to *Eq1*, and updating the values of μ and σ. Then, the time interval (t_0, $t_0 + t$) is renewed given that t_0 counts from the origin time of *Eq1* instead of *Eq0*.

The conditional probability, $P(T_c)$, for the tsunami generation in the time interval from t_0 to $t_0 + t$ can be independently determined from (5) by considering only the tsunami time series in the zone. This is also the conditional probability for the generation of tsunamigenic earthquakes. However, as mentioned this approach suffers in that the numbers of tsunami events in the catalogues usually are relatively low.

Data and Application

Three characteristic zones of high seismicity and tsunamicity were selected for the application of the methodology described above. They are the regions of South

America, Kuril-Kamchatka and Japan (Fig. 1). Earthquake and tsunami events occurring in these regions were selected from the database of GUSIAKOV (2001) for the time interval from 1900 to 2000. This database indicates that the tsunamis listed in the three regions were produced by strong ($M_s \geq 7.0$), shallow ($h < 50$ Km) earthquakes, where M_s and h are the surface-wave magnitude and focal depth, respectively. Therefore, in our analysis we consider only strong, shallow earthquakes that occurred in the time interval from 1900 to 2000. Completeness analysis based on the magnitude-frequency or Gutenberg-Richter relationship showed that in the three earthquake data sets the completeness magnitude threshold is well below 7.0.

In the database used the tsunami size is quantified in terms of tsunami intensity, I, following the scale introduced by SOLOVIEV (1970)

$$I = \log_2 \sqrt{2}(H), \tag{6}$$

where H (in m) is the mean tsunami height in the coast. Completeness analysis based on the intensity-frequency relationship demonstrated that the tsunami data are complete for $I \geq 0$, $I \geq 2$ and $I \geq -2$ for the regions of South America, Kuril-Kamchatka and Japan, respectively (Fig. 2). Only tsunami complete data sets were introduced in our calculations. The number of earthquake and tsunami events inserted in the data sets are shown in Table 1. As one may expect in the three regions, the number of earthquakes exceeds significantly the number of tsunamis. On the other hand, the

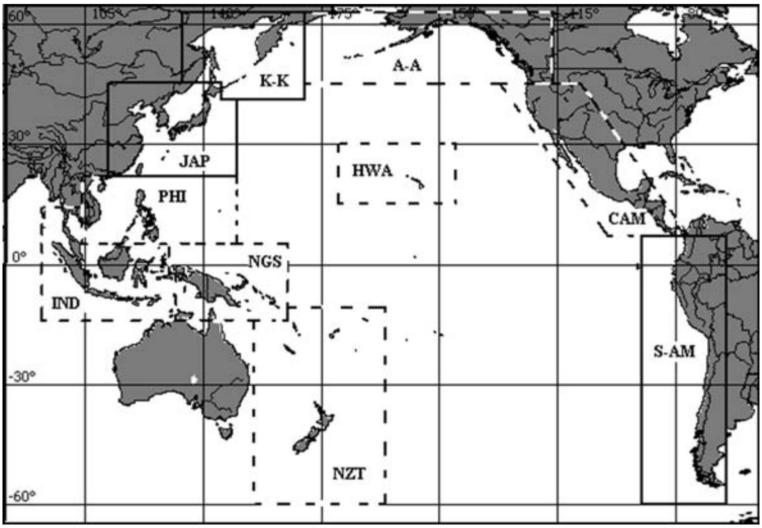

Figure 1
The ten tsunamigenic regions in the Pacific as defined by GUSIAKOV (2001). The solid lines show boundaries of the three regions examined in this paper, while the dashed lines show the boundaries of the remaining regions. *Key*: S-AM = South America, K-K = Kuril-Kamchatka, JAP = Japan, A-A = Alaska-Aleutians, CAM = Central America, NZT = New Zealand-Tonga, NGS = New Guinea-Solomon I., IND = Indonesia, PHI = Philippines, HAW = Hawaii.

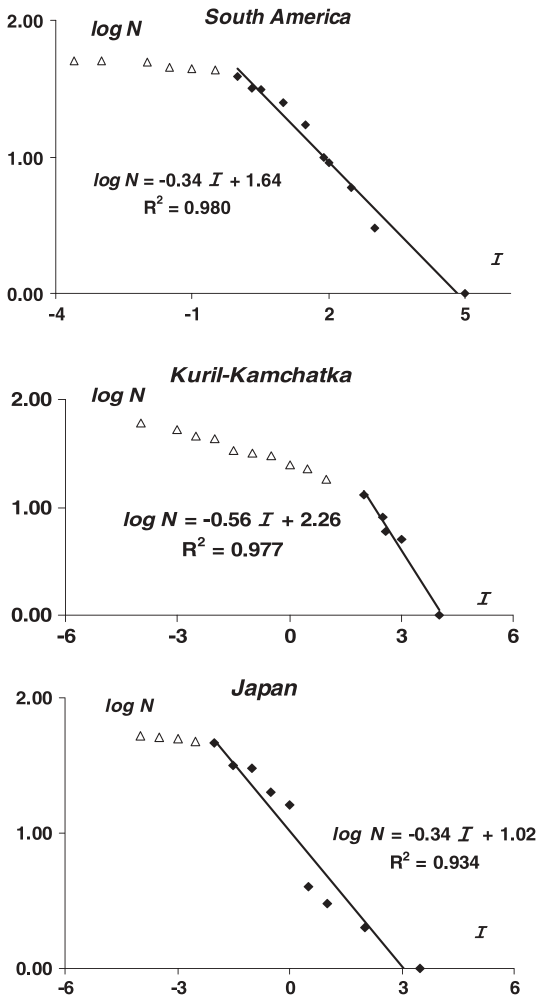

Figure 2
Intensity-frequency relationships for tsunamis occurring in the three zones examined. *Key*: N = cumulative number of events in the time interval from 1900 to 2000, I = tsunami intensity in the SOLOVIEV (1970) scale, R = correlation coefficient. The intensity cut-off for the completeness of tsunami reporting has been determined by the linear regression best-fit at the right side of the diagram. Solid diamond and open triangle indicate the complete and incomplete parts of the data set, respectively.

ratio, r, of tsunami generating earthquakes over the total number of earthquakes varies from 0.39 in South America to 0.12 in Kuril-Kamchatka to 0.26 in Japan.

Results

For each one of the three regions examined, Table 2 lists the conditional probabilities for the occurrences of (i) strong, shallow earthquakes and (ii) strong,

Table 1

Numbers of strong ($M_s \geq 7.0$), shallow (h < 50 Km) earthquakes and tsunamis inserted in the data sets analyzed for the time interval 1900–2000. Tsunami intensity (I) completeness threshold varies

Region	Number of earthquakes	Number of tsunamis
South America	77	30 ($I \geq 0$)
Japan	116	30 ($I \geq -2$)
Kuril-Kamchatka	85	10 ($I \geq 2$)

shallow tsunamigenic earthquakes, which is also the conditional probability for the tsunami occurrence, as well as the probability for the tsunami occurrence as calculated by the total probability theorem. The parameters of the log-normal distribution used for the evaluation of the conditional probabilities of earthquake and tsunami occurrence are shown in Table 3.

Tsunami probabilities concern tsunami events with minimum tsunami intensity as shown in Table 1. The conditional probabilities for the occurrence of the next tsunami are systematically smaller than the respective probabilities for the earthquake occurrence (Fig. 3) since earthquakes occur more frequently than tsunamis. However, the difference in probabilities is not significant for long-time intervals. In the three regions examined, the conditional probabilities for the

Table 2

Parameters introduced and results obtained

Region	t_E	t_{Ts}	d_E	d_{Ts}	t	$P(E_c)$	$P(T_c)$	$P(T)$
South America	31.3.1999	21.02.1996	5.52	8.63	1	0.14	0.07	0.05
					5	0.46	0.27	0.18
					20	0.80	0.61	0.31
					50	0.93	0.81	0.36
					100	0.97	0.91	0.38
Kuril-Kamchatka	04.8.2000	04.10.1994	4.17	10.0	1	0.11	0.05	0.01
					5	0.39	0.20	0.05
					20	0.76	0.48	0.09
					50	0.92	0.68	0.11
					100	0.97	0.81	0.12
Japan	06.1.1995	28.12.1994	9.75	9.78	1	0.15	0.10	0.04
					5	0.48	0.37	0.13
					20	0.80	0.74	0.21
					50	0.92	0.91	0.24
					100	0.97	0.97	0.25

Key: t_E = date of the last earthquake, t_{Ts} = date of the last tsunami, d_E = time interval (in years) since the occurrence of the last earthquake, d_{Ts} = time interval (in years) since the occurrence of the last tsunami, t = time window for the occurrence of the next event (in years), $P(E_c)$ = conditional probability to observe one earthquake in t years from t_E given that no earthquake occurred after the last event, $P(T_c)$ = conditional probability to observe one tsunami in t years from t_{Ts} given that no earthquake occurred after the last event, $P(T)$ = probability by the total probability theorem.

Table 3

Parameters of the log-normal distribution used for the evaluation of the conditional probabilities of earthquake and tsunami occurrence

Regions	*Earthquake data*		*Tsunami data*	
	μ	σ	μ	σ
South Amerika	5.18	1.93	6.43	1.62
Kuril-Kamchatka	4.96	2.09	6.60	2.44
Japan	4.18	1.99	6.18	2.03

occurrence of the next tsunami exceed 0.5 when the time interval exceeds 20 years. The conditional probabilities of tsunami occurrence in Japan are slightly higher than in the South America region. Although the number of events in the data sets of both regions is equal ($n=30$), in Japan the last tsunami occurred earlier than in South America. On the contrary, in Kuril-Kamchatka the conditional probabilities of tsunami occurrence are clearly lower than in Japan and South America due to the relatively low number ($n=10$) of tsunami events inserted in the data set.

The probabilities, $P(T)$, obtained by the total probability theorem are not comparable to the conditional probabilities calculated for a particular time interval. This is because the maximum value these probabilities can attain is not 1, as for conditional probabilities, but r, the ratio of tsunami-generating earthquakes. For instance, in the Kuril-Kamchatka region the probability of tsunami occurrence in the next 20 years based on the total probability theorem is only 0.09. This value appears very low as compared to 1, but is actually high enough with respect to 0.12, the value of r in that region. It is worthnoting, however, that the variation of probabilities $P(T)$ calculated by this method follows the variation of the ratio r from one region to another. In fact, probabilities $P(T)$ are systematically higher in South America than in Japan because the ratio r is higher in South America ($r=0.39$) than in Japan ($r=0.26$). Contrastingly, $P(T)$ are significantly higher in Japan than in Kuril-Kamchatka because in this region r is only equal to 0.12.

Conclusions

The stochastic description of the potential for tsunami generation in a seismic zone has been approached by two probabilistic techniques. The first calculates conditional probabilities for the occurrence of tsunamigenic earthquakes as a function of time by assuming that the inter-arrival times of the past earthquake events are lognormally distributed and by taking into account the time of occurrence of the last event in the time series. Because of the relatively low number of tsunamis in the complete parts of the existing tsunami catalogues, this approach suffers reduced reliability, particularly for short time windows. The second approach is

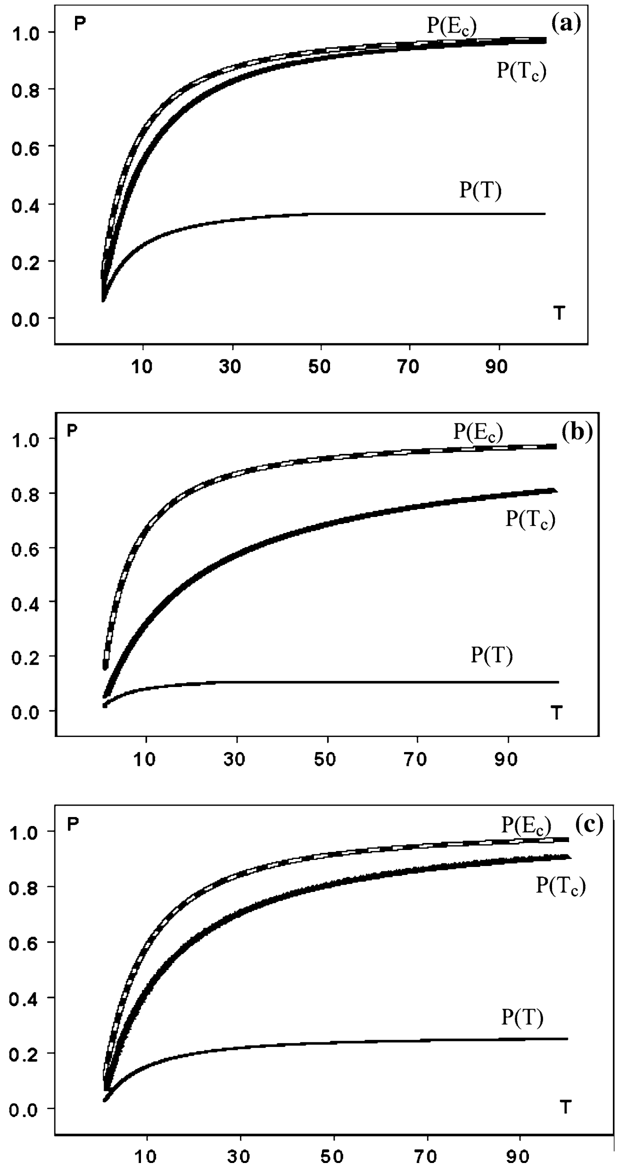

Figure 3

Probabilities, P, for the occurrence of earthquakes and tsunamis in South America (a), Kuril-Kamchatka (b) and Japan (c) as a function of time T (in years). Time starts at the time of occurrence of the last event in each of the zones examined. *Key*: *P(E_c)* (dashed line) = the conditional probability to observe one earthquake given that no earthquake occurred after the last event, *P(T_c)* (thick solid line) = the conditional probability to observe one tsunami given that no tsunamigenic earthquake occurred after the last event, *P(T)* (thin solid line), = probability to observe one tsunami calculated by the total probability theorem.

based on the total probability theorem. Then, the probability for the tsunami occurrence equals the product of the ratio, *r*, of tsunami-generating earthquakes over the total number of earthquakes, by the conditional probability for the occurrence of earthquakes. The probabilities obtained by the total probability theorem are bounded upwards by the ratio *r* and, therefore, they are not comparable with the conditional probabilities.

The two methods were successfully tested in three characteristic seismic zones of the Pacific Ocean: South America, Kuril-Kamchatka and Japan. For time intervals of about 20 years and more the probabilities become high in the three zones. However, it has been found that the results depend on the approach applied. Factually, conditional probabilities of tsunami occurrence in Japan are slightly higher than in the South America region, and in Kuril-Kamchatka they are clearly lower than in South America. Probabilities calculated by the total probability theorem are systematically higher in South America than in Japan, and in Kuril-Kamchatka they are significantly lower than in Japan. The stochastic techniques tested in this paper are promising for the tsunami potential assessment not only in the rest of the Pacific Ocean but also in other tsunamigenic regions of the world.

Acknowledgement

This is a contribution to the EU TRANSFER tsunami research project, contract n, 037058–2006.

References

ADAMS, W.M. (1972a), *A hiatus in the energy range of tsunamis generated off Northeastern Honshu*, Bull. IISEE *9*, 11–25.

ADAMS, W.M. (1972b), *Statistical significance of the energy gap observed for tsunamis generated off the Sanriku coast*, Bull. IISEE *9*, 137–141.

GEIST, E.L. and PARSONS, T. (2006), *Probabilistic analysis of tsunami hazards*, Natural Hazards *37*, 277–314.

GO, C.N., KAISTRENKO, V.M., and SIMONOV, K.V. (1985), *A two-parameter scheme for tsunami hazard zoning, Marine Geodesy 9,* 469–476.

GUSIAKOV, V.K. (2001), *Basic Pacific tsunami catalogs and database, 47 BC – 2000 AD: results of the first stage of the project*, Proc. Internat. Tsunami Symp. August 7–9, 2001, Seattle, USA, PMEL/NOAA, pp. 263–272.

LIN, I.-C. and TUNG, C.C. (1982), *A preliminary investigation of tsunami hazard*, Bull. Seismol. Soc. Am. *72*, 2323–2337.

NISHENKO, S.P. (1985), *Seismic potential for large and great interplate earthquakes along the Chilean and Southern Peruvian margins of South America: A quantitative reappraisal*, J. Geophys. Res. *90*, 3589–3615.

PAPADOPOULOS, G.A. (2003), *Tsunami hazard in the Eastern Mediterranean: Strong earthquakes and tsunamis in the Corinth Gulf, Central Greece*, Natural Hazards *29*, 437–464.

RABINOVICH, A.B., SHEVCHENKO, G.V., and SOKOLOVA, S.E. (1992), *An estimation of extreme sea levels in the northern part of sea of Japan*, Proc. 6th Japan and East China Seas Study Workshop, 22–27 April 1991, Fukuoka, Japan, 30, pp. 179–190.

RIKITAKE, T. and AIDA, I. (1988), *Tsunami hazard probability in Japan*, Bull. Seismol. Soc. Am. *78*, 1266–1278.

SANCHEZ, A.J. and FARRERAS, S.F. (1987), *Tsunami flooding extension for coastal zones of Mexico*, Marine Geodesy *11*, 127–135.

SOLOVIEV, S.L., *Recurrence of tsunamis in the Pacific*. In *Tsunamis in the Pacific Ocean* (ed. Adams, W.M.) (Honolulu, East-West Center Press 1970) pp. 149–163.

TINTI, S., *Evaluation of tsunami hazard in Calabria and eastern Sicily, Italy, Tsunamis in the World*. In (ed. Tinti, S.) *Advances in Natural and Technological Hazards Research* (Kluwer Academic Publishers, Dordrecht 1993) pp. 141–157.

WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES, Probabilities of Large Earthquakes Occurring in California on the San Andreas Fault (ed. U.S. Government Printing Office) (U.S.Geological Survey Open-File Report, Washington, D.C. 1988) pp. 88–398.

(Received March 10, 2006, accepted June 20, 2006)