Heavy-Tailed Distributions in Disaster Analysis

Advances in Natural and Technological Hazards Research

Volume 30

For other titles published in this series go to
http://www.springer.com/series/6362

V. Pisarenko • M. Rodkin

Heavy-Tailed Distributions in Disaster Analysis

Springer

Dr. V. Pisarenko
International Institute of Earthquake Prediction
Theory and Mathematical Geophysics
Russian Academy of Sciences
Moscow, Russia
pisarenko@yasenevo.ru

Dr. M. Rodkin
International Institute of Earthquake Prediction
Theory and Mathematical Geophysics
Russian Academy of Sciences
Moscow, Russia
rodkin@mitp.ru

Revised and updated edition of the book in Russian by V. Pisarenko and M. Rodkin, “Raspredeleniya s tyazelym khvostom: prilozeniya k analizu katastroph”, as part of the book series Vichislitelnaya Seismologiya, Issue 38, Moscow, GEOS, 2007, 240 pp.

ISBN 978-90-481-9170-3 e-ISBN 978-90-481-9171-0
DOI 10.1007/978-90-481-9171-0
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010931864

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Introduction

The end of the twenty-first century saw a dramatically increased interest in safety and reduction of losses from natural and manmade disasters. The cause of this was the combination of a strongly felt social need and the appearance of new theoretical approaches that have significantly progressed in this important interdisciplinary area of study. We wish to emphasize that the increased social need was caused, not only by the (conventionally stressed) growth of the losses due to natural and manmade disasters, but also by the growing interrelationships between different regions of the world. To take one example, the seismic disasters which occurred in China in 1920 and 1927 and entailed the loss of about 200,000 lives each, went almost unnoticed in Europe and in the rest of the world. The similar-sized seismic disasters of 1976 and 2004 which occurred in China again and in Sumatra reverberated throughout the world and stimulated the development of geophysical research and major national and international measures taken to reduce the losses due to possible reoccurrence of such events.

The energy of even a moderate-sized natural disaster and the associated losses is very large. The number of casualties and the loss due to significant natural disasters are comparable with those resulting from regional military conflicts. The energy of natural disasters still exceeds the energy potential of mankind. For example, the yield of the largest nuclear bomb (58 megatons of TNT) detonated in 1961 was 24×10^{23} ergs, while the energy of a major hurricane (recurring at a rate of about two events per year) is estimated to be about 3×10^{25} ergs. The energy of the elastic waves excited by an earthquake with an average rate of occurrence once a year is about 6×10^{23} ergs, while its total energy is about two orders greater, and the total energy of rare great earthquakes is approximately two orders greater still.

The study of disaster statistics and disaster occurrence (here and below, we mean natural and manmade disasters, not catastrophes in a strictly mathematical sense) is a complicated interdisciplinary field involving an intimate interplay of new theoretical results from several branches of mathematics, physics, and computer science, as well as some important applied problems, including socio-economic ones. It is usually thought [VVM] that this research area "is a connecting link between natural, technical, and social sciences". This interdisciplinary character of the area is reflected in the present monograph, which discusses both the limit theorems in mathematical statistics and a possibility of practical realization of the sustainable development of mankind.

There has been little progress so far in the study of disasters (catastrophes) which are sudden and rare processes, and which therefore are little amenable to analysis. The progress that can be seen today in that field crucially depends on the new theoretical approaches that have been developed in several areas of physics and mathematics during the later half of the twentieth century and on the advanced systems now available for environmental monitoring. Understanding the nature of catastrophes essentially relies on new theoretical approaches such as the mathematical theory of catastrophes by Thom and the theory of dissipative structures due to I. Prigogine; other important contributions include P. Bak's concept of self-organized criticality, M.A. Sadovsky's concepts of hierarchical structure and inner activity of the geophysical medium, as well as several other new approaches and conceptual innovations. The statistical studies in the mode of occurrence of natural disasters largely rely on fundamental results in the statistics of rare events derived in the twentieth century. In this connection one can mention R. Fisher, D. Tippet, R. von Mises, E. Gumbel, B.V. Gnedenko, J. Pickands, K. Pickands, and Ya. Galambos.

In relation to natural disasters, it is not so much the fact that the importance of this problem for mankind had been realized during the last third of the twentieth century (the myths one encounters in ancient civilizations show that the problem of disasters has always been urgent), but the realization of mankind's possessing the necessary theoretical and practical tools for effective studies of natural disasters with consequent effective, major practical measures taken to reduce the respective losses. The realization of this situation found its expression in the International Decade for Natural Disaster Reduction adopted by the UN General Assembly in 1989 and in numerous national programs for loss reduction.

All the above factors combine to facilitate considerable progress in natural disaster research. The accumulation of factual material relating to various kinds of natural disasters and the use of advanced recording techniques have expanded possibilities for the analysis of empirical distributions of disaster characteristics. The necessary terminological basis had been developed by N.V. Shebalin and his associates in terms of geophysical magnitude, intensity, and disaster category [She, RS1, RS2].

However, despite the considerable progress reached, the situation with the study of disasters is still far from desirable. It was noted in a review [VMO] that sufficiently complete catalogs of events are still not available for many types of disaster, and the methodological and even terminological bases of research need further development and unification. Not also that the methods of theory of catastrophes and corresponding mathematical approaches are of limited applicability in the very majority of natural disasters because the corresponding potential function is unknown.

The present monograph summarizes our long-continued work in the field of disaster statistics and related questions. We provide a brief description of the terminology and of several modeling approaches and use a broad range of empirical data on a variety of natural disasters. Graded by the amount of factual material, the focus is on seismicity and earthquake loss data, less attention being given to hurricane observations. We also use data on the maximum discharge of rivers, volcanic eruptions, sea-level surges, climatic fluctuations, and manmade disasters.

The main focus is on the occurrence of disasters that can be described by power-law distributions with heavy tails. These disasters typically occur in a very broad range of scales, the rare greatest events being capable of causing losses comparable with the total losses due to all the other (smaller) disasters of this type. The disasters of such type are the most sudden and entail great losses of human life. It is this type of disaster which is meant by the word "disaster" or "catastrophe" in mass media and in mass consciousness. The mode of occurrence and statistics of these disasters are considered for seismic disasters, the information on these being the most complete compared with the others. The monograph contains several new results in the statistics of rare large events. One of the most important results is the conclusion about instability to which the "maximum possible earthquake" parameter is subject, this parameter being frequently used in seismic risk assessment. We suggest alternative and robust ways to parameterize the tail of the earthquake frequency–magnitude relation.

Several results derived in analysis of earthquake loss data seem to hold a universally human interest. For example, the analysis of earthquake losses suggests a revision of the well-known pessimistic forecast of rapid increase of losses from natural disasters; according to this forecast, the increase in the economic potential will as early as in the mid-twenty-first century be entirely consumed by increased losses from natural disasters. We show that the increase in the total number of reported earthquakes is caused by enhanced detection capabilities resulting in the reporting of smaller events rather than by a real increase in the vulnerability of society or increased seismic activity. As to the nonlinear time-dependent growth of total earthquake losses, this effect can be explained (at least the bulk of it) by the peculiar power-law distribution of losses and thus is hardly related to a deterioration of the geoecological environment.

Comparison of the earthquake losses in regions with different levels of economic development suggests a decrease of death toll in economically developed countries. This tendency will extend in the course of time to the third world countries. While the trend of increasing absolute material losses will continue, it is to be expected that the normalized losses (divided by per capita income) will be comparatively stable and even decreasing.

On the whole, the analysis of the relationship between earthquake losses and socio-economic conditions suggests that some sort of equilibrium exists, and that the statistics of losses is compatible with sustainable development.

The monograph has the following structure. The first chapter provides a general overview of the problem, quotes data on different kinds of natural disasters, and gives a classification of these. The second chapter discusses conditions that favor the occurrence of distribution laws typical of disaster magnitude values and gives a consistent description of disasters in terms of distributions and in terms of the mode of occurrence for individual events. The third chapter considers methods in use for a nonparametric description of disasters; these methods are of considerable practical interest in those cases where the actual distribution is either unknown or debatable. The fourth chapter discusses the nonlinear growth over time for characteristic values of the total effect caused by events that obey a heavy-tailed distribution. In the fifth chapter we investigate the distribution of earthquake seismic moment and define the notion of the maximum characteristic earthquake. The sixth chapter is concerned with the distribution of rare, extremely large earthquakes, demonstrating that the maximum possible earthquake concept which is widely used in seismic zonation suffers from instability. The alternative robust approach to parameterization of the tail of the earthquake frequency-magnitude relation is suggested and applied to several cases. The seventh chapter discusses relationships between earthquake losses and the socio-economic conditions, also developing a forecast for characteristic loss values.

This monograph aims primarily at specialists in the field of seismology and seismic risk, but could also be useful for those interested in other kinds of natural and manmade disasters. The basic statistical results (which are largely due to the present authors) are set forth both at the professional and the popular level, thus making them accessible to readers with no special mathematical background. This makes the monograph useful for workers in regional and federal governing bodies, as well as for a broad class of readers interested in the problems of natural disasters and in their effect on the development of mankind. We sought to facilitate the understanding of the problems under discussion (also for nonprofessionals in geophysics and mathematical aspects of risk assessment) by summarizing the results of each chapter at the end of that chapter and outlining the relationships between these and other sections of the book.

This book is a revised and enlarged version of the previous Russian edition of the monograph by the same authors "Heavy-Tailed Distributions: Applications to Disaster Analysis", Computational Seismology, Issue 38, Moscow, GEOS, 2007. The authors are grateful to Prof. A.A.Soloviev (Executive Editor of the Issue) for permission to use the materials of the Russian Edition. Some authors' materials published earlier in a number of papers were used also. These materials are reproduced by permission of the "Russian Journal of Earth sciences" and "Fizika Zemli (Izvestiya, Physics of the Solid Earth)". The authors are very grateful to A.L. Petrosyan for his valuable help in making the English more readable.

We are grateful to the late Nikolai Vissarionovich Shebalin, who was our friend and colleague and provided general guidance of, and himself took an active part in the cycle of studies which have led to the writing of this book. The consideration of several important problems would have been impossible without the initiative and participation of G.S. Golitsyn (Institute of Physics of the Atmosphere, Russian Academy of Sciences, RAS), M.V. Bolgov (Institute of Water Problems RAS), I.V. Kuznetsov (International Institute of Earthquake Prediction Theory and Mathematical Geophysics RAS), A.V. Lander (International Institute of Earthquake Prediction Theory and Mathematical Geophysics RAS), A.A. Lyubushin (Institute of Physics of the Earth RAS), D. Sornette and A. Sornette (Federal Institute of

Technology, Switzerland), Yu.A. Shishkov (Institute of Oceanology RAS), and M. I. Yaroshevich (NPO Taifun Rosgidromet). To all these we are sincerely grateful. We also thank O.D. Voevoda, N.P. Grudeva, T.N. Tseplinskaya, L.S. Kadlubovich, O.A. Matsievskaya, and T.A. Rukavishnikova for the aid they gave during the writing of this book and for valuable advice which has improved the style.

Our disaster studies were supported by the Russian Foundation for Basic Research (grants 93-05-09193, 95-05-14815, 97-05-64583, 99-05-64924, 02-05-64379, 05-05-65276, 09-05-01039, 09-05-92655), by Program 16 of the RAS Presidium “Climate and Environment Changes: Natural Disasters”, and by the Program of State Support of Leading Scientific Schools of the Russian Federation, projects no. NSh-5009.2006.5.

Contents

Chapter 1
Distributions of Characteristics of Natural Disasters: Data and Classification

Any useful classification contains
three to six categories.

From scientists' folklore

1.1 The Problem of Parameterization and Classification of Disasters

This chapter is devoted to a review of empirical data on natural disasters, to a discussion of distribution laws for disaster size and to the description of the approaches used for parameterization and classification of disasters. We wish, first, to acquaint the reader with the great variety of statistical and physical characteristics used to describe different kinds of natural and man-induced disasters. As a result of this review, it appears possible to specify the place occupied in the variety of different kinds of disasters and adverse phenomena by a special class of disasters, namely, those described by distributions with "heavy tails". The present book is mainly devoted to results of research in the statistical characteristics of just this kind of disaster. The term "heavy-tailed distribution" is commonly used for a distribution density that decreases at infinity slowly enough, for example, more slowly than any exponent. We will use the term in a narrower sense, namely, for distributions with an infinite mathematical expectation. For such distributions the law of large numbers and the central limit theorem in the theory of probability do not hold, and thus the standard statistical characteristics such as sample mean and variance are inapplicable. The distributions possessing this property – being theoretically with an infinite mean - will be called heavy-tailed distributions in what follows.

The analysis outlined later in this chapter includes not only a review of available data sets on natural disasters (such as the USGS and EM-DAT Data Bases), but also a discussion of the terminology and of the approaches to parameterization and classification of disasters. The need for a discussion of these methodological aspects

V. Pisarenko and M. Rodkin, *Heavy-Tailed Distributions in Disaster Analysis*,
Advances in Natural and Technological Hazards Research 30,
DOI 10.1007/978-90-481-9171-0_1, © Springer Science+Business Media B.V. 2010

arises from the absence of a commonly used and effective theoretical base and terminology in this interdisciplinary field of research. The situation is aggravated by the use of specific ways for description of disasters and of specific functions in the fitting of empirical data for different kinds of disasters. The great variety of the approaches now in use complicates the comparison of data on different kinds of disasters and the development of a unified system of measures to take for loss reduction following various possible natural and man-induced disasters.

Note that the study of distribution laws for different natural disasters (earthquakes, volcanic eruptions, floods, hurricanes, etc.) is necessary not only for risk assessment, but it is useful also for understanding the physical nature of the underlying processes.

The earthquake statistics is the best-studied case among disasters. The distribution laws for the other kinds of natural disasters are much less known. For most kinds of natural disasters (floods, thunderstorms, etc.) the relevant empirical data on disaster occurrence are traditionally fitted by a specific parametric distribution function. These approaches are often injustified statistically (especially in the range of rare greatest events), and why a particular parametric distribution should be chosen is not quite clear. A formal character of the distribution functions in use and the ambiguity of their physical meaning tend to restrict their applicability for extrapolation into the domain of rare large events.

An alternative to the description of empirical distributions in terms of different historically evolved specific parametric distributions is the use of the "classical" physical distribution laws. These "classical" distributions frequently encountered in physical systems are the Gaussian (normal) distribution, the Boltzmann (exponential), the Pareto power-law distribution, as well as modifications of these. Each of these distributions occurs under some set of conditions that are fairly clear and widespread in natural systems. Consequently, one obvious advantage for the use of these "classical" distributions is the possibility of physical interpretation of the results and various physical analogies of the phenomena. It seems reasonable to assume that, if empirical data for a given kind of disasters can be described by one of these distributions, then the conditions generating disasters of the kind in hand are similar to the conditions for the occurrence of the corresponding "classical" distribution.

In view of the prevalence of conditions that favor the occurrence of classical distributions, they can be considered as the most natural and expected. Accordingly, if a given empirical distribution can be described well enough by one of the classical physical distributions, it is reasonable to consider such parameterization as the more preferable compared with other possible laws, historically evolved or chosen more or less formally. The use of classical distribution laws also inspires some hope for deriving correct results when the parameterization is applied to other similar situations or in the domain of rare events. Considering the aforesaid, preference is given below to the description of empirical distributions by one (or a combination) of classical physical distributions. This approach will prove to be productive enough throughout this book.

The empirical distributions for many kinds of natural disasters have not been sufficiently studied because of shortage of statistical data for corresponding kinds of natural disasters. Insufficient development of the parameterization techniques has also affected the situation. In both these directions considerable progress has been made recently. The use of new recording methods has considerably improved data acquisition. In the methodology, the conceptual scheme developed by [She, RS1, RS2] provides a sound methodological basis for disaster parameterization. It includes explicit definitions of the notions of geophysical magnitude of a disaster, its intensity, and its category.

It would be most natural to use the energy released in a disaster as the magnitude (size) of that disaster (earthquake, hurricane, etc.). However, for most cases the energy of an event cannot be calculated directly, but only an indirect energy-related characteristic can be estimated. The magnitude of disasters is usually calculated from an available, easily derived physical parameter that is connected with energy. An example of such an approach is the classical definition of earthquake magnitude as the logarithm of the maximum amplitude of a certain seismic wave corrected for distance from the epicenter.

The intensity of a disaster characterizes its impact at a given site (upon the objects located nearby: people, buildings and various man-made structures and natural objects). The intensity depends on the magnitude of the disaster, on the distance from the epicenter, and on the site effects. A particular disaster is characterized by a set of intensity values describing the impacts at some points of the epicentral area.

The disaster category is characterized by the damage caused by the disaster. The disaster category depends on several parameters: number of casualties, number of injured, direct and indirect loss of property. Exact values of the loss, as well as a full description of different kinds of the damage are very seldom available, however. Even till very recently, the loss was characterized usually only by the death-toll, and sometimes by direct economic losses.

Before passing to a discussion of the data sets characterizing different kinds of disasters, we wish to note a few general points. Smaller events are often less completely reported in the catalogs of disasters. This reduction can stem from several causes. For the majority of cases we have an incomplete reporting of smaller events. In this connection it is unpromising as a rule to search for the unique distribution function that would be valid in the entire range of magnitude of a given kind of disaster. Possible errors in the number of smaller disasters do not however seriously affect the estimation of losses because of the insignificant input of small disasters into the total loss, despite the large number of these events.

We notice, however, that the reduction is sometimes due to the fact that in the range of small events the fitted law (e.g., the Pareto law) is invalid. Such a situation arose, for example, in predicting the number of smaller deposits on the assumption that their distribution is the self-similar Pareto law [KDS, BU1, RGL].

The next point is connected with the very wide range (several orders of magnitude) of disaster size and disaster loss. Because of this, and also because the empirical loss distributions frequently obey power-law, a log scale was suggested for the classification of disasters [RS1]. Then the category of disaster changes from local (weak) accidents to as far as disasters on a planetary scale. The scheme accepted now in the Ministry of Emergencies of the Russian Federation uses a similar classification system for accidents by severity levels of local, territorial, regional, and federal significance [SAK]. For the same reason many empirical scales of natural disasters (earthquake magnitude and intensity, the Iida tsunami scale, scale of volcanic eruptions) are logarithmic. Other widespread scales of natural effects include the Beaufort wind velocity scale, the Saffir-Simpson scale for hurricanes, the Fujita and Parson scale for tornadoes; all of these are compromises between the linear and logarithmic principles. Detailed descriptions of the basic scales and references to such descriptions are given in [Fu, Bl].

1.2 Empirical Distributions of Physical Parameters of Natural Disasters

The empirical distributions of physical parameters describing different natural and man-induced disasters and adverse phenomena are examined below. The analysis is given in the order of possible data fitting using the normal, exponential, and power-law distributions. All data sets are shown to be suitable material to be fitted by one of these classical distributions or by a combination of these.

The empirical cumulative non-normalized distribution functions of temperature deviations from the mean seasonal values for St.-Petersburg are shown in Fig. 1.1 along with the fitted normal law. As can be seen, these data can be described well

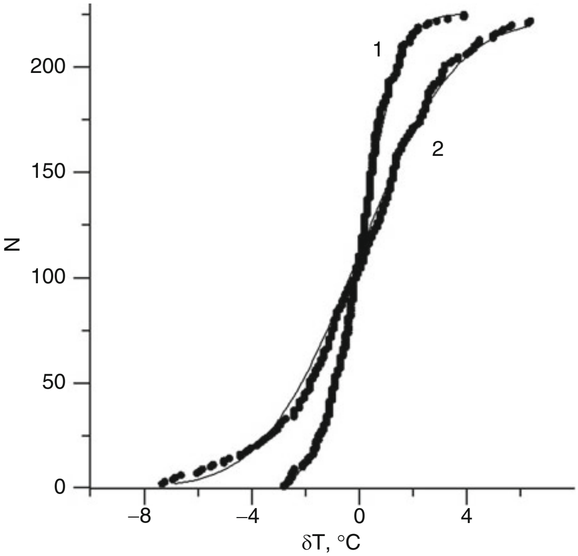

Fig. 1.1 Distribution of δT deviations of seasonal temperature from the means for St.-Petersburg for 1744–1980 (*dots*), data from [Bo], (**1**) summer, (**2**) winter. The *line* gives fitted normal distribution laws

enough by the normal distribution. The chi-square test shows that the data are consistent with the normal distribution at a confidence level greater than 90%. Considering the short time series used (220–240 data points or years), this confidence level appears to be quite acceptable.

The wide prevalence of the normal distribution in application to natural phenomena is due to the conditions required for the occurrence of this law. The normal distribution arises in cases where the parameter of interest depends on a sum of independent factors none of which is prevailing or infinite. Situations such as this are rather common and fulfill the conditions required for the Central Limit Theorem in probability theory on the convergence of distributions of the sum to the normal (Gaussian) law. The distribution function for the normal law is

$$F(x) = \int_{-\infty}^{x} (2\pi\sigma^2)^{-1/2} exp\left[-(t-m)^2/2\sigma^2\right] dt, \tag{1.1}$$

where m, σ are the parameters: mean and variance, respectively.

The probability density of the normal law has the form

$$f(t) = \left(2\pi\sigma^2\right)^{-1/2} exp\left[-(t-m)^2/2\sigma^2\right]. \tag{1.1a}$$

It can be seen that the probability of deviation $\delta\varsigma = (t-m)$ quickly decreases as $\delta\varsigma \to \infty$. The probability of deviation $\delta\varsigma$ exceeding 3σ can often be disregarded in practice (its value is about 0.002).

The next widespread distribution is the Boltzmann exponential law. This distribution law is also typical of catastrophic processes. Its occurrence in physical systems is treated as the distribution of energy values for a set of particles being in thermal equilibrium with a thermostat. Empirical non-normalized complementary distribution functions for sea level variation amplitudes as observed at a number of points along the east coast of Canada for 1965–1975 are presented as examples in Fig. 1.2. Distributions of wind velocity recorded at weather stations are similar in character.

The distribution function for the exponential law is

$$F(x) = \int_{-0}^{x} 1/x_0 \; exp(-t/x_0) dt, \tag{1.2}$$

where the parameter x_0 is the mean. According to a popular physical interpretation, x_0 is equal to the mean energy of particles at temperature T. More exactly, for the case of ideal gas model, we have $x_0 = kT$ where k is the Boltzmann constant. The associated probability density has the form

$$f(t) = 1/x_0 \; exp(-t/x_0). \tag{1.2a}$$

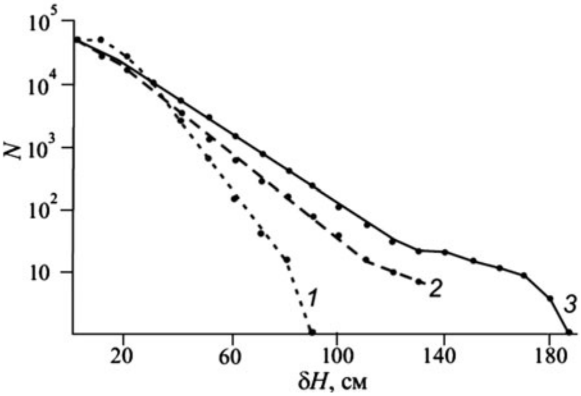

Fig. 1.2 Distribution of the number N of sea level changes with amplitude not less than δH at the east coast of Canada for 1965–1975, data from [EMB]: (**1**) sea tides for the Saint Johns station, (**2**, **3**) tides and ebbs for the Rivier-du-Lop station, respectively

The observed sea level variations and wind velocity distributions [Ku, GD, SY] corroborate the possibility of a good exponential fitting, though the authors do not always use this fit. The distribution parameters based on different data sets vary widely depending on the season and registration conditions (for example, site elevation above the ground surface in measurements of wind velocity).

The case of sea level variations can evidently be interpreted as variations in the potential energy of a system and the empirical distribution can thus be treated as an analogue to the classical Boltzmann distribution. The interpretation is less evident for the case of wind velocity. Naturally, wind velocity itself is not the same thing as energy of weather phenomenon. It seems a plausible hypothesis, however, that wind velocity and thermal energy of the relevant weather process can be connected by a linear regression. Actually, air pressure and wind velocity correlate closely (see below an example of the evolution of hurricanes). At the same time, according to the universal gas law, the pressure change in a gas is proportional to the change in temperature. It is reasonable to suppose that wind velocity is related statistically to temperature variations and consequently to the energy of the associated weather phenomenon. It thus appears that the exponential distribution of wind velocity could be treated as an example of a distribution of the Boltzmann type.

The distributions of wind velocity when measured in hurricanes (also referred to as typhoons and tropical cyclones) and tornadoes can differ from the distribution of wind velocity under ordinary conditions. The statistics of wind velocity for tropical cyclones occurring in the western part of the Pacific Ocean were studied in [GPRY], where the cyclone hazard for a few large cities in Southeast Asia and the Far East was evaluated. According to the United Nations data for 1962–1992, the tropical cyclones cause more than 20% of all loss of life due to natural disasters [O1]. We are going to consider the case of typhoons in more detail.