UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL ATOMIC ENERGY AGENCY
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/1011 - 4

Fourth Workshop on Non-Linear Dynamics and Earthquake Prediction

6 - 24 October 1997

Catalogs of Earthquakes

V.G. KOSSOBOKOV

**International Institute of Earthquake Prediction Theory and Mathematical Geophysics
Russian Academy of Sciences
Moscow 113556, Russian Federation**

Catalogs of Earthquakes

V.G. Kossobokov

Introduction

Are earthquakes real? Some people have "felt" the ground motion from small earthquakes, and some people have experienced a large damaging earthquake, but most of the hundreds of earthquakes that occur every day are not "felt" or otherwise experienced by a human being. The only evidence for their existence are wiggles on a seismogram and this evidence is usually lost unless a number of seismograms are analyzed to determine a location and a size for that earthquake. These data are often lost unless they are included in some standard catalog that survives. So the question "Are earthquakes real?" becomes "Are the catalogs real?". The reality or usefulness of a catalog must be judged by the predictions we can make from the data in the catalog.

There are many indications that earthquakes are not random events: (1) there are more small earthquakes than big ones, however, the energies released in earthquakes are dominated by the largest events; (2) when averaged over a large space and time interval, earthquakes follow the *magnitude-frequency law*, i.e., Gutenberg-Richter relationship; (3) earthquakes have an inhomogeneous spatial distribution (for small earthquakes this distribution is well-defined in statistical sense); (4) earthquakes do cluster in time and are often classified into *foreshocks, mainshocks, and aftershocks*; e.t.c.

So far the nature of earthquake sequences is neither completely understood nor sufficiently studied. Most of seismologists believe that earthquake occurrences are chaotic and non-predictable. They usually cite the late Charles F. Richter, the "curmudgeonly doyen of American earthquake seismology", who used to say that "*only charlatans and fools predict earthquakes*". There is good reason for this scepticism because the subject attracts quite a number of "oracles" which claim prediction of a disaster without any systematic analysis and/or research. Although the Richter's opinion does not reject a possibility that earthquakes are predictable, its influence on formation of earthquake prediction community was sort of a crucial one putting the subject beyond the limits of respectable scientific study.

Many "precursory" and other phenomena described in terms of earthquake sequences have been reported, however, a few of them are defined in a reproducible manor and even less are confirmed by independent testing. A simple question arises: *Can one part of a catalog be used to predict another part of a catalog?* The answer is clearly "*Yes*", if the catalog was compiled in a standard systematic way. When we plot the locations of earthquakes in any given year from the NEIC Hypocenter Data Base, most of the earthquakes are located in the bands of activity where the earthquakes occur in previous years. When we count the annual number of earthquakes above magnitude 7.5 or greater worldwide, we notice that this number never exceeded 10 during this century: thus, we can be absolutely sure that next year the number will be less than 20. These kinds of obvious predictions encourage us to apply the methods of statistics and pattern recognition to the catalogs in attempt to predict the largest earthquakes. These notes and a supplementary package of computer programs intend to facilitate a researcher who wants to retrieve information from an earthquake catalog available.

Hypocenter Data Bases

Although instrumental observation of earthquakes started more than a hundred years ago, the seismic instrumentation and data acquisition procedures keep changing in time. This raises the question how uniform the earthquakes have been observed. There are many earthquake catalogs that were published and are being published by regional, national, and international agencies operating seismic networks and/or compiling data from primary sources. An earthquake catalog usually contains a list of earthquakes described by the five basic parameters: *origin time, latitude and longitude of the epicenter, focal depth, and magnitude.* These parameters are determined from readings of seismograms recorded by seismographs. Although there is a general agreement on the reading procedure, values of the readings may differ between analysts. Furthermore, it is easy to make mistakes, manually and/or automatically. However, this problem could be resolved if many readings are available from different independent sources and the parameters' determination is careful. There are many different location programs and for a "well-observed earthquake" they give similar results. But sometimes the sets of parameters for the same earthquake may be very different when reported by different agencies. The nightmare of observational seismology is the "magnitude" problem. There are several magnitude scales which are commonly used. But even within the same magnitude scale the actual practice is often different for different agencies. Furthermore, the practice of a single agency may change in time.

The following three data bases are usually considered.

The NEIC Global Hypocenter Data Base is produced by the United States Geological Survey/National Earthquake Information Center at Denver, Colorado. It consists of the *Global Hypocenter Data Base CD-ROM* and the Extended Preliminary Determination of Epicenters (PDE) Catalog. The *Global Hypocenter Data Base CD-ROM* (version 3.0) provides information on 915,655 earth tremors (natural and artificial) from 2100 B.C. through December 1992. This data set was assembled over a period of several decades. Many institutions contribute their information to it. The EPIC rretrieval software of the NEIC GHDB makes use of forty-six source catalogs. The USGS has not been able the sources for some catalogs. These are retained in the data base because they clearly represent the cataloging efforts and may include significant events that are not included in other catalogs. Each earthquake in the data base is detailed according to source, date, time, latitude, longitude, magnitude, depth, intensity, and associated phenomena. Two or more records for the same quake may result when searching the data set.

In the NEIC catalog up to four magnitudes may be reported. These are the average body wave magnitude, mb, the average surface wave magnitude, Ms, and two authority magnitudes. In the past these authority magnitudes were mainly reported by Cal Tech and Berkeley. In recent years, after 1992, these authority magnitudes have been replaced by moment magnitudes reported by Harvard University and the Geological Survey. The moment magnitudes seem to average about 1/3 unit higher than surface wave magnitudes.

The ISC Bulletin and Source Parameters was recently recompiled by International Seismological Centre at Newbury, United Kingdom. Unfortunately, we did not manage to get the new version of the Data Base yet. The previous version published on CD-ROM in Denver, Colorado, report on earthquakes from January 1,1964 through August 31, 1987. The primary goal of this cataloging was to provide the original data on phases. Although the data set reports more seismic events than NEIC GHDB, many of these are lacking determination of magnitude (surface wave magnitude, Ms, in particular).

The Harvard University Centroid-Moment Tensor Catalog is compiled in the Department of Earth and Planetary Sciences, Harvard University, and is widely available by FTP from Goran Ekstrom (ekstrom@geophysics.harvard.edu). The original description of the CMT method can be found in (Dziewonski, Chou, and Woodhouse, JGR, 86, 2825-2852, 1981; Dziewonski and Woodhouse, JGR, 88, 3247-3271, 1983). All the moment tensor solutions are being published in Physisc of the Earth and Planetary Interior as a quarterly reports. The reports also contain descriptions of various enhancements of the original algorithm, including the incorporation of laterally heterogeneous models of the Earth and the use of broad band data for depth determination.

Tools to analyze an earthquake catalog

It is ultimately the user's responsibility to assess the accuracy and completeness of a data-set and to determine if these are sufficient for the purposes to which the data set would be applied. To help the user in analysis of a seismic catalog the scientists from the International Institute for Earthquake Prediction Theory and Mathematical Geophysics have developed a problem-oriented software. The core of the package consists of the following programs - EDCAT, CATAL, HIST, and AFT.

The program EDCAT transforms a catalog of earthquakes written in any of more than twenty optional formats into one of the two IIEPT&MG standards, i.e., the ASCII 41-bytes format and the binary 20-bytes format. The program permits to check for erroneous records, like February 30 or 100 degrees latitude, as well as to search out duplicate entries and make corrections either automatically or manually. It can sort the catalog by ascending time. This kind of natural ordering is assumed in other applications from the package.

The program CATAL performs a number of operations on catalogs. These optional operations permit to select a subcatalog with parameters from given ranges, to compare two catalogs, to merge two catalogs into a composite one, etc. The set of operations appears to be very useful and powerful in catalog processing.

The program HIST makes a wide variety of histograms based on events from a catalog, e.g, magnitude-frequency distribution of events by year, distribution of epicentral locations sliced by depth, and many other.

The program AFT eliminates obvious clustering of events in a seismic catalog into mainshocks and aftershocks. It uses a simple rigorous definition of an aftershock as a lower magnitude event following the mainshock in its vicinity. The user has a wide variety of optional limitations which specify space and time span of aftershocks as functions of the mainshock magnitude. If requested the program outputs the catalog of mainshocks and counts the number of aftershocks for each mainshock.

These basic tools permit to make a reasonable assessment of the accuracy and completeness of a catalog of seismic events in subject, and, accordingly, to conclude if they are sufficient for further investigation. It's a good point to remind that no chef with the best cooking technique can prepare a consumable meal if only rotten ingredients are available.

Catalog errors

As mentioned above there are many reasons to believe that practically all catalogs of earthquakes have errors. Of course the presence of errors may render invalid a study based on such a catalog. In general, the errors may be avoided in two ways - (1) by postponing the analysis until the data are refined, and (2) by a robust analysis of existing data within the limits of its applicability. The danger of pursuing the first way only was long ago indicated by a founder of non-linear mechanics and computational methods Professor N.Krylov: "Undue precision of computations is the first symptom of mathematical illiteracy".

Errors can be divided roughly into two categories of "sporadic" and "systematic". Sporadic errors occur as a result of mistakes of different nature, e.g., mistyping when entering data manually (which could be pretty systematic, in some cases). Systematic errors are usually associated with changes in a data acquisition system, e.g., enhancement of the original definition of parameters. Such an enhancement could easily affect the integrity of the data set and any decision to introduce it must be taken with a great precaution.

Most of the errors could not be corrected unless we refer to the original source records. In many cases this opportunity to correct errors has been lost already. However, many errors could be detected when several catalogs available. We strongly recommend to match catalogs available for equivalent and non-equivalent entries, and, for a set of equivalent entries, we recommend to compare the values of parameters (e.g., magnitudes) from different sources.

Many catalogs of earthquakes are composite ones where a number of local and/or regional data merged together. These may be heavily populated by duplicate entries. For obvious reasons, such catalogs are of limited use and usually would require a removal of duplicates.

What can we learn analyzing a catalog of earthquakes?

There are two extreme opinions on this subject. The pessimistic one states that "in the case of seismicity data, most of the observed variations are, in fact, related to changes in the system for detecting and reporting earthquakes and not to actual changes in the Earth". Thus implying that earthquake catalogs in their present form are of very limited if any use. The optimistic opinion is based on the fact that among existing data seismic catalogs remain the most reliable record on distribution of earthquakes in space and time. This apparently nonuniform distribution displays many features of predictability. In general sense, this was proved long ago by statistical studies of the occurrence and spatial distribution of earthquakes (see, for example, papers by L. Knopoff and Y. Kagan).

We can predict the location of most strong earthquakes. They occur at locations where there is a high level of moderate or low earthquake activity, and the locations of high or moderate earrthquake activity tend to be associated with the active boundaries of tectonic plates. Actually the entire field of Plate Tectonics would hardly have been possible without seismic evidence.

We can put estimates on the future average rate of occurrence of strong earthquakes. The history of the earthquake cataloging as one can easily observe from the global annual magnitude frequency graph reflects - 1) an overall increase in the number of reported earthquakes from the beginning of the century to the early 1960's, which corresponds to the improvement of seismographs; 2) sharp changes at the times of installation of California (1932) and the World Wide Seismic Standard Seismograph Networks (1963); 3) stability of the number of the magnitude 7.0 and larger events from the beginning of the century; 4) uniform width of bands since 1963, in agreement with the Gutenberg-Richter relationship, down to magnitude 5.0. From a statistical viewpoint, since about 1963 the catalog appears to be surprisingly consistent in reporting magnitude 5.0 and above earthquakes: the magnitude bands have almost the same width in logarithmic scale, in agreement with the Gutenberg-Richter relationship, i.e., log N ~ (8 - M). One can see also that the list of earthquakes above 7.0 in the NEIC GHDB is probably complete from the beginning of the century. Such a remarkable stability suggests that global underlying processes in a time scale of decades stationary and encourages research aimed at intermediate-term prediction of earthquakes.

Predictions of the time of future large earthquakes, in a sense that we identify periods of time when they are most likely to occur, are often very uncertain. Many methods for earthquake prediction have been proposed and some of them may be useful. The problem is that most of these methods cannot be adequately tested and evaluated in our lifetime. The methods we present at this Workshop are rare exclusions. They are reproducible algorithms, each having the ultimate formulation as a computer code and the list of fixed parameters. The predictions issued by these algorithms are intermediate-term, i.e., the time-span of a prediction lasts from several months to several years while the territorial uncertainty ranges from about one to ten source zones of the incipient earthquake. At present the M8 algorithm which has predicted several large earthquakes is subject of a rigid testing in real-time research prediction of the greatest earthquakes in Circum Pacific.

- At the moment a milestone at which the statistical significance of the M8 algorithm has been justified is behind us: the algorithm predicted five out of five earthquakes of magnitude 8.0 and above since the beginning of our testing in 1992. We hope that our approach will lead eventually to more objective estimates of the earthquake risk.

***

I conclude citing from p.603 of Numerical Recipes by W.H. Press et al. some key principals of the data analysis:

> **Data consist of numbers, of course. But these numbers are fed into the computer, not produced by it. These numbers to be treated with considerable respect, neither to be tampered with, nor subjected to a numerical process whose character you do not completely understand. You are well advised to acquire a reverence for data that is rather different from the "sporty" attitude that is sometimes allowable, or even commendable, in other numerical tasks.**

The analysis of data inevitably involves some trafficking with the field of *statistics*, that gray area which is not quite a branch of mathematics - and just as surely not quite a branch of science. In the following sections, you will repeatedly encounter the following paradigm:

- **apply some formula to the data to compute "a statistic"**
- **compute where the value of that statistic falls in a probability distribution that is computed on the basis of some "null hypothesis"**
- **if it falls in a very unlikely spot, way out on a tail of the distribution, conclude that the null hypothesis is *false* for your data set**

If a statistic falls in a *reasonable* part of the distribution, you must not make the mistake of concluding that the null hypothesis is "verified" or "proved". That is the curse of statistics, that it can never prove things, only disprove them! At best, you can substantiate a hypothesis by ruling out, statistically, a whole long list of competing hypotheses, every one that has ever been proposed. After a while your adversaries and competitors will give up trying to think of alternative hypotheses, or else they will grow old and die, and *then your hypothesis will become accepted.* Sounds crazy, we know, but that's how science works!

***

References

Gabrielov, A. M., O. E. Dmitrieva, V. I. Keilis-Borok, V. G. Kossobokov, I. V. Kouznetsov, T. A. Levshina, K. M. Mirzoev, G. M. Molchan, S. Kh. Negmatullaev, V. F. Pisarenko, A. G. Prozoroff, W. Rinehart, I. M. Rotwain, P. N. Shebalin, M. G. Shnirman, and S. Yu. Schreider (1986). *Algorithms of long-term earthquakes' prediction*, CERESIS, Lima (Peru), 61 p.

Gutenberg, B., and C.F. Richter (1944). Frequency of earthquakes in California, *Bull. Seism. Soc. Am.*, **34**, 185-188.

Gutenberg, B., and C.F. Richter (1944). Magnitude and energy of earthquakes in California, *Annali di Geofisica*, **9**, 1-15.

Global Hypocenters Data Base CD-ROM (1989). NEIC/USGS, Denver, CO

Global Hypocenters Data Base CD-ROM. Version 3.0 (1994). NEIC/USGS, Denver, CO

Habermann R. E. and F. Creamer (1994). Catalog Errors and the M8 Earthquake Prediction Algorithm, *Bull. Seism. Soc. Am.* **84**, 1551-1559.

Healy, J. H., V. G. Kossobokov, and J. W. Dewey (1992). A test to evaluate the earthquake

prediction algorithm, M8, *U.S. Geol. Surv. Open-File Report* **92-401**, 23 p. with 6 Appendices.
Healy, J. H., V. G. Kossobokov, and J. W. Dewey (1994). The ongoing test of the M8 earthquake prediction algorithm, *EOS Transactions* **75**, No. 44, 1994 AGU Fall Meeting, S31B-5.
International Seismological Centre (1990). Associated Phase and Comment Data, ISC Bulletins, 1964-Aug. 1987, ISC Bulletin and Source Parameters, version 1.0, U.S. Geological Survey National Earthquake Information Center CD-ROM, World Data Center-A for Seismology, Box 25046, MS 967, Denver Federal Center, Denver, CO 80225.
Keilis-Borok, V. I. (Editor) (1990). Intermediate-Term Earthquake Prediction: Models, Algorithms, Worldwide Tests, *Phys. Earth Planet. Inter.* **61** (Special Issue), Nos. 1-2.
Keilis-Borok, V. I. and V. G. Kossobokov (1987). Periods of High Probability of Occurrence of the World's Strongest Earthquakes, *Computational Seismology* **19**, Allerton Press Inc., 45-53.
Keilis-Borok, V. I. and V. G. Kossobokov (1990a). Premonitory activation of seismic flow: algorithm M8, *Phys. Earth and Planet. Inter.* **61**, 73-83.
Kossobokov, V. G. (1994). Intermediate-term changes of seismicity in advance of the Guam Earthquake on August 8, 1993, *EOS Transactions* **75**, No. 25, AGU 1994 Western Pacific Geophysics Meeting, Additional Abstracts, SE22A-10.
Kossobokov, V. G., J. H. Healy, J. W. Dewey, and I. N. Tikhonov, 1994. Precursory changes of seismicity before the October 4, 1994 Southern Kuril Islands earthquake, *EOS Transactions* **75**, No. 44, 1994 AGU Fall Meeting Addendum, S51F-11.
Kossobokov, V.G., J.H. Healy, J.W. Dewey, and L.L. Romashkova, (1996).Intermediate-Term Changes of Seismicity in Advance of the 10 June 1996 Delarof Islands Earthquake, *EOS Transactions* 77, AGU 1996 Fall Meeting, S31A-08.
Nishenko, S.P., C. Bufe, J. Dewey, D. Varnes, J. Healy, K. Jacob, V. Kossobokov,1996. 1996 Delarof Islands Earthquake - A Successful Earthquake Forecast/Prediction? *EOS Transactions* 77, AGU 1996 Fall Meeting, S71C-04.
Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical recipes in C: the art of scientific computing* - 2nd ed.. New York: Cambridge University Press, 994 p.
Sadovsky, M. A. (Editor) (1986). *Long-Term Earthquake Prediction: Methodological Recommendations*, Moscow: Inst. Phys. Earth, 127 p. (in Russian).
Shebalin, P. N. (1987). Compilation of earthquake catalogs as the task of clustering with learning, *Doklady Ac. Sci. USSR* **292**, 1083-1086.

V.G. Kossobokov
International Institute for Earthquake Prediction Theory
and Mathematical Geophysics, Russian Academy of Sciences
Moscow 113556, 79-2 Warshavskoye Sh.
RUSSIAN FEDERATION
Phone: 7(095)1102657
Fax: 7(095)3107032
E-Mail Address: volodya@mitp.rssi.ru