INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

H4.SMR/303 - 5

WORKSHOP
GLOBAL GEOPHYSICAL INFORMATICS WITH APPLICATIONS TO RESEARCH IN EARTHQUAKE PREDICTIONS AND REDUCTION OF SEISMIC RISK

(15 November - 16 December 1988)

ANALYSIS OF THE CATALOGS OF THE EARTHQUAKES

V.I. KEILIS-BOROK
&
I.V. KUZNETSOV

Institute of Physics of the Earth
Academy of Sciences of the U.S.S.R.
Bolshaya Gruzinskaya 10
123 242 Moscow
U.S.S.R.

ANALYSIS OF THE CATALOGS OF THE EARTHQUAKES

GENERAL OUTLINE AND SOFTWARE

CATALOG is a FILE of EARTHQUAKES (an "earthquake flow").

Usualy each earthquake is described by a VECTOR:
(T(I), LA(I), LO(i), DEP(I), mb, ms, ml, mp)

where I - sequence number;
T - origin time;
LA,LO - latitude and longitude of the epicenter (degrees and decimals, western longitude and southern latitude are negative);
DEP - focal deph, in kilometers;
mb,ms,ml,mp - four versions of magnitude for different magnitude scales.
Magnitude is a logatithmic measure of the energy of seismic waves, generated by the earthquake.

The earthquakes in a catalog should be ordered in increasing time:
T(I+1) >= T(I)

AN EXAMPLE of organisation of the catalog:

File:

```
198712 1 0 954 5841-14267 10  0  0  0  00
198712 1 01421 5874-14268 10  0  0390  00
198712 1 02632 5859-14275 10  0  0  0  00
198712 1 03347 5848-14273 10  0  0370  00
198712 1 03754 5850-14269 10  0  0  0  00
198712 1 04632 5833-14293 10  0  0350  00
198712 1 05539 5876-14282 10  0  0  0  00
198712 1 1 738 5806-14288 10420  0380  00
198712 1 11313 5859-14228 10460  0410  00
198712 1 114 4 1897 -6512 33  0  0430  00
```

Printout:

nn	year	mo	da	ho	mi	lat	lon	dep	mb	ms	ml	mp
1	1987	12	1	0	9	58.41	-142.67	10	0.00	0.00	0.00	0.00
2	1987	12	1	0	14	58.74	-142.68	10	0.00	0.00	3.90	0.00
3	1987	12	1	0	26	58.59	-142.75	10	0.00	0.00	0.00	0.00
4	1987	12	1	0	33	58.48	-142.73	10	0.00	0.00	3.70	0.00
5	1987	12	1	0	37	58.50	-142.69	10	0.00	0.00	0.00	0.00
6	1987	12	1	0	46	58.33	-142.93	10	0.00	0.00	3.50	0.00
7	1987	12	1	0	55	58.76	-142.82	10	0.00	0.00	0.00	0.00
8	1987	12	1	1	7	58.06	-142.88	10	4.20	0.00	3.80	0.00
9	1987	12	1	1	13	58.59	-142.28	10	4.60	0.00	4.10	0.00
10	1987	12	1	1	14	18.97	-65.12	33	0.00	0.00	4.30	0.00

INTRODUCTORY REMARKS

At this Workshop we analyze a catalog in order to DIAGNOSE THE APPROACH OF A STRONG EARTHQUAKE.

WHY CATALOGS ONLY?

(i) They are very RELEVANT.

(ii) They constitute MOST COMPLETE DATA set for our purpose:
cover most of the world for longest time-period.

MOST COMPLETE GLOBAL CATALOGS:

US Geological Sarvey (called also NOAA or NEIS);

International Seismological Center (ISC) -
more complete, but compiled with 2 years delay.

Other data may be happen to be even more relevant to prediction.
Residuals of arrival times of seismic waves are already available.

Coming soon:
digital seismograms;
sattelite geodesy (GPS - global positionig system);
water level;
.....

THE ANALYSIS OF CATALOG.

The analysis of catalogs during this Workshop is LOGICAL ("heuristic") and not statistical, sice strong earthquakes are too rare for usual statistical approach.

Please consider our analysis as

-- A way to look for earthquake prediction.
-- An example of logical analysis of time-series in general.

MAJOR TRAIT OF SUCH ANALYSIS - IT IS ROBUST: we consider integral properties of earthquake flow smoothing away many details. THAT MAKES OUR RESULTS STABLE, though limited.

PREPARATION OF A CATALOG

THIS LECTURE DEALS WITH PREPARATION OF CATALOGOG TO DIAGNOSIS OF TIPs BY EXISTING ALGORITHMS AND TO SEARCH FOR NEW ALGORITHMS.

Practicaly ALL CATALOGS ARE
INCOMPLETE
INPRECISE
CONTAIN ERRORS

WE WILL:

-- ELIMINATE OBVIOUS ERRORS
i.g. multiple repetitions of the same earhquakes (duplicates); absurd values of parameters (like 13th month).

-- EVALUATE COMPLETENESS without trying to improve it.

-- SINGLE OUT that PART OF THE CATALOG, WHICH seems SUFFICIENTLY COMPLETE for our purpose (e.g. eliminate early years, smallest magnitudes, marging of territory etc.).

Except in extreme cases, WE WILL NOT POSTPONE THE ANALYSIS UNTIL CATALOG IS PERFECTED. But WE will COMPENSATE THIS BY FANATICAL TESTS OF THE RESULTS.

More traditional alternative :
postpone analysis and return to original data in order to improve the catalog.

This is offen wrong on two count:
- it never ends;
- befor you try to analyse the data, you don't know what improvement are actually necessary.

" FIRST SYMPTOM OF MATHEMATICAL ILLITERACY IS
UNNECESSARY PRECISION OF COMPUTATIONS"
(Academician A.Krilov)

PROGRAMS

EDCAT – REORGANIZATION OF CATALOGS with different formats into one of the two standard formats (20 or 41 bytes); IDENTIFICATION OF obvious ERRORS and DUPLICATES; CORRECTION of errors.

CATAL – MERGING of two CATALOGS; SELECTION of SUBCATALOGS from a given catalog; COMPARISON of two CATALOGS.

HIST – CONSTRUCTION of various HISTOGRAMS of some parameters of the catalog.

AFT – identification of the main shocks and aftershocks in the catalog; CREATION OF THE CATALOG OF MAIN SHOCKS.

STAGES OF ANALYSIS

Suppose YOU HAVE received some CATALOG from somewhere.
Following is the sequence of actions:

1. FIND and ELIMINATE obvious ERRORS.
   (Program EDCAT)

2. FIND roughly WHAT TERRITORY and TIME period
   are COVERED BY the CATALOG.
   (Program EDCAT)

AN EXAMPLE of the printout from EDCAT:

```
1: Reorg. of catal ex1.dat                format 4
    No output catalog
2: Transformation of magnitude - n
3: D Min,Sec,   La,   Lo,  H,  Mb,  Ms,  Ml,   Mp for dupl.:
       1   1  0.01  0.01   1 0.01 0.01 0.01 0.01
4: Standart range for nonsense was used
BOUNDARY!-y-*m--d--h-mi-sc----la-----lo--de--MB--MS--ML--MP
     4 1940 15 19  4 36  0  3273 -11550   0 670  20   0   00
CHECK TI !---m--d--h-mi-sc----la-----lo--de--MB--MS--ML--MP-
     3 1940  5 19  4 36  0  3273 -11550   0 670  20   0   00
     4 1940 15 19  4 36  0  3273 -11550   0 670  20   0   00
DUPLICAT--y--m--d--h-mi-sc----la-----lo--de--MB--MS--ML--MP-
     8 1956  2  9 14 32  0  3175 -11592   0 680  46   0   00
     9 1956  2  9 14 33  0  3174 -11591   0 681  45   0   00
CHECK TI !---m--d--h-mi-sc----la-----lo--de--MB--MS--ML--MP-
    11 1979 10 15 23 16  0  3260 -11532  12 660 359   0   00
    12 1971  2  9 14  0  0  3440 -11840   8 640  45   0   00
Read:    19 , Could be written:    19 events

Min,Max range in input catalog:
Y,Mon,Day,Hour,Min from  1940 15 19  4 36
                    to 1987 12  1  1  1
Lat.  from    31.74 to    37.60
Long. from  -120.25 to  -114.75
Depth from        0 to    12
Magn. :
from   6.40 to   7.70
from   0.00 to   3.59
from   0.00 to   0.00
from   0.00 to   0.00
```

3. CHOOSE MAGNITUDE SCALE.

No formal recepees. Usual ways:

Just assume maximal value of magnitude among all indicated for each earthquake or choose definition of "COMMON" magnitude and recomute all other into it.

If necessary - find relation between different magnitude scale.
(program EDCAT, CATAL, HIST)

Now you are ready for a general review of seismicity, as it is represented by the catalog.

4. TERRITORIAL DISTRIBUTION of epicenters:
histograms N (LO,LA) and Mmax (LO,LA).

AN EXAMPLE of histogram N (LO,LA):

	-124.00	-120.00	-116.00	-112.00			
40.00	.	.	.	.	.	.	.	0
	.	.	.	.	.	.	.	0
39.00	.	1	12	7	.	.	.	20
	.	3	195	8	3	.	.	209
38.00	.	25	1559	34	13	.	.	1631
	.	35	1033	143	18	.	.	1229
37.00	.	483	33	59	27	.	.	602
	1	751	128	178	12	.	.	1070
36.00	.	296	520	574	28	.	.	1418
	.	43	1327	169	.	.	.	1539
35.00	.	216	374	436	16	3	.	1045
	.	50	1080	1473	150	3	.	2756
34.00	.	33	806	1141	159	1	.	2140
	.	1	124	1241	703	.	.	2069
33.00	.	.	91	305	1307	1	.	1704
	.	1	22	237	728	6	.	994
32.00	.	.	3	296	561	6	1	867
	.	.	4	37	272	22	.	335
31.00	.	.	.	3	21	16	.	40
	.	.	.	.	.	.	.	0
30.00	.	.	.	.	.	.	.	0
		1	7	6	4			
		9	3	3	0			
		3	1	4	1	5		
	1	8	1	1	8	8	1	

19668 events

5. DISTRIBUTION OF focal DEPTH:
histograms DEPmax (LO,LA),
N (LO,DEP) or N (LA,DEP),
Mmax (LO,DEP) or Mmax (LO,DEP).

After that you may wish to repeat 4. for some interval of depth.

6. "COMPLETENESS" :
minimal magnitude, for wich the catalog is complete. Usually it changes with time.
Histogram N (M,T).

AN EXAMPLE of histogram N (M,T):

	2.00	4.00	6.00	8.00			
32 1 1 0	5	294	163	14	1	1	. 478
	.	418	114	11	4	1	. 548
36 1 1 0	.	335	78	4	1	.	. 418
	.	326	103	13	.	.	. 442
40 1 1 0	.	349	152	20	4	.	. 525
	.	330	126	11	1	.	. 468
44 1 1 0	44	257	52	8	.	.	. 361
	135	523	97	18	2	.	. 775
48 1 1 0	122	513	70	7	1	.	. 713
	94	350	62	7	.	.	. 513
52 1 1 0	173	557	222	20	4	1	. 977
	143	505	118	20	3	.	. 789
56 1 1 0	74	303	164	27	5	.	. 573
	62	306	104	6	1	.	. 479
60 1 1 0	73	316	68	9	.	.	. 466
	84	346	70	7	.	.	. 507
64 1 1 0	77	261	76	7	.	.	. 421
	71	250	49	4	1	.	. 375
68 1 1 0	166	571	103	44	1	.	. 885
	191	654	79	6	1	.	. 931
72 1 1 0	153	372	43	3	.	.	. 571
	293	489	55	3	.	.	. 840
76 1 1 0	293	422	43	2	.	.	. 760
	477	834	83	10	1	.	. 1405
80 1 1 0	663	1357	131	13	4	.	. 2168
	458	843	96	12	1	.	. 1410
84 1 1 0	280	528	59	2	1	.	. 870
		1					
	4	2	2				
	1	6	5	3			
	3	0	8	0	3		
	1	9	0	8	7	3	0

19668 events

Above this magnitud lgN should increase linearly with decrease of M, according to frequency-of-occurence law.
But don't make magnitude intervals nor time-interval too short.

Most offen this minimal magnitude is larger for earlier years, unless the author of the catalog cut off the magnitudes, represented incompletely.

You can try to improve completeness by merging different catalogs for the same region.

CATALOGS OF MAIN SHOCKS

THEY DO NOT INCLUDE AFTERSHOCKS.
Forshocks are not distinguished from main shocks.

An earthquake is identified as an aftershock of some preceding main shock if the following conditions are satisfied:

its magnitude does not exceed the magnitude M of the main shock;
the distance between their epicenters does not exceed R(M);
time difference does not exceed T(M);
modulus of depths difference does not exceed H(M).

Here T(M),R(M) and H(M) are empirical functions.

If the conditions are satisfied for several preceding main shocks the earthquake is assigned to the latest among the strongest main shocks.

The catalog of main shocks is used in many algorithms.
Elimination of aftershocks allows to avoid the situation, when relatively strong earthquakes are represented twice:
by themselves and by large number of aftershocks.

P R O G R A M E D C A T

The following modes are possible:

R - reorganization of format of catalog with checking for obvious errors (nonsense) in the input catalog (incorrect records, time, disordering, impossible values of T,LA,LO,DEP,mb, ...,mp) and duplicates.
Transformation of magnitude is also possible.

C - correction of errors in input catalog with editor options (search, change, delete, insert or add events).

T - sorting input catalog by time.

Errors and duplicates for mode R.

The program in mode R can eliminate events if their parameters are outside some range entered in the dialog. These obvious errors of the catalog are called "nonsense". The following order of testing for nonsense is used for each entry:

Each record is tested, whether its format corresponds to the type of the catalog. If program can not recognize this record it is displayed. This entry will be skipped.

Year - should be within given range; month should be within 1 - 12; day - 1-31; hour - 0-23; minute, second - 0-59; latitude,longitude,depth,input magnitudes and transformed magnitude - should be wihin given ranges.

Events should be ordered by time.

When the first nonsense in input entry is found the program displays this entry marking its incorrect position. This entry is not tested further.

EDCAT can also identify "DUPLICATES", i.e. the entries which are very similar, so that they probably correspond to the same earthquake. We assume that two entries, E1 and E2, are duplicates if their corresponding parameters are sufficiently close, e.g. $|T(E1)-T(E2)| \le dT$, $|LA(E1)-LA(E2)| \le dLA$, ..., $|M(E1)-M(E2)| \le dM$. The thresholds dT,dLA,...,dM are entered in the dialog.

PROGRAM CATAL

The following modes of operation are possible:

t: Type/print input catalog starting from some event.
s: Select a subcatalog from input catalog.
p: Select a subcatalog from input catalog and add this subcatalog at the end of second catalog.

The next two modes of operation are used to eliminate "duplicates" i.e. different entries which correspond to the same earthquake (the same definition as for EDCAT). For each event of the first catalog the program finds the duplicates in the second catalog.

a: Eliminate these duplicates; then merge the catalogs into a single output catalog with events ordered in time.
q: Print these duplicates, if any, for contemplation together.

d: Print all events from both input catalogs which have no duplicates.
h: Help.

Epicentral area.

The program selects the earthquakes with the epicenters situated within some area on the map.

This area can be specified as rectangle, polygon or a set of circles. The boundaries are included.

P R O R R A M H I S T

Input is a catalog.
Output is a histogram A(P1,P2,P3).
Here P1, P2, P3 are some parameters of an earthquake;
A is one of characteristics of an earthquake flow.

An earthquake flow can be specified by space, time and magnitude limits: certain rectangle of (LA,LO) plane, intervals of focal depth, origin time, and magnitude.

The histogram is represented by a sequence of two-dimensional tables. Each table corresponds to a certain interval of the values of P3. Each column corresponds to a certain interval of P1(horizsontal axis). Each line corresponds to a certain interval of P2(vertical axis).

The following parameters can be considered as P1,P2,P3:

Parameter	Code for entering
Time - year,month,day,hour	ti ¦ can not be
Days of month(1-31)	dm ¦ used for
Hours(0-23)	hr ¦ horizontal
Days of year(1-365)	dy ¦ axis P1
Latitude	la
Longitude	lo
Depth	de
Magnitude	mc (see EDCAT),or mb, or ml, or ms, or mp

The cells of tables are filled with the observed values of A(P1,P2,P3).

The following characteristics can be chosen as A :

Characteristic	Its code for entering
Number of earthquakes	nu
Weighted sum of earthquakes (Sigma)	sg
Average weight of an earthquake	sn
Variation of magnitude	vm
Average magnitude	av
Maximal depth	de
Maximal magnitude	MC, mb, ms, ml, mp

sg - E [c*10**(d(M(i)-f))] is the sum of weights. Summation is carried over all earthquakes in a cell. Here M is selected magnitude (MC or mb, ... or mb,...) and c,d,f are entered coefficients; 10**x is 10 in power x.

sn - sg/(N**g), where N is the number of earthquakes in a cell and g is the entered coefficient.

vm - E |M(i+1)-M(i)| is the sum of magnitude jumps in a cell. Here M(i) is magnitude of earthquake with number i from a cell.

av - [E M(i)]/N is an averege value of magnitude in a cell.

P R O G R A M A F T

The program separates earthquakes into main shocks and aftershocks.

For each main shock the following statistics are given in printout:
- the number of aftershocks b(e(i)) during the time interval e(i) after the main shock;
- the weighted sum Sigma (sg) counted for these aftershocks.

Here $e(1) < e(2) < \ldots < e(j) \leq T(M)$, $j \leq 10$.

After a strong earthquake (i.e. main shock with $M \geq M0$) the count of b and sg is terminated for all preceding main shocks. However, the identification of aftershocks is carried on.