Istituto Nazionale
di Geofisica

Seismological Analysis with Automatic Processing for the Italian Telemetered Network

E. Boschi, A. Basili, R. Console, F. Mele e G. Smriglio

Pubblicazione n. 518

Seismological Analysis with Automatic Processing for the Italian Telemetered Network

E. Boschi, A. Basili, R. Console, F. Mele e G. Smriglio

Istituto Nazionale di Geofisica
Via di Villa Ricotti 42
00161 Roma

Aprile 1990

SEISMOLOGICAL ANALYSIS WITH AUTOMATIC PROCESSING FOR THE ITALIAN TELEMETERED NETWORK

E. Boschi, A. Basili, R. Console, F. Mele e G. Smriglio

Istituto Nazionale di Geofisica

Abstract

Digital data from more than 70 seismological stations spread over the Italian territory are recorded and processed at the Data Center of the Istituto Nazionale di Geofisica (Rome).

The acquisition system is provided with programs for fully automatic phase picking and epicentral parameters estimation of local and teleseismic events in real time. Standard seismological procedures are also applied to magnitude estimation for both kinds of events. The waveforms are stored for on-line analysis based on interactive procedures. Both the temporary real time information and the seismological permanent data-base are available to external users through computer-to-computer communication.

More of two years of operation of the system allow a comparison between automatic and interactive procedures used on the same digital waveforms. Most of the local and regional events detected by at least three stations are automatically located within a distance of 20 km from the epicenter determined by the human analyst by means of standard programs. The similarity is very high for events occurring within or close to the network and recorded by six or more stations. Only a few teleseismic events are not correctly identified while most of them are generally located and identified with reasonable accuracy. The correct phase association has proved to be by far the most delicate problem, due to the variety of situations that take place during the routine activity.

1. Introduction

In the last ten years the Istituto Nazionale di Geofisica (ING) has been developing its National Telemetered Seismological Network (NTSN). It was designed to provide national agencies with fast and reliable information about seismic events for disaster relief purposes and to improve the national system collection of data for seismological studies and earthquake prediction.

In the late 1970's this project was in an initial stage. The disaster caused by the Southern Italy earthquake of 23 November 1980 made its completion to appear necessary and urgent. In August 1982, the new national network (even though it only counted, at that time, 15 stations) reached the operational level with the 24-hours a day service at the Data Center of Rome. Since then, the ING system continued expanding and the number of stations reached 70 in 1989 (fig.1).

The growing mass of data available from the network increased the load of work for the analysts of the Center forcing the substitution of a traditional way of reading seismograms to a new method using modern digital techniques.

Since the beginning of 1983, tests of automatic phase picking were carried out with an off-line digital acquisition system. Later, in 1984, in co-operation with the U.S. Geological Survey, a new system, based on the VAX computer of the Center, began to work in an experimental way for on-line and automatic processing of digital waveforms.

The automatic system became fully operational in the first months of 1987. It was created by joining the experience of both specialists in seismology and in informatics. The data collected automatically and checked by means of interactive procedures are routinely used both for rapid information to Civil Protection Agencies and for the preparation of bulletins and catalogues.

2. Features of an Automatic Seismic System

The purpose of an automatic seismic data acquisition and analysis system is to double the performance of a human operator as far as possible. The first step is inspecting the data streams coming from individual stations of the network with the following purposes in mind:

- detecting seismic phases, distinguishing them from instrumental and ground noise;
- picking the arrival times of onsets;
- measuring maximum amplitude and period of wavelets; and
- measuring the seismic signal duration.

After phase parameters are collected from a number of stations in a given time window, they must be checked and processed by means of suitable algorithms in order to get information on the seismic events that have generated them, taking the following steps:

a) associating phase parameters belonging to the same event;

b) distinguishing between local/regional and teleseismic events;

c) locating such events and determine their magnitude; and

d) calculating azimuth, distance and magnitude of teleseismic events.

The state-of-the-art of real time seismic processing had been limited, until a few years ago, to dense (50-200 stations), local (about 2 degrees diameter) networks. The success of these

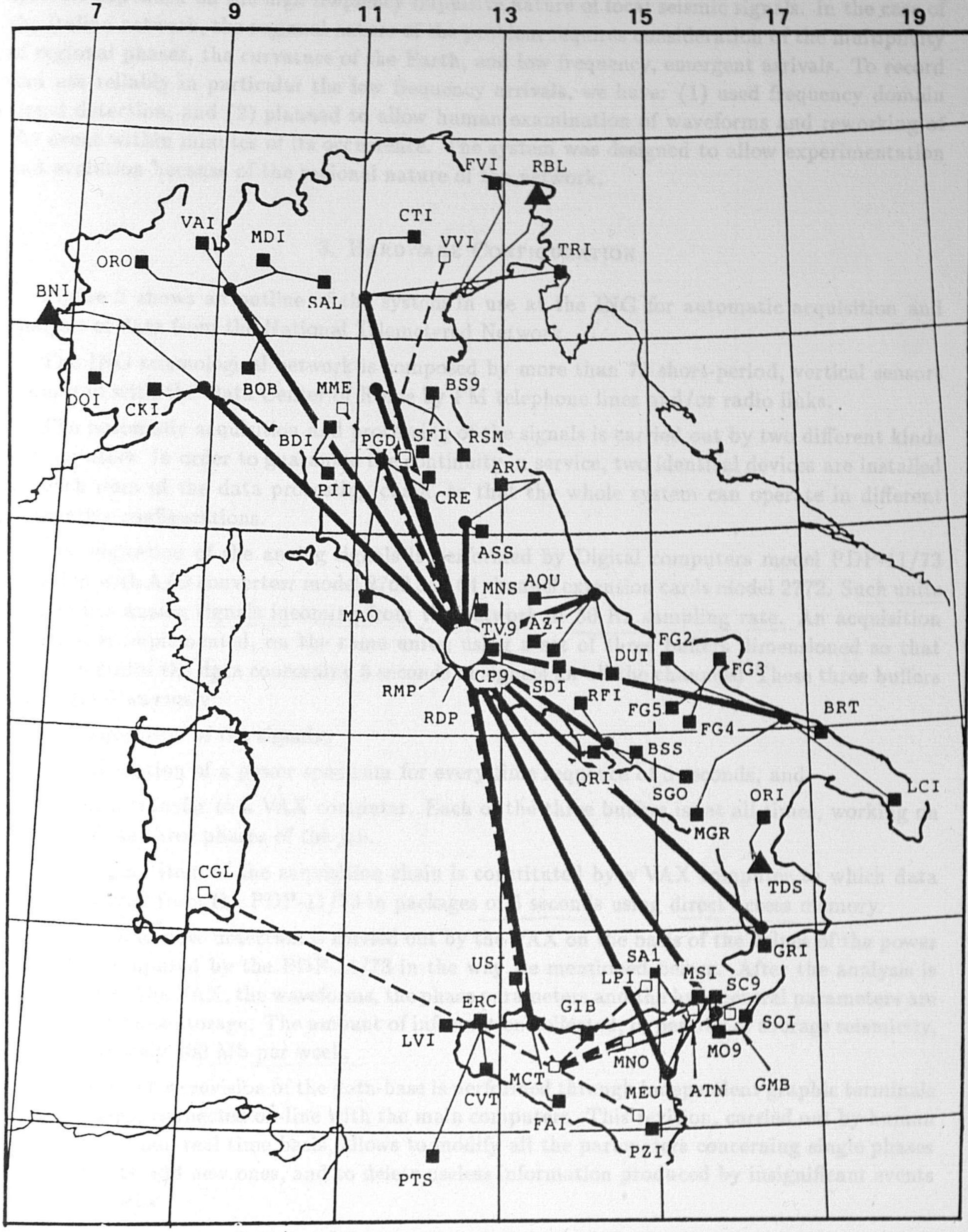

Fig. 1 - The Italian Telemetered Seismological Network (April, 1988). Full lines: public telephone links. Hatched lines: other links.

systems depended on the high frequency impulsive nature of local seismic signals. In the case of the Italian network, the regional nature of the problem requires consideration of the multiplicity of regional phases, the curvature of the Earth, and low frequency, emergent arrivals. To record and use reliably in particular the low frequency arrivals, we have: (1) used frequency domain signal detection, and (2) planned to allow human examination of waveforms and reworking of the event within minutes of its occurrence. The system was designed to allow experimentation and evolution because of the regional nature of the network.

3. Hardware Configuration

Figure 3 shows an outline of the system in use at the ING for automatic acquisition and analysis of data from the National Telemetered Network.

The ING seismological network is composed by more than 70 short-period, vertical sensors connected with the Data Center of Rome by FM telephone lines and/or radio links.

The automatic acquisition and processing of the signals is carried out by two different kinds of computers. In order to guarantee the continuity in service, two identical devices are installed for each item of the data processing chain, so that the whole system can operate in different alternative configurations.

The digitation of the analog signals is performed by Digital computers model PDP-11/73 provided with A/D converters model 2762 and 64 channel extension cards model 2772. Such units sample the analog signals incoming from the network at 50 Hz sampling rate. An acquisition program is implemented, on the same units, using a set of three buffers dimensioned so that each one stores the data concerning 5 seconds of signals for all the channels. These three buffers allow simultaneously:

1) the acquisition of the signals;
2) the calculation of a power spectrum for every time sequence of 5 seconds, and
3) the data transfer to a VAX computer. Each of the three buffers is, at all times, working on one of the three phases of the job.

The second item of the acquisition chain is constituted by a VAX computer to which data are transferred from the PDP-11/73 in packages of 5 seconds using direct access memory.

The seismic wave detection is carried out by the VAX on the basis of the values of the power spectrum computed by the PDP-11/73 in the way we mentioned before. After the analysis is completed by the VAX, the waveforms, the phase parameters and the hypocentral parameters are archived on mass storage. The amount of information collected, in periods of average seismicity, is approximately 100 Mb per week,

The interactive revision of the data-base is performed through independent graphic terminals (workstations) connected on-line with the main computers. This revision, carried out by human analysts on a non-real time basis, allows to modify all the parameters concerning single phases and events, to add new ones, and to delete useless information produced by insignificant events or cultural noise.

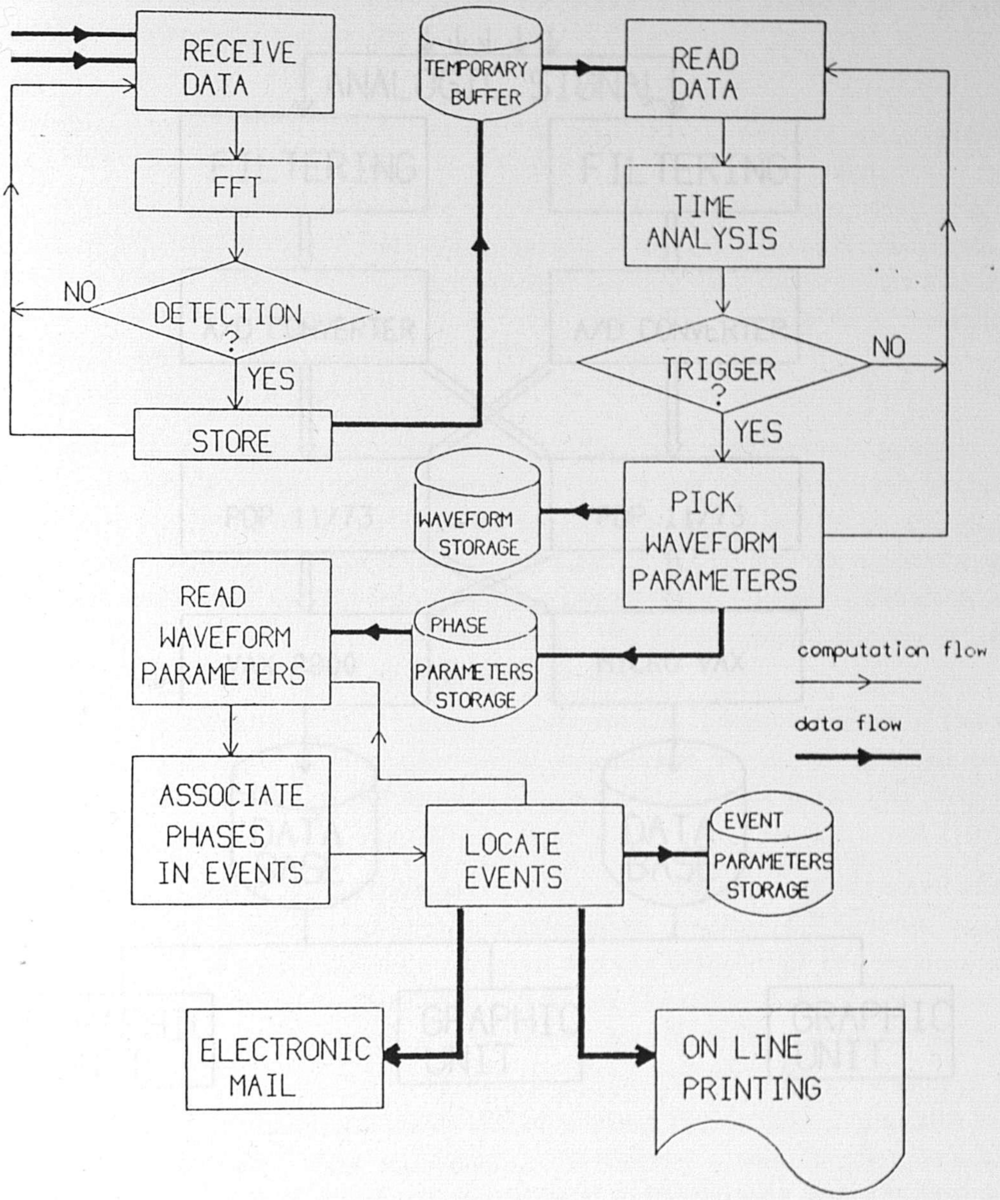

Fig. 2 - Simplified flow diagram of the automatic acquisition system.

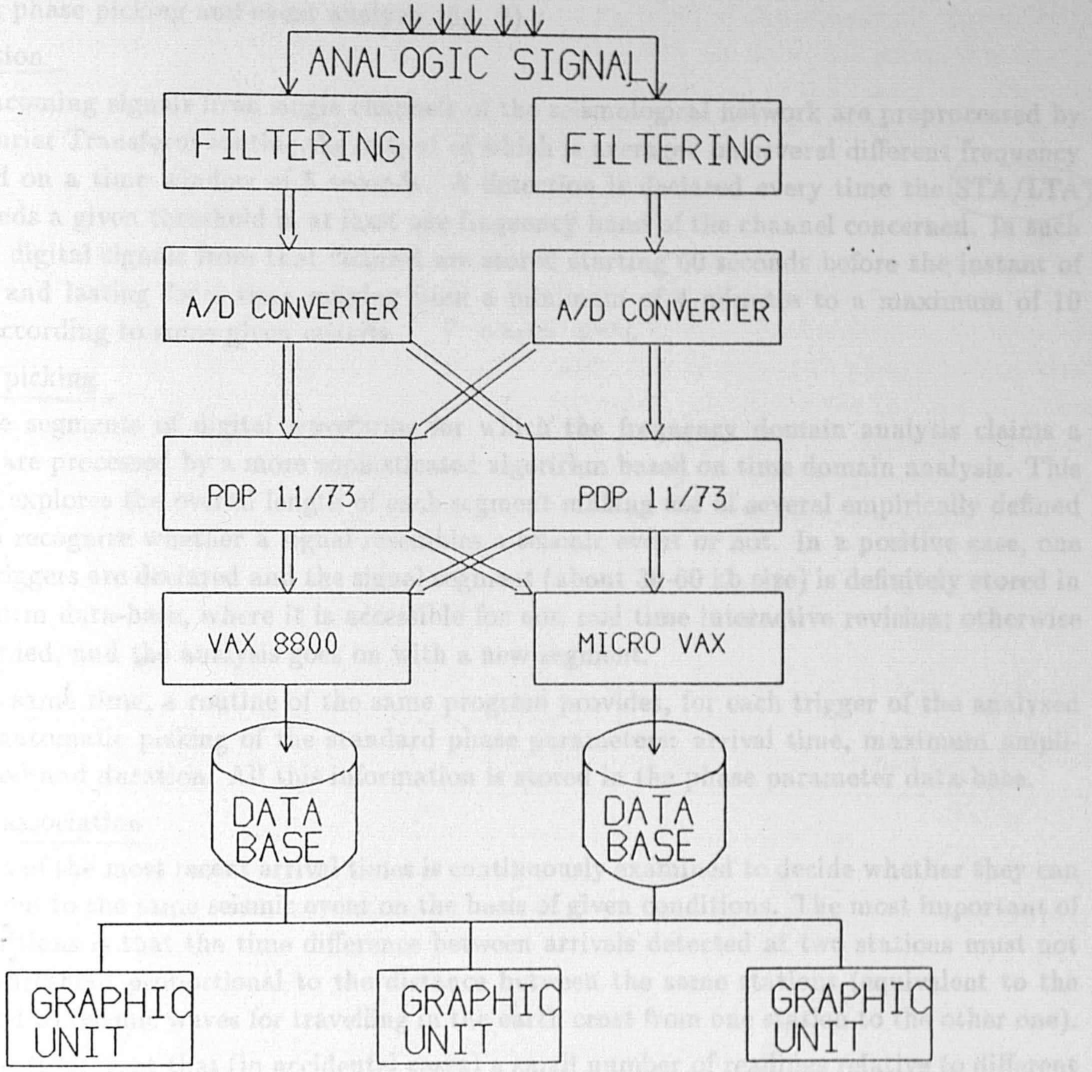

Fig. 3 - Hardware configuration of the automatic acquisition system.

4. Organization Criteria of the Data Flow

To achieve the above mentioned features, the software of the system was structured in three different groups of programs, each running in real time with different tasks, receiving different data and producing separate data-bases. These programs deal respectively with seismic wave detection, phase picking and event analysis (fig. 2).

- Detection

The incoming signals from single channels of the seismological network are preprocessed by a Fast Fourier Transform routine the output of which is averaged on several different frequency bands and on a time window of 5 seconds. A detection is declared every time the STA/LTA ratio exceeds a given threshold in at least one frequency band of the channel concerned. In such cases, the digital signals from that channel are stored starting 60 seconds before the instant of detection and lasting for a time ranging from a minimum of 4 minutes to a maximum of 10 minutes according to some given criteria.

- Phase picking

All the segments of digital waveforms for which the frequency domain analysis claims a detection are processed by a more sophisticated algorithm based on time domain analysis. This algorithm explores the overall length of each segment making use of several empirically defined criteria to recognize whether a signal resembles a seismic event or not. In a positive case, one or more triggers are declared and the signal segment (about 30-60 kb size) is definitely stored in the waveform data-base, where it is accessible for non real time interactive revision; otherwise it is discarded, and the analysis goes on with a new segment.

At the same time, a routine of the same program provides, for each trigger of the analysed segment, automatic picking of the standard phase parameters: arrival time, maximum amplitude, period and duration. All this information is stored in the phase parameter data-base.

- Phase association

The list of the most recent arrival times is continuously examined to decide whether they can be associated to the same seismic event on the basis of given conditions. The most important of these conditions is that the time difference between arrivals detected at two stations must not exceed a threshold proportional to the distance between the same stations (equivalent to the time needed to seismic waves for travelling in the earth crust from one station to the other one).

In order to prevent that (in accidental cases) a small number of readings relative to different sources on stations very far from each other should produce the definition of a false event, the associations are tested against their significance. We defined a sum function to which each station contributes with a weight inversely proportional to its distance from the first station in the group. The event then is rejected if the value of the sum function doesn't reach a threshold assigned previously.

We also found it necessary to limit the cases in which, still because of accidental circumstances, a weak seismic event detected by a single station is associated with a stronger one recorded simultaneously by other stations in distant areas. That is how the program was also provided with a criterium for discriminating data coming from stations distanced, more than 1.8 times the mean square root of all the distances, from the gravity center of the whole group of associated stations.

The procedures set up in this way proved to be efficient in reducing the number of false or mislocated events.

• Event analysis

Every time three or more phases are associated, the program calls a routine for local or regional event location or a routine for teleseismic event location according to the relative delay of the seismic phases, also taking into account the longer period of the teleseismic waves. Both routines are provided with criteria for automatically discarding the data that produce too large P-arrival residuals. After the location is determined, the magnitude of the events is calculated from the amplitude and/or duration of the recorded signals according to standard seismological procedures.

• Outputs

The final results of the automatic determination of the event parameters (origin time, epicentral coordinates, magnitude ml, md or mb) together with the station readings (arrival times, amplitudes, periods and durations) are stored in the event data-base. Such information, available a few minutes after the event origin time, is automatically sent in output to a letter printer used by the staff in charge of the real time monitoring of the seismic activity on the national territory. The same data are simultaneously sent, via electronic mail, to international data centers such as NEIC or EMSC. A specific area, in which the information concerning the last recorded events is deposited, is available to any other user through computer-to-computer communication.

5. Interactive Revision and Editing

All the data automatically recorded and processed by the system (i.e. digital waveforms, phase parameters and event parameters) are checked by the analysts in charge of the interpretation, who use the programs offered by the system operating interactively on graphic terminals and select those waveforms which according to their experience are actually due to seismic events distinguishing them from those deriving from instrumental or cultural noise.

In their work of revision, the analysts check arrival times and give every phase its proper code according to the standard seismological practice. They dispose of a group of programs that allow, among other things:

- to filter waveforms within given frequency bands,
- to magnify any waveform window both vertically and horizontally,
- to relocate epicenters and check the modified or new arrival times,
- to modify manually the associations previously carried out by the computer, and
- to delete false events or to create new ones.

The main task of the revision work is to set up a data-base to be used later for editing the ING seismological bulletin. The procedures used in editing the bulletins are performed by programs written in FORTRAN at the ING (Smriglio and Valensise, 1985). They operate off-line and only need a list of phase parameters, including station and phase code, arrival time, maximum amplitude, period and duration. The output of each program is taken as an input in the process which follows.

The procedure makes a preliminary selection of the data, discarding those that were not revised by the analyst on the basis of the phase code. The selected data are temporarly stored in the computer, to be, further on sorted chronologically, and processed by a program of automatic association, and then stored definitely on disk in the file concerning the current year. The hypocentral parameters of the associated events are redetermined by means of an automatized version of the location program used at the ING (Basili et al., 1984). Finally, also the magnitude (ml and md) is computed averaging the values obtained by the amplitudes and durations reported by the various stations.

In the yearly data-base, data from other Italian or foreign stations, collected by ING from NEIC (Golden, USA) and EMSC (Strasbourg, France), are also entered automatically. These data are sent through X.25 communication links. They are selected before entering into the data-base. The first selection is based on the epicentral coordinates, considering only the events which took place in the Italian territory or the "surrounding area" defined in the "Foreward to the National Seismic Catalogue" (Console et al., 1979). The next selection operates on the magnitude, on the basis of the expected detectability of the event by the stations of the Italian NTSN. Furtherly, the single phases of the events that are accepted are selected, discarding those readings that in the NEIC location present a residual which exceedes 5 seconds.

Recently, an automatic data manager was implemented to allow the access, in reading only, to the data-base of seismic data stored in the computer of the ING. Such procedure allows external users to request the automatic issue of electronic mail messages containing the hypocentral location and arrival times of a chosen event.

6. Results

As we said before, in creating an automatic system for the recognition and location of seismic data, and for the storage of recorded data, we aimed to process the enormous data set incoming from the NTSN. Our aim, however, will only be reached when the reliability of the entire system will be comparable with the one of a human analyst's work.

In order to verify 'a posteriori' the reliability of the system, a test was carried out over more than 3 months of systematic observation (15 June - 26 September 1988). The locations obtained during this period by the automatic procedure were compared with those obtained by means of interactive procedures on identically stored waveforms. The locations obtained by the analysts were taken as reference data sets.

The system collected 476 events detected by at least 3 stations. Of these, 319 (67%) were located within 20 km from the reference epicenter; 49 more events (10%) were located at a distance between 20 and 50 km from the reference epicenter. The remainder (23%) of the events were located at a distance of more than 50 km from the reference epicenter; this percentage of "mislocated" earthquakes would have seemed rather high if we hadn't proceeded to a further test.

The accuracy of all locations and more so of the automatic ones depends on a few different elements: a good location is obtained not only with a high number of stations recording the event, but it is also very desirable for the recording stations to surround the event so as to reach a good azimuthal distribution. Furthermore, every automatic acquisition system records, in addition to signals actually produced by earthquakes, a variety of other signals not associable to earthquakes. It may so happen that one of these signals be recorded together with a real seismic event, in these cases the automatic procedure may erroneously associate them. The weight of this kind of false data in a location procedure, and of the resultant error, decreases the higher is the amount of data assigned correctly to the event.

Going back to the analysis of mislocated earthquakes, we must point out that most of them are weak events recorded only by a few stations. Taking the events recorded by at least 6 stations, the number of automatic locations with a difference of more than 50 km from the data set of reference is only around 40, most of which are not whithin the national territory and as such don't have a good azimuthal distribution. Taking only the Italian events with at least 6 recordings, the number goes down to 7 minor ones.

Fig. 4 shows the epicenters of the events for which the automatic location differs less than 50 km from the final interactive one. The middle of each square corresponds to the position of the interactive location; the arrow indicates the difference with the automatic location.

Fig. 5 contains all the events recorded by at least 6 seismographic stations of NTSN, the automatic location of which differs more than 50 km from the interactive one. As previously pointed out only 7 of these took place in the National Territory and the automatic location can be considered totally wrong only for two of them. It is obviously very difficult to locate events taking place outside the network.

We have also verified the procedure used to recognize and locate automatically the teleseismic events. This procedure uses the NTSN as an antenna to record distant earthquakes. Statistics regarding 32 locations, reported by NEIC, of earthquakes of Mb greater than or equal to 5.5 and less than 11.000 km far from Italy gave the following results: 26 of them were recognized correctly as teleseismic events, of these, 21 were located correctly. Two of the 32 reported earthquakes were erroneously considered local; and 4 of them were not recorded at all but their Mb didn't reach 6.0.

In the same period NEIC reported 34 more events with Mb. 5.5 and more than 12.000 km far from Italy; it is impossible to locate these earthquakes using only Italian data. The automatic system however recognized 13 of them, classifying them correctly as "not locateable" ones. Of the other 21, not recorded by the aquisition system, only one had Mb greater than 6.0.

This study proves the choice of the new methods to be good, and at the same time it shows the need of improving the procedures for the system to become faster and more reliable.

7. Comments and Conclusions

The most evident result of our test is that a high percentage of locations performed through the automatic procedures falls within 20 km from those calculated by interactive analysis. This also applies to the calculation of the magnitude.

The false locations obtained are restricted to the events characterized by low magnitude and consequently detected by very few stations. False events are due mainly to electronic disturbances occurring randomly on the transmission lines.

The cases of two or more events originating almost simultaneously in different zones whose first arrivals are considered by the system as belonging to the same event are fairly rare.

One of the problems when designing and creating an automatic system is its complexity. The whole procedure not only needs a high performance computer facility, but also a large number

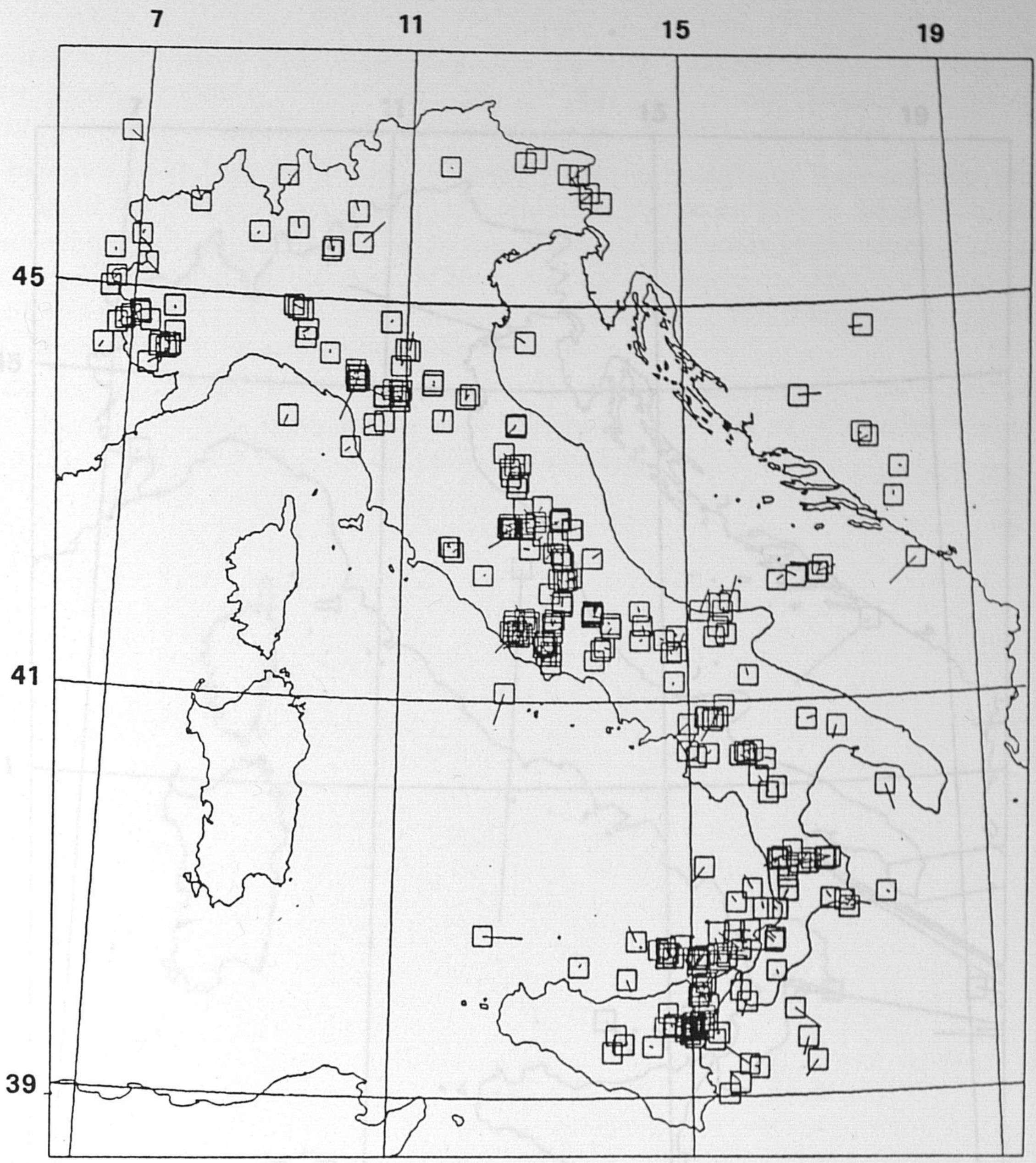

Fig. 4 - Comparison between interactive and automatic locations for the seismic events detected from 15 June to 26 September 1988. The bars represent the distance between the two locations in the scale of the map. Only the events for which the discrepancy doesn't exceed 50 kms (77% of the total) are represented.

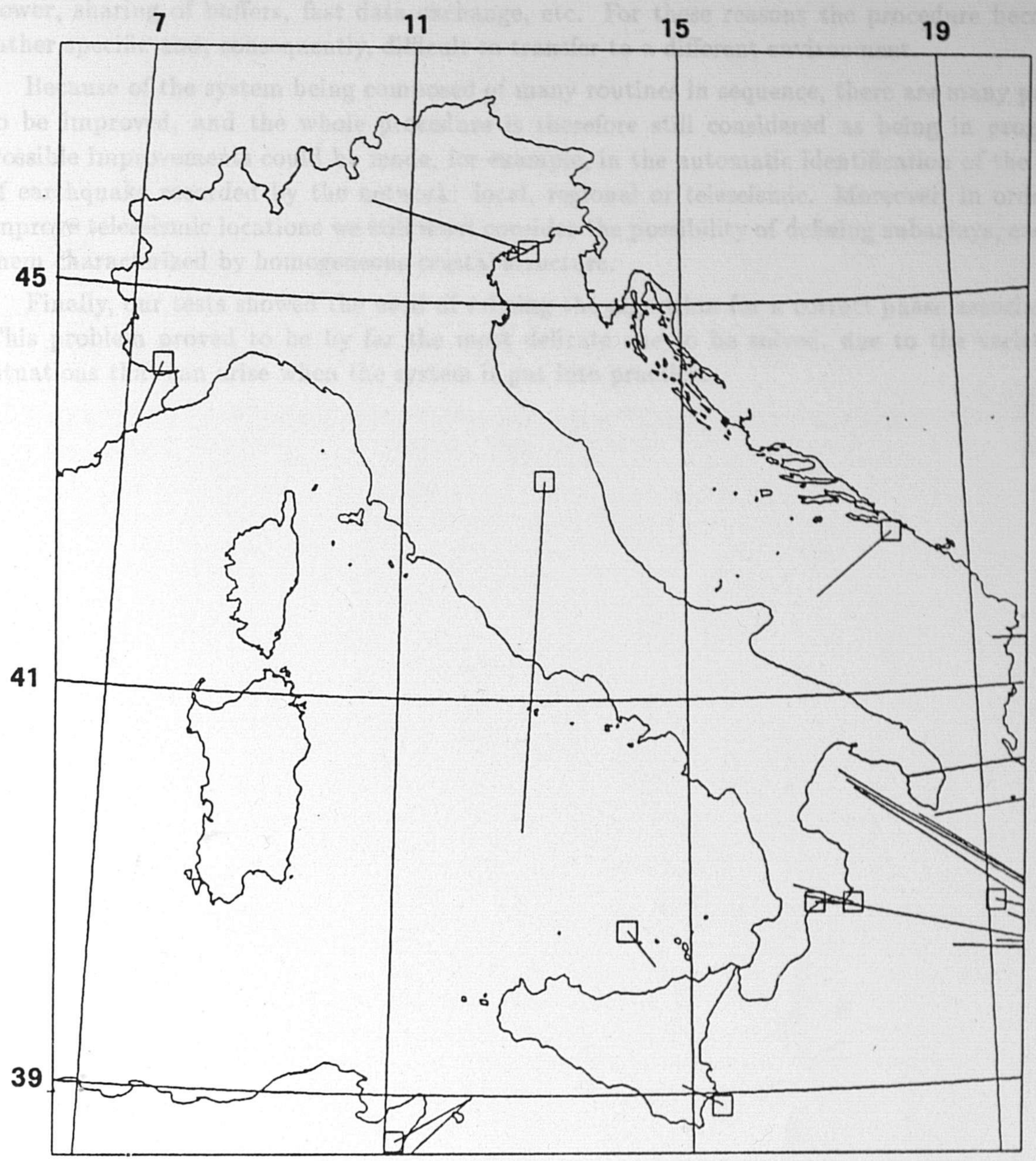

Fig. 5 - As fig. 4, for the events recorded by at least 6 stations, and for which the discrepancy between the epicenters obtained by interactive procedures and by fully automatic procedures exceeds 50 kms. Of these, only 7 (less than 2% of the total) are included in the national territory.

of sophisticated routines which make heavy use of hardware resources in terms of computation power, sharing of buffers, fast data exchange, etc. For these reasons the procedure becomes rather specific and, consequently, difficult to transfer to a different environment.

Because of the system being composed of many routines in sequence, there are many points to be improved, and the whole procedure is therefore still considered as being in progress. Possible improvements could be made, for example, in the automatic identification of the kind of earthquake recorded by the network: local, regional or teleseismic. Moreover, in order to improve teleseismic locations we still must consider the possibility of defining subarrays, each of them characterized by homogeneous crustal structure.

Finally, our tests showed the need of refining the algorithm for a correct phase association. This problem proved to be by far the most delicate one to be solved, due to the variety of situations that can arise when the system is put into practice.

References

A. Basili, G. Smriglio, G. Valensise (1984). Procedure di determinazione ipocentrale in uso presso l'Istituto Nazionale di Geofisica. *Atti del 3 Convegno Annuale del Gruppo Nazionale di Geofisica della Terra Solida*, 875-884.

R. Console, B. De Simoni, C. Gasparini, L. Marcelli, M.C. Spadea, M. Vecchi (1979). Preambolo al Catalogo Sismico Nazionale. *Annali di Geofisica*, Vol. XXXII, 37-77.

G. Smriglio and G. Valensise (1985). Procedura per l'analisi di un data-base annuale di dati sismometrici. *ING Internal report* no. 506.