ROBUST PROCEDURES FOR AUTOMATIC PRECISE LOCATION OF REGIONAL SEISMIC EVENTS

Vladimir Pinsky and Avi Shapira

Seismology Division, Geophysical Institute of Israel
P.O. Box 2286
Holon 58122, Israel

Sponsored by the U.S. Department of Defense
Defense Threat Reduction Agency
Contract No. DSWA01-97-C-0151

ABSTRACT

Coordinates YY,H of seismic source and origin time T are routinely determined through P and S phases manual picking and an optimization procedure minimizing the discrepancy between model arrival times and network observations. The proposed procedure has two main steps: preliminary rough location and final, accurate location. Preliminary location is based on the computation of **seismogram envelopes** for each station and finding the envelope maximums **Pmax** and **Smax**. This is followed by estimation of the epicenter X,Y and the origin time T, using travel time curves (lines) for propagation of P and S seismic energy and a **grid search robust optimization** procedure. The method has already been presented and its accuracy demonstrated, DX, DY~6 km for the epicenter coordinates, and T~1 Sec for the origin time. Values of X, Y, T estimated in the first step are used together with a local velocity model to approximate P and S on-set times for each station and improved in the second step using **likelihood-maximum estimation**. For increased reliability additional P and S pickings are provided due to intersection of the corresponding envelope fronts with noise level. This last procedure is applied to the parts of the seismogram corresponding to the intervals containing the hypothetical P and S arrivals. Finally, a grid-search **robust optimization** procedure (60X60X25km3, ±10 sec), based on the local P and S velocity model, is used to provide accurate location. This procedure is based on maximization of the target function equal to the sum of the "bell-like" functions (equals 1when discrepancy is 0 and equals ~ 0 at infinity), accounting for good pickings and ignoring bad ones. Due to the robustness, the package does not require pre-selection of seismograms for events recorded with relatively good quality by at least several stations, even if the others are noisy or do not contain any signal at all. The quality of location is indicated by the value of the target function (≈ equal to the number of "good" stations) and/or the usual confidence ellipsoid estimation. The procedure was checked using records of the Israel Seismic Network from four calibration quarry blasts at the Arad quarry close to the Dead Sea, providing location accuracy of 0.3 to 1.7 km and, in all the cases, better than a qualified analyst. This was then applied to a data set of 61 earthquakes and quarry blasts occurring in Israel with an overall 80% accuracy of DX, DY ~3 km, DH~ 7 km and DT~0.7 sec.

Key Words: seismic event location, envelopes, maximum likelihood on-set estimation, grid search optimization.

OBJECTIVE

Within the general framework of improving the monitoring capabilities of the CTBT in the Middle East, this research is focused on implementation of innovative, automatic and robust location procedures intended to improve accuracy and enhance efficiency of regional network performance on the basis of data collected by the Israel Seismic Network (ISN).

RESEARCH ACCOMPLISHED

Automation of regional event location using ISN recordings is one of the tasks of our diverse research program initiated under the DSWA contract. This paper describes the continuation of the research first reported at the previous symposium presents the main aspects of our approach and the initial results of the prototype automatic location procedure.

INTRODUCTION

The routine source location of regional and local events is based on picking the onsets of P (first) and S (second) phases on the seismogram. These are then converted into the corresponding geographical source coordinates, depth and time, using the given travel time model as determined by the velocity structure. In doing this, we need to remove noise, spikes and sporadic simultaneous events. In such cases, analysts must rely on their view of the seismogram and previous experience. Our task was to imitate the performance of the analyst.

In the first stage of our investigation, we simplified the problem by picking not first arrivals, but maxima of energy of the corresponding phases having a better signal-to-noise (SNR) relation. The physical basis for this is that the energy of the P and S seismic phases obeys the law of diffusion and propagates in the stratified medium at a constant velocity. Based on this, Ryzhikov et al. (1996) derived a theoretical function over the observed envelopes of seismic phases which has a maximum pointing to the source epicenter. Further to this premise, and in parallel with Shoubik et al. (1997) and Husebye et al. (1998), we developed a procedure for estimating source epicenters, X,Y (no depth) and time, T, using specific envelopes of the seismogram (Pinsky and Shapira, 1998). This procedure utilized automatic picking of maxima Pmax and Smax from the envelopes and a robust search inversion method. The automatic algorithm yielded an accuracy for X,Y of ~6 km (local coordinates) and T ~1 sec on the data set from pre-selected seismograms (with clear seismic phases) from the ISN data base of 150 earthquakes and quarry blasts. However, it did not achieve the location accuracy (~2 km for X,Y; ~3 km for H and T ~0.3 sec) as set out in the bulletin. This is probably due to the relatively large variation in Pmax and Smax times around theoretical values, while the P and S first arrivals seem to be more stable, at least for earthquakes.

In view of this, we returned to the location method based on first arrival time estimation, but with reasonable preliminary estimates of X,Y and T from the envelope location procedure. There is a variety of automatic methods for finding first phase arrivals on seismograms. We tested one method based on maximum likelihood ratio (Pisarenko, et al (1987)) which was shown convincingly by Kushnir, (1996) to be equivalent to, or even better than, manual picking and another inspired by the envelope processing algorithm of Fedorenko and Husebye (1999). Both require a specified time interval, assumed to comprise phase arrival. We placed it on the time axis according to the preliminary X,Y,T and theoretical travel time computations for P and S phase first arrivals. Finally, source coordinates, time and depth are estimated through a grid search using a robust optimization procedure based on M-estimates developed by Huber (1972) and tested for seismic location by Andersen (1982).

The data used in this study are presented in the short period seismograms (50 Hz sampling rate with efficient bandwidth 0.5-12 Hz) of the Israel Seismic Network (~30 stations) telemetered to a Hub via radio FM link. The data include 61 earthquakes, M_L=1.0-2.8 at distances of 15-310 km and four calibration quarry blasts exploded in the Arad quarry with precisely known GMT coordinates and ignition time (see Table 1.). Event epicenters, station and quarry locations are presented in Fig. 1.

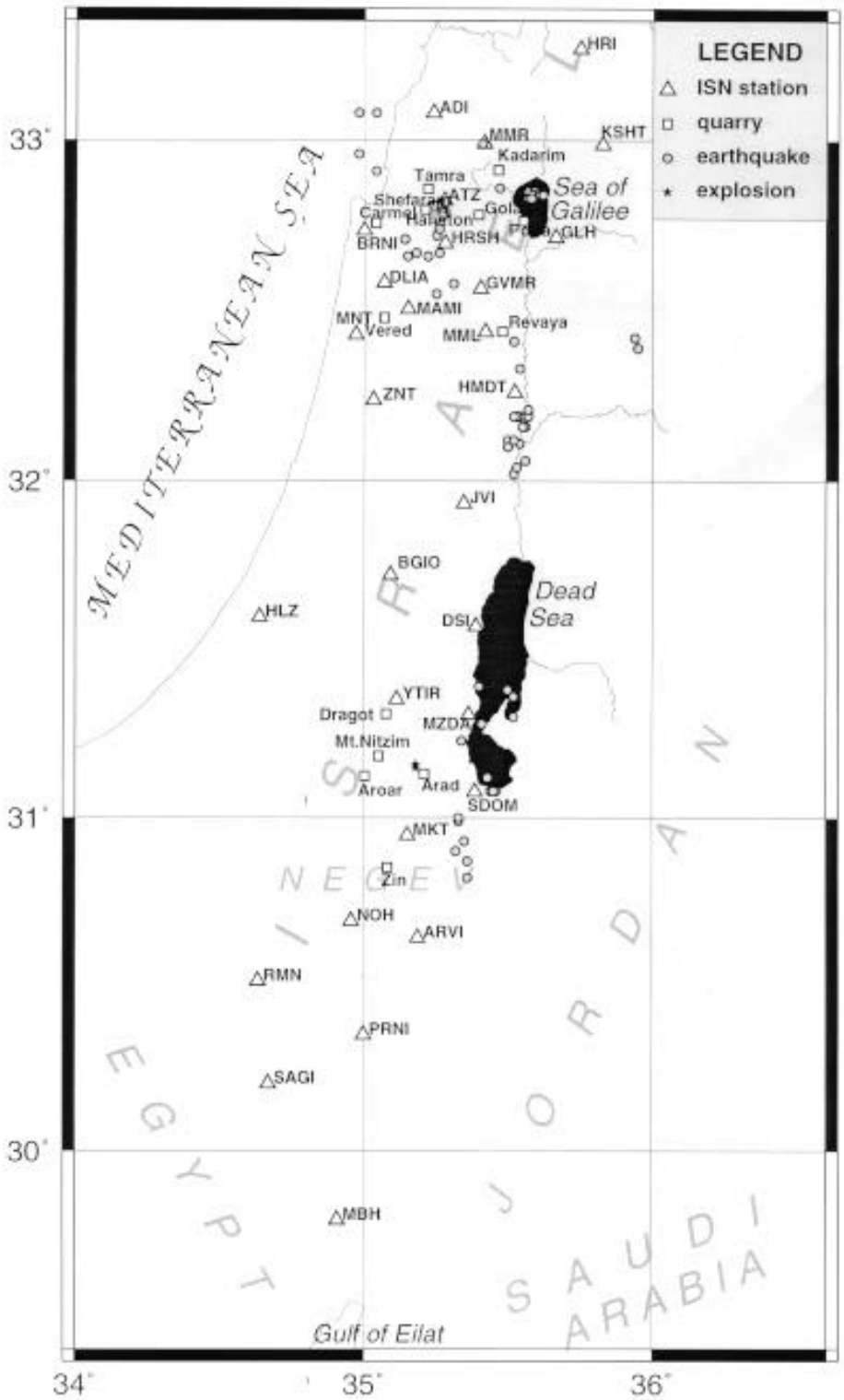

Fig. 1 Map of Israel and adjacent area, showing the data base region, earthquakes, quarries and explosions.

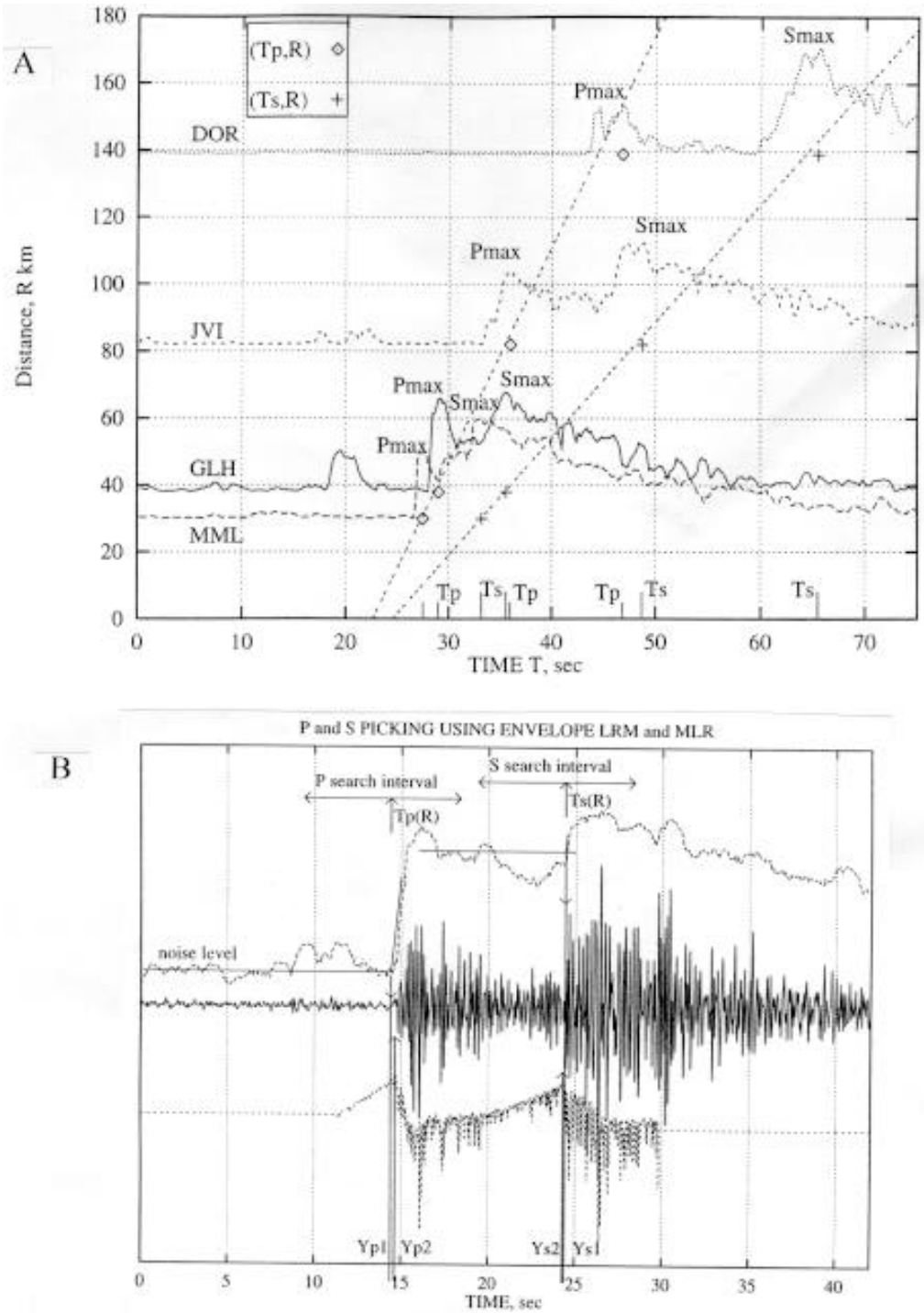

Fig 2. Automatic P and S phases picking using a) time picks Tp and Ts, corresponding to the local maximums of Pmax and Smax of the envelope curves and using b) time picks Ypi, Ysi, i=1,2 provided by the maximum likelihood ratio (lower curve) and linear regression method of the envelope (upper curve) front. The Tp and Ts time picks are compared to the travel time curves of the maximum of energy for the P and S phases. The Yp and Ys time picks are found inside P and S search intervals, guided by Tp(R) and Ts(R) theoretical travel times, where R is the distance from preliminary epicenter to the given station.

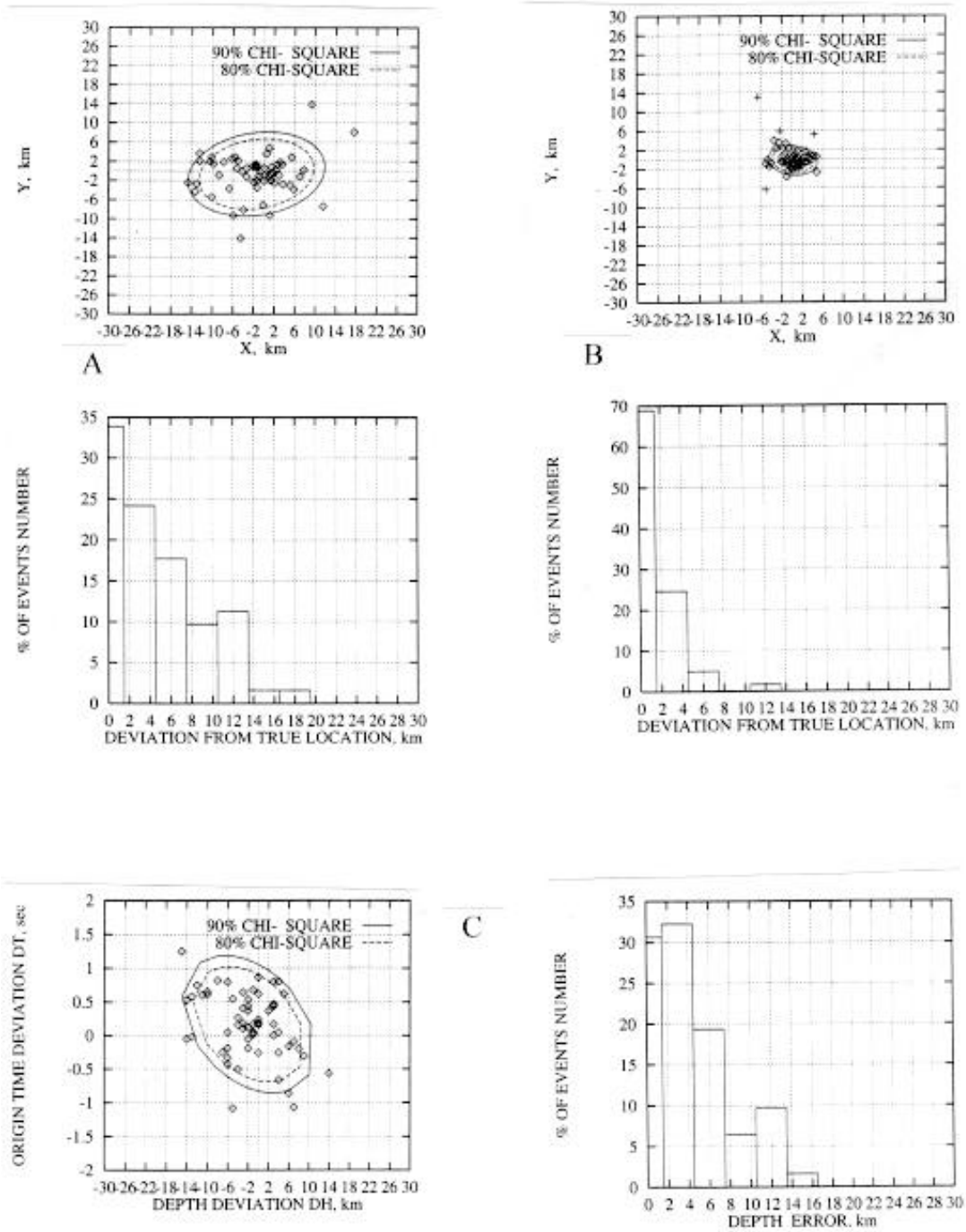

Fig. 3. Maps and histo grams of the 61 earthquake automatic locations relative ISN bulletin, provided by the a) envelope method of preliminary epicenter estimations due to the automatic on-set pickings.

METHOD

The method used comprises four main stages; (i) preprocessing; (ii) preliminary epicenter location; (iii) P and S onset picking and (iv) final hypocenter location.

Preprocessing: this includes despiking and band pass filtering of the seismograms in an appropriate frequency band and automatic selection of the traces with appropriate signal-to-noise ratio (SNR). The records used are often marred by spikes or noise outbursts caused by radio pulses or industrial and micro-seismic noise that cause inaccurate interpretation. It is, therefore, crucial to the success of a procedure based on detecting seismic signals, that records are clean. A spike is assumed to be "thin" and "high", thus we perform a simple de-spiking by cutting the whole trace into a series of intervals (2 sec. each) and comparing average noise level with its current value inside the interval. Naturally, some of the channels may suffer partial distortion of the signal front as a result of this procedure. The next step is to calculate seismogram envelopes for each station which are used for preliminary **epicenter location** by the **envelope location method** (Pinsky, 1999). Two different types of envelope are computed, both characterized by abrupt fronts and sharp energy maxima, even for rather weak seismic phases; thus providing, in most cases, a clear separation between P and S phases even at short distances. The envelopes have local maximums P_{max}, S_{max} at time points T_p and T_s which are coincident with P and S arrivals in the original non-transformed records. The quality of a trace is characterized by the SNR=P_{max}NL ratio, where NL is the noise level of the envelope. The traces with low SNR are omitted to give a clearer result, thus providing automatic selection of channels for further processing. Additional selection criteria are - closeness of T_p or T_s picks between the two types of envelopes, checking whether $dt_{\alpha\beta} = |T_\alpha - T_\beta| < dt0$, at least for one α, β pair, where α, β {p, s}; dt0 meanwhile is chosen as equal to 2 sec. Those traces that are left after trace selection as described above, are arranged according T_p time, the first having the minimum T_p among M channels, having the largest SNR. In this way it is possible to estimate the station which is probably closest to the source.

<u>*PRELIMINARY EPICENTER LOCATION*</u> is provided by the mentioned above **envelope location method**, based on fitting time-distance (T_p,R) , (T_s, R) pairs to the travel time curves presented in this case by the two lines: $T_{\mathbf{p}} = T_{0\mathbf{p}} + R/V_{\mathbf{p}}$, $T_{\mathbf{s}} = T_{0\mathbf{s}} + R/V_{\mathbf{s}}$, which characterize propagation of maximum of energy for the regional P and S seismic phases (see Fig. 2a) correspondingly. The parameters of the two equations, estimated through a least squares fitting for a set of earthquakes and quarry blasts, are as follows V_p=6.0, V_s=3.5, T_{0p}=2.5, T_{0s}=3.5. The $T_{\mathbf{p}}$ and $T_{\mathbf{s}}$ estimates, found by triggering the envelopes with the following maximum determination are not especially reliable. However, for the case where we have sufficient stations (NS > 10) with good SNR, true $T_{\mathbf{p}}$ and $T_{\mathbf{s}}$ are the expected majority and, therefore, a robust optimization that ignores false observations may provide a satisfactory source location. The optimization procedure of such a class was specified by Huber, 1972 as M-estimators. The task is to maximize (minimize) target function

$$F(\theta) = \sum_r w_r f(\Delta_r(\theta)) \qquad (1)$$

with a gain (loss) function **f**(Δ) of residuals Δ (Δ=[observations minus model]). The usual and, in many cases efficient, procedure for location is a least-squares technique in which the solution is due to minimization of a sum of squares of the residuals $\Delta_{ai} = T_a - T_a(R_i) - T_0$, a = P,S. for a number of stations, where T_0 is the source time. If, however, the observations include random outliers of the time T_a then this procedure is inefficient and, besides, may cause large errors in location. We observe, that many P identifications give S travel time and vice versa by mistake; so a random T_a variable for the fixed a should bi-modal or even multi-modal distribution. This feature should be taken into account in the specialized location scheme since the conventional least-squares procedure is not applicable in this case.

In such cases residuals appear as

$$\Delta_{aci} = T_a - T_c(R_i) - T_0 \qquad (2)$$

Next, instead of one station's residuals Δ_{aci} we shall consider two station residuals $\Delta_{abcij} = \Delta_{aci} - \Delta_{bcj}$, ij=1,N, *a*, *b*, *c* = P, S, which do not depend on origin time T_0, thus, reducing the number of unknown variables from 3 to 2. From the "picker" above we obtain two time picks T_p and T_s for each channel, but the "picker" is not aware to which phase: P, S, or Rg they belong. To distinguish between correct and incorrect identifications we use a gain function in Eq. (1)

$$f(\Delta) = (|\Delta| + \varepsilon)^{-1} \qquad (3)$$

deviations Δ (instead of the usual $\mathbf{f(\Delta)=\Delta^2}$), "voting" for true pickings for its large values and against false pickings for its small values respectively. Thus it is reasonable is to estimate epicenter X,Y by maximization of the $F(\theta)$ function (1), which is provided by computing the $F(\theta)$ value at each of the grid points (say, 60X60 km of 1 km step).

Consequently, the procedure unites in one the two tasks: the location with the phase identification, where majority of "good" pickings, close to P or S track, would guarantee successful solution. The control parameter $\varepsilon > 0$ helps to avoid possible zeroes of the residuals Δ_{abcij}, thus providing robustness of the X_{max}, Y_{max} estimator, which maximize the $F(\theta)$. After that we estimate origin T_0 providing a maximum to $F(\theta)$ with $f(\Delta)$ due to (3) and fixed $X= X_{max}$, $Y= Y_{max}$. and Δ determined by (2). The parameter ε is selected automatically as ε=0.1+k*2., k=l,..,5 according to the empirical criterion that for more than N/2 stations there is at least one combination of *a*, *b*, *c*, *i* and *j* that ($|\Delta_{abcij}|+\varepsilon$)<3.

<u>*P and S ONSET PICKINGS*</u> are provided by estimating P and S first arrivals $\mathbf{Y_p}$ and $\mathbf{Y_s}$, which seems more stable than T_p and T_s for good SNR conditions. For the automatic $\mathbf{S_p}$ and $\mathbf{Y_s}$ estimation we chose a statistically optimal algorithm proposed by Kushnir (1994) and shown to be efficient for seismic phase picking. The procedure is based on statistical testing of the hypothesis, that time series before and after signal arrival are two different autoregreessive time-series in the chosen time interval [T1,T2], possibly containing the point of interest T0. The two autoregressive models are approximated using "noise" and "signal" recordings corresponding to the opposite edges of the interval. The point T0 is estimated as a maximum likelihood ratio (MLR) by scanning through the interval (see Fig. 2b). Another approach, demonstrated in Fedorenko & Husebye (1999) is to look for the point where signal energy starts to grow. This comprises two steps: (1) linear regression of the envelope front in the neighborhood of a trigger point **Tr** and (2) computation of the intersection point **Ti** of the line obtained with the noise level, which provides us with an estimate of a seismic phase onset time (see Fig. 2b) and the inclination angle **A** . Let us call this the Linear Regression Method (LRM). Both methods require determination of the time interval [T1,T2] inside which the onset time T0 is expected. The interval is computed as T1=T(R)-dtl, T2=T(Ri)+dt2, where T(R) denotes P or S travel time determined by the preliminary epicenter location and by the local velocity model. The values of dt1 & dt2 are set at 5 and 4 sec, correspondingly for both P and S phases.

<u>*HYPOCENTER LOCATION*</u> is a robust grid search procedure based on maximization of a type (1) function equal to the linear combination of the "bell-like" gain functions $\mathbf{f}(\Delta)$ (equal to 1 when the discrepancy is 0 and equal to ~0 at infinity). The algorithm is comprised of two steps: first, the P arrivals $\mathbf{Y_p}$ provided previously by the two onset estimators, MLR and LRM, are used in computing:

$$\Delta_{ip} = Y_p - Y_p(R_i,H) - T_0 \qquad (4)$$

where i=1,...,N is a channel index, $\mathbf{Y_p(R_i,H)}$ - is a local travel time model for the first P phase arrivals at station number **i**, $\mathbf{T_0}$ -origin time,

$$\begin{aligned} f(\Delta) &= \exp(-|\Delta|), & |\Delta| &< 0.5 \\ f(\Delta) &= 0. & |\Delta| &> 0.5 \end{aligned} \qquad (5)$$

The scanning grid is (60x60x25 kM^3, ±10 sec) with a relatively large step dx,dy,dh=2 km, dt=0.5 sec. The algorithm using gain function (5), ignores all observations **S**, deviating more than 0.5 seconds from the model. When the optimum is found, it tries to reach higher accuracy. For this purpose it continues computations on a smaller grid (5x5x25 km^3, ±3 sec), dx,dy=0.3 km, dh=l km, dt=0.l sec., using (4) for computation of Δ_{ip} and Δ_i. (with index s substituting p) but using gain function due to

$$f(\Delta) = \exp(-|\Delta|), \qquad |\Delta| < 0.5$$
$$f(\Delta) = 0. \qquad |\Delta| > 0.7 \tag{6}$$

Weighting coefficients $\mathbf{w_r}$ in (1) are set $\mathbf{w_r = l. -exp(-Tg(A_r))}$, where $\mathbf{A_r}$ is the envelope inclination angle determined above. Thus, abrupt onsets are highlighted and overly slow onsets neglected. Reliability of location is determined by the maximum value of the F(θ) function close to the number of donating channels. Accuracy is determined by its sharpness in the $R^?$ space, or as usual by the error ellipsoid.

RESULTS

Under the supervision of the Geophysical Institute of Israel a series of four calibration explosions was performed at the Arad quarry (see Fig. 1) with controlled source parameters: coordinates, time of ignition, charge, delays etc. The seismograms of the explosions were recorded by the ISN stations and processed by the analysts.

In Table 1 we compare the results of the analysts' location with those provided by the fully automatic procedure. From the table we see that though preliminary estimates provided by the envelope processing algorithm are less accurate than those provided by the analysts, the final automatic decision is the best and is very close to the real source location. We applied the developed automatic algorithm to a data set of 62 earthquakes from the Galilee, Gilad and Dead Sea areas. Our epicenter locations results in Figs. 4 and 5 1D by histograms of deviations R from X_0,Y_0 and 2D by plots of points (X-X X_0,Y- Y_0). Here X_0, Y_0 are presumed true source coordinates as reported in bulletins. The 80% and 90% error ellipsoids are presented on the corresponding plots. From Fig. 3a we see that preliminary epicenter estimates provided by the envelope processor STA+OD are sometimes rather rough. Nevertheless, epicenters demonstrated in Fig. 3b are significantly improved by the MLR+RLM procedure and in many cases are equivalent to the bulletin solutions, showing 80% of epicenters "mislocation" dR < 3 km (assuming bulletin solution true) and standard deviation α(dR)=2.1 km. Fig. 3c depicts 2D estimates of origin time T0 and depth by MLR+RLM procedure and shows that T0 is well done; the depth estimates, however, often deviate considerably from those in the bulletin.

CONCLUSIONS AND RECOMNLUENDATIONS

The above results appear promising for atomization of local seismicity monitoring. Our new procedure performed well providing robust and accurate solutions for earthquake epicenters in the different geological regions of Israel; the Galilee, Gilad and Dead Sea areas, as well as for the four calibration explosions in the Arad quarry. There are some locations which deviate from the results of standard processing , but usually these events are qualified in the bulletin as having "low reliability" and, as such, are difficult to judge as ground truth. The same remark is true for the depth estimates which are known to be poor in the bulletin quite often due to the low resolution of the existing system of surface observations. We also tried to apply this technique to the local quarries. For many of them the results were very accurate. But almost 10% of the events proved problematic for analysis most of them due to the failure on the preliminary location stage. The reason is that this kind of signal source causes strong excitation of the Rg surface waves, propagating with low velocities in the range of 1-3 km/sec. This may, in turn, result in too late pickings of the true P and S phases. Another factor is scattering and multi-pathing due to heterogeneity of the uppermost part of the crust. The local seismicity is 99.9% represented by quarry blasts and earthquakes, hence our suggestion that earthquakes/quarry blast automatic discrimination be carried out before the automatic location as shown to be reliable in Gitterman et al. (1998) with around 98% reliability. If an event is classified as an earthquake it can automatically be located by the procedure developed. On the other hand, if it is a quarry blast, it is

possibly easier to identify the source quarry among the set of known quarries using pattern recognition approach as in Joswig, 1990 or, Fedorenko et al, 1998.

Table 1. Results of Location of the Four Calibration Explosions by Routine and Automatic Techniques

			ROUTINE PROCESSING			AUTOMATIC PROCESSING					
			SEIS			Envelope Processing			Onset Estimation		
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Event N date	Estim. Coord	GPS	Analyst	**Error km**	N St.	STA+ OD	**Error km**	N St.	MLR+ LRM	N St.	**Error km**
N1 30.07.98	X Y	167.9 57.0	167.3±2.0 55.0±2.3	**2.1**	15	167.0	**3.1**	13	167.2	15	**1.737**
N2 26.01.99	X Y	168.0 57.0	167.1±1.3 57.1±0.8	**0.9**	21	167.0 59.80	**3.0**	15	167.8 56.4	19	**0.63**
N3 26.01.99	X Y	168.0 57.6	166.3±1.2 58.1±1.5	**1.77**	37	166.5 55.8	**2.34**	24	167.7	35	**0.36**
N4 27.01.99	X Y	168.0 57.5	170.3±5.3 56.4±2.7	**2.55**	10	168.6 62.1	**4.64**	11	167.2 57.7	16	**0.82**

ACKNOWLEDGMENT

We would like to thank Eystein Husebye for suggesting this research and careful assistance in the period of study, Boris Gurevich and Yefim Gitterman for fruitful discussions. This study is supported by the U.S. Department of Defense, Contract Number DSWA01-97-C-0151.

REFERENCES

Andersen K.R.(1982). Robust earthquake location using M-estimates. *Phys. Earth Plan Inter*., .**30**, 119-130.

Fedorenko, Yu., E.S. Husebye, B. Heincke and B.O. Ruud (1998) Recognizing explosion sites without seismogram readings: Neural network analysis of envelope transformed multistation SP recordings 3-6 Hz. *Geoph J Int*., **133**, F1.

Fedorenko, Yu.V., E.S. Husebye, 1999. First breaks-Automatic pickings of P- and S- onsets in seismic records. Geoph. Res. Let., submitted for publication.

Y.Gitterman, V. Pinsky & A. Shapira.(1998). Spectral Classification Methods in Monitoring Small Local Events by the Israel Seismic Network. J. Seismology 2, p. 237-256.

Huber, P.J. (1972). Robust statistics: a review. Ann.Math.Stat.,**43**:1041-1067.

Husebye, E.S., B.O.Ruude.and A.M. Dainty (1998). Fast Robust and Reliable Epicenter Determinations: Envelope Processing of Local Network Data. . *Bul. Seism. Soc. Am*., **88**, pp. 284-290.

Joswig,. M. (1990). Pattern recognition for earthquake detection. *Bul. Seism. Soc. Am.* **80**, pp. 170-186.

Kushnir, A.F. (1995). Algorithms for adaptive statistical processing of seismic array data. In Monitoring Comprehensive Test Ban Treaty Ed. E.S. Husebye abd A.M. Dainty, NATO ASI Series. Series E: Applied Sciences - Vol. 303.

Ryzhikov, G.A., MS. Birulina and E.S. Husebye. (1995). Automatic event location using local network data. pp. 393-399 Proceedings of the 17th Annual Seismic Research Symposium on Monitoring a Comprehensive test Ban Treaty, 12-15 September 1995.

Shoubik, B.M., A.F.Kushnir, L.M. Haikin, A. Dainty (1996). SCANLOC: Automatic seismic event location based on local seismic network data. Proceedings of the 18th Annual Seismic Research Symposium on Monitoring a Comprehensive test Ban Treaty, 4-6 September 1996, pp.7 74-781.

Pinsky, V. and A. Shapira. Regional Events Location Using Seismic Energy Time-Space Distribution, Proceedings of the 20th Annual Seismic Research Symposium on Monitoring a Comprehensive Nuclear Test Ban Treaty, Santa Fe, September 21-23, 1998, p.265-274.

Pinsky V. I. (1999). An Approach to Automatic Location of Regional Events. PEPI ELSEVIER- Accepted for publication 7 July 1998.

Pisarenko, V.F., A.F. Kushnir and I.V. Savin (1987) Statistically adaptive algorithms for estimations of onset moments of seismic phases, Phys. Earth. And Planet. Interiors, 47, 888-900.