Bull. Nov. Comp. Center, Math. Model. in Geoph., 3 (1998), 51–58

On accuracy of estimation of the envelope maximum

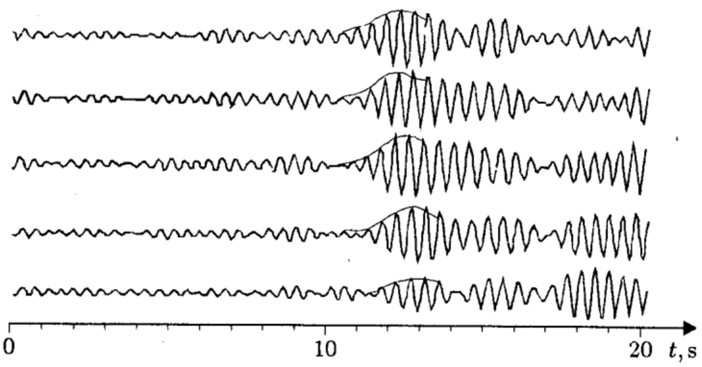
B.M. Glinsky and V.I. Znak

A specific character of "the wave arrival time" concept with reference to seismorecords taken from a vibrating source is considered. The arrival time is defined as the envelope maximum on the grouped waves of the corresponding convolution. The accuracy of estimation of the envelope maximum depending on the degree of approximating polynomial in conditions of various *signal-to-noise ratio* is investigated.

1. Statement of problem

The accuracy of estimation of the wave arrival time depends on a number of factors: definition of the current concept, criteria, and procedures. We will formulate a problem arising here, with reference to a seismorecord taken from the vibrating source of fluctuations.

First of all, various kinds of errors are imposed on a seismorecord of signals in the field conditions. For this reason one may speak about a casual nature of the function, reflecting a real oscillatory process at a recording point. Therefore for processing and analysis of such a stochastic function let us address to the appropriate techniques of mathematical statistics. In the most general form, the problem in question is considered as estimation of the moment of changing the statistical characteristics discussed in many works, in particular, [1, 2]. However, it seems that the multi-criteria approach given in work [3], permits obtaining the higher accuracy of estimation. However, the offered algorithm is limited by the VSP, i.e., the corresponding criteria in our case should be accepted critically. Not dwelling on opportunities of each of the considered criteria in work [3], we address to the correlation criterion, as the most universal. In particular, the convolution of the original seismorecord with a certain basic signal permits us to pass to the appropriate correlation function or to the correlogram of a signal, where the *signal-to-noise ratio* essentially increases. As a support signal either a seismorecord of the original signal near to a source or its mathematical model is used. For realization of transformation the technique offered in work [4] can be used. Correlograms of some real signals are presented in the figure (convolution of seismograms with the model $y = x \sin x$ has been kindly given by A.P. Grigoryuk).



Convolution of sweep-signals with the model $y = x \sin x$

We shall further assume that the necessary transformation of the original signal has been done, and it is required to estimate the wave time arrival on the corresponding correlogram.

First of all note that a signal from the vibrating source of fluctuations is transformed to a new quality and now it represents, in essence, a correlation function. Specific fetures of the latter are determined by both the quality of the basic signal (by accuracy of its conformity with the original), and by the length of its time interval. In any case the specific character of the function is such that it represents a variable harmonic signal, whose maxima distribution, as we believe, corresponds to the distribution of the wave arrivals time on the original seismorecord. Here we will not dwell on the aspects related to identification of waves. We only assume that at any case in the correlogram under study there is distinguished an area containing an attribute of the wave arrival time (or the corresponding maximum). It is required with rather a high accuracy to estimate the time corresponding to the maximum of the picked out area of the correlogram, which is taken as moment of the wave arrival time.

2. Algorithm of estimation of wave arrival time

Correlograms, for which the envelopes in the areas of interest are picked out are shown in the figure. We offer to make an estimation of the time, at which the maximum of the picked out area of a correlogram is located, in two stages: at the first stage we are building up envelope (components are listed) in the positive half-plane, limited by the ends of *a priori* picked out area; at the second stage, analysis of the envelope is made, and the time is estimated, at which there is a maximum of the picked out area of a correlogram.

As was marked, the considered correlogram represents a stochastic function (by virtue of values of the original vibrogram). It means that the value of a maximum of each period of such a correlation function inside the picked out area represents a random value. Thereby the envelope of interest is a stochastic function, and the position of a maximum on it obeys to a law of random numbers, i.e., special methods are demanded, enabling us to approach an estimation of a position of a maximum to the true position (to minimize the corresponding error or dismatching of the values "the arrival time – the estimation").

One of such methods, verified in practice, is a random function approximation by the least square method. The essence of the method is described in a number of works. Here we shall refer only to one of them, [5]. The question is only in the choice of a degree of approximating polynomial. To answer this question we should note that the increase of the degree of such a polynomial brings about (starting with some moment) the decrease of its smoothing effect (with approaching of a polynomial degree to the number of terms of approximate functions the polynomial begins "to follow" all the singularities of the latter). At the same time a polynomial of small degree can result in the constant dismatching of a true position of a maximum and its estimation. Note, the number of envelope terms, as a rule, is about ten.

For the practical analysis of an envelope we offer to obtain a number of estimations of the wave arrival time (estimations of a position of a maximum) with attraction of approximate polynomials of various degrees, in particular, of degrees from the second to the fifth order. Then the algorithm of estimation of the wave arrival time in the correlogram of a signal from the vibrating source of fluctuations is of the following form:

- building up an envelope in the picked out area, whose terms will be maxima of the positive half-periods of the harmonic component of the correlogram in question;
- picking out such a moment of time, which would correspond to the maximum of values from the terms of the envelope, which is further used as datum mark for subsequent estimations;
- calculating the factors for each of the offered polynomials, which would permit us to present an approximating function in the form

$$y(n,t) = \sum_{i=0}^{n} a_i t^i, \qquad n = 2, \ldots, 5; \tag{1}$$

- searching for the roots of the functions

$$y'(n,t) = \sum_{i=1}^{n} i a_i t^{i-1}, \qquad n = 2, \ldots, 5; \tag{2}$$

- picking out such roots of the functions $y'(n,t)$ as estimations of the position of a maximum, which would be the closest to the datum mark;
- numerical estimation of positions of the corresponding maxima of the functions $y(n,t) = \sum_{i=0}^{n} a_i t^i$, $n = 2, \ldots, 5$;
- comparison, relatively to the datum mark, of the above numerical estimations with the previously selected analytical estimations (check of the analytical estimations with their subsequent rejection and replacement by numerical estimations in case of an unsatisfactory quality);
- on the basis of the analysis of the obtained estimations (including the datum mark) the desicion is taken about the value of the final estimation of a position of the envelope maximum.

For the choice of a procedure, which would supply the final estimation, we compare the accuracy of the estimation, obtained as mathematical expectation, to the accuracy of the estimation obtained as the median of intermediate results.

We further investigate the accuracy of the estimation of the position of the envelope maximum as dependence on the degree of an approximating polynomial and on the accuracy of the final estimation.

3. Accuracy of estimation of the position of the envelope maximum

Components of a few envelopes, shown in the figure, are given in Table 1. Analysis of a model of such a signal will be made with the use of the parabola $y = -20t^2 + 1680t - 4000$, which is modified in one way or another. This parabola is distorted by a random noise with zero mathematical expectation and with dispersion satisfying the required *signal-to-noise ratio*.

As one can see from the envelopes of real correlograms (see the figure), the corresponding function can have more than one extremum or a point of inflexion, where a derivative of the function changes its sign. According to the above-said an additional extremum on the left (see "Model 1" in Table 1) is introduced into the model. The next form of the model is presented in Table 1 as "Model 2". Here, instead of the point of inflexion a steep side, which is also present on the left, is introduced. Basically, there is an innumerable number of modifications of formz. However for studying the quality of estimation of a position of the maximum it is sufficient to consider a limited number of them. As one can see from Table 1, the maxima of the functions, given by Models 1 and 2, are localized at $t = 42$ *discrete time*. A hundred experiments have been carried out for each particular case to investigate the connection between the accuracy of estimations and the degree of approximating polinomial of an envelope. The results of experiments for

Table 1

N	Envelope 1		Envelope 2		Model 1		Model 2	
	t_1	y_1	t_2	y_2	t_1	y_1	t_2	y_2
1	0	0	0	0	0	15600	18	−4000
2	7	−2973	7	2788	7	6780	19	6780
3	13	−7034	15	8434	14	15600	20	15600
4	19	−1087	22	16062	21	22460	21	22460
5	26	5262	29	22460	28	27360	28	27360
6	33	8709	36	23659	35	30300	35	30300
7	41	9894	43	19832	42	31280	42	31280
8	48	6119	50	12556	49	30300	49	30300
9	55	−3409			52	27360	56	27360

the case of the *signal-to-noise ratio* equal to 5 are given in Tables 2 and 3 according to Models 1 and 2. Here the zero line contains the results of estimation for the case of a clean signal, i.e., under conditions of the absence of distortions. The results of experiments for the case of the *signal-to-noise ratio* equal to 2 are given in Tables 4 and 5 according to Models 1 and 2. Thus, in all the tables Line M represents the appropriate mathematical expectations; Line σ – the quadratic mean deviations of estimations from the true value; Column "Datum mark" – the values of a moments of time, taken as datum mark values; Column "E" is an average value of the corresponding experiment, Column "Med" represents the median values.

A sample of 100 experiments is not completely representative. However the further increase of the number of experiments would lead to changes in the decimal digits after third, i.e., the presented values may be used for the qualitative analysis and subsequent conclusions.

4. Analysis of results. Conclusions

As is seen from Tables 2–5, neither of the considered modes of estimation of the envelope maximum is sufficiently reliable. As this takes place the dependence of dismatching on the *signal-to-noise ratio* is not linear (see the quadratic mean of the estimation deviations from the true value in Tables 2, 4, and 3, 5). For a more accurate answer to the given question additional studies are required.

Comparison of the estimation results, obtained by means of averaging and by the median, permits us to speak about some advantage of the latter in terms of accuracy. However, as a whole, the known result [6] about their closeness in the statistical sense proves to be true.

It seems that the accuracy of estimations can be improved by means of attraction of the weighted-order sratistics [7, 8]. However, additional inves-

Table 2. Approximations of Model 1; signal-to-noise ratio = 5

N	Datum mark	Polynomial degree				Estimation	
		2	3	4	5	E	Med
0	42	44.80	42.52	38.02	47.17	42.90	42.52
1	42	44.80	42.66	38.25	38.08	41.16	42.00
...	...	...	...	...	...	...	...
M	40.53	44.92	42.53	38.11	41.25	41.47	41.60
σ	3.21	2.95	0.54	4.00	6.05	1.79	1.68

Table 3. Approximations of Model 2; signal-to-noise ratio = 5

N	Datum mark	Polynomial degree				Estimation	
		2	3	4	5	E	Med
0	42	42.05	35.27	28.47	42.49	38.06	42.00
...	...	...	...	...	...	...	...
M	40.18	42.06	36.05	28.53	40.29	37.42	40.02
σ	3.57	0.33	6.61	13.48	4.11	4.82	3.63

Table 4. Approximations of Model 1; signal-to-noise ratio = 2

N	Datum mark	Polynomial degree				Estimation	
		2	3	4	5	E	Med
1	35	45.11	42.39	37.51	33.31	38.67	37.51
...	...	...	...	...	...	...	...
M	39.06	45.00	42.53	38.42	39.07	40.82	41.12
σ	5.24	3.10	0.59	4.12	6.49	2.25	2.16

Table 5. Approximations of Model 2; signal-to-noise ratio = 2

N	Datum mark	Polynomial degree				Estimation	
		2	3	4	5	E	Med
1	42	43.45	49.29	27.50	45.93	41.63	43.45
...	...	...	...	...	...	...	...
M	39.69	42.09	38.80	28.51	39.55	37.73	39.35
σ	4.48	0.92	6.86	13.52	5.49	4.94	4.65

tigations are also required. Finally, the main question of interest here is: the connection of the dismatch "arrival time – estimation" with a degree of approximating polynomial of envelopes. Such a connection exists. As this takes place the value of the estimation deviation from the true position of the envelope maximum is determined not only by a degree of approximating polynomial and the *signal-to-noise ratio*, but also by the shape of the envelope and distribution of its singularities.

First of all, rare, but significant misses are intrinsic of this datum mark. The frequency of such misses grows with the fall of the *signal-to-noise ratio.* As consequence it reduces the accuracy of the corresponding values of the mathematical expectation E.

To the second degree polynomial there corresponds rather stable estimations, which are characterized by a constant displacement in the case of an additional (except the point of a required maximum) inflexion of the envelope. Such a displacement takes plase even in the case of the absence of noise (see Line 0 in Table 2). In the other considered case the second degree polynomial permits obtaining estimations of suffitiently high accuracy. Hence, if a method of estimating the value of constant displacement in each particular case is developed, this polynomial can be widely used. However, even with such features the estimation obtained here can be used as datum mark. In spite of its possible displacement the corresponding estimation is ensured from miss, whose value essentially surpasses a displacement which is possible in the latter case.

As in the previous case, in the estimation by the third degree polynomial a constant displacement is present, if the envelope contains an additional inflexion. However the value of displacement in this case is much smaller. If an envelope contains a steep side (Model 2), the accuracy of the estimation obtained sharply falls, being practically constant low independent of *signal-to-noise ratio.*

The estimations obtained with the help of the fourth degree polynomial are of the poorest quality of all the considered cases, the steep side of the envelope being of even more importance than in the previous case.

Finally, the fifth degree polynomial has not got appreciable advantages as compared to the former. However here a stable displacement of an estimation is absent even in the case of the presence in the envelope of an additional inflexion. The disorder of estimations seems to be casual. As this takes place a steep side does not result in the growth of an error, as in the two previous cases. A more accurate characteristic of polynomial requires an additional study of properties of an estimation, i.e., its parameters in the statistical sense.

As summary, let us outline the direction of the further research and note that the variety of the attracted estimations of various character is a unique guarantee of obtaining an appropriate final result.

In case of attraction of the weighted-order statistics for the analysis of estimations obtained it is desirable to investigate an opportunity of the dynamic change of weights of an estimation depending on the presence in the envelope of additional inflexions and steep sides.

As recommendation for the practical use of results of the given work we should note that: the fifth degree polynomial can be used in the analysis of data, obtained by means of seismic groups. The techniques of such an analysis are considered in book [9].

References

[1] S.M. Prigarin, *The analysis of estimations of a moment of disorder*, Problemno-orientirovannye vychislitel'nye kompleksy, Novosibirsk, 1985, 87–92 (in Russian).

[2] I.V. Nikiforov, *Consecutive detection of change of properties of time series*, Nauka, Moscow, 1983 (in Russian).

[3] V.I. Znak and S.M. Panov, *Algorithm of interactive processing of vertical seismic profiling data*, Geology and Geophysics, No. 8, 1992, 121–127 (in Russian).

[4] S.I. Dryakhlov, *Line convolution of large sequences*, Problemno-orientirovannye vychislitel'nye kompleksy, Novosibirsk, 1990, 54–58 (in Russian).

[5] N.S. Bahvalov, *Numerical methods*, Nauka, Moscow, 1975 (in Russian).

[6] B.I. Justusson, *Median filtering: statistcal properties: two-dimentional digital signal processing II*, in: T-transforms and Median Filters, Springer–Verlag, Berlin, 1981, 161–196. (Russian translation: Radio i Svjaz, Moscow, 1984).

[7] V.I. Znak, *Weighted median filter and VSP result processing*, Problemno-orientirovannye vychislitel'nye komplekcy, Novosibirsk, 1992, 78–85 (in Russian).

[8] V.I. Znak, *Some models of noise signals and heuristic search for weighted-order statistics*, Math. Comput. Modelling, **18**, No. 7, 1993, 1–7.

[9] K. Aki and P.G. Richards, *Quntative Sesmology. Theory and Methods*, Vol. II, 1980. (Russian translation: Mir, Moscow, 1983).