Bull. Nov. Comp. Center, Math. Model. in Geoph., 3 (1998), 111–116

Percentile filtering: some generalizations; geophysical data processing

V.I. Znak

We present some definitions of the percentile filter and the weighted-order statistics. Then we offer the probability interpretation of the weighted-order statistics and the definition of the percentile filter for this case. To conclude, we give some examples of data processing, characteristic of some fields of geophysics.

Introduction

At the present time a large number of publications deal with the median filters. Now we only refer to works [1–3]. In their essence the median filters are based on the order statistics or the notion of a variational series

$$x_{1,n} \leq x_{2,n} \leq \ldots \leq x_{n,n}, \tag{1}$$

where the terms are ordered in size. In turn, the components of series (1) are the terms of some sequence of independent and equally distributed random values

$$x_1, x_2, \ldots, x_n = X. \tag{2}$$

The terms of the variational series $x_{r,n}$ $(r = 1, \ldots, n)$ are referred to as order (rank) statistics, i.e., the r-th order statistics $x_{r,n}$ of the sequence X is the r-th in size term of the given sequence. Let a ratio

$$\alpha = \frac{r-1}{n-1} \tag{3}$$

be a measure of the size of the corresponding sample on the sequence X. Then an idea of the order statistics is transformed to the following percentile form. A sample of the sequence $X = \{x_1, \ldots, x_n\}$ is called the α-order statistics, denoted as $x_{(\alpha,n)}$, if there exists a number

$$n_\alpha = (n-1) \cdot \alpha \tag{4}$$

of $x_{i(\alpha)}$ quantities and a number

$$n_\beta = (n-1) \cdot (1-\alpha) = (n-1) \cdot \beta \tag{5}$$

of $x_{i(\beta)}$ quantities under the conditions

$$x_{i(\alpha)} \leq x_{(\alpha,n)} \leq x_{i(\beta)}, \qquad x_{(\alpha,n)} \cup \left(\bigcup_{i=1}^{n_\alpha} x_{i(\alpha)} \right) \cup \left(\bigcup_{i=1}^{n_\beta} x_{i(\beta)} \right) = X, \tag{6}$$

and

$$\alpha+\beta=1. \tag{7}$$

This latter definition of α-order statistics in the sequence will be further used. If $\alpha=\beta=0.5$, then the sample $x_{(0.5,n)}$ is called the median.

Let us have a set K of the integers k_i, $i=\overline{1,n}$, $k_i\in\{0,1,2,\ldots\}$, and the elements k_i are assotiated with the samples $x_i\in X$. Assume each k_i from the set K or the weight of the corresponding sample is defined as the number of copies of the corresponding sample $x_i\in X$. It means the substitution of each x_i for a subset Y_i such that the number of its elements is k_i under the condition

$$y_l\equiv y_j,\quad \forall y_l,y_j\in Y. \tag{8}$$

Weights are introduced for giving more weight to the central points and for emphasizing some other elements of the sequence. The extended sequence X is transformed thereby to a new set $\overline{Y}=\{Y_1,\ldots,Y_n\}$ with the number of elements $N=\sum_{i=1}^{n}k_i$.

By analogy with the previous case we call $y_{(\alpha,N)}$ the α-order element if there exists a number $N_\alpha=(N-1)\times\alpha$ of $y_{i(\alpha)}$ quantities and the number $N_\beta=(N-1)\times\beta$ of $y_{i(\beta)}$ quantities under the conditions

$$y_{i(\alpha)}\le y_{(\alpha,N)}\le y_{i(\beta)},\qquad y_{(\alpha,N)}\cup\left(\bigcup_{1}^{N_\alpha}y_{i(\alpha)}\right)\cup\left(\bigcup_{1}^{N_\beta}y_{i(\beta)}\right)=\overline{Y},$$

and $\alpha+\beta=1$. These expressions are formally the same as expressions (2)–(5), and the previous remarks are valid for this case. Thus, we have an opportunity to use expressions (2)–(5) in the most general case.

In this paper we offer the probability interpretation of the weighted-order statistics, then we give the definition of the percentile filtering in the most general form and consider some possibilities of its application for processing of data that are characteristic of some fields of geophysics.

Further, the size of a series will be called the aperture of the filter.

1. Probability treatment of weighted-order statistics; procedure of percentile filtering

Let there be a vector of weights $W=(w_1,\ldots,w_n)$. The value $p_i=w_i/S$, where $S=\sum_{i=1}^{n}w_i$, is considered as probability of the choice of an element $x\in X$ as a result of processing of the sequence X, i.e., the output of the filter of some $x\in X$ is considered as an event to which a certain probability $p_i\le 1$ $(i=1,\ldots,n)$ is associated. For the percentile filter in case of the ordinary order statistics the equality $p_i=1/n$ holds.

We now address to the histogram Z of the initial set of data and associate it to the adequate histogram of the integrated event probabilities. For this purpose we break the elements $z \in Z$ to a number of sets $r_1, \ldots, r_m$, where r_1 represents the association of the minimal and the equal elements $z \in Z$, r_2 is the association of elements of the set r_1 and the next in size elements (also equal) $z \in Z$, etc. To each set r_i $(i = 1, \ldots, m;\ m \leq n)$ we associate the joint probabilities $P_i = \sum_{x \in r_i} p_j$, to be included in the histogram of the association of probabilities (Figure 1).

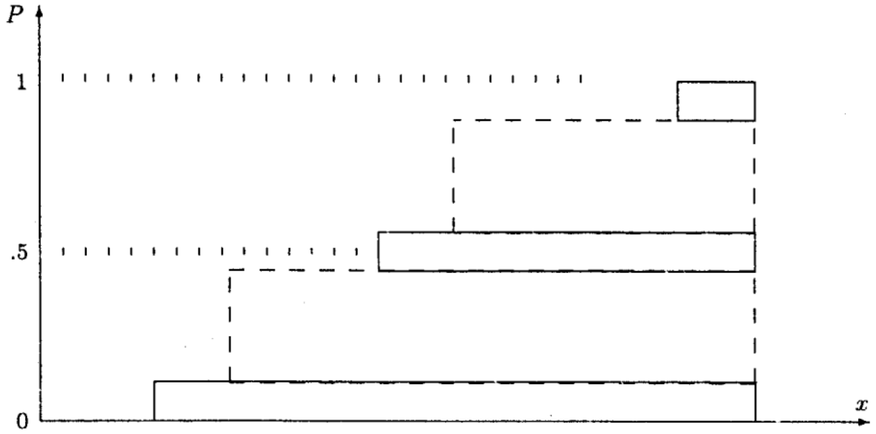

Figure 1

Now the procedure of the percentile filtering (which we call the procedure of the weighted percentile filtering) takes the following form.

Let there be a sequence $X = x_1, x_2, \ldots, x_n$, to which corresponds a vector of weights $W = (w_1, \ldots, w_n)$ and the corresponding vector of probabilities $P = (p_1, \ldots, p_n)$. Then:

- Following the ranking of elements $x \in X$, the increasing summation of the corresponding probabilities $S_j = \sum_{i=1}^{j} p_i$, $j = 1, 2, \ldots$ is made;
- Checking of the increasing sum is made;
- As output of the filter such $x \in X$ is chosen, an addition of probability of which to the increasing sum results in $S_j \geq \alpha$ $(j \leq n)$ provided that $S_{j-1} < \alpha$ $(1 \leq j \leq n)$.

It is known, that it is difficult to use an exact distribution of the order statistics in the statistical analysis, since it essentially depends on the initial distribution of the event probabilies. To even more extent the difficulties increase when dealing with the weighted-oder statistics, since even at a fixed vector of weights the output of the filter essentially depends on a relative locus of elements of the sequence. Therefore modeling of appropriate processes is so widely used here.

2. Processing of harmonic signals

Let us consider a signal, obtained in the case of the seismic sounding of a geological medium. Here a vibrating source of fluctuations generates a harmonic or a sweep signal. An example of seismorecords processing is presented in work [4]. Let it be necessary to pick out the harmonic motion of a signal whose period is R sec. The standard practice of processing of such signals is the realization of the convolution operation. Here we shall consider opportunities of using the percentile filters. For this purpose we refer to a model, representing a sine-shaped signal, on which random noise with zero mathematical expectation and variance, satisfying the required *signal-to-noise ratio*, is applied. As shown in the example in work [5], the use of the median filters, with attraction of a specific kind of the weighted-order statistics, can improve the quality of a harmonic signal.

We consider the case, when the noise to some extent surpasses a signal. Figure 2 presents: a sine-shaped signal (SIN); a sine-shaped signal plus random noise (SIN + NOIS), satisfying the condition of *signal-to-noise ratio* $\approx 0,3$ and the results obtained after processing of the latter by the weighted percentile filters (PROCESSING). The frequency of discrete time for a signal is equal to 0.1(6) of the period of a simple harmonic motion, the aperture of the filter $n = 17$, and the vector of weigts consists of numbers $\{0, 1\}$, where the period of distribution of units corresponds to the period of the harmonic motion of a signal (SIN).

The scheme of processing for this case includes the four steps:

- moving averaging of a signal over three samples;
- four independent percentile filterings of a signal with $\alpha = 1, 0.65, 0.3, 0$, two iterations for an individual case;
- averaging of the results of the percentile filtering;
- moving averaging of the above results on three samples;

Basically, as is seen here, the signal is distinguished from noise. However the results of the latter processing are not sufficiently stable, and the corresponding scheme demands an additional attention. Nonetheless, the latter results have importance for processing of seismorecords and estimation of wave arrival times. The essence of the matter is that the percentile filter belongs to the class of robust filters. Here, as opposed to the linear filters, an exact knowledge of parameters and characteristics of distortions of a signal is not required. Thus, the percentile filters permit one to save moments of changes in a signal or its occurrence. It means that in addition to the known techniques of estimation of the wave arrival times (stated, for example, in [4, 6]), one more approach can be employed which is based on the analysis of the moment of the wave arrival, instead of the analysis of

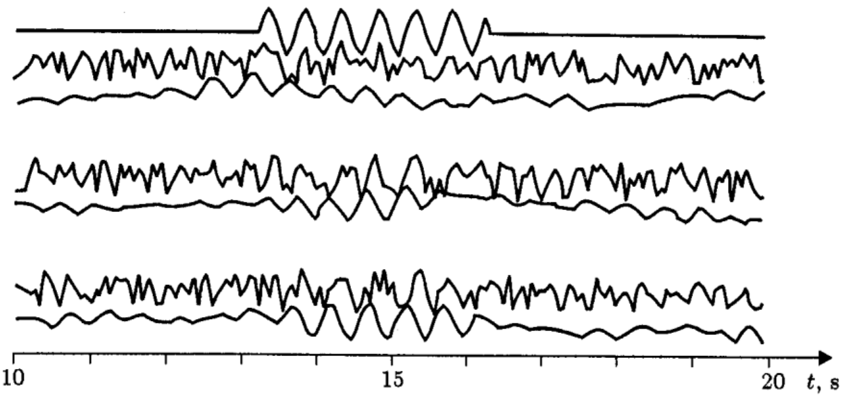

Figure 2. Weighted percentile filtering

its correlation function (see paper [4]). Thus we increase the accuracy of estimation of the disorder moment, if we use the technique, stated in work [7, 8] after percentile filtering.

3. Processing of hodograph as a set of interconnected values

Actually, a hodograph (or a travel time curve) is a curve which reflects the dependence of the arrival time of one or another type of the elastic wave at a given point, as a function of distance up to a source of the forced oscillations $t = f(s)$. In the ideal case a hodograph represents a broken line, where the segments linearly or quasi-linearly depend on time. Here, moments of breakdowns correspond to the medium interfaces and the inclination of segments is determined by the appropriate travel times. However, in practice the process of building-up the dependence $t = f(s)$, takes place with various sorts of distortions and errors. On the other hand, as follows from [9, 10], the characteristics and parameters of a wave are also a function of distance, i.e., even in conditions of the ideally exact restoration of records we should speak about a limited accuracy of evaluations of arrival times. Thus, at building-up the empirical dependence $t = f(s)$ both during the recording, restoration of records, and during the analysis or estimations of arrival times, there take place various sorts of distortions and errors having a casual character. It permits us to consider a hodograph as a random function.

In an informal sense we must approach the experimental data to the initial or to the ideal kind of a travel time curve, i.e., we must save the information base (moments of breakdowns and inclination of segments) by processing of a hodograph. In work [11], some particular cases of processing of a model of the noised hodograph by the median filters on condition of using the weighted-order statistics are considered. The results of the work do not permit us to speak about appreciable advantages of the percentile filters in the given cases. The systematic approach, with statement of the corresponding problems is required, as it is offered in work [12].

References

[1] J.W. Tuky, *Nonlinear (nonsuperposable) methods for smoothing data*, EASCON Conf. Rec., 1974.

[2] D.R.K. Brownrigg, *The weighted median filter*, Communications of the ASM, **27**, No. 8, 1984, 807–818.

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

[4] B.M. Glinskii and V.I. Znak, *On accuracy of estimation of a maximum of the envepope*, The present issue.

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

[6] 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).

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

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

[9] W.S. Ross, *The velocity-depth ambiguity in seismic traveltime data*, Geophysics, **59**, No. 5, 1994, 830–843.

[10] S.H. Bickel, *Velocity-depth ambiguity of reflection traveltimes*, Geophysics, **55**, 1990, 266–276.

[11] V.I. Znak and S.Iu. Iakimov, *Processing of hodograph by median filters: research of some partiqular cases*, Works of Computing Center of RAN SB, series: Mathematical modeling in geophysics, Novosibirsk, 1996, 146–152 (in Russian).

[12] V.I. Znak, *About some problems, arising on ways increases of quality of hodograph representation*, The Second Siberian Congress on Applied and Industrial Mathematics (INPRIM-96), thesis of the reports, Sobolev Institute of mathematics RAN SB, 1996 (in Russian).