Algorithmization of Histogram Comparing Process. Calculation of Correlations after Deduction of Normal Distribution Curves

Vadim A. Gruzdev
Russian New University
E-mail: 2801218@gmail.com

A newly established computer program for histogram comparing can reproduce main features of the "macroscopic fluctuations" phenomenon: diurnal and yearly periodicity of histogram shapes changing; synchronism of their changing by local and absolute times and "palindrome" phenomenon. The process is comparing of histogram shapes by correlation coefficients and figure areas resulting reducing of picked normal curves from histograms.

1 Introduction

Discovery of macroscopic fluctuations in stochastic processes provided actuality of a computer able to compare shapes of histograms releasing an expert from this labor [1]. Fuzziness of examined shapes and difficulties in their grouping, that is, forming of similar shapes "clusters" made a computer comparison of histogram shapes a rather hard task [1].

Our paper presents a brief of Histogram Comparer (HC), a computer program replacing an expert essentially. Calculation of correlation coefficients of curves resulting deduction of an appropriate normal distribution from a smoothed histogram is taken as a basis for the algorithm. To compare such curves, the same as with expert comparison, maximal correlation coefficients are obtained after the correlations are shifted relatively each other and mirrored, if necessary. The idea of such a transformation of histograms has been used in N. V. Udaltsova's PhD theses [2].

Main effects revealed at visual expert comparison could be reproduced with the help of the HC program [4, 5].

The HC should be run together with E. V. Pozharsky Histogram Manager program (GM) as a whole complex [3]. In this complex GM performs operations of conversion of time series into histograms and construction of distribution of intervals between histograms marked in results of HC comparison as similar. A histogram massif obtained with GM is exported into HC, which performs their comparison. The result is reloaded into GM for construction of interval distributions.

2 Main stages of histograms comparison with the GM-HC program complex

Fig. 1 shows a GM conversion of a time series of results of successive measurements of ^{239}Pu alpha-activity into series of correspondent histograms, illustrating the work of GM program.

Further the histograms are exported into HC. After the histograms are loaded, preprocessing starts — a corresponding normal distribution is calculated for each histogram. Calculation is made according to equation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \sum_{i=1}^{L} a_i \,. \qquad (1)$$

Conversion of histograms following deduction of appropriately picked normal distribution is shown at Fig. 2.

As one can see from Fig. 2, histogram structural features essential for our analysis in "replicas" remain unchanged and become more distinct. Comparison of "replicas" in a suggested program is realized in two versions: simple and detailed. A simple comparison implies relative shift and mirroring in a pair of "replicas".

Detailed comparison implies additionally compressing-stretching of one of the "replicas" — from 0.5 to 1.5 of an initial length in 10% increment. Consequently, a detailed comparison requires higher consumption of computing time.

Fig. 3 demonstrates process of replicas coinciding necessary for following determination of correlation coefficient maximal achievable for this pair.

2.1 Picking of correlation coefficients range

Results of comparison of each pair are entered into a table as values of maximal achievable correlation coefficient and curve areas ratio. A pair is regarded as similar when values of its correlation coefficient and curve areas ratio overshoot corresponding values of a threshold filter. Threshold values are set by a user. Criterion of threshold meanings is presence or absence of expressed intervals of reoccurrence of similar pairs in a result of comparison. Experience of using the program tells there are not more than 2 versions of combinations of threshold values allowing expressiveness of correspondingly 2 alternately expressed intervals, or expressed intervals are absent.

2.2 Analysis of comparison results and construction of similar pairs numbers distribution according to values of time intervals separating them in GM

A result of program comparison is entered into a binary file of a histograms similarity table in GM-supported GMA format

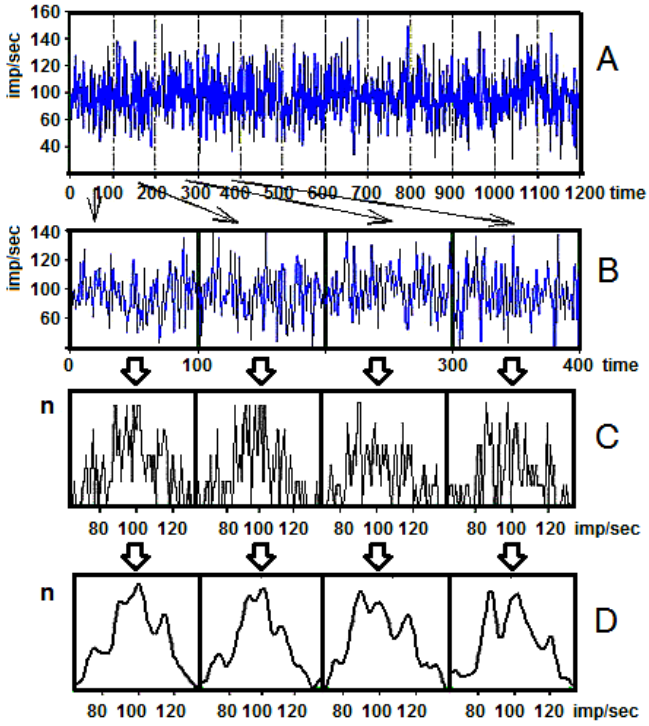

Fig. 1: Steps of histograming with GM program [1] exemplified by measurements of radioactivity a — a fragment of time series of measurement results. X-axis is time (sec.); Y-axis is numbers of alpha-decays per a second b — a time series is divided into non-overlapping sections, 100 successive numbers each c — each section is followed by a histogram (X-axis is value of activity (imp/sec); Y-axis is numbers of measurements corresponding to a value) d — the histograms from 1-c are seven-times smoothed with "moving summation" or with "a window" equal, for example, to 4; typical histogram shapes can be seen.

(description of GMA format ia a courtesy of the GM author, E. V. Pozharsky). GM calculates time interval separating each histogram pair marked as similar in the table and constructs a graphical display of intervals occurrence, i.e. histogram.

3 Examples of GM-HC complex use at determination of near-a-day periods of similar shape histograms reoccurrence and examination of "palindrome phenomenon"

3.1 Near-a-day periods

Fig. 4 presents an example of visual ("expert") comparison of histograms resulting measurements of 239Pu alpha-activity. Each histogram was constructed by 60 results of one-minute measurements. Comparison with total mixing (randomization) was made by T. A. Zenchenko. A whole series contained 143 one-hour histograms. 1,592 similar pairs were picked. The figure shows distribution of numbers of similar pairs according to values of time intervals separating them.

There are sharp extremes at the intervals equal to 1, 24 and 48 hours at the figure. These extremes correspond to a

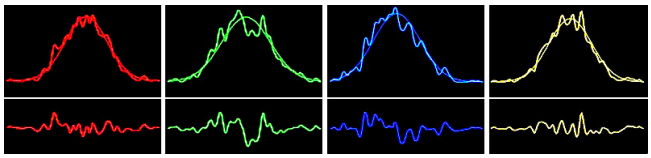

Fig. 2: The upper line is histograms with applied correspondent normal curves; the lower line is results of normal curves deduction from histograms; the resulting curves are, in fact, "replicas" of fine structures of fluctuation amplitudes spectrum.

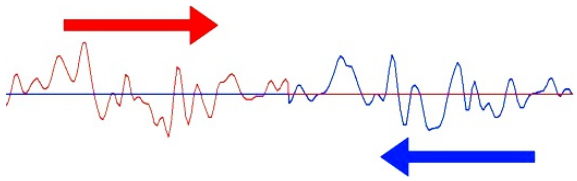

Fig. 3: Illustration of "simple" comparison. Direction of shift is pointed by arrows. Two these experiments were performed on April 8, 2011 (the left histogram arc) and April 9, 2011 (the right histogram arc).

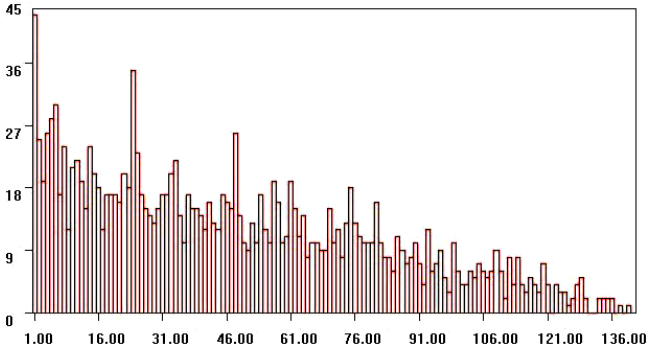

Fig. 4: Results of comparison of one-hour histograms constructed by results of ^{239}Pu alpha-activity measurements from July 7 to July 15, 2000, in Puschino. X-axis is intervals (hours); Y-axis is number of similar pairs corresponding to value of interval. (Taken from [1].)

"near zone effect" — maximal probability of realization of similar histograms in nearest, neighboring, intervals and their realization with near-a-day periods. Total mixing (randomization) of histogram series guaranties reliability of regularities revealed in expert comparison [1].

Fig. 5 presents result of automatic comparison, performed by HC computer program in the same task. It is clear that in reproduction of main effects the program is rather inferior to the expert in quality of histograms comparison.

3.2 "Palindrome effect" [1]

Figs. 7 and 8 show one of the main phenomena of "macroscopic fluctuations" — a palindrome effect — reproduced with HC program.

A "palindrome effect" is conditioned by dependence of a histogram shape from correlation of Earth motion directions: at its axial rotation and circumsolar movement. In a day-time axial rotation of Earth is antisircumsolar. In night-time the

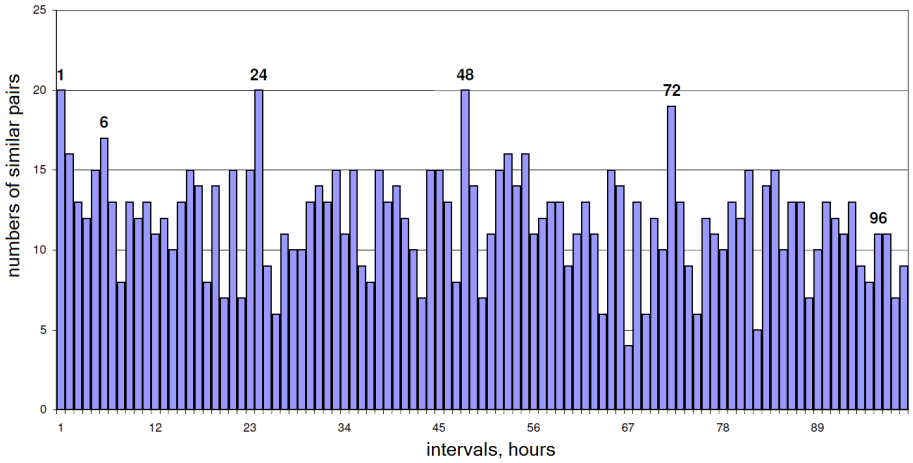

Fig. 5: Diurnal periods revealed at comparison of non-smoothed one-hour histograms by HC program. The histograms are computed by measurements of ^{239}Pu alpha-activity with a collimator-free counter since February 10 till March, 1, 2010.

both movements are co-directed. Succession of "day-time" histograms shapes was shown to repeat inversely in nights. In other words, one the same "text" is "read" forward and backward forming a palindrome. The moments of day-night transitions are 6 pm by local time and of night-day transitions are 6 am. Comparison of forward and backward sequences gives a valuable possibility to verify objectiveness of obtained distributions. The same histograms are compared. Result of comparison depends on direction of sequences only.

Figs. 6–8 present results of "palindrome effect" examination at measurements of fluctuations of intensities of light beam and ^{239}Pu alpha-decay in two ways: at expert (visual) estimation of histograms similarity (Fig. 6) and at HC comparison (Figs. 7 and 8). Fig. 6 is an expert comparison of histograms constructed by 60 one-second measurements (that is, during 1 minute). Figs. 7 and 8 show the results of comparison of 10-minute histograms with the HC program.

As one can see from Figs. 6–8, when histograms constructed by measurements of a light beam fluctuations [4] or fluctuations of ^{239}Pu alpha-decay intensities [1] are compared, a distinct palindrome effect can be observed. There is high probability of synchronous similarity of "day-time" histograms series with inverse series of night-time histograms. There is no similarity of synchronous histograms when a day-time series is compared with a direct (non-inverted) night-time series. Nevertheless comparison of histograms with HC programs gives "coarser" results, with 10 minutes interval, versus one-minute intervals at expert comparison.

The illustrations show principal availability of the HC program for examination of fine structure of histograms. But expert comparison determines similar histograms more specially, HC gives much higher "background" of stochastic shapes. Besides, these figures show similarity of palindrome effects at measurements of fluctuations of alpha-decay intensity, that is, independence of a macroscopic fluctuations phenomenon from nature of processes under examination [1].

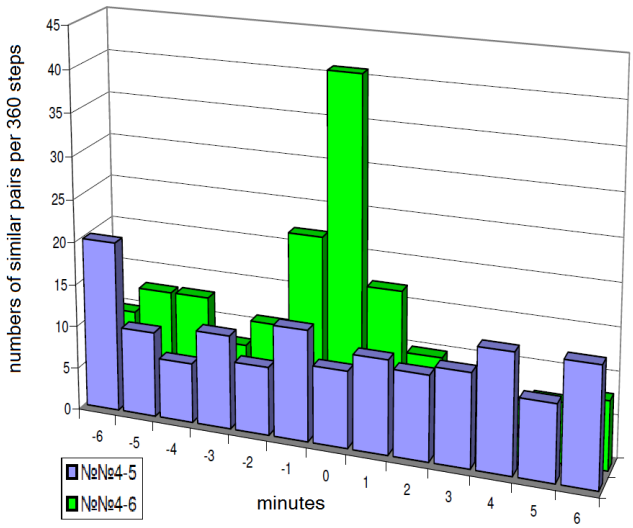

Fig. 6: Expert comparison. A palindrome effect in an experiment with measurements of fluctuations of light-diode light beam intensities. High similarity of synchronous one-minute histograms at comparison of "day-time" and "night-time" histograms when one series is inverted and absence of similarity at the absence of inversion. Measurements on April 6, 2011. X-axis is one-minute intervals; Y-axis is number of similar pairs corresponding to an interval.

4 Conclusions

1. Application of HC program allows reproduction of main effects of "macroscopic fluctuations" phenomenon.
2. Nevertheless, range of correlation coefficients values is to be picked up each time complicating the work.
3. The optimal way is combination of expert analysis with estimation of confidence of main conclusion with GM-HC combination.

You can get text and manual of the program from its author after e-mail request via: 2801218@gmail.com

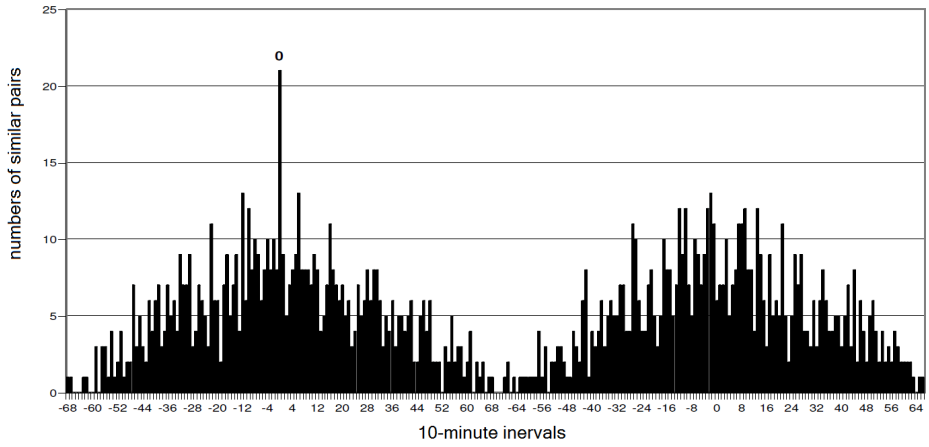

Fig. 7: Comparison of 10-minute histograms by the HC computer program. A palindrome effect in experiments with measurements of fluctuations of light-diode light beam intensities. Left is distribution of numbers of similar histogram pairs at comparison of "day-time" (since 6 am till 6 pm April 6, 2011) series histograms with inverse "night-time" (since 6 pm April, 27 till 6 am April 7, 2011). Right — the same at comparison of series without inversion.

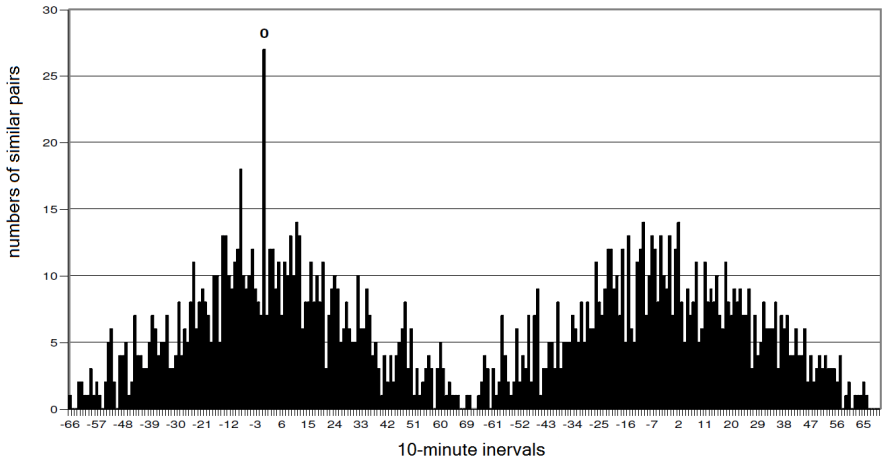

Fig. 8: Comparison of 10-minute histograms by computer HC program. Palindrome effect in experiments on measurements of ^{239}Pu alpha-activity with West-directed collimator. Left — distribution of numbers of similar histograms at comparison of histograms of "day-time" series (from 6 am to 6 pm May 27, 2005) with inverse "night-time" series of histograms (from 6 pm May 27 to 6 am May 28, 2005); right — the same at comparison of series without inversion.

Acknowledgements

Author is grateful to S. E. Shnoll for formulation of the problem, provided materials and essential discussion; V. A. Kolombet for essential discussion and his attention to the work; E. V. Pozharsky for consulting in work with GM interfaces and Professor V. I. Maslyankin ("Programming technology", Russian New University) for tutoring.

Submitted on March 11, 2012 / Accepted on March 16, 2012

References

1. Shnoll S. E. Cosmic Physical Factors in Random Processes. Svenska Fisikarkivet, Stockholm, 2009, (*in Russian*).
2. Udaltsova N. V., Possible cosmophysical casuality of parameters of biochemical and physic-chemical processes. *PhD theses Inst. Biophysics, USSR Acad. Ssi.*, 1990, (*in Russian*).
3. Shnoll S. E., Kolombet V. A., Pozharsky E. V., Zenchenko T. A., Zvereva I. M., Konradov A. A. Illustration of synchronism of changes of fine structure of measurement results distribution exemplified by radioactive decay of radium family isotopes. *Biofizika*, 1998, v. 43 (4), 732–735 (*in Russian*).
4. Rubinshtein I. A., Shnoll S. E., Kaminsky A. V., Kolombet V. A., Astashev M. A., Shapovalov S. N., Bokalenko B. I., Andreeva A. A., Kharakoz D. P. Dependence of changes of histogram shapes from time and space direction is the same when intensities of fluctuations of both of light-diode provided light flux and ^{239}Pu alpha-activity are measured. *Progress in Physics* , 2012, v. 3, 17–24 (this issue).
5. Shnoll S. E., Astashev M. A., Rubinshtein I. A., Kolombet V. A., Shapovalov S. N., Bokalenko B. I., Andreeva A. A., Kharakoz D. P., Melnikov I. A. Synchronous measurements of alpha-decay of ^{239}Pu carried out at North Pole, Antarctic, and in Puschino confirm that the shapes of the respective histograms depend on the diurnal rotation of the Earth and on the direction of the alpha-particle beam. *Progress in Physics*, 2012, v.3, 11–16 (this issue).