EFFECTIVENESS OF ROBUST PROCEDURES OF AN ESTIMATION AT RADIO METEOR COMPARISON OF TIME SCALES

Obelchenko Veronika V

Kharkov National University Radio Electronics; pr. Lenin 14, Kharkov, Ukraine, 31726; Obelchenko@yahoo.com

1. GENERAL

Reduce of a radio meteor comparison error of time scales on the radiometeoric channel at implantation of automation of gathering and information processing is the actual task. The solution of this task is ensured with application of robust procedures of an estimation of a position and scale of sampling of measurement outcomes, effective in conditions of incompleteness of the a priori information.

Comparison accuracy, time scale, robust estimation

2. PROPERTIES OF SAMPLINGS OF THE RADIO METEOR COMPARISON RESULTS

Data obtained from of regular radio meteor comparison of scales of the Russian UTC (SU) and Ukrainian UTC (UA) time standards within 2000 on special meteoric equipment complexes "METKA - 6", developed at Kharkov Technical University of Radio Electronics was used in the present work. – Ref. 1. Systems of spatial-temporal provision of Ukraine and Russia are equipped with such complexes. UTC (SU) standard was located in Moscow, and UTC (UA) standard was located in Kharkov. Distance between them is equal to approximately 750 km. The relative instability of these standards is within $10^{-15}...10^{-14}$. Large volume of the fact data have allowed to analyze statistical properties of result sample of measurement of shift of time scales.

It is impossible to consider an array of radio meteor comparison results of the remote time measurement standard scales on the radio meteor channel per a session owing to singularities intrinsic to physics of meteoric appearances, modifications of the ration a signal - parasite, probable malfunctions in operation of the equipment as the having homogeneous, uniformly precise and is normal distributed sampling.

The views of typical realizations of the measurement outcomes allow to make statistically provided a conclusion, that, it is the essential nonstationary processes related to meteoric phenomenon, which is stipulated by effects of forming and corrupting of a meteoric track, and also by modification of influx of meteoric substance within day and within year. The figure 1 gives the representation about the character of modifications of the measurement outcomes in time for meteoric tracks existing a long time. There are typical realizations, namely magnification of dispersion by the extremity of a radioecho from a meteoric track and a modification of the average value in time in some cases. The probable reasons of such behaviour are stipulated by singularities of physics of meteoric appearances: by a diffraction on a formed meteoric track, wind transition of mirroring area, diffuse extension of a meteoric track, resonance in a meteoric track, multipath effect propagation of radio waves and other similar reasons.

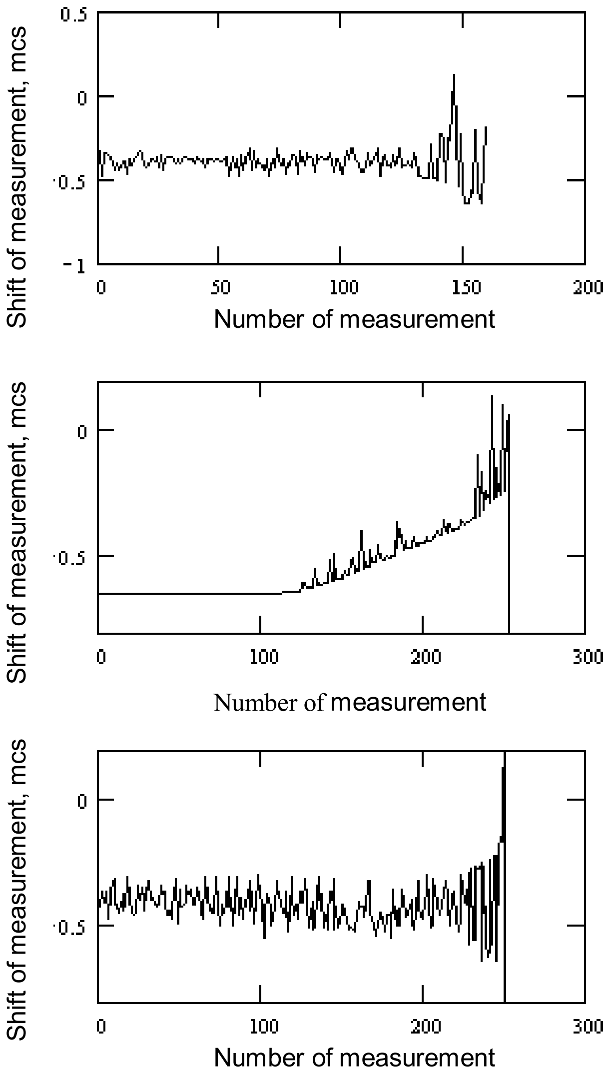


Figure 1 - The examples of realizations from long meteoric tracks

Based on Pearson, Bartlett, Student and Fisher statistics, the statistical hypotheses about the hypothetically the normal distribution, homogeneity and equal accuracy of the samplings were tested on samplings of measurement results for all sessions by appropriate goodness-of-fit tests.

Usage of probability logic of a hypothesis acceptance or rejection allows to make the following conclusions: it is impossible to consider the data to be equal precise of groups in 26 % of cases; groups can not be referred to the same distribution law in 6 % of cases because the hypothesis about equality average does not prove to be true, and in 8 % of cases both the hypotheses about equality average and about equality of variances should be rejected.

The outcomes of a statistical analysis consist in the following: the hypothesis about the normal distribution function of measurement outcomes for a session does not contradict the experimental data in 24,5 % of cases; the tests of such hypothesis can not be carried out because sizes of samplings were small in 35,8 % of cases, and the observable sample values will not be coordinated to hypothetical distribution in 39,7 % of cases. The testes of statistical hypotheses has shown, that the measurement outcomes represented by 53 sessions for 2000 year, can not be considered as were normal distributed approximately in 40 % of cases, homogeneous and equal precise samplings. These statistical properties of the sample data, obtained by a radio meteor comparison of scales of the time standards, justify necessity of searching of the noise-resistant procedures for their processing

These statistical properties of sampling radio meteor comparison data justify necessity of searching of noise-resistant procedures for their processing.

3. RESEARCH OF ROBUST PROCEDURES OF RESULT PROCESSING

The effectiveness of robust procedures at radio meteor comparison was estimated by data processing obtained from of regular radio meteor comparison of scales of the Russian UTC (SU) and Ukrainian UTC (UA) time standards and by methods of statistical simulation.

The statistical model of comparison accuracies of time scales on a meteoric radio channel is offered.

The probability density (PD) of vector $\vec{x}_n$ formed from compound entire assembly $\Gamma = \Gamma_1 \bigcup \Gamma_2$ by a random sampling with one-dimensional densities $p(x)$, $f(x,m,\sigma)$, and $h(x,m_h,\sigma_h)$ accordingly was esteemed as the basic model of measuring inaccuracies. Is suspected, that distribution of measuring inaccuracies belonging to the class ε - contaminated distributions.

The procedure of selection of readouts with distribution is determined by the following stochastic rule:

$$x_i = (1-\xi_i)x_{1i} + \xi_i x_{2i}, \quad i = \overline{1,n}, \qquad (1)$$

where ξ_i - random variable taking on a values 0 or 1 with probabilities $1-\varepsilon$ and ε accordingly; x_{1i} and x_{2i} - random variables with PD $f(x,m,\sigma)$ and $h(x,m_h,\sigma_h)$.

Thus, at $\varepsilon \neq 0$ to readouts x_{1i} from the basic entire assembly Γ_1 the readouts, "polluting" sampling x_{2i}, from foreign entire assembly Γ_2 are adulterated, we shall call which one as abnormal. In practice abnormal are the lets (or zerofilling), conditioned by operating of an intermittent interference of an industrial or atmospheric parentage, malfunctions of synchronization in switchboards of receivers, short-duration generator noise created by radiations of radio electronic systems, including deliberately.

One-dimensional PD of readouts, conforming to a procedure (1), will be written down as

$$p(x/m,m_h,\sigma,\sigma_h,\varepsilon) = (1-\varepsilon)f(x,m,\sigma) + \varepsilon h(x,m_h,\sigma_h)$$

Let's consider cases, when models basic and contaminated of distributions are the symmetrical laws: Gaussian, uniform, Laplace, Simpson, and arcsine. Here m, m_h – mathematical expectations, σ, σ_h – standard deviations of basic and contaminated distributions accordingly.

The capability of level variation of contamination ε and the parameters basic and of contaminated distributions is stipulated.

For a selected set of conditions: distribution laws of measuring inaccuracies, their parameters and level of contamination of sampling was conducted of r tests: r sampling of measuring inaccuracies for a session by a volume n was generated. Was received, that in each sampling $(1-\varepsilon)n$ of members belongs to the basic distribution, and εn members contaminated.

Each r th sampling from n independent outcomes of measuring inaccuracies $x_1, x_2, ..., x_i, ..., x_n$ was ranked, i.e. it was introduced by variational series $x'_1 \le x'_2 \le, ..., \le x'_i, ..., \le x'_n$.

The estimator of location were determined for each sampling by the following procedures:

1) sample mean–

$$m_s = \frac{1}{n}\sum_{i=1}^{n} x'_i ;$$

2) truncated mean–

$$m_t = \frac{1}{n-2[\alpha n]}\sum_{i=1}^{n-[\alpha n]} x'_i ;$$

3) winsorised mean –

$$m_W = \frac{1}{n}\left\{\sum_{i=[\alpha n]+2}^{n-[\alpha n]-1} x'_i + ([\alpha n]+1)\left(x_{[\alpha n]+1} + x'_{n-[\alpha n]}\right)\right\}$$

4) quantile estimation –

$$m_q = 0{,}5\left(x'_{0,25} + x'_{0,75}\right),$$

where $x'_{0,25}$ and $x'_{0,75}$ – 25 % and 75 % quantiles of an empirical distribution;

5) sample median –

$$m_m = \begin{cases} x'_{k+1} & \text{при} \quad n = 2k+1; \\ 0{,}5\left(x'_{k+1} + x'_{n-k}\right) & при \quad n = 2k \end{cases}$$

6) center of the sample range –

$$m_r = \frac{(x'_1 + x'_n)}{2};$$

7) M – the Huber estimation – m_H;

8) M – the Hampel estimation – m_{Ha};

8) M – the Andrews estimation – m_A;

10) M – the Tukey estimation – m_T.

Found on procedures 7 … 10 M – estimations are not expressed in an obvious kind. The most developed generalized maximum-likelihood technique (Huber's min-max approach) was used for construction these of robust estimations. – Ref. 2. Thus the estimations receive from a condition

$$\min Q(\Phi) = \sum_{i=1}^{n} \Phi\left(\frac{x_i - \hat{a}}{\sigma}\right),$$

where x_i - sample units, $\hat{a}$ - steady estimation of a centroid of distribution.

M - the estimation is the decision of the equation

$$\sum_{i=1}^{n} \phi\left(\frac{x_i - \hat{a}}{\sigma}\right) = 0, \quad \text{where} \quad \phi(x) = \frac{d}{dx}\left[\Phi(x)\right]. \tag{2}$$

The iterative methods are used for the decision of the equation (2).

The decision of the equation (2) was by application of the Gauss-Newton iterative circuit:

$$a_{l+1} = a_l + \frac{\sum_{i=1}^{n} \phi\left(\frac{x_i - a_l}{\sigma}\right)}{\sum_{i=1}^{n} \phi'\left(\frac{x_i - a_l}{\sigma}\right)} \cdot \sigma, \tag{3}$$

where l - number of iteration, and sampling median – $a_0 = m_m$, being a steady simple estimation of a centroid of distribution, was accepted as initial approximation.

M - the Huber, Hampel, Andrews, Tukey estimations differ by a kind of weight functions $\Phi(x)$ [4].

As under the same conditions for the chosen distribution of measuring inaccuracies, their parameters and level of a contamination was carried out r tests, i. e. sessions of measurement of scales shift of the time standards, it has allowed carry out the statistical estimation analysis of each of estimations considered above. The homogeneous, of equal accuracy and hypothetically normally distributed samples of estimations of volume r everyone were received: $\{m_{ji}\}$, where m_{ji} – a i th sampling unit of a j th estimation of a centroid of distribution, where $j = \overline{1{,}10}$, $i = \overline{1, r}$. The sample processing of the estimations was carried out by classical statistical procedures: bias of estimators of a centroid of distribution and their standard deviation were determined.

$$\overline{m_j} = \frac{1}{r}\sum_{i=1}^{r} m_{ji}, \; s_{mj} = \sqrt{\frac{1}{r-1}\sum_{i=1}^{r}(m_{ji} - \overline{m_j})^2}. \tag{4}$$

The dependences bias of estimator of centroid of distribution from level of contamination was shown on a figure 2

On the first investigation phase was accepted, that one parameter of the distribution function – centroid of distribution requires specification only, and second parameter - deviation is known. Such assumptions reflect a situation of processing of the measurement results received on certificated Radio Measuring Complexes of the Equipment (RMCE), as the accuracy characteristic of RMCE were repeatedly specified at passage of certification.

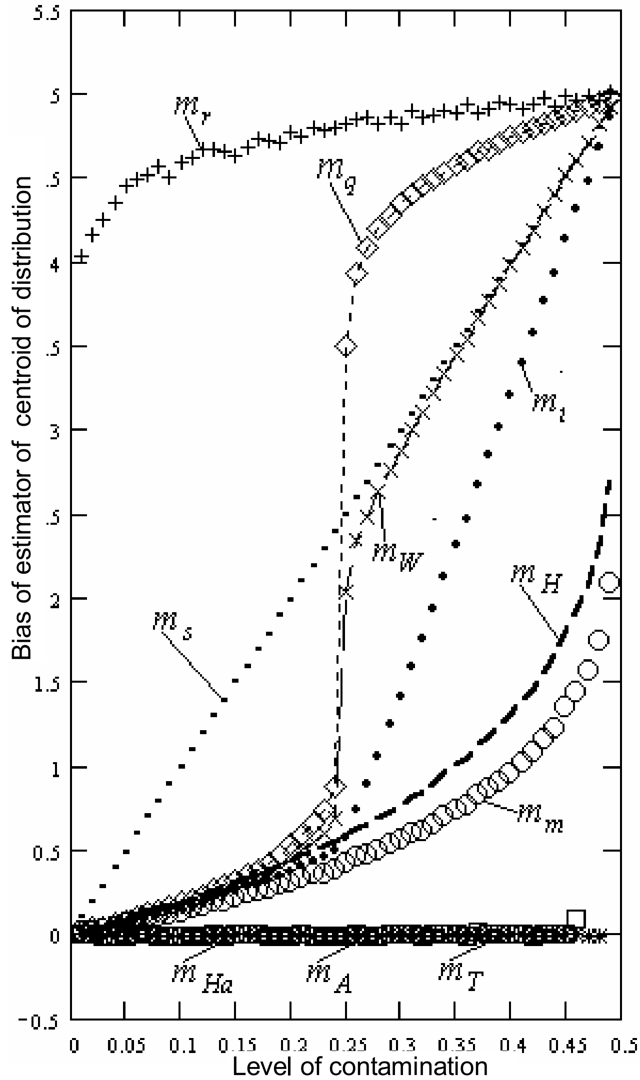


Figure 2 – The dependences bias of estimator of centroid of distribution from level of contamination

The efficiency of steady estimations was quantitatively characterized as follows.

At absence of abnormal errors ($\varepsilon = 0$) the effective algorithm of an estimation of a distribution center on sample is synthesized by the maximum-likelihood method and is given by statistics selective mean m_{sa}. Quality of information estimations $\hat{\theta}_n$ is determined PD of an absolute measurement error $\delta = \hat{\theta}_n - m$. This is Gauss PD with zero mathematical expectation and deviation $D_\delta = \sigma^2 / n$, i.e. the procedure 1 at $\varepsilon = 0$ provides unbiasedness of information estimate, and it effectiveness corresponds(meets) to the minimal border variance, determined by an Rao–Cramer inequality. In our case selective standard deviation of an error δ will make $s_\delta = \sigma / \sqrt{n}\Big|_{\sigma=1;\quad n=400} = 0{,}05$.

The relative efficiency of conditional estimations was, as the ratio of a square standard deviation of an effective estimation from estimated parameter to a square standard deviation of the considered biased estimator

$$e_{n,j}(\theta) = \frac{E\{(\hat{\theta}_{n,eff} - \theta)^2\}}{E\{(\hat{\theta}_{n,j} - \theta)^2\}}$$

(5)

where - operation of statistical averaging. In our case estimated parameter is $\theta = m$ It is clear, that $0 \le e_n(\theta) \le 1$. The meaning of relative statistical efficiency is equal to unit, when the estimation is effective. After simple transformations is received a design parity for the comparative analysis of efficiency of researched procedures

$$e_j = \frac{\sigma^2}{n \cdot s_{mj}^2 + (\overline{m_j} - m)^2}$$

(6)

The dependences of relative efficiency of the considered estimations from a level of a contamination of the samples are submitted in the table 1. The results of computational investigation shown in figure 2 and in the tables 1, are carried out under identical conditions: The dependences of relative efficiency of the considered estimations from a level of a contamination of the samplings are submitted in the table 1. The results of computational investigation shown in figure 2 and in the tables 1, are carried out under identical conditions: $n = 400$, $r = 100$,

$$p(x) = (1-\varepsilon)N(0,\quad 1) + \varepsilon N(10,\quad 1),$$

where $N(m,\ \sigma^2)$ – Gaussian density, m – expected value, σ^2 – variance.

The basic conclusions consists in the following: M-estimations determined by procedures 8, 9 and 10 appear steady to a contamination at known variance; at unknown variance it is recommended to estimate σ on sampling by a median deviation

$$s = \frac{med\,|x_i' - med\,(x_i')|}{0{,}675};$$

in these conditions best appears an estimation Tukey, maintaining 28 % a contamination without sharp deterioration of quality; measurement results for a session in approximately 40 % of cases could not be considered as homogeneous and equal accurate and normally distributed sampling; we propose to approximate the distribution of the measurement results of the shift of the scales by the epsilon–contaminated normal distribution with the normal distributed contamination;

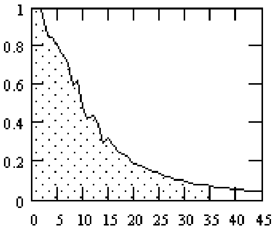
it is recommended to use the sampling median and the sampling median deviation as the elementary estimates for the parameters of a location and a scale, and the Tukey estimate as the basic estimate for a center value of distribution at secondary processing of the measurement results.

REFERENCE: 1. Кащеев Б.Л., Коваль Ю.А., Горбач В.И. и др. Метеоры сегодня. Киев: Техника, 1991.196 с. **2.** Хубер П.Дж. Робастность в статистике: Пер. с англ. – М. : Мир, 1984. – 320 с.

Table 1 – The dependences of the relative effectiveness of the estimations

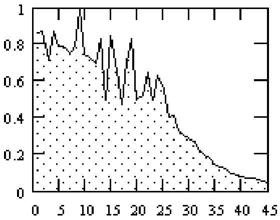
The relative effectiveness of the estimations

1. Sample mean



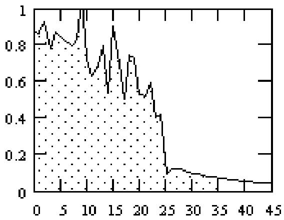
Level of contaminated, %

2. Truncated mean



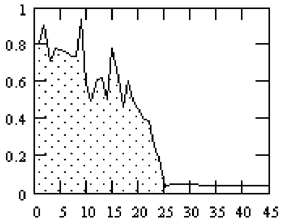
Level of contaminated, %

3. Winsorised mean



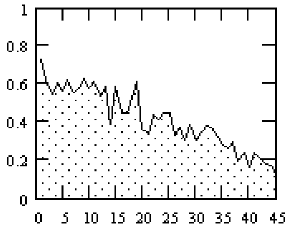
Level of contaminated, %

4. Quantile estimation



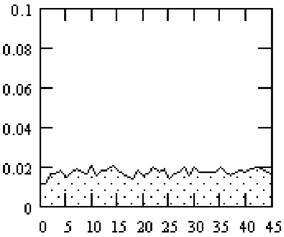
Level of contaminated, %

5. Sample median



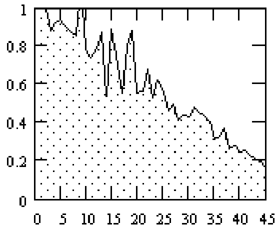
Level of contaminated, %

6. Center of sample range



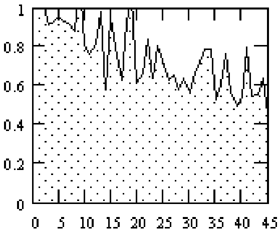
Level of contaminated, %

7. Huber estimation



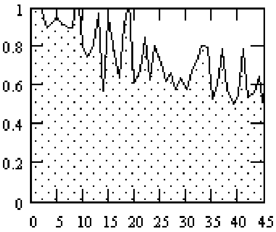
Level of contaminated, %

8. Hampel estimation



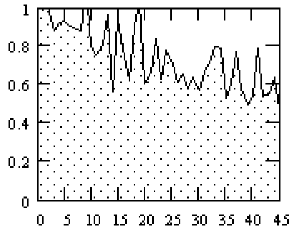
Level of contaminated, %

9. Andrews estimation



Level of contaminated, %

10. Tukey estimation



Level of contaminated, %