NEW ADAPTIVE METHODS FOR PARTICLES FLUX INTENSITY MEASUREMENT REDUNDANCY REDUCTION AND THEIR EFFICIENCY

V.P. EVDOKIMOV AND V.M. POKRAS
Institute for Space Research, USSR

Summary. Particles flux intensity measurements redundancy reduction algorithms are proposed. Accuracy criteria consists in limiting of a samples relative error maximum value. The algorithms are based on prediction or interpolation operations with a variable threshold, adaptive to a changing flux intensity. A formula for computation of an adaptive threshold zero order predictor compression ratio is deduced. Computed values show good coincidence with those received by signal and algorithm computer simulation. Adapter threshold zero order predictor (AT-ZOP) and first order interpolator (AT-FOI) algorithms applied to real telemetry data reveal their high efficiency as relating to attainable compression ratios. Algorithms compression ratio comparison results in predictor advantage against interpolator and unsignificantly small predictor loss when preliminary data smoothing is applied. Compression ratios for joint application of background removal [2] and adaptive predictor algorithms are also evaluated. AT-ZOP simplicity and high efficiency allow to recommend it for use in particle flux intensity measurements redundancy reduction systems.

1. Redundancy reduction algorithms. Particle flux intensity measurements are an important part of many scientific experiments on board the spacecraft. It is often supposed that particle flux with certain characteristics is described by a single parameter, mean number of particles per time unit (intensity) λ . Flux-measuring gauges usually consist of sensors where an electric pulse appears at the output if a particle with certain characteristics arrives at the input and of counters for registering number of pulses during accumulation time T. number of pulses registered during successive time intervals T has Poisson probability distribution

$$P(x_i) = e^{-M} M^{x_i} / x_i!$$

where $M = \lambda T$ is a mean number of particles for T. We suppose here that accumulation time T is chosen to be of such a value that with a small error λ may be considered constant over T. Particles number dispersion over T is equal to its mean value, $\sigma_x^2 = M = \lambda T$. Having successive particle counts over T_i it is possible to determine values of λ that the experimenter is interested in. Number of particles registered at each time interval is the maximum likelihood estimate for its mean value M [1]. Flux intensity estimate over T_i is equal then to $\hat{\lambda}_i = X_i / T$. Statistical scatter, determining measurement's accuracy for M is then

$$\frac{\delta_x}{M} = \frac{\sqrt{\lambda T}}{\lambda T} = \frac{1}{\sqrt{\lambda T}}$$

If λ = Const and T is increasing, then estimate accuracy for may be very high. The upper limit of T depends on rate of change of λ with time that is of interest for the experimenter. He usually sets lower and upper limits for intensity change $\lambda_{min} + \lambda_{max}$. The accumulation time is chosen according to the rule that statistical scatter δ_x /,M for $\lambda = \lambda_{min}$ must not exceed a certain preset value. If accumulation time is constant and intensity is increasing, than measurements accuracy becomes higher. If the experimenter does not get additional information owing to higher accuracy for $\lambda > \lambda_{min}$ then part of successive counts becomes abundantly accurate. If one plans to have constant accuracy for any λ from a given range then there appears a possibility to trade this statistically abundant accuracy for experiment data redundancy reduction. To make it possible one has to apply intentional error insertion into these counts that correspond to $\lambda > \lambda_{min}$. Inserted error absolute value must be a variable that becomes bigger for increasing λ . Algorithms that have such quality of making absolute error grow in accordance with growth of λ are based on insertion of relative error, not exceeding the preset value E, into the abundantly accurate counts. Such a requirement of having the relative error constant throughout the whole change range of λ is usual for the experimenter.

$$\varepsilon = \frac{|\lambda_i - \hat{\lambda}_i|}{\lambda_i} \leq E \quad , \qquad 0 \leq E \leq 1$$

If we substitute λ in this expression by it's maximum likelihood estimate we have then inequality which forms the background for possible redundancy reduction algorithms

$$|X_i - \hat{X}_i| \leq E X_i$$

Let's consider an algorithm that makes use of prediction operation to form the estimate $\hat{X}_i$ .It is a modified zero order floating aperture predictor.

1 step - The reference count X_i is transmitted

2 step - Adjustable threshold for counts comparison is formed $Z = E X_i$. It depends on reference count.

3 step - Absolute value of difference between count X_{i+n}, , where at first n = 1, and reference count X_i , $|X_{i+n} - X_i|$ is compared with the threshold Z .

4 step - If $|X_{i+n} - X_i| \leq Z$, then count X_i is believed to be abundantly accurate and set aside, n changes for n+1 and 3 step is repeated. If $|X_{i+n} - X_i| > Z$, then X_{i+n} is taken as the following reference count and 1 step is repeated, where i changes for i+ n. Let's consider another algorithm with interpolation operation to receive an estimate $\hat{X}_i$.

1 step - The reference count X_i is transmitted.

2 step - The initial tolerance area is set around X_i, $X_i - Z_i$ and $X_i + Z_i$, where $Z_i = EX_i$.

3 step - The final tolerance area is set around X_{i+n}, with n = 2 at first, $X_{i+n} - Z_{i+n}$ and $X_{i+n} + Z_{i+n}$, where $Z_{i+n} = EX_{i+n}$.

4 step - The straight line for intermediate counts upper estimates is drawn between $X_i + Z_i$ and $X_{i+n} + Z_{i+n}$, for lower estimates - between $X_i - Z_i$ and $X_{i+n} - Z_{i+n}$

5 step - All intermediate counts X_{i+k}, where i < K < n) are to be checked for being located between upper and lower estimates straight lines, i.e. the following inequality fulfillment is being checked

$$|X_{i+k} - \hat{X}_{i+k}| \leq Z_{i+k}, \quad Z_{i+k} = E\hat{X}_{i+k}$$

If this inequality is correct then all the intermediate counts are set aside, the 2 step is repeated, where n changes for n+1.
If one or more counts are located beyond the estimates straight lines then the 1 step is repeated where i changes for i + n - 1.

Inasmuch as counts X_i are acting estimates for M = λT and the threshold Z is proportional to the flux intensity, both algorithms become adaptive to the flux intensity. As the intensity estimate is defined at each step by only one count there exists no other constraints for the λ change rate except for those that were taken into consideration at choosing accumulation time T.

2. Algorithms efficiency for simulated signal.

Let us define the maximum value of relative error inserted into abundantly accurate counts. If estimates are taken instead of passive counts then relative error will have its maximum for those counts that are located on the lower estimate line. For the predictor the reference count X_i becomes an estimate and

$$E_{max} = \frac{X_i - (X_i - Z)}{X_i - Z_i}, \quad Z = EX_i \quad .$$

For the interpolator $\hat{X}_{i+k} = X_i + K\frac{(X_{i+n} - X_i)}{n}$
and

$$E_{max} = \frac{\hat{X}_{i+k} - (\hat{X}_{i+k} - Z_{i+k})}{\hat{X}_{i+k} - Z_i}, \quad Z_{i+k} = E\hat{X}_{i+k}$$

or

$$E_{max} = \frac{E}{1 - E}$$

If E_{max} is given then it is possible to define the threshold relative value.

$$E = \frac{E_{max}}{1 + E_{max}} \cdot 100\,\%$$

To calculate possible predictor compression ratios let us take Poisson independent counts as a flux model. The probability of the event that the count is active is equal to

$$P_a = P(|X_{i+n} - X_i| > EX_i) =$$
$$P[(X_{i+n} - X_i(1+E)) > 0] + P[(X_{i+n} - X_i(1-E)) < 0]$$

It is defined by the difference of two Poisson random values. For this difference a probability is know [1]

$$P(A - B = z) = e^{-(\lambda+\nu)}\left(\frac{\lambda}{\nu}\right)$$

where υ and λ are intensities.

Direct calculation with such a formula for λ and υ values of practical interest is impossible. If λ and υ are close to each other and large, E is small then difference probability distribution may be taken as approximately normal. We have then the following formula for active count appearance probability

$$P_a = \frac{1}{2}\left[2 - \varphi\left(\frac{E\sqrt{\lambda}}{\sqrt{1+(1+E)^2}}\right) - \varphi\left(\frac{E\sqrt{\lambda}}{\sqrt{1+(1-E)^2}}\right)\right] \quad (2)$$

where

$$\varphi(x) = \frac{2}{\sqrt{2\pi}} \int_0^\infty \exp\left(-\frac{t^2}{2}\right) dt$$

Compression ratio (without additional timing information) is equal to

$$K_c = \frac{1}{P_a}$$

Fig. 1 shows compression ratios for different and E = 5, 10, 15% of the count's magnitude. K_c values fastly grow for $\lambda > 100$. To define relative error dispersion σ_ε^2 , reconstructed counts flux intensity and it's dispersion, formula (2) approximation accuracy the Poisson counts flux with given intensities was simulated on a digital computer. Simulation results are shown in Table 1. Compression ratio K_c and maximum relative error E_{max} values received by simulation show good coincidence with those that were calculated with the help of formulas (1) and (2).

3. Algorithms efficiency for real telemetry data.

For algorithms efficiency estimation real telemetry data obtained on "Prognoz-3" satellite in May 1973 were used. The scientific load on board includes devices for electrons, protons and X-ray emission measurement for different energy ranges,

Processed communication sessions contained information about rise and fall of a solar burst. At this period the particles flux intensity increased in great extent, but there were no big changes of λ during considerable part of the session.

Logarithmic counters with reset to zero and analog logarithmic intensimeters were applied for measurements. 3 binary digits were used for exponent of measured number and 4 - for mantissa.

The accumulation time was equal to sampling period and was about 41 sec. Analog intensimeters time constant was considerable less than sampling period that resulted in additional counts error. To decrease this error the counts of analog intexisimeter output were smoothed before compressions.

The linear sliding smoothing for 3 and 5 points were used.

To define the influence of parameters smooth on the compression efficiency the counts at the logarithmic counters output were also smoothed.

To illustrate the results of proposed algorithms application following signals were considered:

Channel 1 - logarithmic electron flux counter for energy range E > 25 kev.

Channel 2 - analog logarithmic electron flux intensimeter for energy range E >500 keV.

The intensity could change from 10 to 10^4 particles per sec.

The above mentioned algorithms were applied to calculate compression ratio and relative errors.

The values of data compression ratios and mean squared relative error are given in Table 2 for the predictor and in Table 3 for the interpolator.

It is interesting to note that the compression ratio values for interpolator without smoothing are smaller than those for predictor.

If extent of smoothing is increased application of interpolator have some advantage, but so insignificant that interpolators application is not advantageous.

The affect of smoothing on compression ratio is more considerable for the channel with analog intensimeter. 5 points smoothing gives only a small compression ratio gain than 3 points one.

Taking into account the fact, that the contribution of the events with great intensity increase above background level in total measurements set is rather small, efficiency of combined application of zero predictor with adaptive threshold algorithms and background removal algorithm described in [2] was also studied. This algorithm is based on counts summing and particles counter reset only if accumulated number of particles exceeds a threshold B , which is given by experimenter and depends on background level. Usually this threshold is defined by the particles number when statistical scatter becomes less than some fixed value. The sums which exceed the threshold form a new sequence of counts. This one is then processed by means of zero order with adaptive threshold. The total data compression ratio for both algorithms application is equal to:

$$K_{ctotal} = K_{cbgr} \times K_{cAT\text{-}ZOP}$$

It should be noted that for calculation of real compression ratio an additional data about counts number when the counts some exceeds the threshold is added.

The results of compression ratio calculation (without additional data) and mean squared relative error are given in Table 4. The thresholds are equal to 100 particles per .accumulation time T for logarithmic counter and 10 particles per second for analog intensimeter.

Combined application of the zero order predictor with adaptive threshold and the background removal algorithm leads to sufficient increase of total compression ratio.

Simplicity and high efficiency of zero predictor with adaptive threshold for real telemetry data allow to recommend their application in data compression systems for particles fluxes intensity measurements.

REFERENCES

[1] F. Haight, Handbook of the Poisson distribution, New York, 1967.

[2] J. Khodarev, V. Nicolayev, I. Skobkin, Yu. Shtarkov, E. Vassiliev, On-board registration and redundancy reduction method for quasi-stationary Poisson processes, Proceedings of the International Telemetering, Conference, vol. VIII, 1972, Los-Angeles, USA.

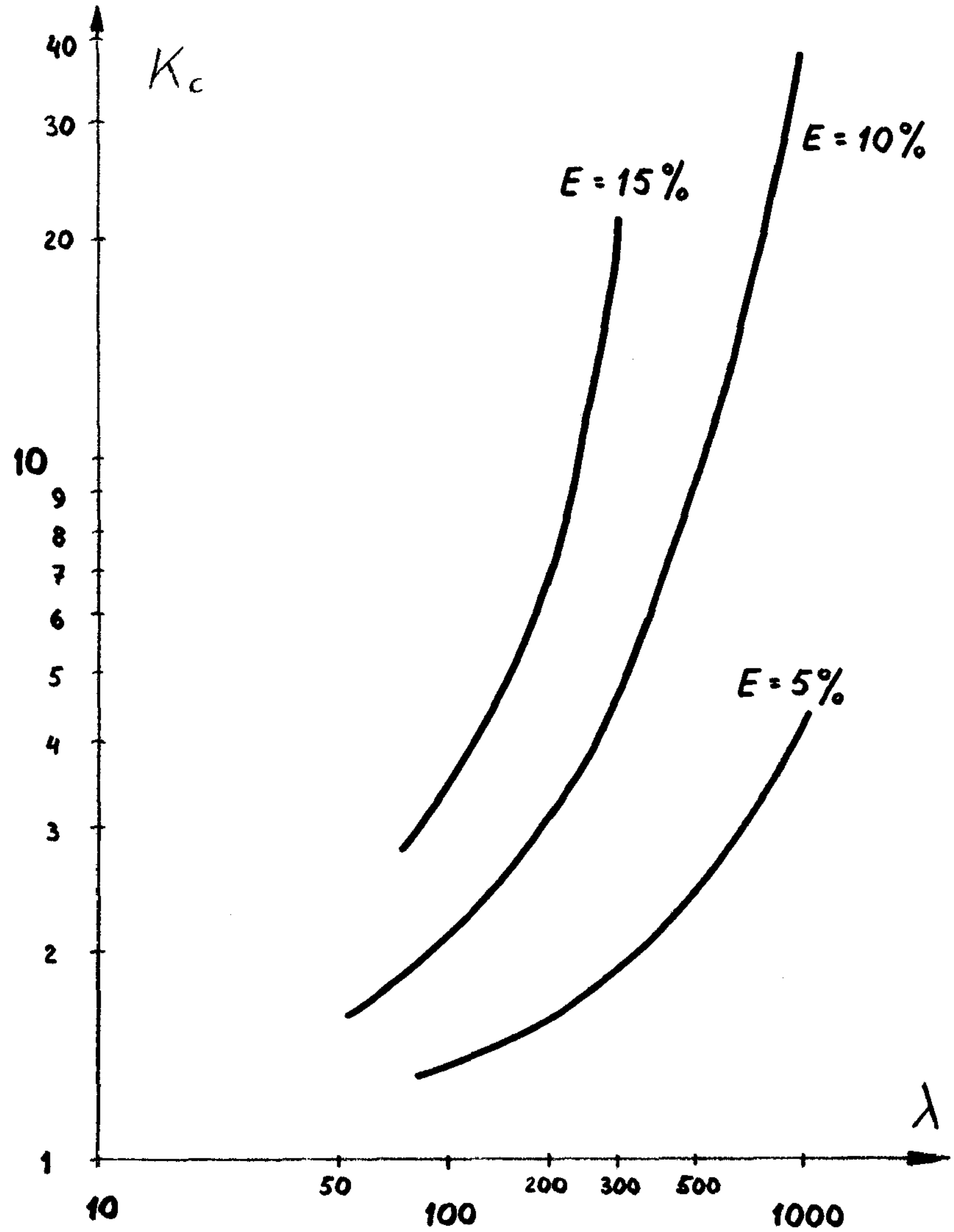

Fig. 1. AT-ZOP compression ratio / flux intensity for simulated signal.

Table 1. AT-ZOP for simulated signal.

$E\%$	K_c	$\gamma_{compr.}$	$\sigma_{compr.}$	σ_ε %	E_{max} %
$\gamma = 100$, $\gamma_{sim.sign} = 99,24$, $\sigma_{sim.sign} = 9,69$					
5	1,38	99,3	9,69	1,51	5,2
10	2,1	99,4	9,67	4	11
15	3,7	100,2	9,54	7	17,6
$\gamma = 289$, $\gamma_{sim.sign} = 287,6$, $\sigma_{sim.sign} = 16,2$					
5	1,87	288	16,3	1,9	5,24
10	4,6	289	15,74	4,6	11,1
15	20	292	13,4	6,65	17,6
$\gamma = 400$, $\gamma_{sim.sign.} = 396,7$, $\sigma_{sim.sign.} = 18,9$					
5	2,15	397	18,9	2	5,25
10	6,9	399	18,3	4,6	11,1
15	40	407	17,7	6,1	17,6

Table 2. AT-ZOP for real data.

	Without smoothing		3-points smoothing		5-points smoothing	
$E\%$	K_c	σ_ε %	K_c	σ_ε %	K_c	σ_ε %
Channel 1. Logarithmic counter.						
5	3,34	0	7,5	2	7,83	1,9
10	9,7	4,35	9,8	3,3	13,1	4,4
15	11,9	6,4	14	6,2	19,3	6,6
20	12,5	7,8	19,5	9,6	28,3	10,1
Channel 2. Intensimeter.						
5	2,64	2,1	4,74	2	6,25	4,3
10	3,8	3,8	9,7	4,1	12	8
15	5,1	5,3	15,1	5,6	18,3	12,1
20	8,4	9	23	8	23	18,6

Table 3. AT-FOI for real data.

	Without smoothing		3-points smoothing		5 points smoothing	
E %	K_c	σ_ε %	K_c	σ_ε %	K_c	σ_ε %
Channel 1. Logarithmic counter.						
5	2,9	1,5	6,7	2,2	8,5	2
10	6,9	4,2	10,3	3,7	15,4	3,9
15	8,2	4,8	14,2	6,1	23,1	6,2
20	8,8	5,8	18,4	7,7	30	8,2
Channel 2. Intensimeter.						
5	2,24	1,7	4,6	2	6,4	2,1
10	3,14	3,5	9	3,8	13,2	3,9
15	4,25	5,4	14,5	5,5	20,4	5,3
20	5,7	7,6	21	6,6	26,7	6,8

Table 4. Background removal + AT-ZOP.

	Without smoothing		3-points smoothing		5-points smoothing	
E %	K_c AT-ZOP	K_c total	K_c AT-ZOP	K_c total	K_c AT-ZOP	K_c total
Channel 1. Logarithmic counter. $K_{c\,BGR} = 2,62$						
5	4,1	10,8	4,7	12,4	4,8	12,6
10	6,5	17	6,6	17,4	7,14	18,7
15	9,4	25	10,7	28	11,1	29
20	18,6	49	19,9	52	24,1	63,3
Channel 2. Intensimeter. $K_{c\,BGR} = 2,3$						
5	2	4,6	2,9	6,6	3,3	7,5
10	2,6	6	4,9	11,2	5,5	12,7
15	3,5	8,1	6,9	15,8	7,7	17,6
20	5,2	12,2	11,9	27,3	12,2	28,1