



## Finite resolution and serial correlations in mass metrology

Viktor Vabson<sup>a,b\*</sup>, Riho Vendt<sup>b,c</sup>, Toomas Kübarsepp<sup>a</sup>, and Mart Noorma<sup>b,c</sup>

<sup>a</sup> Research and Development Division, Metrosert Ltd, Central Office for Metrology, Riia 181c, 51014 Tartu, Estonia

<sup>b</sup> Institute of Physics, Faculty of Science and Technology, University of Tartu, Tähe 4, 51010 Tartu, Estonia

<sup>c</sup> Tartu Observatory, Tõravere, 61202 Tartu County, Estonia

Received 13 October 2013, revised 9 December 2013, accepted 10 December 2013, available online 20 May 2014

**Abstract.** In this study the influence of the digital resolution on the properties of a data sample is experimentally determined in mass measurements. A mass comparator with adjustable digital resolution interval, also known as the quantization step size, was used in well controlled repeatable conditions. The same measurement procedure with the resolution differing by up to a factor of 200 was repeatedly carried out, and at least 150 mass differences were recorded with every resolution setting. A clear relationship was observed between the digital resolution and the type of random process characteristics for the data sample. Analysis shows that a white noise process dominates for data sets measured with the smallest digital resolution step from 0.001 mg and up to 0.005 mg. For resolutions from 0.01 mg to 0.2 mg random walk noise is observed for which variance of the sample mean can increase proportionally with sample size. We demonstrate that instrumental resolution is a strictly limiting factor in mass measurements only for the data sets with significant positive correlations, such as those having random walk noise. Otherwise for the white noise process, the smallest possible variance is inversely proportional to averaging time, like in time-frequency metrology, and not limited by the instrumental resolution. Our measurement results show that in this case sample standard deviation of the mean (0.00003 mg) can be more than ten times smaller than that of a single result (0.0005 mg) or the Type B component of the digital resolution (0.00029 mg).

**Key words:** finite resolution, minimum uncertainty, optimal signal quantization procedure.

### 1. INTRODUCTION

Specially designed statistical methods are recommended for the uncertainty analysis of randomly varying repeated measurements that may be correlated [1]. Such special methods include Allan variance, autocorrelation function, power spectral density, etc., and are well developed in the time and frequency domain [2–4]. They are commonly used for electrical and radiation standard measurements [5–9], but in well established methods such as mass measurement [10,11] they are still rather rare. Nevertheless, in all metrology areas the influence of correlations on the uncertainty evaluation of repeated measurements can be important [1,2,5].

Random processes peculiar to the time series measured at uniform intervals can be of many different

types [2,3]. In the time and frequency field random processes are modelled by five integer power law spectrums  $S_y(f) \sim f^\alpha$ , where the appropriate exponent  $\alpha$  varies from  $-2$  to  $+2$  depending on the instrument used and the region of Fourier frequency  $f$  or the averaging time  $\tau$  under consideration. For electrical quantities two spectrums are considered [5], with  $\alpha = -1$  and  $\alpha = 0$ . However, in some other fields usually only a white noise process with  $\alpha = 0$  (uncorrelated measured quantity values) is assumed [2,5]. Although at first glance the measurement sequences may imply different types of random process, it is impossible to clearly distinguish a white noise process from a  $1/f$  noise process purely by looking at a plot of a time series or at its histogram [5]. These distinctions are important, as the uncertainty associated with these processes will be much different: the variance of the sample mean of a white noise process is inversely proportional to the

\* Corresponding author, [viktorv@metrosert.ee](mailto:viktorv@metrosert.ee)

number of values in the sample, whereas for a  $1/f$  noise process the variance of the sample mean is independent of sample size.

Digital data processing is widely used in modern measurements with rather different sampling frequencies and digital resolutions applied. It is well known that quantization of the analogue signal is a highly nonlinear process, but for the probability densities of the initial and quantized signals the process can be linear [12]. Therefore quantizing conditions are focused on the probability density functions (PDF) of data samples rather than on the particular values. These conditions are formulated in quantizing theorems (QT). According to QT I [12], the PDF of initial signal  $x$  can be derived from the PDF of the output digital signal  $x'$  if the quantization step size  $q$  is suitably small in comparison with the breadth of the PDF. According to QT II, the moments of  $x$  can be calculated from the moments of  $x'$  if  $q$  is respectively small. For QT II the size of  $q$  can be up to twice as large as needed for QT I. So, among digitally measured time series there are quite obviously those fulfilling the requirements of QT I and/or QT II, but also those measured with larger  $q$  and not fulfilling either QT. The situation can be even more complicated when the PDF of an initial signal is a combination of different processes. For example, a white noise can be present together with some low frequency noise, which may substantially burr the limits for the validity of QTs. Thus, specifying the dominant noise of the sample may be helpful for the further data analysis.

Different methods including conventional frequentist [1,13–16], Bayesian [17–20], and fiducial inference [21,22] have been developed for the estimation of the measurand and the uncertainty of repeated measurements affected by finite resolution. For the theoretical modelling of the problem Monte Carlo simulation [23–27] has been used. Two significant factors contributing to the measurement outcome and uncertainty are always assumed in [13–26] to be present: finite resolution and Gaussian noise. In the majority of these publications, it is generally concluded that if the sample standard deviation  $s$  is sufficiently large in comparison to the resolution or quantization step size  $q$  of the measurement instrument (about  $s \geq 0.5q$ ), the effect of the finite resolution will be insignificant, and conventional statistical inference will be valid. Therefore these publications focus on measurement series with relatively small variance ( $s \leq 0.5q$ ) without considering the dominant random process present. However, identification of the dominant noise type of the data series, especially for samples with relatively small variance where the requirements of QT I and QT II may be not met [12], should be of primary importance for uncertainty evaluation.

Some methods presently used for the identification of particular random-noise processes are given in

[2,5,28–31]. All these methods are firmly applicable only if a sufficiently large data set with over a hundred values is available, which was not the case for many of publications ([15–19,21,22]) treating finite resolution. As a kind of criterion of the applicability of conventional statistics, sometimes ([24,32]) the ratio of the standard deviation  $s$  to the resolution  $q$  is used, but its use is not sufficiently justified.

In our experimental study the influence of the digital resolution on the statistical properties of the data sample is experimentally determined. The same mass difference was extensively measured under repeatable conditions. Different resolution settings of the automatic mass comparator from 0.001 mg up to 0.200 mg were applied, and at least 150 mass differences with every resolution setting were recorded. The digital resolution interval  $q$  was the single parameter intentionally changed during the experiment process. From 600 up to 1600 readings at uniform time intervals with every resolution setting were obtained; this is a large data set of 150 to 400 mass differences sufficient for our statistical analysis. Data sets with different statistical properties were recorded and, as recommended in [1,2,5], special methods of accounting for possible correlations were used for the statistical analysis. Our paper is focused on the effects that cannot be observed in routine practice when only relatively small data samples are available. Nevertheless, knowing possible threats due to digitization is helpful also for routine practice.

## 2. STATISTICAL TOOLS FOR AUTO-CORRELATED MEASUREMENTS

One of the central tasks in time–frequency metrology is frequency stability analysis in the time domain based on an array of data points  $y_i$  with constant time intervals between successive measurements. The statistical noise characteristics of a frequency source are usually analysed only after eliminating factors like drift and environmental effects. Data sampling or measurement is carried out during time interval  $\tau_0$ , the analysis or averaging during time  $\tau$ , which is usually a multiple of  $\tau_0$ .

$$\tau = m\tau_0, \quad (1)$$

where  $m$  is the averaging factor.

More than ten different statistical variances have been used for frequency stability analysis in time–frequency metrology [3,4]. Amongst the others, the most widely used are Allan variance and its later overlapping version which, for the same data set, can provide an extension to longer averaging times and better confidence than the original version.

**Table 1.** Frequency and time domain exponents for some statistical power-law noises

Type of noise	Exponent $\alpha$	Exponent $\mu$
White (WFM)	0	-1
Flicker (FFM)	-1	0
Random walk (RWFm)	-2	1

The stability of a frequency source in the time domain is usually shown by using the so-called logarithmic  $\sigma$ - $\tau$  plot that presents some variance  $\sigma^2$  (or standard deviation  $\sigma$ ) of data  $y_i$  as a function of the time  $\tau$  over which the points  $y_i$  are averaged. At the same time, the  $\sigma$ - $\tau$  plot shows the stability of the signal and the type of statistical noise. The slope of the curve,  $\mu$ , is characteristic for a particular power law noise. The most relevant types of noise related to the frequency and time domain exponents  $\alpha$  and  $\mu$  are given in Table 1. In the frequency domain, white noise is a random signal with a constant power spectral density  $S_y(f)$  not depending on frequency, thus  $\alpha=0$  means that  $S_y(f) \sim f^0$ . In the time domain, the well-known law for white noise is valid, stating that the variance of the mean is adversely proportional to the number of averaged values  $N$ :  $\sigma_y^2(\tau) \sim N^{-1}$ , where respective exponent  $\mu=-1$ . For the data set with frequency modulation (FM) the relation between  $\alpha$  and  $\mu$  is:

$$\alpha = -\mu - 1. \quad (2)$$

An important application of the  $\sigma$ - $\tau$  plot is the determining of the flicker floor of the frequency signal or standard [3,33]. This is a point where a white FM noise with  $\mu=-1$  will turn to a flicker FM noise with  $\mu=0$ . This point defines the principal stability and/or uncertainty limit achievable with this particular signal source. Determination of this point requires a lengthy measurement series, which may take several months, and also depends on the analytical method (type of variance) used for analysis. Continued averaging after the flicker floor is achieved will not further improve stability or uncertainty estimates of the signal.

### 2.1. Variances suitable for correlated data

The estimated variance for  $N$  independent random variables  $y$  following [1] is calculated as:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2, \quad (3)$$

where  $y_i$  are the  $N$  values of the data set and  $\bar{y}$  is their arithmetic average.

In the case of auto-correlated and therefore non-independent random data,  $s^2$  is not applicable because

it is non-convergent but (3) shows that  $s$  depends on  $N$ . This problem of adequate estimation will arise if the average value of measurements is not stationary. However, with the Allan variance this sample-dependent unstable behaviour is normally avoided. The Allan variance AVAR (called also the 2-sample variance) is calculated from the data set  $y_i$  as follows [2]:

$$\sigma_y^2(\tau) = \frac{1}{2(M-1)} \sum_{i=1}^{M-1} (y_{i+1} - y_i)^2, \quad (4)$$

where  $y_i$  is the  $i$ th of  $M$  values averaged over the sampling interval,  $\tau$ .

Stability is often expressed as the square root of variance,  $\sigma_y(\tau)$ , the Allan deviation, ADEV. For white FM noise the Allan variance is the same as the variance  $s^2$ . For more divergent noise types such as flicker noise, the Allan variance, as distinct from  $s^2$ , converges to a value that is independent of the size of averaged samples.

At present, the most common choice in time-frequency metrology is the overlapping Allan variance defined as [3,4]:

$$\sigma_y^2(\tau) = \frac{1}{2m^2(M-2m+1)} \sum_{j=1}^{M-2m+1} \left[ \sum_{i=j}^{j+m-1} (y_{i+m} - y_i) \right]^2, \quad (5)$$

where according to (1)  $m$  is the averaging factor,  $\tau$  the averaging time, and  $M$  is the full sample size. In comparison with the original Allan variance, the overlapping version has much better reproducibility and allows for larger averaging times  $\tau$  to be used for the same data set of  $M$  observations. The original and overlapping versions of the Allan variance will give exactly the same result for the smallest time  $\tau_0$  during which a single value is measured, and for longer averaging times the expectations of both versions are the same. The overlapping Allan variance (or estimated sample variance) and respectively Allan or standard deviations are mostly used in our study. As an analogue of the quantity ‘‘frequency’’ commonly analysed in time-frequency metrology our study has the quantity ‘‘mass difference’’, see formula (12).

### 2.2. Power law noise identification

Knowing the power law noise type can considerably improve the planning of measurements; for example, deciding on the optimal averaging time, determining the uncertainty intervals, the equivalent number of degrees of freedom, and correcting for different biases. In the practice of time and frequency metrology, three methods have been used for power law noise identification:

1. Barnes  $B_1$  bias function [2,3,34], which is the ratio of  $s^2$  to the Allan variance;
2. Slope on the logarithmic  $\sigma^2$ - $\tau$  plot [2,3];

3. Use of the lag 1 autocorrelation coefficient [3,28] including differencing and recalculation of autocorrelation for the noises with more divergent data.

The Barnes  $B_1$  bias function shows the non-convergence of standard variance, and is defined as:

$$B_1 = \frac{s^2(\tau)}{\sigma_y^2(\tau)}. \quad (6)$$

Here  $s^2(\tau)$  is the estimated sample variance for points with averaging time  $\tau$  and  $\sigma_y^2(\tau)$  is the Allan variance for averaging time  $\tau$ . An approximate relation of  $B_1$  with lag 1 autocorrelation coefficient  $r_1$  for the same data set is:

$$B_1 \cong \frac{1}{1-r_1}. \quad (7)$$

For similar purposes, the Durbin–Watson statistic  $d$  is used [29–31] instead of the  $B_1$  bias function in many other fields of statistics. Formula (8) gives the Durbin–Watson statistic  $d$  and its relation with  $B_1$ :

$$d = \frac{\sum_{j=2}^N (z_j - z_{j-1})^2}{\sum_{j=1}^N z_j^2} = \frac{2}{B_1}, \quad (8)$$

where the vector  $z$  of the residuals from the regression  $z = \{z_1, z_2, \dots, z_n\}$  of the time series  $y_i$  is defined by  $z_i = y_i - \hat{y}_i$ ; here  $\hat{y}_i$  is a least squares estimate for the value  $y_i$ .

Method 2 for the identification of power law noise is based on the slope of the line fitted through the logarithmic plot of the Allan variance and averaging time. For this method, obviously at least two different averaging times are needed, and estimation is valid for all points used. For practical purposes a single point estimate of method 1 or 3 is preferable.

Method 3 makes use of the lag 1 auto-correlation coefficient calculated from:

$$r_1 = \frac{\sum_{i=1}^{N-1} (y_i - \bar{y})(y_{i+1} - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (9)$$

where  $y_i$  are the successive results of the recorded data set.

According to the algorithm of Method 3, the exponent  $\alpha$  can be calculated from the expression:

$$\alpha = -2(\delta + d), \quad (10)$$

where  $d$  is the order of the differencing of data  $y_i$  ( $d = 0; 1; 2$ ), and  $\delta$  is defined as:

$$\delta = \frac{r_1}{1+r_1}. \quad (11)$$

If  $\delta < 0.25$ , we assume the result of (10) to be valid and final. Otherwise, the lag 1 auto-correlation coefficient is recalculated from the first differences  $y_{i+1} - y_i$  of the frequency data, and (10) is applied with  $d = 1$ . Method 3 can give reliable results for samples starting with  $N \geq 64$ , where for  $\alpha = 0$  the expanded uncertainty covering 94% is approximately from  $-0.4$  to  $+0.7$ . For samples with 128 points the expanded uncertainty covering 99% will be from  $-0.3$  to  $+0.5$  [28].

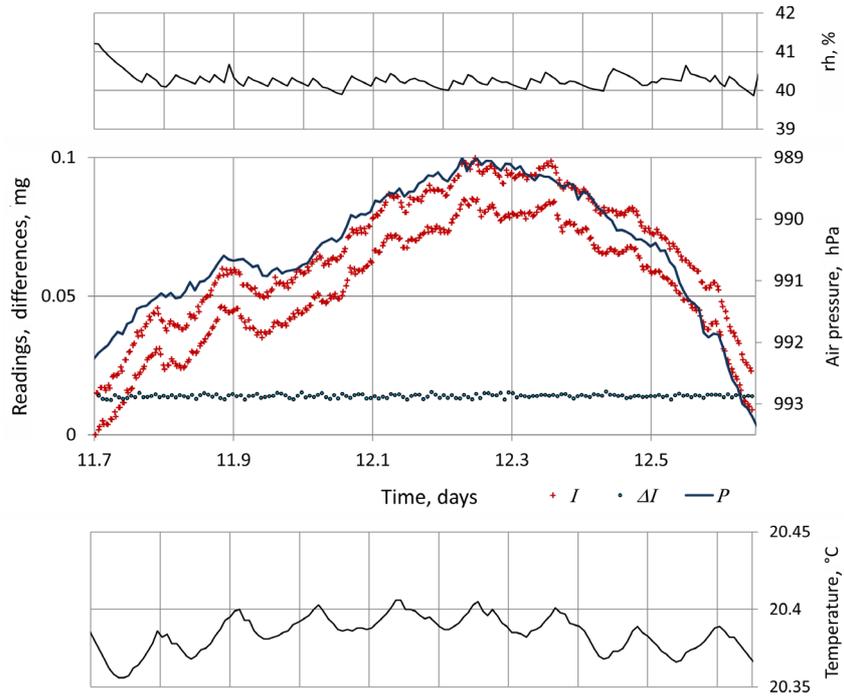
### 3. MEASUREMENT PROCEDURE AND DATA SETS USED FOR ANALYSIS

In our study, the RTTR type measurement cycle is used for the high accuracy mass comparisons with an automatic digital Mettler AT106H comparator operated in an air conditioned room [35]. The comparison cycle starts off with loading the reference weight  $R$ , and after stabilization time the first indication  $I_1$  is recorded. Then test weight  $T$  is loaded and the second indication  $I_2$  is recorded. After a short wait with an unloaded comparator the test weight  $T$  is loaded again, and the third indication  $I_3$  is recorded. Finally, after loading the reference weight  $R$ , indication  $I_4$  is recorded. Weights are automatically placed on the load receiver and removed from it between each measurement. The time interval between the sequent indications is 2.1 min. The full variability range of indications caused by comparator zero drift over one day is normally within  $\pm 0.2$  mg. A typical set of one-day indications and the mass differences calculated from the results is presented in Fig. 1. Air temperature [36], pressure, and relative humidity measured inside the comparator chamber concurrently with mass indications are also shown. The variability of air pressure seems to be the main cause for the zero drift of the comparator seen in Fig. 1. The influence of temperature variations can be noticed as well.

In order to eliminate comparator zero drift from each of the four sequent readings, the mass difference between the weights  $T$  and  $R$  is calculated as

$$\Delta I_i = \frac{-I_1 + I_2 + I_3 - I_4}{2}. \quad (12)$$

Normally the full range of differences is within  $\pm 3$  least significant digital units around the mean; the time required for determining one difference is 8.4 min. Thus, during a one-day experiment usually 150 differences were determined, see Table 2. Because the compared weights were of the same density, air buoyancy correction was not relevant.



**Fig. 1.** One-day measurement sequence: readings (+) and mass differences (•). Solid lines show relative humidity (rh), air pressure, and temperature inside the comparator chamber.

**Table 2.** List of the samples used for the statistical analysis. PDF – probability density function

Resolution, mg	Sample size	Type of PDF & number of values	Sample mean, mg	$s$ , mg	$\sigma_y^2(\tau_0)$ , mg
0.001	100	.i↑↑↑i.	0.014017	0.00043	0.00050
0.001	100	.i↑↑↑i.	0.013953	0.00047	0.00050
0.001	150	.i↑↑↑i.	0.014017	0.00065	0.00069
0.001	120	.i↑↑↑i.	0.014074	0.00029	0.00031
0.001	140	.i↑↑↑i.	0.014071	0.00049	0.00057
0.001	150	.i↑↑↑i.	0.014125	0.00051	0.00057
0.001	110	.i↑↑↑i.	0.014098	0.00029	0.00030
0.001	120	.i↑↑↑i.	0.014073	0.00042	0.00043
0.001	400	.i↑↑↑i.	0.014045	0.00051	0.00055
0.001	160	.i↑↑↑i.	0.014051	0.00036	0.00040
0.001	160	.i↑↑↑i.	0.014077	0.00031	0.00035
0.001	150	.i↑↑↑i.	0.014077	0.00052	0.00055
0.002	150	.i↑↑↑i.	0.014207	0.00080	0.00076
0.005	150	.i↑↑↑i.	0.014083	0.0014	0.0014
0.01	150	↑↑↑	0.0141	0.0045	0.0036
0.02	150	↑↑↑	0.0116	0.0091	0.0061
0.05	150	↑↑↑	0.0082	0.018	0.0061
0.1	150	↑↑↑	0.0083	0.027	0.016
0.2	150	↑↑↑	0.0087	0.040	0.014
0.5	150	↑			

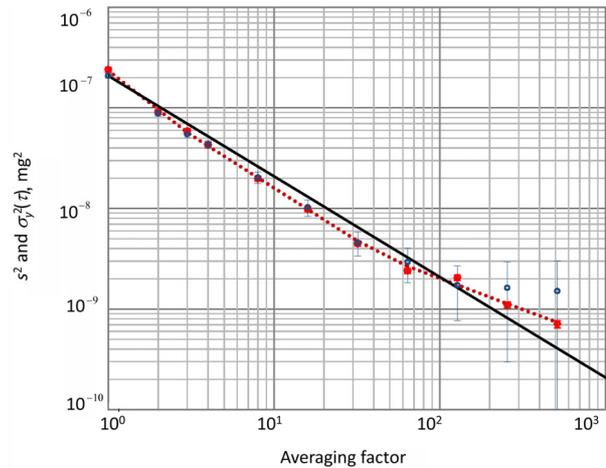
It is interesting to note that random walk noise was dominant for the initial readings of the comparator shown in Fig. 1. According to [3], calculating the first difference of a data set makes it less divergent, and the absolute values of the power law exponents will change by 2. That is exactly the case for mass differences  $\Delta I_i$  calculated from initial readings by using (12), and characterized by the white noise. Indeed, (12) is rather close to the usual differentiating operation.

In the present study the mass differences of the same two 20 g weights were determined over a month (about 20 one-day series with  $N > 100$ ) in repeatable conditions [35]. A list of the samples used for the statistical analysis and some of their key parameters are given in Table 2. Time intervals between the successively measured samples lasted from some hours to some days. A sample with 400 differences was measured without any time gaps during more than two days. About half of the measurements were carried out using the highest digital resolution of comparator  $q = 0.001$  mg; for the rest (below the line in Table 2) a different resolution setting was used for each of the one-day series.

According to [12], the probability density function (PDF) of quantized series  $x'$  consists of a string of Dirac impulses located on a scale of  $q$  and weighted by the samples of the PDF of  $x+n$ , where  $n$  is “pseudo-quantization noise” added to the initial signal  $x$ . In column 3 of Table 2, for  $q$  resolutions from 0.001 mg to 0.005 mg digital distributions with bell-type envelope and with more than 4 different values are typical, and following [12] they likely fulfil at least the requirements of QT II. Starting from  $q \geq 0.01$  mg, the PDFs have only three different values with dominating side columns, which probably are less suitable for the evaluation of the shape or moments of an initial signal.

### 3.1. Statistical properties of mass differences

Data over ten days that contained about 1860 measured differences and were obtained using the highest possible resolution (0.001 mg) of the comparator were analysed together. Figure 2 shows logarithmic plots of  $s^2$  and the Allan variance of mass differences versus averaging factor  $m$  according to (1). Here, factor  $m = 1$  means that the variance is calculated from all successive values and corresponds to measurement time  $\tau_0$ ,  $m = 2$  means that the groups of two successive values are averaged and the variance calculated from them corresponds to time  $2\tau_0$ , and  $m = 3$  means that groups of three values with measurement time  $3\tau_0$  are used, and so on. Both variances have a similar dependence on averaging but different uncertainties (shown at the level of  $k = 2$ ). The small uncertainty of the overlapping Allan variance calculated from (5) makes the curve in Fig. 2 suitable as a basis of reference for results measured with a larger digital step. The estimated sample variance  $s^2$  is in



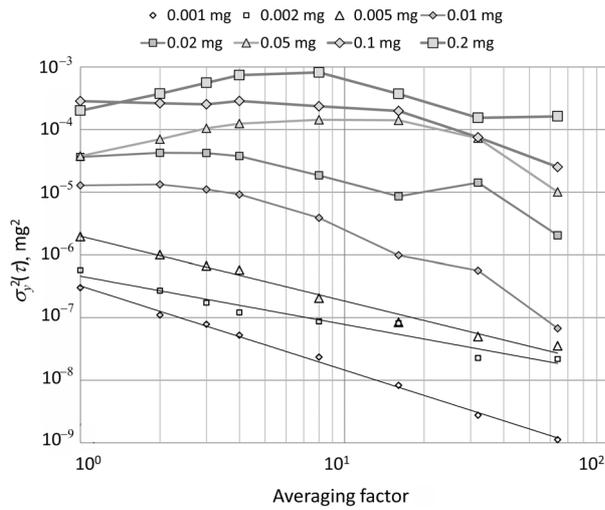
**Fig. 2.** Determination of the lowest limit for uncertainty from the  $\sigma^2$ - $\tau$  plot. The Allan variance is shown with squares,  $s^2$  with open circles, and both together with the expanded uncertainty interval;  $k = 2$ . The white noise  $1/n$  dependence of variance is shown as a straight line.

good agreement with the Allan variance, but has much larger uncertainty, thus both imply that the white noise process is peculiar to the combined large data sample.

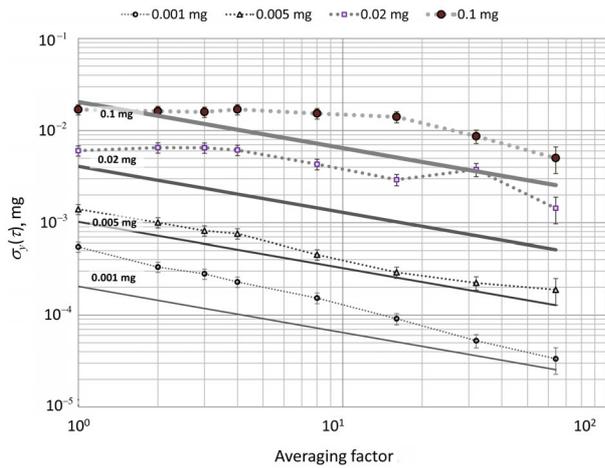
The slope of this plot also specifies the random noise type as a function of  $m$  or  $\tau$ . For all 11 data sets measured with the highest resolution of 0.001 mg, the dominant power law noise process for  $\tau_0$  was identified:  $\alpha = 0.45 \pm 0.23$  and  $\mu = -1.45 \pm 0.23$ . These values are valid for the first few points in the series. In Fig. 2 the slope for the central part of the curve is close to white noise ( $\mu = -1.05$ ), and starting from an averaging factor of 64 the curve turns from white to flicker noise and has a significantly smaller slope. Nevertheless, the actual flicker floor is not reached and, for samples with  $N < 100$ , averaging still reduces the uncertainty. In Fig. 2 the Allan deviation for a single observation  $\sigma_y(\tau_0 = 8.5 \text{ min})$  is 0.0005 mg, and  $\sigma_y(\tau = 4300 \text{ min})$  equals 0.000027 mg, which is about ten times smaller than the assumed Type B contribution of the digital resolution  $q = 0.001$  mg, i.e. of  $(q/12)^{1/2} \cong 0.00029$  mg. Data obtained with  $q = 0.001$  mg most likely fulfil the requirements for QT I and QT II [12].

The Allan variance of mass difference as a function of averaging factor  $m$  as depending on the digital resolution of the comparator is presented in Fig. 3. The effect of digital resolution is evident. For resolutions from 0.001 mg to 0.005 mg the Allan variance of the mass differences measured decreases with averaging time, as expected for white noise with  $\mu \approx -1$ . For resolutions from 0.01 mg to 0.2 mg, slopes  $\mu \geq 0$ , and by using averaging uncertainty cannot be firmly improved, at least for the first ten points in a measurement series.

Some of the  $\sigma^2$ - $\tau$  curves given in Fig. 3 are modified to  $\sigma$ - $\tau$  in Fig. 4. The straight lines in the



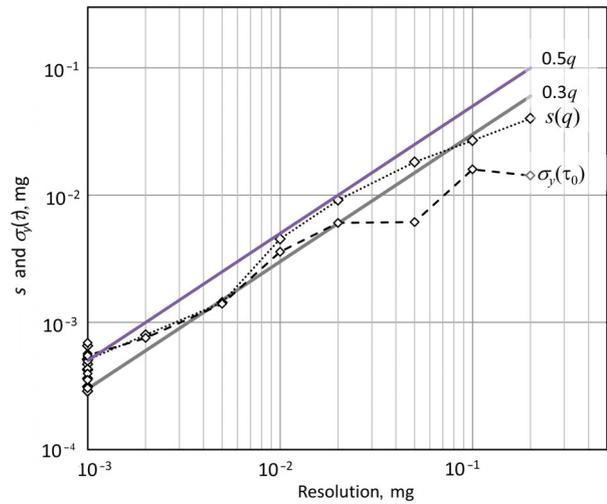
**Fig. 3.** Allan variance of measured mass differences as a function of averaging factor  $m$  and digital resolution  $q$ .



**Fig. 4.** Allan deviation as a function of averaging factor  $m$  and digital resolution  $q$ . Straight lines represent only the uncertainty of the digital rounding effect due to the corresponding  $q$ .

figure show the contribution of the digital rounding effect due to corresponding digital resolution step  $q$  in respect to the measured Allan deviation of the mass differences. For a resolution of 0.001 mg, the Allan deviation of a single result will be reduced more than ten times if a series of 64 points is averaged. At the same time, for a resolution of 0.1 mg the Allan deviation will be reduced only 3.4 times, and averaging up to the first 16 points will not show any reduction. Therefore, a common Type B estimate of 0.029 mg for  $q = 0.1$  mg is quite adequate and this estimate cannot be reliably further improved by means of averaging.

In Fig. 5 the experimental standard deviation and the Allan deviation are presented as functions of the



**Fig. 5.** Standard deviation and the Allan deviation as a function of digital resolution  $q$ . Straight lines show the range from  $0.3q$  to  $0.5q$  of the respective resolution.

digital resolution applied, including a number of values for  $q = 0.001$  mg, see also Table 2. As stated in [13,23,24,32], resolution  $q$  will not affect the conventional statistical inference if  $s \geq 0.5q$ ; and conventional statistical inference cannot be applied if  $s \leq 0.3q$  [32]. These limits are depicted as straight lines in Fig. 5. The major part of all experimental standard deviations determined over a month (see Table 2) fit in the range specified by those limits, and only two for the resolutions of  $q = 0.1$  mg and  $q = 0.2$  mg are smaller, implying that conventional statistical inference cannot be applied. At the same time, as confirmed by Figs 2 and 3 and Table 3, data sets with dominating white noise suitable for conventional analysis are obtained only for the three first points with resolutions from  $q = 0.001$  mg to  $q = 0.005$  mg; for data with resolutions from 0.01 mg to 0.2 mg apparently more advanced tools are required.

**Table 3.** Determination of the power law noise exponent as a function of resolution

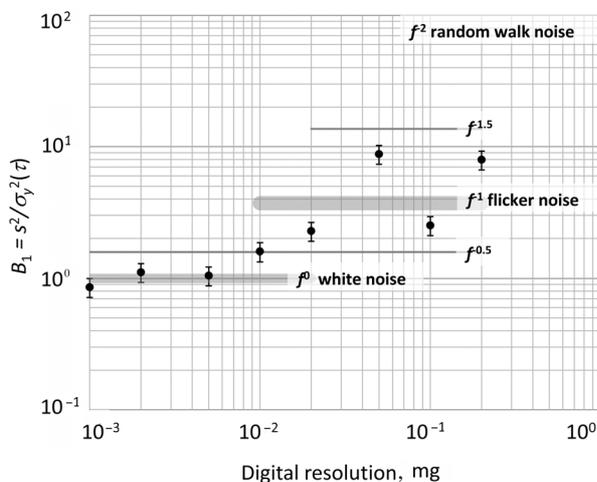
Resolution, mg	Method 1: from $B_1$		Method 2: from slope		Method 3: from AC	
	Exponent $\alpha$	Type of noise	Exponent $\alpha$	Type of noise	Exponent $\alpha$	Type of noise
0.001	0.3	WFM	0.2	WFM	0.4	WFM
0.002	-0.2	WFM	0.1	WFM	-0.2	WFM
0.005	-0.1	WFM	-0.0	WFM	-0.0	WFM
0.01	-0.5	FFM	-0.9	FFM	-1.6	RWFM
0.02	-0.8	FFM	-1.1	FFM	-1.8	RWFM
0.05	-1.4	FFM	-1.9	RWFM	-2.3	RWFM
0.1	-0.8	FFM	-0.9	FFM	-1.6	RWFM
0.2	-1.3	FFM	-1.9	RWFM	-2.3	RWFM

The minimum value of the experimental standard deviation observed with  $q = 0.001$  mg was  $s = 0.00029$  mg and  $\sigma_y(\tau_0) = 0.00030$  mg. Nevertheless, conventional analysis of these data will lead to a moderate overestimation of uncertainty, as the determined exponents ( $\alpha = 0.54$  and  $\mu = -1.54$ ) surpass the white noise values, and thus the dependence of the variance of the sample mean on the number  $N$  of the values in the sample is somewhat stronger than inversely proportional to  $N$ . The ratio of standard deviation to resolution  $q$  is rather insensitive to the type of random process (Fig. 5). A significantly better choice for the distinction of data sets with zero auto-correlation is the Allan deviation, but still it is less reliable than identifying the type of power law noise.

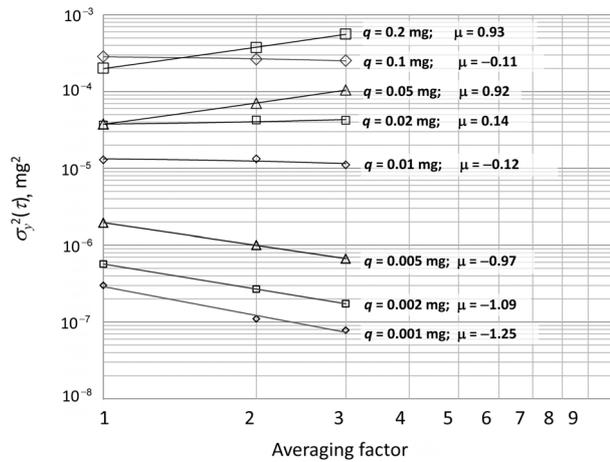
**3.2. Results of power law noise identification**

In Method 1, the Barnes  $B_1$  bias function is applied to data sets measured with different digital resolutions. The results are shown in Fig. 6. Comparing the actual value of  $B_1$  with the values expected for the various noise types allows identification of the dominant noise type: see Table 3. It is evident that dominant random noise types depend on the resolution used and change from the white noise process for  $q$  ranging from 0.001 mg to 0.005 mg to the combined  $1/f$  and  $1/f^2$  noise processes for  $q$  ranging from 0.02 mg to 0.2 mg.

In Method 2, the Allan variance as a function of averaging factor  $m$  (logarithmic  $\sigma^2-\tau$  plot) calculated for different digital resolution units from the same data sets as used in Fig. 6 is presented in Fig. 7. The slopes of the curves obtained for larger values of  $q$  show that the type of random noise is different. The slope of the three smallest values of  $q$  is close to  $-1$ , which is characteristic of the white noise process (WFM); the



**Fig. 6.** Random noise type identification with Method 1 as a function of digital resolution.



**Fig. 7.** Random noise identification with Method 2: slope of the  $\sigma^2-\tau$  curves.

slopes of the three largest values of  $q$  are more typical of the  $1/f$ -noise process (FFM).

For larger averaging factors the slopes of  $\sigma^2-\tau$  curves tend to be more similar (see Fig. 3), which is consistent with an increasing influence of the comparator zero drift for the dominating noise in this case<sup>1</sup>. Nevertheless, for the first 20 differences shown in Fig. 3 the noise processes are clearly different, and for larger resolutions a predictable reduction of uncertainty with increasing averaging time cannot be assumed.

For Method 3, the intermediate calculation results are presented in Table 4. In the first step lag 1 correlation coefficients are calculated from mass differences obtained by using (9). For the results where  $\delta > 0.25$  a

**Table 4.** Random noise identification with Method 3: from lag 1 autocorrelation

Resolu- tion, mg	First step		Second step		Final result
	Correla- tion $r_1$	$\delta$	Correla- tion $r_1$	$\delta$	
0.001	-0.172	-0.213			0.4
0.002	0.091	0.084			-0.2
0.005	0.017	0.016			-0.0
0.01	0.372	0.271	-0.175	-0.206	-1.6
0.02	0.560	0.359	-0.103	-0.115	-1.8
0.05	0.886	0.470	0.167	0.143	-2.3
0.1	0.587	0.370	-0.156	-0.186	-1.6
0.2	0.874	0.466	0.167	0.143	-2.3

<sup>1</sup> Although before the calculation of  $s^2$  or the Allan variance the zero drift has been eliminated, before signal quantization it is fully present and so certainly its amplitude has an influence on the outcome of quantization.

second step is required (see column 3 of Table 4). After differentiating the initial mass differences, the lag 1 correlation coefficients are calculated again. A third step is not needed as  $\delta < 0.25$  for all  $q$ .

In Table 3, the exponent  $\alpha$  is determined as a function of resolution  $q$  by using all three methods described in Section 2.2. As shown in [28], Method 3 has a better capability for noise identification than the other methods. According to this method, two types of random noise are evident in connection with digital resolution. Our measurements indicate that white noise (WFM) dominates for  $q$  in the range from 0.001 mg to 0.005 mg and random walk type noise (RWFM) is typical for  $q$  in the range from 0.01 to 0.2 mg.

#### 4. CONCLUSIONS

We measured mass differences obtained from a high-accuracy comparator set to different digital resolutions to study the effect of resolution or digital quantization error on the statistical properties of measured data. For the resolution of  $q = 0.001$  mg, the difference between the estimated sample variance and the Allan variance is insignificant (Fig. 2). This was expected because the statistical tools suitable for correlated random data would give similar results with conventional uncertainty measurement (GUM) practice [1] if the relevant for Type A estimation assumptions were fulfilled. Furthermore, the tools intended for random measurements that may be correlated can reveal features not considered in conventional analysis. For example, as shown in Fig. 2, due to a small negative autocorrelation the dependence of uncertainty on averaging factor or time is stronger ( $\mu = -1.45$ ) than for the pure white noise ( $\mu = -1$ ) often assumed in practice. However, after averaging a certain number of points, about 50 to 100, the curve tends to the flicker noise with  $\mu \approx -0.5$ , and further averaging obviously will be less effective.

The data sample measured with  $q = 0.001$  mg most likely fulfils the requirements of QT I and QT II [12], samples with  $q = 0.002$  mg and  $q = 0.005$  mg at least the requirements of QT II. For resolutions in the range 0.02–0.2 mg significant differences between the sample variance estimated using (3) and the Allan variance calculated from (4) or (5) are evident (Fig. 6). This indicates that the assumptions needed for Type A estimation – randomness and independence of sequent observations – are not met and that the tools of Section 2 are more convenient but would require much larger data sets and more complicated calculations. Distinction between data sets suitable for conventional analysis and data sets requiring more advanced tools cannot be based on the ratio of standard deviation to resolution  $q$  as proposed previously [32]. Power law noise identification or the Allan deviation are much more effective for that.

Nearly proportional dependence of a standard deviation on digital resolution  $q$  presented in Fig. 5 confirms that, among the factors causing variability of indications in a sample, the digital rounding is often dominating. This is true until the comparator would yield the same value during the measurements as in the case of  $q = 0.5$  mg. The other components arising from internal (Gaussian) noise and from the zero drift of the comparator are substantially contributing to uncertainty for  $q = 0.001$ – $0.005$  mg.

By using the Allan variance or the Allan deviation of measured mass differences as a function of averaging time, the relation between finite resolution and the type of random process of the sample became very evident (Figs 3 and 4). Setting  $q$  from 0.001 mg up to 0.01 mg, or even larger, the apparent power law noise exponent is switched from white with the slope of  $\mu \approx -1$  to the random walk type with  $\mu \approx 1$  (see Figs 3 to 7, Tables 2 and 3). Concurrently the shape of the PDF of the sample changes from bell-type envelope to a distribution with strongly dominating side columns.

We think that for  $q = 0.001$  mg the requirements for QT I are fulfilled, and the shape of the probability density function of quantized data is similar to that of the initial data. For  $q = 0.002$  mg and  $q = 0.005$  mg the requirements for QT II are fulfilled, and the moments of the probability density function of quantized data can be reliably used. Otherwise, the uncertainty of the mean value of the data set measured by using  $q > 0.02$  mg cannot be effectively improved by means of averaging, and correlation between results must be accounted for. Our analysis confirms that the type of random process present will affect the estimation of measurement uncertainty. Knowing the noise behaviour will facilitate the planning of experiment: the optimal averaging or integrating time (useful number of repetitions) and the best resolution achievable with the particular measurement procedure can be estimated.

#### ACKNOWLEDGEMENTS

The authors are grateful to the two anonymous referees for carefully reading the manuscript and for their useful suggestions, which helped to improve this work.

#### REFERENCES

1. *Guide to the Expression of Uncertainty in Measurement*. BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, OIML, 1993.
2. Allan, D. W. Should the classical variance be used as a basic measure in standards metrology? *IEEE T. Instrum. Meas.*, 1987, **IM-36**(2), 646–654.
3. Riley, W. J. *Handbook of Frequency Stability Analysis*. Hamilton Technical Services, 2007.

4. IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology – Random Instabilities. IEEE Std 1139-2008, 2009.
5. Witt, T. J. Practical methods for treating serial correlations in experimental observations. *Eur. Phys. J. Spec. Top.*, 2009, **172**(1), 137–152.
6. Tang, Y., Solve, S., and Witt, T. J. Allan variance analysis of Josephson voltage standard comparison for data taken at unequal time intervals. *IEEE T. Instrum. Meas.*, 2011, **60**(7), 2248–2254.
7. Witt, T. J. and Fletcher, N. E. Standard deviation of the mean and other time series properties of voltages measured with a digital lock-in amplifier. *Metrologia*, 2010, **47**(5), 616–630.
8. Witt, T. J. Using the Allan variance and power spectral density to characterize DC nanovoltmeters. *IEEE T. Instrum. Meas.*, 2001, **50**(2), 445–448.
9. Clare, J. F. and White, D. R. Variance in the mean of a sequence of partially correlated measurements. *Metrologia*, 1990, **27**(4), 193–200.
10. Georgakaki, D., Mitsas, C., and Polatoglou, H. Spectral analysis and Allan variance calculation in the case of phase noise. *arXiv:1202.6627*, 2012.
11. Georgakaki, D., Mitsas, C., and Polatoglou, H. Time series analysis of the response of measurement instruments. *arXiv:1202.6599*, 2012.
12. Widrow, B. and Kollár, I. *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*. Cambridge University Press, Cambridge, UK, 2008.
13. *ISO/TC 14253-2:2011, Geometrical Product Specifications (GPS) – Inspection by Measurement of Workpieces and Measuring Equipment. Part 2: Guide to Estimation of Uncertainty in GPS Measurement in Calibration of Measuring Equipment and in Product Verification*. ISO, 2011.
14. *International Recommendation OIML R 111-1, ed. 2004 (E), Weights of classes E1, E2, F1, F2, M1, M1-2, M2, M2-3, M3, Part 1: Metrological and technical requirements*. OIML, Paris, France, 2004.
15. Lira, I. H. and Wöger, W. The evaluation of standard uncertainty in the presence of limited resolution of indicating devices. *Meas. Sci. Technol.*, 1997, **8**(4), 441–443.
16. Taraldsen, G. Instrument resolution and measurement accuracy. *Metrologia*, 2006, **43**(6), 539–544.
17. Elster, C. Evaluation of measurement uncertainty in the presence of combined random and analogue-to-digital conversion errors. *Meas. Sci. Technol.*, 2000, **11**(9), 1359–1363.
18. Willink, R. On the uncertainty of the mean of digitized measurements. *Metrologia*, 2007, **44**(1), 73–81.
19. Lira, I. Resolution revisited. *Metrologia*, 2006, **43**(3), L14–L17.
20. Lira, I. Assigning a probability density function for the value of a quantity based on discrete data: the resolution problem. *Metrologia*, 2012, **49**(6), 765.
21. Hannig, J., Iyer, H. K., and Wang, C. M. Fiducial approach to uncertainty assessment accounting for error due to instrument resolution. *Metrologia*, 2007, **44**(6), 476–483.
22. Frenkel, R. B. Fiducial inference applied to uncertainty estimation when identical readings are obtained under low instrument resolution. *Metrologia*, 2009, **46**(6), 661–667.
23. Frenkel, R. B. and Kirkup, L. Monte Carlo-based estimation of uncertainty owing to limited resolution of digital instruments. *Metrologia*, 2005, **42**(5), L27–L30.
24. Phillips, S. D., Toman, B., and Estler, W. T. Uncertainty due to finite resolution measurements. *J. Res. Natl. Inst. Stan.*, 2008, **113**(3), 143.
25. Burr, T., Hamada, M., Cremers, T., Weaver, B., Howell, J., Croft, S., and Vardeman, S. Measurement error models and variance estimation in the presence of rounding error effects. *Accredit. Qual. Assur.*, 2011, **16**(7), 347–359.
26. Witkovský, V. and Wimmer, G. Interval estimation of the mean of a normal distribution based on quantized observations. *Mathematica Slovaca*, 2009, **59**(5), 627–645.
27. Burr, T., Croft, S., Hamada, M. S., Vardeman, S., and Weaver, B. Rounding error effects in the presence of underlying measurement error. *Accred. Qual. Assur.*, 2012, **17**(5), 485–490.
28. Riley, W. J. and Greenhall, C. A. Power law noise identification using the lag 1 autocorrelation. In *Proc. 18th European Frequency and Time Forum (EFTF), 5–7 Apr. 2004*, pp. 576–580.
29. Durbin, J. and Watson, G. S. Testing for serial correlation in least squares regression. I. *Biometrika*, 1950, **37**(3–4), 409–428.
30. Durbin, J. and Watson, G. S. Testing for serial correlation in least squares regression. II. *Biometrika*, 1951, **38**(1–2), 159–178.
31. Durbin, J. and Watson, G. S. Testing for serial correlation in least squares regression. III. *Biometrika*, 1971, **58**(1), 1–19.
32. Cordero, R. R., Seckmeyer, G., and Labbe, F. Effect of the resolution on the uncertainty evaluation. *Metrologia*, 2006, **43**(6), L33–L38.
33. Predehl, K., Grosche, G., Raupach, S. M. F., Droste, S., Terra, O., Alnis, J., et al. A 920-Kilometer optical fiber link for frequency metrology at the 19th decimal place. *Science*, 2012, **336**(6080), 441–444.
34. Barnes, J. A. *Tables of Bias Functions, B1 and B2, for Variances Based on Finite Samples of Processes with Power Law Spectral Densities*. NBS Technical Note 375, Jan. 1969.
35. Vabson, V., Kübarsepp, T., Vendt, R., and Noorma, M. Traceability of mass measurements in Estonia. *Measurement*, 2010, **43**(9), 1127–1133.
36. Vendt, R., Vabson, V., Kübarsepp, T., and Noorma, M. Traceability of temperature measurements in Estonia. *Proc. Estonian Acad. Sci.*, 2013, **62**, 116–121.

## Piiratud lahutus ja tulemuste autokorrelatsioon massimõõtmistel

Viktor Vabson, Riho Vendt, Toomas Kübarsepp ja Mart Noorma

Eksperimentaalselt on uuritud massikomparaatori digitaalse lahutuse mõju mõõteseeria statistilisele jaotusele. Kasutati muudetava digitaalse lahutusega massikomparaatorit hästi kontrollitud ja korratavates tingimustes. Sama mõõteprotseduuri korrati kaheksa erineva digilahutusega alates komparaatori piirlahutusest 0,001 mg kuni lahutuseni 0,2 mg. Suurima ja väikseima lahutuse suhe oli 200 korda. Iga erineva lahutuse korral mõõdeti seeria, milles oli vähemalt 150 üksiktulemust. Kasutatud digilahutuse ja mõõteseeriast iseloomustava juhusliku protsessi vahel ilmnes selge seos. Analüüs näitas, et nn valge müra tüüpi protsess domineeris väiksemate digisammudega (0,001 mg kuni 0,005 mg) saadud mõõteseeriast korral. Tavapäraselt eeldatava valge müra tüüpi protsessi puhul on aritmeetilise keskmise dispersioon pöördvõrdeline keskmistatud tulemuste arvuga. Digisammude 0,01 mg kuni 0,2 mg korral domineeris madalsageduslik  $1/f^2$  müra, mil aritmeetilise keskmise dispersioon oli võrdeline keskmistatud tulemuste arvuga. Seega ilmnes, et mõõtevahendi digilahutus piirab massimõõtmiste lahutust vaid tugeva positiivse korrelatsiooni korral mõõteseerias ja see on iseloomulik suuremate digisammude (0,01 mg kuni 0,2 mg) korral. On ilmne massimõõtmiste analoogia aja ja sageduse metroloogiaga, kus allika stabiilsushinnangut piirab protsessist sõltuv optimaalne keskmistamisaeg. Komparaatori lahutuse 0,001 mg korral võib massierinevuste seeria keskmine standardhälve (0,00003 mg) olla üle kümne korra väiksem üksiktulemuste standardhälbest (0,0005 mg) või tavapärasest digilahutuse määramatuse B-tüüpi hinnangust (0,00029 mg).