UDC 006.91:53.088

Estimation of the uncertainty bands of regression line for correlated data of variable *Y* using the GUM rules

Z.L. Warsza [1], J. Puchalski [2]

[1] *Industrial Research Institute for Automation and Measurements PIAP-Ł (Łukasiewicz Research Network), Aleje Jerozolimskie 202, 02-486 Warszawa, Poland*
zlw1936@gmail.com
[2] *Central Office of Measures (GUM), Elektoralna 2, 00-137 Warszawa, Poland*
jacek.puchalski@gum.gov.pl

Abstract

The estimation of the accuracy of the linear regression method used for measurements based on the GUM recommendations is considered. The impact of correlation and autocorrelation of the variable ***Y*** data together with type A and type B uncertainties, not provided in statistical literature about regression methods, is discussed. The theoretical backgrounds are given. The simulated examples of determining the uncertainty bands of the regression line fitted to the measured points with different cases of correlated values and absolute and relative uncertainties of type A and type B of the dependent variable ***Y*** are considered.

Keywords: measurement; autocorrelation; regression line; type A and type B uncertainty bands.

Received: 03.11.2022 Edited: 08.12.2022 Approved for publication: 14.12.2022

Introduction

The estimation of the accuracy of measurement results and instrumentation used in national metrology services, science, industry and many other fields is now conducted according to the uncertainty terms recommended in the international GUM Guide [1–3]. It is also described in a few books and other literature. In the literature on the linear regression method, e.g. [4–5], the accuracy of its straight-line parameters is statistically estimated only from the measured data and from their random uncertainties in the experiment. The impact of type B uncertainty obtained from the Maximal Permissible Error (MPE) of the used instruments on the uncertainty bands of the straight-line, when a measurement result is obtained using the regression method, and their use over an acceptable period and conditions of the validation data is not considered [4, 6–10]. Even in the latest research works on the regression method based on the Bayes, suggested for the new version of the GUM (GUM 2), the impact of type B uncertainty on the accuracy of the regression function was not separately considered.

The determination of equations for the borders of the uncertainty bands of the regression line for the uncorrelated *y*-ordinates of the measured points was analysed earlier in [7–9]. The equations of the regression line uncertainty of the autocorrelated and correlated coordinates of the variable ***Y***, based on works [6–9], are presented below. In the estimation of the accuracy, the influence of type B uncertainty is also included.

1. Linear regression method for correlated data of variable *Y*

The linear regression method is used below for a dependent random variable ***Y*** (x, y) to determine a straight-line $y = ax + b$ and borders of its expanded uncertainty band $\pm U_{y.95}$. This line is determined according to the specified criterion for measuring y_i-ordinates (for $i = 1, ..., n$) of n tested points $Q_i(y_i, x_i)$ – see Fig. 1, i.e.:

$$y = ax + b \pm U_{y.95}(x). \quad (1)$$

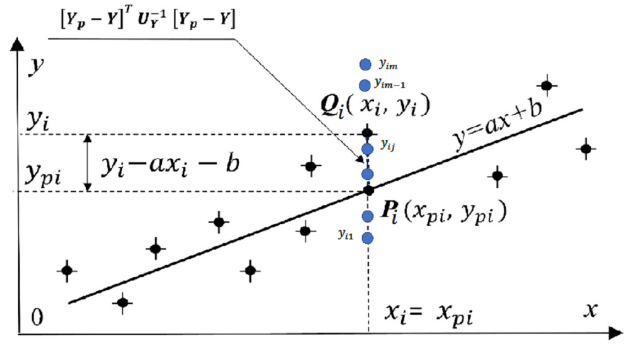


Fig. 1. Regression line for points with measured y_i-values of variable ***Y*** and known x_i-values

For most scientific and technological experiments, it can be assumed that the measured results of y_i for

the dependent variable $\boldsymbol{Y}(x)$ are random, and the x_i-values (for $i = 1, ..., n$) of an independent variable $\boldsymbol{X}(x)$ are known and precisely determined. Then according to the GUM [1], the y_i-values have combined uncertainties $u(y_i)=\sqrt{u_{iA}^2+u_{iB}^2}$, and $u(x_i)=0$. If for every given x_i, the y_i-values are stable during the experiment, it is enough to measure them once. If they are randomly changed, then can be measured repeatedly as (y_{i1}, ..., y_{ij}, ... y_{im}) to find the best estimators of the y_i-values and their uncertainties $u(y_i)$.

To describe the accuracy of correlated values y_i, ..., y_n, the covariance matrix is used:

$$\boldsymbol{U_Y}=\begin{bmatrix} u^2(y_1) & \cdots & \rho_{y1n}u(y_1)u(y_n) \\ \vdots & \ddots & \vdots \\ \rho_{y1n}u(y_1)u(y_n) & \cdots & u^2(y_n) \end{bmatrix}, \quad (2)$$

where ρ_{yij} is the correlation coefficient of y_i, y_j of the variable $\boldsymbol{Y}$; and $u(y_i)$, $u(y_j)$ are their uncertainties, $i, j = 1, ...n$.

For a regression line of real data, the correlation of the y-ordinates of close points is usually much stronger than that of distant points. The elements with the highest values of the correlation coefficients are located around the main diagonal of the $\boldsymbol{U}_Y$ covariance matrix, and other elements can be treated as equal to zero.

For the least squares' criterion of such a variable $\boldsymbol{Y}$, the multivariate Gauss distribution is used:

$$f(\boldsymbol{Y})=\frac{1}{(2\pi)^{n/2}\sqrt{\det(\boldsymbol{U}_Y)}}\exp\left(-\frac{1}{2}[\boldsymbol{Y_p}-\boldsymbol{Y}]^{\mathrm{T}}\boldsymbol{U}_Y^{-1}[\boldsymbol{Y_p}-\boldsymbol{Y}]\right), \quad (3)$$

where $\boldsymbol{Y}=[y_1, ... y_i, ... y_n]^{\mathrm{T}}$ is the n-dimensional vector of the y_i-ordinates of the measured points (observed points), $\boldsymbol{Y}_p$ is the vector y_{ip} of their parallel projections to the regression line in the 0Y-direction (fitted points), $\boldsymbol{U}_Y^{-1}$ is the inverse of the covariance matrix.

A solution to the linear vector equation is:

$$\boldsymbol{Y_p}=a\boldsymbol{X_p}+b\boldsymbol{1}=a\boldsymbol{X_p}+\boldsymbol{b}, \quad (4)$$

where $\boldsymbol{X}_p$ is the cut-off vector x_i of the measured points, $\boldsymbol{1}=[\boldsymbol{1},...,\boldsymbol{1}]^{\mathrm{T}}$ is the unit vector size of n.

If the variable $\boldsymbol{Y}$ is only measured, then the target function $P=f(\boldsymbol{Y})\rightarrow\max$ is applied. This will occur for a minimum of the criteria function $SK(a, b)$

$$SK(a,b)=-\ln(f(\boldsymbol{Y}))\rightarrow\min. \quad (5)$$

It is assumed that the sum of the normalized squares of the distance of the measured points with coordinates (x_i, y_i) for $i=1,2,...n$ from the specified points (x_i, y_{pi}) on a regression straight-line (Fig. 1) will reach a minimum of zero of the sum of the squares of the sub-derivatives after the parameters a and b of this line. In general, i.e. when the values of both coordinates are measured, the sets of the values y_i and x_i may be autocorrelated, as well as the ordinates y_i and x_i of the measured points may be correlated. Then the determination of the linear regression method requires two covariance matrices, $\boldsymbol{U}_Y$ and $\boldsymbol{U}_X$, for both random variables $\boldsymbol{Y}$ and $\boldsymbol{X}$.

When the x_i-values of the measured points do not have a random dispersion, they are defined precisely, and their uncertainty $u(x_i)=0$ and the equation resulting from criterion (5) is simplified, i.e.

$$SK=\Delta\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\Delta\boldsymbol{Y}=[\boldsymbol{Y_p}-\boldsymbol{Y}]^{\mathrm{T}}\boldsymbol{U}_Y^{-1}[\boldsymbol{Y_p}-\boldsymbol{Y}]. \quad (6)$$

Dependencies of the coordinates of points P_i lying on the regression line $y=ax+b$ with abscissas $x_{pi}=x_i$ and y_{pi}-ordinates are described by the vector equation

$$\boldsymbol{Y_p}=a\boldsymbol{X_p}+\boldsymbol{b}. \quad (7)$$

The distance of the measured point $\boldsymbol{Q}_i$ from point P_i is equal to the coordinate differences between these two points (see Fig. 1). In the regression method with parallel projections of the measured points Q_i in the 0y-direction on a designated straight-line, the measured point Q_i is therefore at the distance of $y_{pi}-y_i=ax_i+b-y_i$ from point $\boldsymbol{P}_i$.

The formulas of the regression line are simpler due to the use of auxiliary parameters

$$S=\boldsymbol{1}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{1}=\sum_{i=1}^{n}\sum_{j=1}^{n}[\boldsymbol{U}_Y^{-1}]_{ij}, \quad (7a)$$

$$S_x=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{1}=\boldsymbol{1}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{X}, \quad (7b)$$

$$S_{xx}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{X}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{X}, \quad (7c)$$

$$S_{xy}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{Y}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{X}, \quad (7d)$$

$$S_y=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{1}=\boldsymbol{1}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{Y}, \quad (7e)$$

$$S_{yy}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{Y}=\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{Y}. \quad (7f)$$

Then for $\boldsymbol{X}_p=\boldsymbol{X}$ and $\boldsymbol{Y}_p=a\boldsymbol{X}_p+\boldsymbol{b}$ equation (6) takes the form:

$$\begin{aligned} SK&=[a\boldsymbol{X}+\boldsymbol{b}-\boldsymbol{Y}]^{\mathrm{T}}\boldsymbol{U}_Y^{-1}[a\boldsymbol{X}+\boldsymbol{b}-\boldsymbol{Y}]= \\ &=a^2S_{xx}+2abS_x-2aS_{xy}+b^2S-2bS_y+S_{yy}. \end{aligned} \quad (8)$$

The condition for minimizing expression (8) as the sum of the squares of the distance of n points Q_i from the regression line in the 0y-direction (Fig. 1) results from the solution for the system of two equations for its partial derivatives with respect to the parameters a and b of the regression line, i.e. $\frac{\partial SK}{\partial a}=0$, $\frac{\partial SK}{\partial b}=0$:

$$\begin{cases} aS_{xx}+bS_x=S_{xy} \\ aS_x+bS=S_y \end{cases}. \quad (9)$$

Solving the system of equations (9) we get:

$$a = \frac{SS_{xy} - S_x S_y}{SS_{xx} - (S_x)^2} = \frac{\Delta_a}{\Delta};\ b = \frac{S_y S_{xx} - S_x S_{xy}}{SS_{xx} - (S_x)^2} = \frac{\Delta_b}{\Delta}, \quad (10\,\text{a, b})$$

where $\Delta_a = SS_{xy} - S_x S_y$, $\Delta_b = S_y S_{xx} - S_x S_{xy}$ and $\Delta = SS_{xx} - (S_x)^2$.

The criterion SK_{MIN} from (5) will be met by parameters a i b where:

$$SK_{\text{MIN}} = -aS_{xy} - bS_y + S_{yy}. \quad (11)$$

The borders of the uncertainty band $U(y)$ lies symmetrically on both sides of the regression line. They will be determined considering the uncertainties u_a of slope a and the uncertainty u_b of intercept b. The parameters a, b from formulas (10a, b) are correlated. The partial derivatives of the regression straight-line parameters are:

$$\boldsymbol{c} = \frac{\partial S_{xy}}{\partial \boldsymbol{Y}} = \boldsymbol{U}_{\boldsymbol{Y}}^{-1}\boldsymbol{X},\quad \boldsymbol{d} = \frac{\partial S_y}{\partial \boldsymbol{Y}} = \boldsymbol{U}_{\boldsymbol{Y}}^{-1}\boldsymbol{1}. \quad (12\,\text{a, b})$$

The sensitivity matrix for parameters a, b has the form $\boldsymbol{AB} = [\boldsymbol{A}, \boldsymbol{B}]$ of the elements defined by:

$$\boldsymbol{A} = \frac{\partial a}{\partial \boldsymbol{Y}} = \frac{S\,\boldsymbol{c} - S_x\,\boldsymbol{d}}{\Delta};\quad \boldsymbol{B} = \frac{\partial b}{\partial \boldsymbol{Y}} = \frac{S_{xx}\,\boldsymbol{d} - S_x\,\boldsymbol{c}}{\Delta}. \quad (13\,\text{a, b})$$

The covariance matrix of parameters a, b follows from the matrix equation for the variance propagation, i.e.:

$$\boldsymbol{U}_{ab} = \boldsymbol{AB}\,\boldsymbol{U}_{\boldsymbol{Y}}\,\boldsymbol{AB}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{A}^{\mathrm{T}} & \boldsymbol{A}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}} \\ \boldsymbol{A}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}} & \boldsymbol{B}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}} \end{bmatrix}, \quad (14)$$

where ρ_{ab} is the correlation coefficient between parameters a and b.

As the matrix $\boldsymbol{U}_{\boldsymbol{Y}}^{-1}$ is symmetrical, $\boldsymbol{U}_{\boldsymbol{Y}} = \boldsymbol{U}_{\boldsymbol{Y}}^{\mathrm{T}}$, and $\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{U}_{\boldsymbol{Y}}^{-1} = \boldsymbol{U}_{\boldsymbol{Y}}^{-1}\boldsymbol{U}_{\boldsymbol{Y}} = \mathbf{I}$ – the unit matrix, then:

$$u_a^2 = \boldsymbol{A}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{A}^{\mathrm{T}} = \frac{S}{\Delta},$$

$$u_b^2 = \boldsymbol{B}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}} = \frac{S_{xx}}{\Delta},$$

$$\rho_{ab}u_a u_b = \boldsymbol{A}\boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{B}^{\mathrm{T}} = -\frac{S_x}{\Delta}. \quad (14\,\text{a–c})$$

The standard uncertainty of the regression line $y = ax + b$ for one-dimensional variable $\boldsymbol{Y}$ can be obtained from the formula:

$$u_{ab}^2(y(x)) = \left[\frac{\partial y}{\partial a}, \frac{\partial y}{\partial b}\right]\boldsymbol{U}_{\boldsymbol{ab}}\left[\frac{\partial y}{\partial a}, \frac{\partial y}{\partial b}\right]^{\mathbf{T}} = \quad (15)$$

$$= [x, 1]\begin{bmatrix} u_a^2 & \rho_{ab}u_a u_b \\ \rho_{ab}u_a u_b & u_b^2 \end{bmatrix}\begin{bmatrix} x \\ 1 \end{bmatrix}.$$

and after both multiplications:

$$u_{ab}^2(y(x)) = u_a^2 x^2 + 2x\rho_{ab}u_a u_b + u_b^2 = \frac{1}{S} + \frac{S}{\Delta}\left(x - \frac{S_x}{S}\right)^2. \quad (15\text{a})$$

This equation determines type A uncertainty of the variable $\boldsymbol{Y}$ as a function of two correlated quantities ax and b. In the GUM [1], the function of combined uncertainty $u(y)$ is the geometric sum of the functions of type A and B uncertainties:

$$u(y) = \sqrt{u_{ab}^2(y) + u_B^2(y)}. \quad (16)$$

Formula (16) is used, where the y_i-coordinates of all points are measured using the same instrument and under the same conditions. In these cases, it is not necessary to separately consider type B uncertainties for y_i of each measured point, and $u_{regA}(y) = u_{ab}(y)$. The function of type B uncertainty, i.e. $u_B(y)$ in the y-range of the regression line, shall be determined from the Maximum Permissible Error, $E_{\max\,\%}$, of the meter. It is the standard deviation of uniformly distributed errors with a width of $\pm E$, which is

$$u_B(y) = u_B(y_0) + (y - y_0)\,E_{\max}/\sqrt{3}. \quad (16\,\text{a})$$

In most cases of measurement practice $u_B(y_0) \ll u_B(y - y_0)$.

The parameters a, b of a regression line depend on the y_i-coordinates of n measured points. The effective number of the degrees of freedom is n-2. The expanded uncertainty U_P of a regression line, e.g. when the confidence level is about $P = 0.95$, is determined multiplying the combined uncertainty $u(y)$ by the coverage factor $k_{0.95}$. For few n of points, the Student's t-distribution is used, and

$$U_{0.95} = k_{0.95}u(y) = t_{0.95,n-2}\sqrt{u_{ab}^2(y) + u_{By}^2}. \quad (17)$$

In rare cases, the y_i-coordinates and some or all controlled points are measured with different measures and/or under different influenced conditions. Then the sequence of numerical operations in determining the expanded uncertainty is

$$u_i = \sqrt{u_{Ai}^2 + u_{Bi}^2},\ \ i = 1, \ldots, n, \quad (17\text{a})$$

$$u_{ab} = u_{ab}(ax + b), \quad (17\text{b})$$

$$U_y(ax + b) = t_{0.95,n-2}u_{ab}(ax + b). \quad (17\text{c})$$

Therefore, equation (1) of the regression line with its expanded uncertainty band finally is:

$$y = ax + b \pm k_{0.95}\sqrt{u_{ab}^2(y) + u_b^2(ax + b)}. \quad (18)$$

The regression straight-line parameters for the correlated values of the variable $\boldsymbol{Y}$ and their special cases for the data of the uncorrelated values of the variable $\boldsymbol{Y}$ are summarized together as Table 1 in [9].

2. Kinds of correlation for measured ordinates of measured points

2.1. Points with uncorrelated ordinates and equal uncertainties

For the same value of absolute uncertainties $u_i = u$ of y_i of the measured points, the covariance matrix is:

$$\boldsymbol{U}_{\boldsymbol{Y}} = u^2 \, \boldsymbol{R}_{\boldsymbol{Y}}, \tag{19}$$

where the correlator $\boldsymbol{R}_{\boldsymbol{Y}}$ is a square symmetrical matrix size $n \times n$, elements of which are the correlation coefficients ρ_{yij} between the values y_i and y_j for $i, j = 1, \ldots, n$.

In this case, for given correlation coefficients, the equation for a regression line will always be the same, since the values of the parameters a i b from formulas (9a,b) do not depend on the uncertainty u, but only on the correlator matrix. The regression line for the points with given coordinates and the same relative uncertainties y_i of all points gets other parameters than for the same absolute uncertainties. For all unknown but equal absolute uncertainties u, the estimator of the variance $\hat{u}^2$ is defined as a minimum of the distance squares of divided by n-2 degrees of freedom:

$$\widehat{u^2} = u^2 \frac{-aS_{xy} - bS_y - S_{yy}}{n-2} = \\ = \frac{-a\,\boldsymbol{X}^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{Y}}^{-1} \boldsymbol{Y} - b\,\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{Y}}^{-1} \boldsymbol{1} + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{R}_{\boldsymbol{Y}}^{-1} \boldsymbol{Y}}{n-2}. \tag{20}$$

2.2. Correlation in a series of repeated measurements of y_i of single points with the same x_i

In a series of m repeated measurements y_{i1}, ..., y_{ik}, ... y_{im} ($k = 1, \ldots m$) of the y_i-ordinate of a single point with abscissa x_i, the autocorrelation may occur. With the same values of uncertainties $\sigma_{ik} = \sigma = \text{const}$ for each of the obtained y_{ik}-ordinate values and their correlated results with the correlation coefficients, from ρ_{i1} for the adjacent points to $\rho(i, m\text{-}1)$ for the farthest points, the local covariance matrix size $m \times m$ is obtained, given in equation (21)

$$\boldsymbol{U}_{yi} = \sigma^2 \begin{bmatrix} 1 & \rho_{i1} & \rho_{i2} & \cdots & \cdots & \rho_{im-1} \\ \rho_{i1} & 1 & \rho_{i1} & \rho_{i2} & \cdots & \cdots \\ \rho_{i2} & \rho_{i1} & 1 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \rho_{i2} \\ \cdots & \cdots & \cdots & \cdots & 1 & \rho_{i1} \\ \rho_{im-1} & \cdots & \cdots & \rho_{i2} & \rho_{i1} & 1 \end{bmatrix}. \tag{21}$$

In the covariance matrix $\boldsymbol{U}_{yi}$, the greatest positive correlation occurs for the terms close to the mean y_i-value. Real type A uncertainty u_{yi} of this point is greater than from the classical formula for m repeated measurements of y_i.

In estimating the uncertainty u_{Ai} of the mean value of the results of y_{im} of the point with abscissa x_i, instead of the number of measurements m, the effective number $n_{ieff} < m$ must be used described by the following formula [5, 6]:

$$n_{ieff} = \frac{m}{1 + 2\sum_{k=1}^{m-1} \frac{m-k}{m} \rho_{ik}}, \tag{22}$$

where m is the number of autocorrelated measurements, ρ_{ik} is the value of the autocorrelation function for the measurements of y_{ik}.

If the results of repeated measurements of the tested point Q_i are correlated with the coefficient $\rho_{ik} = \rho$, then the effective number of measurements of this series is:

$$n_{ieff} = \frac{m}{1 + \rho(m-1)}. \tag{22a}$$

Table 1 shows an example of the effective uncertainties u_{yi} for m-fold measurements of the y_i-ordinate of the tested point with abscissa x_i and for three values of the correlation coefficient ρ between the measurements.

In each of the three series, $m = 20$ measurements with the same uncertainty σ_i were performed.

Table 1

Effective uncertainties of a measurement series of y_i with three correlation coefficients

Correlation coefficient	ρ	ρ=0	ρ=0.5	ρ=1
Effective number of measurements	n_{ieff}	$n_{ieff} = m = 20$	$n_{ieff} = \frac{2m}{m+1} = 1.9$	$n_{ieff} = 1$
Effective uncertainties of a series of m measurements of y_i	$\frac{\sigma_i}{\sqrt{n_{ieff}}}$	$0.22\sigma_i$	$0.72\sigma_i$	σ_i

Table 1 shows that with the increase in the correlation coefficient ρ, the standard uncertainty of the series increases because the effective number of measurements decreases from 20 to 1.9, and 1 for the correlation coefficients of 0, 0.5 and 1, respectively. The effective uncertainty for uncorrelated observations will be about 5 times less than when they are fully correlated.

2.3. Correlation between ordinates of various measured points

Let us determine the covariance matrix describing measurements of the variable ***Y*** points with correlated ordinates. For example if y_i, y_j of two points $Q(x_i, y_i)$, $Q(x_j, y_j)$ are measured with the same uncertainties $u_i = u_j = u = \text{const}$, the autocorrelation function can also be used. In matrix (21), the uncertainty σ of separate measurements in a series is replaced by the effective uncertainty u of the measured points. Autocorrelation is a symmetrical function rapidly decreasing from the maximum value to zero, and its initial arms can be approximated with a straight line [5]. Usually, it is enough to consider only the neighbouring values of the measured ordinates y_{i-1}, y_i, y_{i+1} of the points with a known abscissa x_{i-1}, x_i, x_{i+-1}, i.e. $\rho_1 = \rho$ and $\rho_k = 0$ for $i = 2, ..., n\text{-}1$. For n equally remote points and of the same uncertainties u and correlation coefficients ρ, the covariance matrix $\boldsymbol{U}_Y$ (23) with its size $n \times n$ gets all correlation coefficients $\rho_k = \rho$ for $i = 1, ..., n\text{-}1$. When the autocorrelation influence is negligible, the matrix $\boldsymbol{U}_Y$ reduces to the main diagonal:

$$\boldsymbol{U}_Y = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \dots & \rho_k & \dots & \rho_{n-1} \\ \rho_1 & 1 & \dots & \dots & \dots & \dots \\ \dots & \dots & 1 & \dots & \dots & \rho_k \\ \rho_k & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 1 & \rho_1 \\ \rho_{n-1} & \dots & \rho_k & \dots & \rho_1 & 1 \end{bmatrix}.$$

The main equations for the regression straight-line parameters and its uncertainty band for the correlated values of the variable ***Y*** are given together in the left column of Table 2, and special cases for the uncorrelated data of the variable ***Y*** are summarized in the right column.

3. Numerical examples

Simulated numerical examples and their relations will be provided.

3.1. Regressions line and uncertainty bands for the same value of absolute or of relative uncertainty

Table 3 shows the coordinates of 10 measured points with the same uncertainty values.

The values y_i are the averages of multiple measurements of the y_i-ordinates of each point with abscissa x_i. To identify the effects of individual factors, simplified cases will be analysed. It was assumed that absolute uncertainties u_i or relative uncertainties δ_{ui} of the y-ordinate values for all points are the same and their abscissas x_i are constant and known exactly. To compare the results with the results obtained for uncorrelated y_i-ordinates, the same data of points were

Table 2

Parameters of regression line for autocorrelated and uncorrelated variable ***Y***

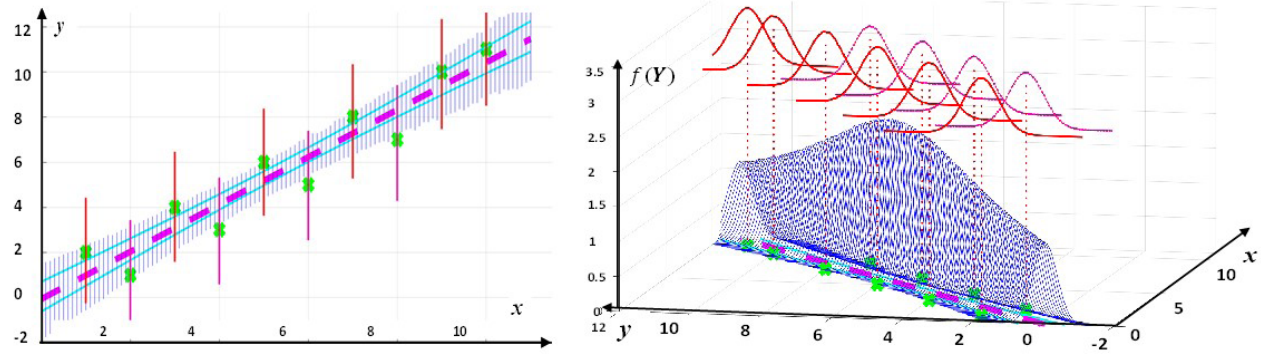
Parameter*	**measured:** $\boldsymbol{Y} = [y_1,...,y_i,...,y_n]^{\mathrm{T}}$ with uncertainty $u_i(y_i) \neq 0$; **known:** $\boldsymbol{X} = [x_1,...,x_i,...,x_n]^{\mathrm{T}}$, $u_i(x_i) = 0$	
Variable *Y*	**Correlated**	**Uncorrelated**
Covariance matrix $\boldsymbol{U}_Y$ and its inverse	$\boldsymbol{U}_Y = \begin{bmatrix} u_1^2 & \dots & \rho_{y1n} u_1 u_n \\ \vdots & \ddots & \vdots \\ \rho_{y1n} u_1 u_n & \dots & u_n^2 \end{bmatrix}$, $\boldsymbol{U}_Y^{-1}$	$\boldsymbol{U}_Y = \begin{bmatrix} u_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & u_n^2 \end{bmatrix}$, $\boldsymbol{U}_Y^{-1} = \begin{bmatrix} u_1^{-2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & u_n^{-2} \end{bmatrix}$ $\boldsymbol{U}_Y, \boldsymbol{U}_Y^{-1}$ only diagonal elements
Criterion	$SK = [\boldsymbol{Y} - a\boldsymbol{X} - \boldsymbol{b}]\boldsymbol{U}_Y^{-1}[\boldsymbol{Y} - a\boldsymbol{X} - \boldsymbol{b}] \to \min$	$SK = \sum_{i=1}^{n} (y_i - ax_i - b)^2 u_i^{-2} \to \min$
Regression line parameters	Slope $a = \frac{SS_{xy} - S_x S_y}{SS_{xx} - (S_x)^2} = \frac{\Delta_a}{\Delta}$, abscissa $b = \frac{S_y S_{xx} - S_x S_{xy}}{SS_{xx} - (S_x)^2} = \frac{\Delta_b}{\Delta}$, where $\Delta = SS_{xx} - (S_x)^2$, $\Delta_a = SS_{xy} - S_x S_y$, $\Delta_b = S_y S_{xx} - S_x S_{xy}$.	

Auxiliary parameters	$\Delta = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(v_i z_j - v_j z_i)^2 \lambda_i^{-2}\lambda_j^{-2}$ *1		$\Delta = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2 u_i^{-2} u_j^{-2}$
	S	$[1,...,1]\boldsymbol{U}_Y^{-1}[1,...,1]^{\mathrm{T}} = \sum_{i=1}^{n}\sum_{j=1}^{n}[u_y^{-1}]_{ij}$	$\sum_{i=1}^{n} u_i^{-2}$ or $\sum_{i=1}^{n}\delta_i^{-2} y_i^{-2}$
	S_x	$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}[1,...,1]^{\mathrm{T}} = [1,...,1]\boldsymbol{U}_Y^{-1}\boldsymbol{X}$	$\sum_{i=1}^{n} x_i u_i^{-2}$ $\sum_{i=1}^{n}\delta_i^{-2} x_i y_i^{-2}$
	S_{xx}	$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{X}$	$\sum_{i=1}^{n} x_i^2 u_i^{-2}$ $\sum_{i=1}^{n}\delta_i^{-2} x_i^2 {}_i y_i^{-2}$
	S_y	$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}[1,...,1]^{\mathrm{T}} = [1,...,1]\boldsymbol{U}_Y^{-1}\boldsymbol{Y}$	$\sum_{i=1}^{n} y_i u_i^{-2}$ $\sum_{i=1}^{n}\delta_i^{-2} y_i^{-1}$
	S_{xy}	$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{Y} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{X}$	$\sum_{i=1}^{n} x_i y_i u_i^{-2}$ $\sum_{i=1}^{n}\delta_i^{-2} x_i y_i^{-1}$
	S_{yy}	$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{U}_Y^{-1}\boldsymbol{Y}$	$\sum_{i=1}^{n} y_i^2 u_i^{-2}$ $\sum_{i=1}^{n}\delta_i^{-2}$ where: $u_i = \delta_i y_i$
Sensitivity Matrix *AB*	$\boldsymbol{AB} = [\boldsymbol{A}, \boldsymbol{B}]$, $\boldsymbol{A} = \frac{\partial a}{\partial \boldsymbol{Y}} = \frac{S\boldsymbol{U}_Y^{-1}\boldsymbol{X} - S_x\boldsymbol{U}_Y^{-1}[1,...,1]^{\mathrm{T}}}{\Delta}$, $\boldsymbol{B} = \frac{\partial b}{\partial \boldsymbol{Y}} = \frac{S_{xx}\boldsymbol{U}_Y^{-1}[1,...,1]^{\mathrm{T}} - S_x\boldsymbol{U}_Y^{-1}\boldsymbol{X}}{\Delta}$		$[\boldsymbol{AB}^{\mathrm{T}}]_i = [a_i, b_i]$ $a_i = \frac{\partial a}{\partial y_i} = \frac{1}{u_i^2}\frac{Sx_i - S_x}{\Delta}$, $b_i = \frac{\partial b}{\partial y_i} = \frac{1}{u_i^2}\frac{S_{xx} - S_x x_i}{\Delta}$.
Covariance matrix U_{ab}	$\boldsymbol{U}_{ab} = \begin{bmatrix} u_a^2 & \rho_{ab}u_a u_b \\ \rho_{ab}u_a u_b & u_b^2 \end{bmatrix} = \boldsymbol{AB}\boldsymbol{U}_Y\boldsymbol{AB}^{\mathrm{T}}$, $u_a^2 = \frac{S}{\Delta}$, $u_b^2 = \frac{S_{xx}}{\Delta}$, $\rho_{ab}u_a u_b = -\frac{S_x}{\Delta}$.		
Type A uncertainties of regression line $y=ax+b$	Standard	$u_{ab}^2 = u_a^2 x^2 + 2x\rho_{ab}u_a u_\partial + u_b^2$, $x_{min} = -\frac{\rho_{ab}u_b}{u_a}$, $u_{ab\,\mathbf{min}}^2 = (1-\rho_{ab}^2)u_b^2$ or $u_{ab}^2 = \frac{1}{S} + \frac{S}{\Delta}\left(x - \frac{S_x}{S}\right)^2$, $x_{min} = \frac{S_x}{S}$, $u_{ab\,min}^2 = \frac{1}{S}$.	
	Expanded	$U_{ab} = t_{0.95,n-2}u_{ab}$	
The same uncertainty $u_i = u$ of y_i of all points	Matrix $\boldsymbol{U}_Y$ is replaced by correlator $\boldsymbol{R}$ $\boldsymbol{U}_Y = u^2\boldsymbol{R}_Y$ and $\boldsymbol{U}_Y^{-1} = u^{-2}\boldsymbol{R}_Y^{-1}$		$u_a^2 = \frac{u^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, $u_b^2 = \frac{1}{n} + \frac{u^2\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $\rho_{ab}u_a u_b = -\frac{u^2\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $u_{ab}^2(x) = u^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$
Variance estimator	$\widehat{u^2} = \frac{-a\boldsymbol{X}^{\mathrm{T}}\boldsymbol{R}_Y^{-1}\boldsymbol{Y} - b\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{R}_Y^{-1}[1,...,1]^{\mathrm{T}} + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{R}_Y^{-1}\boldsymbol{Y}}{n-2}$		$\widehat{u^2} = \frac{-a\sum_{i=1}^{n} x_i y_i - b\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} y_i^2}{n-2}$
Type B uncertainty	$u_B(y) = u_B(y_0) + (y - y_0)E_{\max}/\sqrt{3}$		

Table 3

Coordinates of measured points and equal absolute and relative uncertainties of y_i

Symbol	Coordinates of points (x_i, y_i) and uncertainties of y_i									
x_i	1	2	3	4	5	6	7	8	9	10
y_i	2	1	4	3	6	5	8	7	10	11
$u(y_i)$	0.5									
$\delta(y_i)$	5%									
$u_B(y_i)$	0.01(1+y)%									
Regression line	for u_i=u: y=1.48x – 0.067					for δ_{ui}=δ: y = 0.978x – 0.463				



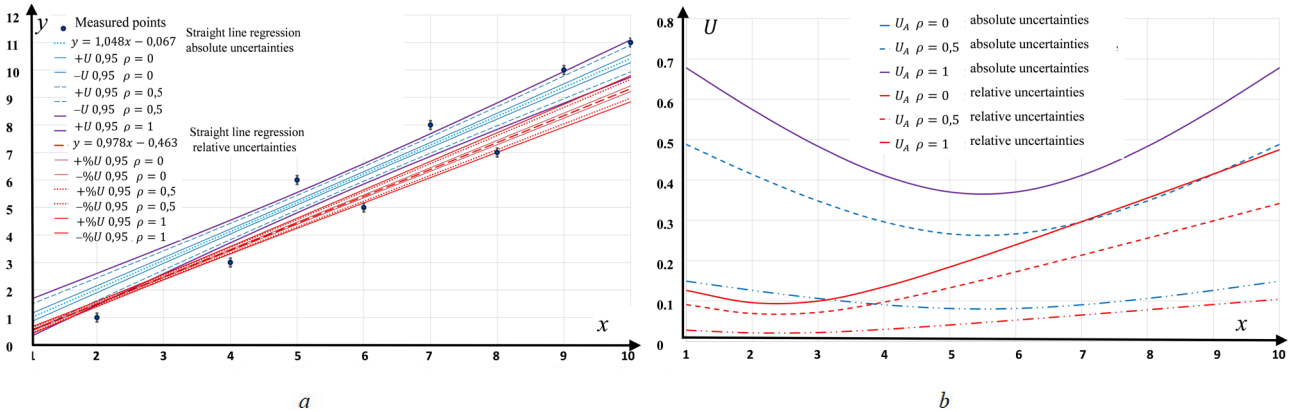
a *b*

Fig. 2. *a* − Graphs of regression line with band of its expanded type A uncertainty; *b* − Gaussian probability density distributions (pdf-s) with uncertainty u = 0.5 equal for all measured points and pdf-s $f(y)$ of y-uncertainties along regression line in 3D view

used as for Experiment I, which is given in our papers [7−9]. All that is illustrated by Fig. 2*a*, *b*.

Fig. 2*a* shows the line and its uncertainty band (blue lines) determined by the linear regression method for the data from Table 2 when the correlation is absent. It also shows the uncertainty of the ordinates of the measured points (red lines). The figure shows that 95% of the expanded uncertainty region changes along the regression line such that its width decreases from the initial value to a minimum near the centre of the regression line, and then widens back to the original initial size.

At the top in Fig. 2*b*, there are the graphs of Gaussian probability density distributions (pdf – red curves) for the ordinates of the measured points located on both sides of the regression line. The pdf of the y_i-ordinates are for 95% coverage with standard deviation u = 0.5. Below, the probability density curves for y around the regression line inside its uncertainty band (blue 3D solid) are shown. In the middle of the range of the regression line, pdf is the narrowest and therefore is the highest. The maximum (narrowest) pdf occurs for the minimum uncertainty bandwidth in the middle part of the straight line. When moving away from this maximum in the opposite directions, the maximum value of the pdf function decreases, and the distribution widens as the uncertainty band boundaries move away from each other.



a *b*

Fig. 3. *a* − Regression lines with expanded type A uncertainty bands; *b* − half height of bands of uncertainty U_A for three values of correlation coefficient ρ for series of 10 points from Table 2 measured 20 times each

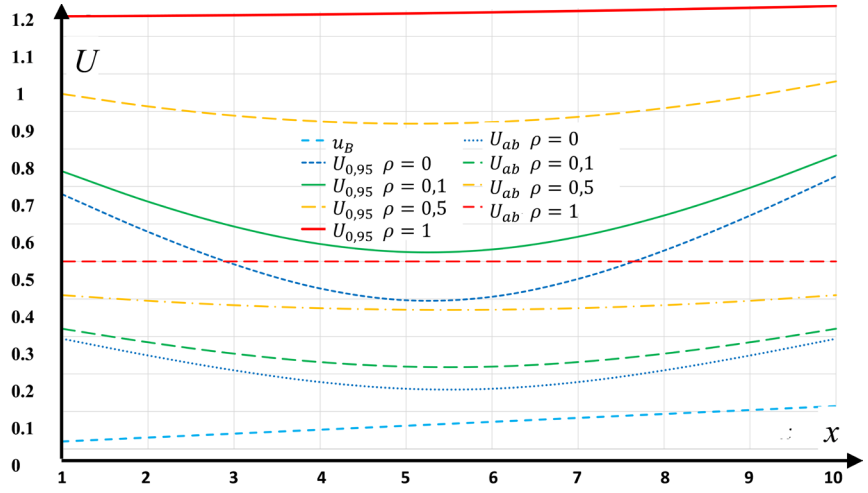


Fig. 4. Uncertainty bands for regression line $y = 1.048x - 0.067$ for correlated and uncorrelated measurement points for three covariance matrixes $\boldsymbol{U}_Y$ (21) for u= 0.5, $\rho_k = \rho$ for i=1, …, 9, ρ = 0, ρ = 0.1, ρ = 0.5 and ρ = 1 with type B uncertainty

Fig. 3*a* shows two regression lines with their extended type A uncertainty bands $\pm U_{0.95}$ calculated for the ordinates (x_i, y_i) of 10 points from Table 3 measured 20 times each with equal absolute uncertainties $u_i = u$ (blue line $y = 1.048x - 0.067$ runs above) or with equal relative uncertainties $\delta_{ui} = \delta_u$ (red line $y = 0.978x - 0.463$ running below) and for three different correlation coefficients for y_i.

Fig. 3*b* shows functions of their $U_{0.95}$. With the increase in the correlation coefficient between the results of repeated measurement observations, the expanded uncertainty U_A increases [5, 6].

In section 2.2, it was shown that the effective number of observations of a sample with 20 measurements for the autocorrelation coefficients 0, 0.5 and 1 decreased from 20 to 1.9 and to 1. The narrowest uncertainty band is obtained for non-correlated observations, i.e. for ρ = 0. If such measurements are made for the y_i-ordinates of all 20 measurements with the same uncertainty, then for the correlation coefficient ρ = 1, the uncertainty band of the regression line is over four times wider than when the correlation is absent. Negative coefficients ρ do not occur in these measurements.

3.2. Influence of the correlation coefficient value between y_i-ordinates on the uncertainty band

Fig. 4 compares two regression lines and uncertainty bands for the covariance matrix $\boldsymbol{U}_Y$ (21) with parameters $u_i = 0.5$, ρ = (0.6; 0.5; 0.4; 0.2; 0.1) and ρ = 0 for others. The regression line for correlated points is $y = 0.981x + 0.417$.

There is a significant increase (over 30%) in the width of an uncertainty band, when the points described by the covariance matrix are autocorrelated. On the other hand, type B uncertainty has a greater impact on the width of a band determined without correlation than on a band determined with correlation because with the data of Experiment I considered here, an increase in uncertainty by 1% is negligible.

The regression lines and their absolute uncertainty bands for full correlation (ρ = 0.99≈1) of the measured y_i-ordinates of the points with the covariance matrix $\boldsymbol{U}_Y = u^2$ [1] and when the correlation is absent (ρ = 0) are given in Fig. 5.

It is shown that with full correlation, i.e. for ρ = 1, the hyperbolic boundaries of the uncertainty band become straight lines parallel to the regression line. The band constriction in the middle of the regression line

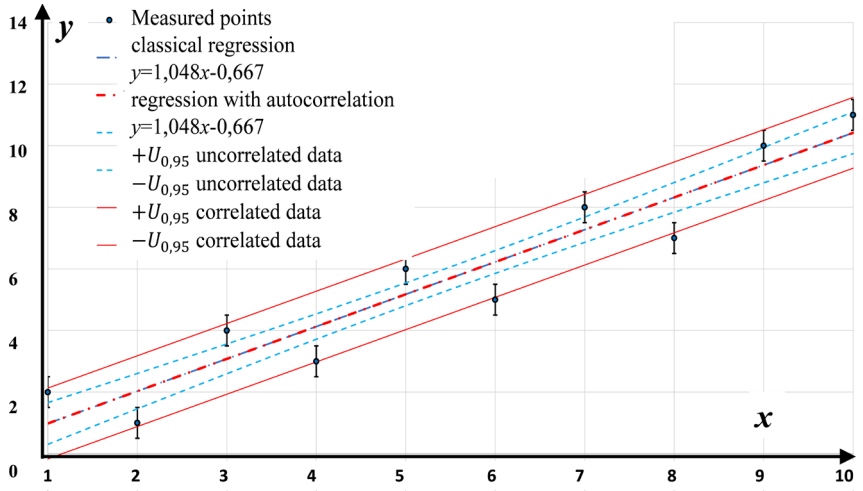


Fig. 5. Regression lines and uncertainty bands for measurement points with uncertainties marked at uncorrelated and correlated points with ρ ≈ 1 for covariance matrix $\boldsymbol{U}_Y$

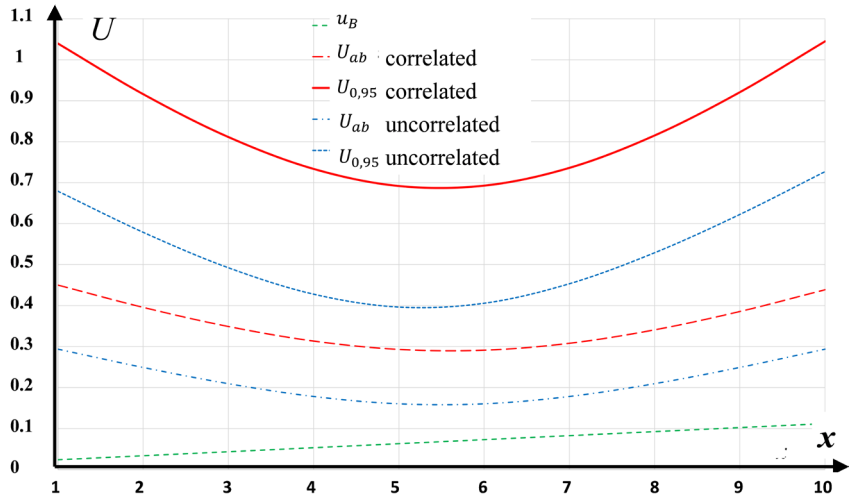


Fig. 6. Uncertainties for uncorrelated and correlated y_i of points with covariance matrix $\boldsymbol{U}_Y$ (21)

range disappear. The parameters of this regression line do not change with the increases of equal coefficients in the covariance matrix.

The graphs in Fig. 6 concern the courses of uncertainty U_{ab} (type A) and uncertainty u_B for the regression line obtained for the same values of the correlation coefficients between the y_{ik}-coordinates of each of the repeatedly measured points of x_i = const, i.e. according to their covariance matrix (21). With the presence of positive correlation, an increase in the width of the uncertainty band is observed.

3.3. Comparison of regression lines for constant absolute or relative uncertainties of measured y_i

In Fig. 7*a*, *b* the regression lines with uncertainty bands for the correlated and uncorrelated y_i-ordinates, measured with relative uncertainty $\delta = 1\%$, were compared.

For small and the same values of correlation coefficients, such a straight line (Fig. 7*a*) deviates from their direction for the same ordinates without correlation. Its slope was reduced from 0.978 to 0.869. The boundaries of the uncertainty band slightly widen for absolute uncertainties, and the boundaries of the band for relative uncertainties remain practically unchanged. The linearly increasing type B uncertainty u_B causes a disturbance of the symmetry of the limits of the expanded uncertainty $U_{0.95}$ band in relation to the centre of the regression line and a shift towards higher values of the independent variable x. As a result, this uncertainty bandwidth increases with x, for both correlated and uncorrelated measuring points.

3.4. Correlation of the variable *Y* values within the uncertainty band

Apart from the correlation between the y_i-ordinates of the measured points $Q_i(x_i, y_i)$, the y_i-values of points $P_i(x_i, y_i)$ on the regression line $y = ax + b$ (see Fig. 1) are also correlated with each other. This is because the covariance matrix of its two points $y_i = ax_i + b$ (for $i = 1, 2$) is determined by the variance propagation equation:

$$U_y(x) = \begin{bmatrix} u_{y_1}^2 & \rho_{y_1y_2}u_{y_1}u_{y_2} \\ \rho_{y_1y_2}u_{y_1}u_{y_2} & u_{y_2}^2 \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial a} & \frac{\partial y_1}{\partial b} \\ \frac{\partial y_2}{\partial a} & \frac{\partial y_2}{\partial b} \end{bmatrix} \begin{bmatrix} u_a^2 & \rho_{ab}u_au_{yb} \\ \rho_{y_1y_2}u_{y_1}u_{y_2} & u_b^2 \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial a} & \frac{\partial y_2}{\partial a} \\ \frac{\partial y_1}{\partial b} & \frac{\partial y_2}{\partial b} \end{bmatrix} =$$

$$= \begin{bmatrix} u_a^2x_1^2 + 2x_1\rho_{ab}u_au_b + u_b^2 & u_a^2x_1x_2 + (x_1 + x_2)\rho_{ab}u_au_b + u_b^2 \\ u_a^2x_1x_2 + (x_1 + x_2)\rho_{ab}u_au_b + u_b^2 & u_a^2x_1^2 + 2x_1\rho_{ab}u_au_b + u_b^2 \end{bmatrix}, \quad (23)$$

where $\frac{\partial y_i}{\partial a} = x_i$, $\frac{\partial y_i}{\partial b} = 1$ and $i = 1.2$.

For two points with ordinates y_1, y_2, variances $u_{y_1}^2$, $u_{y_2}^2$ and correlation coefficient $\rho_{y_1y_2}$ are

$$u_{y_1}^2 = u_a^2x_1^2 + 2x_1\rho_{ab}u_au_b + u_b^2,$$
$$u_{y_2}^2 = u_a^2x_2^2 + 2x_2\rho_{ab}u_au_b + u_b^2 \quad (23a,b)$$

$$\rho_{y_1y_2} = \frac{u_a^2x_1x_2 + (x_1 + x_2)\rho_{ab}u_au_b + u_b^2}{u_{y_1}u_{y_2}}. \quad (24)$$

If the regression line describes the characteristic of the tested instrument, measurement device or process, and under variable influencing conditions, the unknown part of systematic errors cannot be eliminated, then type B uncertainty u_B should also be evaluated. It is assumed that the values of u_B result from maximum permissible errors E_{max} of the applied digital meters as a linear function $u_B(y) = u_{B0} + \delta_B(y - y_0)$ [1], or $u_B(x) = u_{B0} + \delta_B(b - y_0) + \delta_B ax$.

The combined uncertainty u_c of the regression line is larger than type A uncertainty u_{ab} and is described by the formula

$$u_C^2(x) = u_{ab}^2 + u_B^2 = (u_a^2 + a^2\delta_B^2)x^2 + 2(\rho_{ab}u_au_b + a\delta_B(u_{B0} + \delta_B(b - y_0)))x + u_b^2 + (u_{B0} + \delta_B(b - y_0))^2. \quad (25)$$

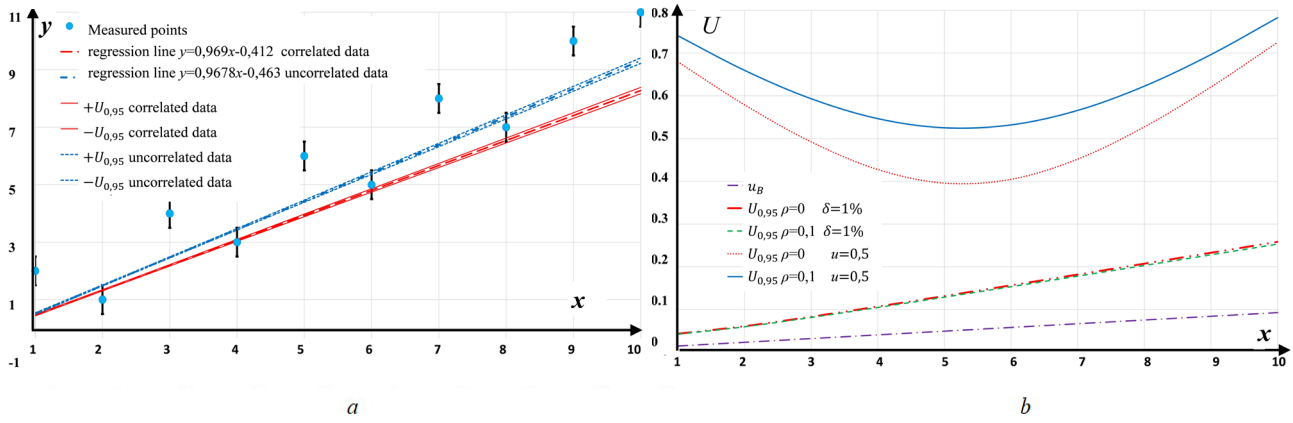


Fig. 7. Cases of equal relative uncertainty δ=1% for uncorrelated and correlated y_i of points for covariance matrix $\boldsymbol{U}_Y$ (21) with δ=1%, ρ_k=0.1 for i = 1, …, 9: a − regression lines; b − uncertainty $U_{0.95}$ together with type B uncertainty

Then with the influence of type B uncertainties, the effective values of variances u_{aC}^2, u_{bC}^2 and the correlation coefficient ρ_{abC}, will be respectively:

$$u_{aC}^2 = u_a^2 + a^2\delta_B^2 \quad u_{bC}^2 = u_b^2 + (u_{B0} + \delta_B(b - y_0))^2 \qquad (26a,b)$$

$$\rho_{abC} = \frac{\rho_{ab}u_a u_b + a\delta_B(u_{B0} + \delta_B(b - y_0))}{u_{aC}u_{bC}}. \qquad (26c)$$

Variances of the combined uncertainty in points x_i for i=1.2 ($u_{B0} = \delta_B = \delta$ for $y_0 = 0$) are:

$$u_{yCi}^2 = (u_a^2 + a^2\delta^2)x_i^2 + 2x_i(\rho_{ab}u_a u_b + a\delta^2(1+b)) + \\ + u_b^2 + \delta^2(1+b)^2. \qquad (27)$$

Fig. 8 shows absolute U and relative uncertainty bands $\delta_U = U/y$ of the regression line. They are obtained from (27) for absolute u=0.5 and relative 1% uncertainties, and the correlation coefficient ρ_k = =0.3 for i =1, ... , 9 with a linearly increasing type B uncertainty $u_{B0} = \delta_B = \delta$ for y_0=0.

In the uncertainty band narrowing area, the uncertainty for the correlated y_i-values increases almost twice as compared to the uncorrelated y_i with the same absolute uncertainties, and the influence of type B uncertainty makes the uncertainty bands slightly asymmetric. The resulting relative uncertainties decrease from several dozen percent at the beginning of the regression line to less than 10% at the end of its range. For relative uncertainties, the bands of correlated and uncorrelated variables almost coincide. They are slightly greater for the values of uncorrelated variables than for correlated variables. The initial relative values linearly increase to the level below 3, while relative values do not exceed the level of 2−3%. The influence of a linearly increasing type B uncertainty is practically imperceptible.

The correlation coefficient $\rho_{Cy_1y_2}$ between the variables y_1 and y_2 from (24) of the regression line, considering the influence of type B uncertainty (27), has the form

$$\rho_{Cy_1y_2} = \frac{(u_a^2 + a^2\delta_B^2)x_1x_2 + (x_1 + x_2)(\rho_{ab}u_a u_b + a\delta_B(u_{B0} + \delta_B(b - y_0)) + u_b^2 + (u_{B0} + \delta_B(b - y_0))^2}{u_{yC_1}u_{yC_2}}. \qquad (28)$$

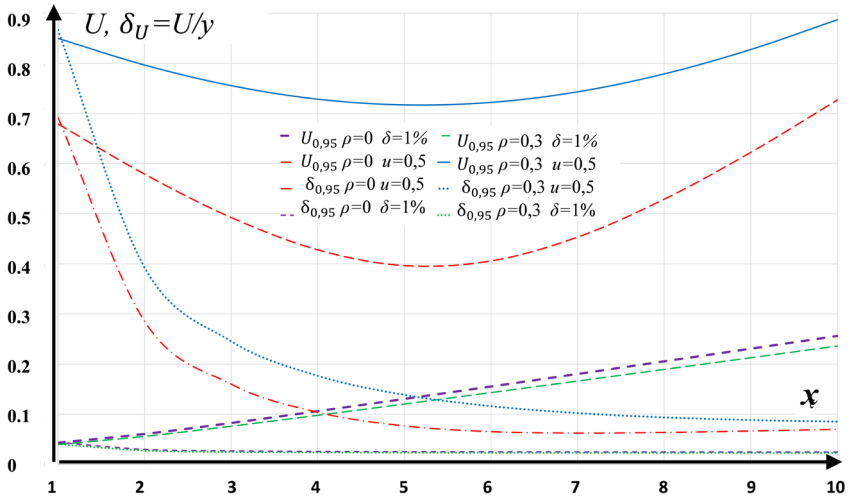


Fig. 8. Absolute U and relative uncertainty bands $\delta_U = U/y$ of the regression line

Due to the linear dependency of y from x, these variables are fully correlated, which means that the same correlation coefficient as (28) have both variables x_1 and x_2.

4. Conclusions

The type A uncertainty bands of the regression line determined from measurements, similarly as for uncorrelated variables, run hyperbolically, and their cut-offs are symmetrical as to this line.

For the same value of relative uncertainties, the limits of the uncertainty band of the regression line are split hyperbolas with tangents diagonal to the line. In the middle part of the band, there is a constriction of width S depending on the parameters of the covariance matrix. With full correlation, i.e. for $\rho = 1$, the hyperbolic boundaries of the uncertainty band become straight lines parallel to the regression line.

With the increase in the correlation coefficient between the ordinates of the tested points, the slope of the regression line decreases significantly. The ordinates of points within the uncertainty band are correlated more strongly if they are closer, and their correlation coefficient tends to 1. This is applied to the entire range of y.

The accuracy of the regression line parameters improves with repeated measurements of the y-ordinate value of each point with the same abscissa x_i. The results of observations in such a series are close to each other, and it is necessary to consider the impact of their autocorrelation. For a given point, the local covariance matrix, or the formula for the effective number of measurements can be used – see works of Warsza and Zieba [4, 5].

Type B uncertainty is evaluated heuristically from the maximum permissible error E (MPB) of the meter. For the uniform distribution of errors with a range $2E$, the uncertainty u_B is the standard deviation $E/\sqrt{3}$. Usually, the error E linearly depends on $(y - y_0)$ inside the measured range $y_{\max} - y_0$ (e g. for digital instruments).

The u_B uncertainty is summed up geometrically with type A uncertainty of the regression line obtained from measurements. The resultant combined and expanded uncertainty bands widen with the increase of y and x.

Оцінка смуг невизначеності лінії регресії для корельованих даних змінної *Y* з використанням правил GUM

З.Л. Варша [1], Я. Пухальський [2]

[1] Промисловий науково-дослідний інститут автоматизації та вимірювань PIAP-Ł (Дослідницька мережа Łukasiewicz), Aleje Jerozolimskie 202, 02-486, Варшава, Польща
zlw1936@gmail.com
[2] Центральний офіс вимірювань (GUM), Elektoralna 2, 00-137, Варшава, Польща
jacek.puchalski@gum.gov.pl

Анотація

Розглянуто оцінювання невизначеності вимірювань при побудові калібрувальної залежності засобів вимірювання за допомогою вимірювального експерименту, у ході якого вимірюються значення змінної ***Y*** спільно з вимірюванням абсциси ***X***. Стаття обмежується дослідженням найпростішого випадку – лінійної регресії, але з урахуванням даних невизначеності вимірювань типу А і типу В та з урахуванням кореляції й автокореляції між ними. Отримано вирази для сумарної стандартної та розширеної невизначеностей у кожній точці вимірювань. Залежність відхилень виміряних значень ординати ***Y*** від значень абсциси ***X*** являє собою досліджувану смугу невизначеності. Наведено основні співвідношення для випадку відсутності невизначеності вимірювання абсциси ***X***. Розглянуто випадок наявності точок із некорельованими ординатами та однаковими невизначеностями. Досліджено кореляцію в серії повторних вимірювань в одній точці ординати з одним і тим самим значенням абсциси. Наведено співвідношення для оцінювання невизначеності вимірювань за наявності кореляції між ординатами різних виміряних точок. Розглянуто змодельовані чисельні приклади. Розраховано лінію регресії та смуги невизначеності для однакового значення абсолютної або відносної невизначеності. Досліджено вплив значення коефіцієнта кореляції для ординат на діапазон невизначеності. Порівняно лінії регресії для постійних абсолютних або відносних невизначеностей виміряних значень ординати. Розглянуто приклад наявності кореляції значень змінної ***Y*** в діапазоні невизначеності.

Ключові слова: вимірювання; автокореляція; лінія регресії; смуги невизначеності типу А та типу В.

References

1. JCGM 100:2008 (GUM 1995 with minor corrections). Evaluation of measurement data – Guide to the expression of uncertainty in measurement. 134 p.
2. JCGM 101:2008. Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method. 82 p.
3. JCGM 102:2011. Evaluation of measurement data – Supplement 2 to the "Guide to the expression of uncertainty in measurement" – Extension to any number of output quantities. 72 p.
4. Warsza Z.L. Methods of Extension of the Measurement Uncertainty Analysis. Monograph. PIAP Warszawa, 2016 (in Polish).
5. Warsza Z.L., Puchalski J. Estimation of Uncertainties in Indirect Multiparameter Measurements of Correlated Quantities. *Proceedings of 12th International Conference on Measurement, 2019*. Slovak Academy of Sciences, Slovakia, 2019, pp. 51–57. doi:10.23919/MEASUREMENT47340.2019.8780042
6. Warsza Z.L., Puchalski J., Idzkowski A. Application of the vector method for estimating characteristic function based on measurements uncertainty at two control points. *Proceedings of International Conference AMSA, 2019*. Novosibirsk, Russia, 2019, pp. 60–73.
7. Warsza Z.L., Puchalski J. Estimation of uncertainty in indirect multivariable measurements. Part 1. Case of correlated quantities. in: R. Szewczyk et al. (Eds.). *Proceedings of Conference Automation 2020*, AISIC series, vol. 1140, pp. 309–325. Springer Nature Switzerland AG 2020. doi: https://doi.org/10.1007/978-3- 030-40971-5_29
8. Warsza Z.L., Puchalski J. Niepewności pomiarów w metodzie regresji liniowej. Część 1. Prosta i jej pasma niepewności dla nieskorelowanych danych pomiarowych. *PAR* (*Pomiary Automatyka Robotyka*), R. 24, Nr 3/2020, 79–91 (in Polish). doi: 10.14313/PAR_237/79
9. Warsza Z.L., Puchalski J. Niepewności pomiarów w metodzie regresji liniowej. Część 2. Niepewności prostej dla zmiennej *Y* o skorelowanych danych. *PAR* (*Pomiary Automatyka Robotyka*), Nr 4/2020, 61–72 (in Polish). doi: 10.14313/PAR_238/61
10. Warsza Z.L., Puchalski J. Uncertainty Bands of the Regression Line for Data with Type A and Type B Uncertainties of Dependent Variable *Y*. *Proceedings of International Conference Automation 2021*: Recent Achievements in Automation, Robotics and Measurement Techniques. AISIC series, vol. 1390, pp. 342–363. Springer Nature Switzerland AG 20210. doi: https://doi.org/10.1007/978-3- 030-74893-7_32