

$$\int_{-\infty}^{\infty} f(x) \cdot g(x) dx$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\int_{-\infty}^{\infty} f(x) \cdot g(x) dx$$

# Messunsicherheit eines Stickstoffdioxid-Jahresmittelwerts aus Passivsammlermessungen mit Passivsammlern des Palmes-Typs

## TEIL I THEORETISCHE GRUNDLAGEN

### MESSUNSICHERHEIT DER KALIBRATION EINES MESSVERFAHRENS AUF DAS ENTSPRECHENDE REFERENZVERFAHREN

Bericht Nr. 20081001  
Aktualisierung Bericht 20040228

**Verfasser/in:**  
Jürg Brunner, Fachbereichsleiter Labor

Zürich, Oktober 2008



$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$
$$C(x) = C_0 + \sum_{i=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot u_{x_i} \cdot x_j + 2 \cdot \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial b}{\partial x_j} \right) \cdot u_{y_i} \cdot y_j$$

## **Impressum**

### **Herausgeberin**

Stadt Zürich  
Umwelt- und Gesundheitsschutz Zürich UGZ  
Abteilung Umwelt  
Fachbereich Labor  
Walchestrasse 31  
8021 Zürich

### **Sachbearbeitung**

Jürg Brunner

## INHALTSVERZEICHNIS

1	Einleitung .....	1
2	Glossar .....	2
3	Lineare Regression (Methode der kleinsten Quadrate) .....	3
4	Orthogonale Regression .....	6
5	Flächen-Regression .....	8
6	Zusammenfassung .....	11
7	Kalibrations- und Analysenfunktion .....	12
8	Literatur .....	13



## 1 Einleitung

Die Messunsicherheit einer linearen Regression (Methode der kleinsten Quadrate), unter der Annahme von unsicherheitsfreien unabhängigen Werten (x-Werte) kann näherungsweise bestimmt werden [1-4].

Etwas anders sieht der Fall aus, wenn die Unsicherheit der x-Werte der linearen Regression nicht vernachlässigt werden kann. Im folgenden kann gezeigt werden, dass für den Spezialfall der Kalibration eines Messverfahren auf das entsprechende Referenzverfahren, unter der Voraussetzung des Nachweis der Gleichwertigkeit und Verwendung einer Flächen-Regression die Unsicherheit der linearen Regression näherungsweise bestimmt werden kann.

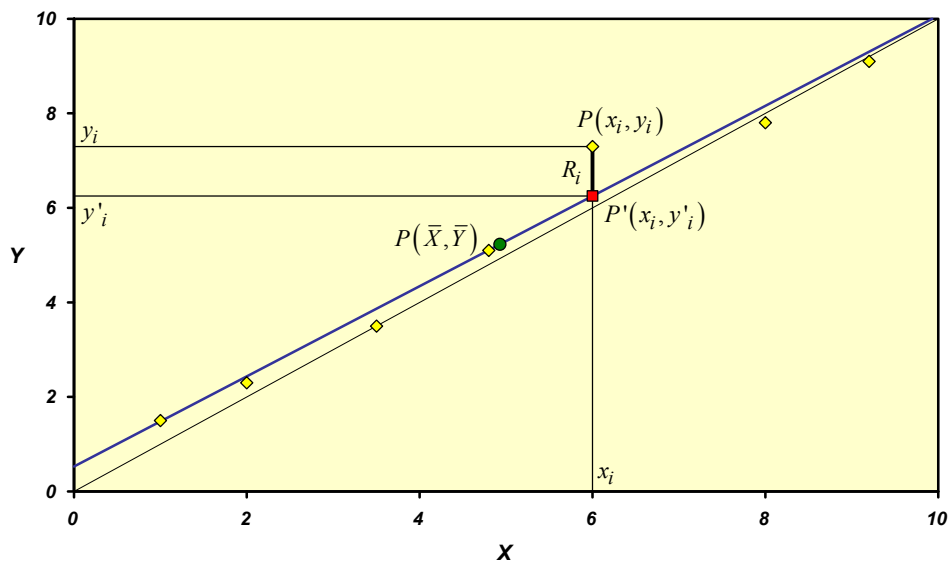
Die Lösung für diesen Spezialfall ist in der Form der Lösung der klassischen linearen Regression sehr ähnlich, berücksichtigt zusätzlich jedoch die Unsicherheit der x-Werte und die Kovarianz der Werte beider Messverfahren.

## 2 Glossar

$x_1, x_2, \dots, x_n$	$n$ x-Werte (unabhängige Werte resp. Werte des Referenzverfahrens)
$y_1, y_2, \dots, y_n$	$n$ y-Werte (abhängige Werte resp. Werte des Vergleichsverfahrens)
$\bar{X}, \bar{Y}$	Mittelwerte $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$
$b, a$	Parameter der linearen Regressionsfunktion (Steigung $b$ , Achsenabschnitt $a$ )
$r$	Korrelationskoeffizient
$r^2$	Bestimmtheitsmass
$S_{xx} = \sum_{i=1}^n (x_i - \bar{X})^2$	Summe der Abweichungsquadrate der einzelnen $x$ -Werte zum Mittelwert $\bar{X}$ (Fehlerquadratsumme)
$S_{yy} = \sum_{i=1}^n (y_i - \bar{Y})^2$	Summe der Abweichungsquadrate der einzelnen $y$ -Werte zum Mittelwert $\bar{Y}$ (Fehlerquadratsumme)
$S_{xy} = \sum_{i=1}^n (x_i - \bar{X}) \cdot (y_i - \bar{Y})$	Summe des Produkts der Abweichungen der einzelnen $x$ - und $y$ -Werte zu den entsprechenden Mittelwerten $\bar{X}$ und $\bar{Y}$ (Fehlerproduktsumme)
$s_R^2$	Varianz der Residuen einer linearen Regression (Methode der kleinsten Quadrate). Restvarianz der Kalibrierpunkte um die Regressionsgerade.
$s_D^2$	Varianz der Differenzen $x_i - y_i$ der Datenpaare
$\text{var}(p)$	Varianz eines Parameters $p$
$\text{cov}(p, q)$	Kovarianz der Parameter $p$ und $q$

Sämtliche folgenden Ausführungen beziehen sich auf ein Signifikanzniveau von 95%.

### 3 Lineare Regression (Methode der kleinsten Quadrate)



Die Parameter Steigung  $b$  und Achsenabschnitt  $a$  der Regressionsgerade  $y = b \cdot x + a$  sollen so bestimmt werden, dass die Summe  $S$  der vertikalen, quadratischen Abweichungen zwischen einem Messwert  $P(x_i, y_i)$  und dem entsprechenden Punkt der Regressionsgerade  $P'(x_i, y'_i)$  in  $y$ -Richtung minimal wird.

$$S = \sum_{i=1}^n R_i^2 = \sum_{i=1}^n (y_i - y(x_i))^2 = \sum_{i=1}^n (y_i - a - b \cdot x_i)^2$$

Für die Parameter  $a$  und  $b$  der Regressionsgerade ergeben sich aus der Minimierungsbedingung für  $S$  über die partiellen Ableitungen:

$$a = \bar{Y} - b \cdot \bar{X} \tag{0.1}$$

$$b = \frac{S_{xy}}{S_{xx}} \tag{0.2}$$

Das Bestimmtheitsmass  $r^2$  ergibt sich aus dem Produkt der Steigungen von Regressionsgerade und ihrer Umkehrfunktion

$$r^2 = b \cdot b' = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}} \quad 0.3$$

Die Unsicherheit der Steigung  $b$  berechnet sich wie folgt

$$u_b^2 = \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial b}{\partial x_j} \right) \cdot u(x_i) \cdot u(x_j) \cdot r_{x_i, x_j} + \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial y_i} \right) \cdot \left( \frac{\partial b}{\partial y_j} \right) \cdot u(y_i) \cdot u(y_j) \cdot r_{y_i, y_j} \\ + 2 \cdot \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial b}{\partial y_j} \right) \cdot u(x_i) \cdot u(y_j) \cdot r_{x_i, y_j}$$

Im Weiteren gilt für voneinander unabhängigen Einzelwerten

$$\left. \begin{array}{l} r_{x_i, x_j} \\ r_{y_i, y_j} \end{array} \right\} = 0 \text{ für } i \neq j \\ = 1 \text{ für } i = j \\ r_{x_i, y_j} = 0$$

und unter der Voraussetzung, dass  $u^2(x_i) = 0$  ist die Standardabweichung der Residuen

$$s_R^2 = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n-1} = \text{var}_y = u^2(y)$$

somit wird die Unsicherheit der Steigung  $b$  zu

$$u_b^2 = \sum_{i=1}^n \left( \frac{\partial b}{\partial y_i} \right)^2 \cdot s_R^2$$

$$\text{var}_b = u_b^2 = \frac{s_R^2}{S_{xx}}$$



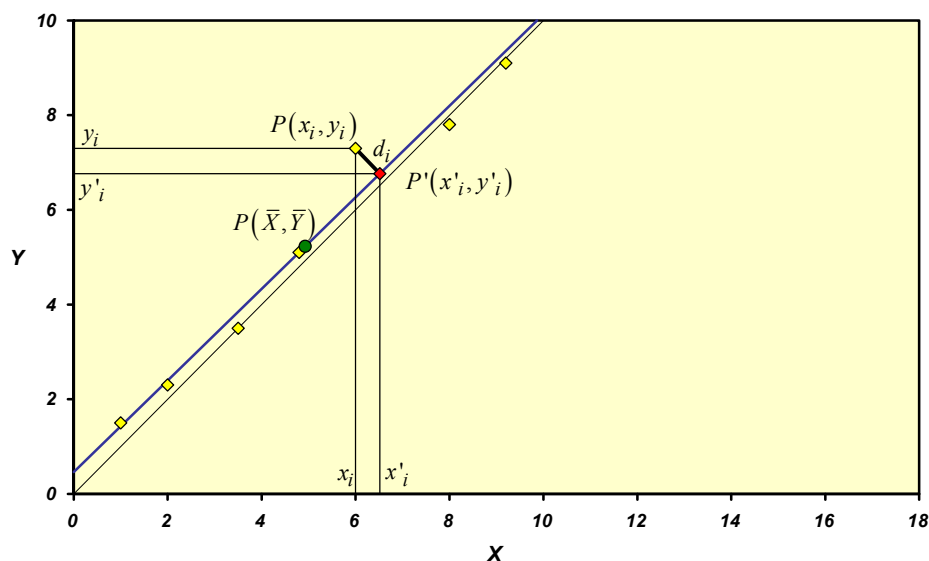
Analoges Vorgehen zur Bestimmung der Unsicherheit des Achsenabschnitts  $a$  und der Kovarianz von Steigung  $b$  und Achsenabschnitt  $a$  ergibt

$$\mathbf{var}_a = u_a^2 = \frac{s_R^2}{S_{xx}} \cdot \left( \bar{X}^2 + \frac{S_{xx}}{n} \right)$$

$$\mathbf{cov}_{a,b} = -\frac{s_R^2 \cdot \bar{X}}{S_{xx}}$$

Für den Vergleich von zwei Messverfahren für dieselbe Messgrösse eignet sich die klassische lineare Regression nicht, da die Unsicherheit der x-Werte als vernachlässigbar angenommen werden, was sie in Wirklichkeit nicht sind.

## 4 Orthogonale Regression



Ein weiteres Regressionsverfahren ist die orthogonale Regression. Dabei wird die Summe  $D$  der Abstandquadrate der Punkte  $P(x_i, y_i)$  auf die Regressionsgerade minimiert.

$$D = \sum_{i=1}^n d_i^2 = \sum_{i=1}^n (y_i - y'_i)^2 + (x_i - x'_i)^2$$

Die Koordinaten eines des Punkts  $P'(x'_i, y'_i)$  sind nicht bekannt, müssen jedoch die Gleichung der Regressionsgerade erfüllen

$$y'_i = b \cdot x'_i + a \quad f_i = y'_i - b \cdot x'_i - a = 0$$

Das Problem eine Funktion unter speziellen Bedingungen zu minimieren, kann mit der Methode der Lagrange Multiplikatoren gelöst werden [5, 6]. Dabei wird eine neue Funktion  $D'$  definiert, welche sich aus der ursprünglichen Funktion  $D$  und allen Einschränkungsbedingungen  $f_i$  mit einem Faktor  $\lambda_i$  skaliert, zusammensetzt:

$$D' = D + \sum_{i=1}^n \lambda_i \cdot f_i$$

Als Lösung ergibt sich für die Steigung  $b$

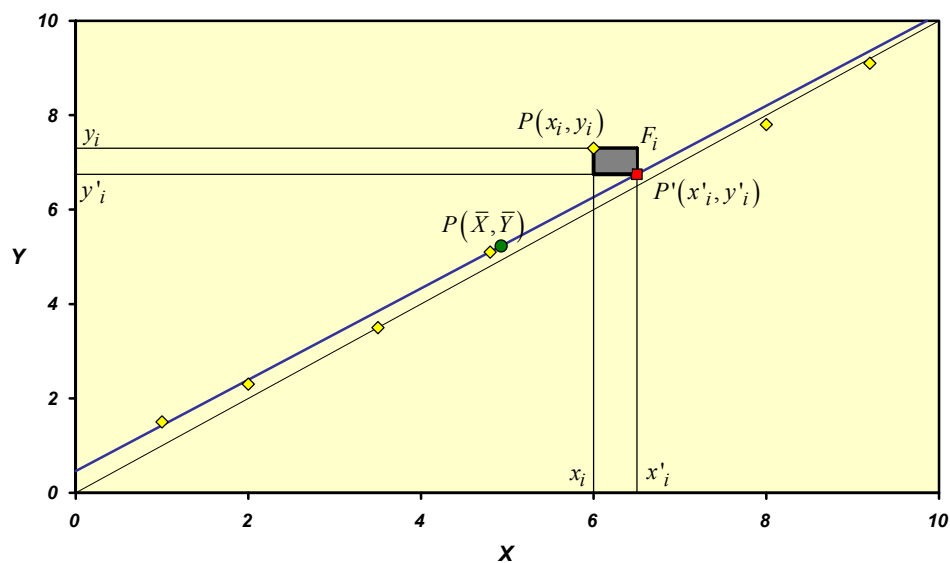
$$b = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4 \cdot S_{xy}^2}}{S_{xy}}$$

und für den Achsenabschnitt  $a$

$$a = \bar{Y} - b \cdot \bar{X}$$

Bereits die Form der Steigung weist auf die Schwierigkeiten hin, die ein Differenzieren der Funktion sehr aufwendig machen. Aus diesem Grund wurde dieser Ansatz nicht weiter verfolgt.

## 5 Flächen-Regression



Eine weitere Möglichkeit der Regression bildet das Minimieren der Summe der durch die Punkte  $P(x_i, y_i)$  und  $P'(x'_i, y'_i)$  aufgespannten Rechtecke mit der Fläche  $F_i$ .

$$F = \sum_{i=1}^n F_i = \sum_{i=1}^n (y_i - y'_i) \cdot (x_i - x'_i)$$

Die Koordinaten eines des Punkts  $P'$  sind nicht bekannt, müssen jedoch die Gleichung der Regressionsgerade erfüllen. Analog der orthogonalen Regression kann das Problem mit der Methode der Lagrange Multiplikatoren gelöst werden.

Als Lösung ergibt sich für die Steigung  $b$

$$b = \sqrt{\frac{S_{yy}}{S_{xx}}}$$

und für den Achsenabschnitt  $a$

$$a = \bar{Y} - b \cdot \bar{X}$$

Die Unsicherheiten von Steigung  $b$  und Achsenabschnitt  $a$  berechnen sich wie folgt

$$\begin{aligned} \text{var}_b &= \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial b}{\partial x_j} \right) \cdot u_{x_i, x_j} + \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial y_i} \right) \cdot \left( \frac{\partial b}{\partial y_j} \right) \cdot u_{y_i, y_j} + 2 \cdot \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial b}{\partial y_j} \right) \cdot u_{x_i, y_j} \\ \text{var}_a &= \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial a}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial x_j} \right) \cdot u_{x_i, x_j} + \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial a}{\partial y_i} \right) \cdot \left( \frac{\partial a}{\partial y_j} \right) \cdot u_{y_i, y_j} + 2 \cdot \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial a}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial y_j} \right) \cdot u_{x_i, y_j} \\ \text{cov}_{a,b} &= \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial x_j} \right) \cdot u_{x_i, x_j} + \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial y_i} \right) \cdot \left( \frac{\partial a}{\partial y_j} \right) \cdot u_{y_i, y_j} + 2 \cdot \sum_{i=1}^n \sum_{j=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial y_j} \right) \cdot u_{x_i, y_j} \end{aligned}$$

Die Unsicherheiten haben die Form

$$u_{x_i, x_j} = u_{x_i} \cdot u_{x_j} \cdot r_{x_i, x_j}$$

Dabei gilt

$$\left. \begin{array}{l} r_{x_i, x_j} \\ r_{y_i, y_j} \end{array} \right\} \begin{array}{l} = 0 \text{ für } i \neq j \\ = 1 \text{ für } i = j \end{array}$$

Da zwei Messverfahren für dieselbe Messgröße miteinander verglichen werden sollen, gilt als Näherung

$$\left. \begin{array}{l} r_{x_i, y_j} \end{array} \right\} \begin{array}{l} = 0 \text{ für } i \neq j \\ \cong 1 \text{ für } i = j \end{array}$$

und die Berechnung der Unsicherheiten vereinfacht sich zu

$$\begin{aligned} \text{var}_b &= \sum_{i=1}^n \left( \frac{\partial b}{\partial x_i} \right)^2 \cdot u_{x_i}^2 + \sum_{i=1}^n \left( \frac{\partial b}{\partial y_i} \right)^2 \cdot u_{y_i}^2 + 2 \cdot \sum_{i=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial b}{\partial y_i} \right) \cdot u_{x_i, y_i} \\ \text{var}_a &= \sum_{i=1}^n \left( \frac{\partial a}{\partial x_i} \right)^2 \cdot u_{x_i}^2 + \sum_{i=1}^n \left( \frac{\partial a}{\partial y_i} \right)^2 \cdot u_{y_i}^2 + 2 \cdot \sum_{i=1}^n \left( \frac{\partial a}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial y_i} \right) \cdot u_{x_i, y_i} \\ \text{cov}_{a,b} &= \sum_{i=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial x_i} \right) \cdot u_{x_i}^2 + \sum_{i=1}^n \left( \frac{\partial b}{\partial y_i} \right) \cdot \left( \frac{\partial a}{\partial y_i} \right) \cdot u_{y_i}^2 + 2 \cdot \sum_{i=1}^n \left( \frac{\partial b}{\partial x_i} \right) \cdot \left( \frac{\partial a}{\partial y_i} \right) \cdot u_{x_i, y_i} \end{aligned}$$

Die Unsicherheit der y-Werte und der x-Werte, kann durch die Varianz der Differenz  $D$  der Wertepaare beschrieben werden.

$$\text{var}_D = s_D^2 = \frac{\sum_{i=1}^n (d_i - \bar{D})^2}{n-1}$$

$$s_D^2 = \text{var}_x + \text{var}_y - \frac{2 \cdot S_{xy}}{n-1}$$

Sollen zwei Messverfahren für dieselbe Messgrösse miteinander verglichen werden, so kann näherungsweise angenommen werden, dass

$$\text{var}_x \cong \text{var}_y = \frac{s_D^2}{2} + \frac{S_{xy}}{n-1}$$

Ob diese Annahme gerechtfertigt ist, kann auf einfache Weise mit einem Vergleich der Varianzen der beiden Datenreihen (F-Test) gemacht werden. Somit ergibt sich

$$\text{var}_b = u_b^2 = \frac{1}{S_{xx}} \cdot \left( \frac{s_D^2}{2} + \frac{S_{xy}}{n-1} \right) \cdot (b^2 - 2 \cdot b^* + 1)$$

$$\text{var}_a = u_a^2 = \frac{1}{S_{xx}} \cdot \left( \frac{s_D^2}{2} + \frac{S_{xy}}{n-1} \right) \cdot \left( \bar{X}^2 + \frac{S_{xx}}{n} \right) \cdot (b^2 - 2 \cdot b^* + 1)$$

$$\text{cov}_{a,b} = -\frac{\bar{X}}{S_{xx}} \cdot \left( \frac{s_D^2}{2} + \frac{S_{xy}}{n-1} \right) \cdot (b^2 - 2 \cdot b^* + 1)$$

mit  $b^*$  der Steigung der klassischen linearen Regression

$$b^* = \frac{S_{xy}}{S_{xx}}$$

## 6 Zusammenfassung

	Lin. Regression	Orthogonale Regression	Flächen-Regression
<b>Steigung b</b>	$b^* = \frac{S_{xy}}{S_{xx}}$	$b = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4 \cdot S_{xy}^2}}{S_{xy}}$	$b = \sqrt{\frac{S_{yy}}{S_{xx}}}$
<b>Achsenabschnitt a</b>	$a = \bar{Y} - b^* \cdot \bar{X}$	$a = \bar{Y} - b \cdot \bar{X}$	$a = \bar{Y} - b \cdot \bar{X}$
<b>var<sub>b</sub></b>	$\frac{s_R^2}{S_{xx}}$	nicht berechnet	$\frac{1}{S_{xx}} \cdot \left( \frac{s_D^2}{2} + \frac{S_{xy}}{n-1} \right) \cdot (b^2 - 2 \cdot b^* + 1)$
<b>var<sub>a</sub></b>	$\frac{s_R^2}{S_{xx}} \cdot \left( \bar{X}^2 + \frac{S_{xx}}{n} \right)$	nicht berechnet	$\frac{1}{S_{xx}} \cdot \left( \frac{s_D^2}{2} + \frac{S_{xy}}{n-1} \right) \cdot \left( \bar{X}^2 + \frac{S_{xx}}{n} \right) \cdot (b^2 - 2 \cdot b^* + 1)$
<b>cov<sub>a,b</sub></b>	$-\frac{s_R^2 \cdot \bar{X}}{S_{xx}}$	nicht berechnet	$-\frac{\bar{X}}{S_{xx}} \cdot \left( \frac{s_D^2}{2} + \frac{S_{xy}}{n-1} \right) \cdot (b^2 - 2 \cdot b^* + 1)$

## 7 Kalibrations- und Analysenfunktion

Die Unsicherheit des Erwartungswerts  $\hat{y}_i$  eines Kalibrierpunktes  $x_i$

$$\hat{y}_i = b \cdot x_i + a$$

ergibt sich mittels Unsicherheitsfortpflanzung

$$u_{\hat{y}_i} = \sqrt{\left(\frac{\partial \hat{y}_i}{\partial x_i}\right)^2 \cdot u_{x_i}^2 + \left(\frac{\partial \hat{y}_i}{\partial a}\right)^2 \cdot u_a^2 + \left(\frac{\partial \hat{y}_i}{\partial b}\right)^2 \cdot u_b^2 + 2 \cdot \left(\frac{\partial \hat{y}_i}{\partial a}\right) \cdot \left(\frac{\partial \hat{y}_i}{\partial b}\right) \cdot u_{a,b}}$$

$$u_{\hat{y}_i} = \sqrt{b^2 \cdot u_{x_i}^2 + u_a^2 + x_i^2 \cdot u_b^2 + 2 \cdot x_i \cdot u_{a,b}}$$

Die Unsicherheit des Erwartungswerts  $\hat{x}_i$  eines Analysenwerts  $y_i$

$$\hat{x}_i = \frac{y_i - a}{b}$$

ergibt sich analog der Kalibrationsfunktion mittels Unsicherheitsfortpflanzung

$$u_{\hat{x}_i} = \sqrt{\left(\frac{\partial \hat{x}_i}{\partial y_i}\right)^2 \cdot u_{y_i}^2 + \left(\frac{\partial \hat{x}_i}{\partial a}\right)^2 \cdot u_a^2 + \left(\frac{\partial \hat{x}_i}{\partial b}\right)^2 \cdot u_b^2 + 2 \cdot \left(\frac{\partial \hat{x}_i}{\partial a}\right) \cdot \left(\frac{\partial \hat{x}_i}{\partial b}\right) \cdot u_{a,b}}$$

$$u_{\hat{x}_i} = \sqrt{\frac{1}{b^2} \cdot u_{y_i}^2 + \frac{1}{b^2} \cdot u_a^2 + \left(\frac{y_i - a}{b^2}\right)^2 \cdot u_b^2 + 2 \cdot \frac{1}{b} \cdot \left(\frac{y_i - a}{b^2}\right) \cdot u_{a,b}}$$

$$u_{\hat{x}_i} = \frac{1}{b} \cdot \sqrt{u_{y_i}^2 + u_a^2 + x_i^2 \cdot u_b^2 + 2 \cdot x_i \cdot u_{a,b}}$$

Für die Unsicherheiten der x- und y-Werte können nun entweder ermittelte Unsicherheiten oder die Näherung über die Differenzen der Wertepaare benutzt werden.



## 8 Literatur

- [1] ISO, Guide to the Expression of Uncertainty in Measurements, 1995
- [2] Eurachem/CITAC, Quantifying Uncertainty in Analytical Measurement (QUAM), Bericht Nr. QUAM:23000.P1, 2000
- [3] V. Meyer, Messunsicherheit in der chemischen Analytik, *SaCH-Kurs Messunsicherheit*, 2001
- [4] M. Rösslein, EMPA, Messunsicherheit in der Analytischen Chemie - Einführung und Grundlagen, St. Gallen, 2000
- [5] Bronstein und Semendjajew, Taschenbuch der Mathematik, Harri Deutsch, Thun, Neubearbeitung 20. Auflage, 1983, 3-8714-4492-8
- [6] M. R. Spiegel, Höhere Mathematik für Ingenieure und Naturwissenschaftler, McGraw-Hill, Nachdruck, 1978, 0-07-092023-0