

Různé pohledy na kalibrační úlohy

Jiří Militký KTM, Technická universita v Liberci, 461 17 LIBEREC, Česká Republika

Milan Meloun, KACH, Universita Pardubice, Česká Republika

Abstrakt

Cílem této práce je ukázat některé problémy vyskytující se při řešení klasických (ne)lineárních kalibračních úloh. Je ukázáno jaké typy úloh se při kalibraci vyskytují a jak postupovat při jejich řešení. Je naznačena obecná definice inverzního odhadu a jejího rozptylu jsou uvedeny základní modely působení poruch a jejich vliv na odhad modelových parametrů. Jako unifikovaný kalibrační model je navržen regresní polynomický spline. Jeho flexibilita je zajištěna vhodnou volbou uzlových bodů. Na příkladu kalibrační přímky jsou demonstrovány různé typy kalibrace.

1. ÚVOD

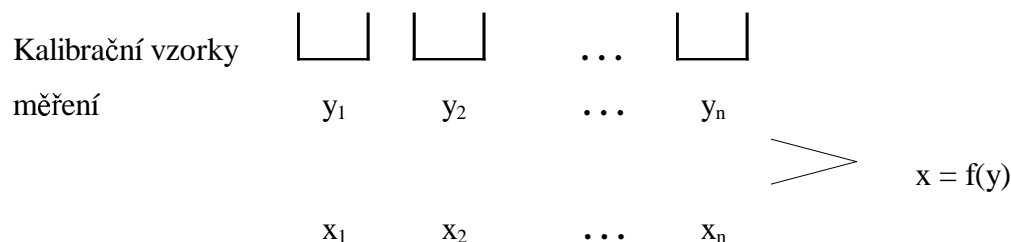
Kalibrace patří mezi základní úlohy řešené nejen v analytické úlohy, ale obecně v technické praxi. Z metrologického hlediska jde o typický problém tzv. nepřímých měření[1]. Obyčejně je účelem stanovit hodnoty nesnadno měřitelné nebo vůbec neměřitelné veličiny y (hledaná veličina, cílová veličina T - target). Tato veličina je funkčně spjata se snadno měřitelnou veličinou x (signál, M - měření). Příkladem T je koncentrace, teplota, vlhkost. Jako M je standardně používáno elektrické napětí nebo proud případně absorbance. Funkční přiřazení

$$y = f(x, \mathbf{a}) \quad (1)$$

je buď známo předem (z fyzikálního principu měření) nebo se hledá metodami regrese. Vždy je však třeba určit odhady modelových parametrů \mathbf{a} aby bylo možno kalibrační model použít. Vlastní kalibrace se skládá ze dvou fází

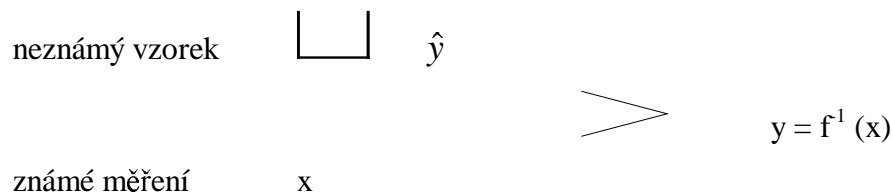
- tvorba kalibračního modelu
- použití kalibračního modelu.

Ve fázi *tvorby kalibračního modelu* je účelem nalézt model $f(x, \mathbf{a})$. Prakticky se provádí pro sadu nastavených hodnot y_i , $i = 1 \dots N$ měření hodnot x_i , $i = 1 \dots N$. Hodnoty y_i jsou typicky nastaveny s využitím velmi přesných metod nebo jsou k dispozici ve formě standardů s definovanou hodnotou. Je tedy možno nalézt vlastně funkci $x = f(y)$ (viz obr.1)



Obr.1 Fáze tvorby kalibračního modelu

Ve fázi **použití kalibračního modelu** je účelem pro neznámý vzorek určit odhad cílové veličiny \hat{y} z měření signálu x . Jde tedy v případě klasické kalibrace o nalezení inverzní funkce $\hat{y} = f^{-1}(x)$ ke kalibračnímu modelu $f(y)$. Schematicky je tato situace znázorněna na obr.2.



Obr.2 Fáze použití kalibračního modelu

Pole toho jakého typu se uvažují veličiny x a y ve fázi **tvorby kalibračního modelu** se pak rozeznávají tyto typy kalibrace[3]

- a) **C - kalibrace** (y se uvažují jako deterministické nebo měřené se zanedbatelnou chybou vůči měření x)

$$x_i = f(y_i, \mathbf{a}) + \varepsilon_x \dots \sigma_x^2 \quad (2)$$

Pro tento typ kalibrace je nutné řešení inverzní úlohy při predikci \hat{y}_{ce}

- b) **I - kalibrace** (x se uvažují jako deterministické nebo měřené se zanedbatelnou chybou vůči nastavení hodnot y)

$$y_i = f(x_i, \mathbf{a}) + \varepsilon_y \dots \sigma_y^2 \quad (3)$$

Pro tento typ kalibrace je možná přímá predikce \hat{y}_{in} .

- c) **O - kalibrace** (obě proměnné jsou náhodné s různými rozptyly)

$$y_i = f(x_i + \varepsilon_{xi}, \mathbf{a}) + \varepsilon_y \dots P = \sigma_y^2 / \sigma_x^2 \quad (4)$$

Pro řešení této kalibrační úlohy postačuje znalost poměru rozptylů P . Pro $P = 1$ jde o minimalizaci kolmých vzdáleností mezi kalibračním modelem a experimentálními body. Pro tento typ kalibrace je možná přímá predikce \hat{y}_{ok} .

Pro případ klasické C-kalibrace je vždy třeba řešit **inverzní úlohu** t.j. nalézt funkci $f^{-1}(x)$. Pro I-kalibraci a O-kalibraci není třeba řešit inverzní úlohu. Ve všech případech je však signál X_0 pro neznámý vzorek měřená veličina (náhodná), takže i predikce \hat{y}_0 je náhodná veličina a je třeba odhadnout její **variabilitu** resp. **interval spolehlivosti**. Zde je v dalším pojednáno o způsobech odhadu parametrů kalibračních modelů v souvislosti s různými typy působení poruch.

Samostatným problémem je volba kalibračního modelu pro případy, kdy nelze nalézt teoretické přiřazení typu rov. (1). Pro tyto účely je v této práci navrženo použití regresních spline.

2. MODELÝ PŮSOBENÍ PORUCH

Ilustrujme si různé modely působení poruch na úloze I - kalibrace. Pro případ C-kalibrace se pouze zamění pořadí proměnných resp. jejich charakter.

Nechť jsou experimentální data $\{x_i^T, y_i\}$, $i=1, \dots, n$, a kalibrační model $f(\mathbf{x}, \mathbf{a})$ je znám. Vysvětlovaná resp. nastavovaná (závisle) proměnná y je chápána jako výsledek měření a je tedy zatížena chybami (nepřesností měření, variabilitou měřeného materiálu, zanedbáním proměnných atd.). Celková chyba ε závisle proměnné je pak kombinací všech dílčích chyb. Obvykle se předpokládá, že celková chyba měření má nulovou střední hodnotu, $E(\varepsilon_i) = 0$, $i=1, \dots, n$.

Pokud je $E(\varepsilon_i) = \text{konst.} \neq 0$ chybí v modelu absolutní člen.

Způsob působení chyb je charakterizován **modelem měření** (aditivní, multiplikatívni, smíšený atd.). Obecný model kalibrace, zahrnující jak vlastní **kalibrační model**, tak i **model měření**, je pak možné vyjádřit ve tvaru

$$y_i = Z_i(x_i, \varepsilon_i, \mathbf{a}) \quad i = 1, \dots, n \quad (5)$$

Konkrétní tvar funkce $Z_i(\cdot)$ závisí na předpokládaném mechanismu působení chyb a konkrétním tvaru kalibrační funkce.

Pokud představují hodnoty y_i výsledky experimentů, předpokládá se obvykle **aditivní model** měření, pro který je

$$Z_i = f(x_i, \mathbf{a}) + \varepsilon_i \quad (6)$$

V řadě případů musí být naměřené hodnoty y_i , $i = 1, \dots, n$, kladné nebo mít alespoň přirozený počátek. Dále rozptyly chyb měření bývají nekonstantní. Konstantní jsou většinou relativní chyby měření. Tomu vyhovuje např. **multiplikatívni model** měření, pro který platí

$$Z_i = f(x_i, \mathbf{a}) \exp(\varepsilon_i) \quad (7)$$

V praxi se vyskytuje také **smíšený model** měření, popsany vztahem

$$Z_i = f(x_i, \mathbf{a}) \exp(v_i) + \varepsilon_i \quad (8)$$

kde o chybách v_i a ε_i se předpokládá, že jsou vzájemně nezávislé.

Pokud není model chyb znám je možno vyjít z obecné třídy modelů zahrnující jak aditivní tak i multiplikatívni model. Jedna taková obecná třída modelů se dá vyjádřit s pomocí Box - Coxovy transformace

$$y^{(\lambda)} = \frac{x^\lambda - 1}{\lambda} \quad \text{pro } \lambda \neq 0 \quad \text{resp. } y^{(\lambda)} = \ln(x) \quad \text{pro } \lambda = 0$$

Obecný model měření pak má tvar

$$y_i^{(\lambda)} = f^{(\lambda)}(x_i, \mathbf{a}) + \varepsilon_i$$

Při identifikaci kalibračního modelu se pak hledá optimální λ a následně modelové parametry \mathbf{a} .

V provozech a laboratořích se často provádí měření na jedné soustavě. Dochází pak ke vzniku kumulativních chyb. Navíc jsou měření, prováděná na přístrojích, zatížena obvykle konstantní relativní chybou $v(y) = \sigma(y) / f(x, \mathbf{a})$, takže rozptyl měřené veličiny $\sigma^2(y) \approx f^2(x, \mathbf{a})$

je úměrný čtverci hodnoty modelované funkce. Celková chyba ε_i je v této situaci vyjádřena vztahem

$$\varepsilon_i = \sum_{j=1}^i u_j + v_i \quad (9)$$

kde v_i jsou *chyby měření* a u_j jsou *procesní chyby*.

Procesní chyby jsou způsobeny fluktuací podmínek experimentů, tj. teploty, tlaku, čistoty surovin a mají kumulativní charakter. Obvykle se předpokládá, že chyby ε_i , vyjádřené rovnicí (9), působí aditivně.

Pro nalezení kritéria regrese a statistickou analýzu je třeba určit rozdělení náhodných veličin y_i . Toto rozdělení bude úzce souviset s rozdělením chyb ε_i , které je dáno sdruženou hustotou pravděpodobnosti $p_{\varepsilon}(\boldsymbol{\varepsilon})$. Ta je také funkcí určitých *distribučních parametrů*, jako je rozptyl σ^2 , atd.

Běžně bývá rozdělení chyb $\boldsymbol{\varepsilon}$ *unimodální*, symetrické s maximem $E(\boldsymbol{\varepsilon}) = 0$. Často se předpokládá, i když někdy nesprávně, že chyby měření ε_i jsou také vzájemně nezávislé. Sdružená hustota pravděpodobnosti $p_{\varepsilon}(\boldsymbol{\varepsilon})$ je pak dána součinem marginálních hustot $p(\varepsilon_i)$. Tedy

$$p_{\varepsilon}(\boldsymbol{\varepsilon}) = \prod_{i=1}^n p_{\varepsilon}(\varepsilon_i) \quad (10)$$

Skupinu rozdělení, zahrnujících rozdělení normální, rovnoměrné, Laplaceovo a lichoběžníkové, je možno vyjádřit hustotou pravděpodobností

$$p_{\varepsilon}(\varepsilon_i) = Q_N \exp\left(\frac{-|\varepsilon_i|^p}{\alpha}\right) \quad (11)$$

kde Q_N je normalizační konstanta a α je parametr úměrný rozptylu. Je-li $p = 1$, jde o Laplaceovo, pro $p = 2$ o normální a pro $p \rightarrow \infty$ o rovnoměrné rozdělení. Nevýhodou rozdělení $p_{\varepsilon}(\varepsilon_i)$, definovaného vztahem (11), je skutečnost, že pro $p < 2$ není okolí počátku lokálně kvadratické. Používají se proto také alternativní systémy hustot pravděpodobnosti, jako je zobecněné Studentovo rozdělení, atd.

Praxi bližší je případ, kdy chyby ε_i nejsou nezávislé, ale jsou charakterizovány kovarianční maticí chyb \mathbf{C}_{ε} . Je-li rozdělení chyb symetrické, unimodální se střední hodnotou $E(\boldsymbol{\varepsilon}) = 0$, volí se sdružená hustota pravděpodobnosti ze třídy eliptických rozdělení

$$p_{\varepsilon}(\boldsymbol{\varepsilon}) = Q_N \sqrt{\det(\mathbf{B})} h(\sqrt{\boldsymbol{\varepsilon}^T \mathbf{B} \boldsymbol{\varepsilon}}) \quad (12)$$

kde Q_N je normalizační koeficient, \mathbf{B} je matice související s kovarianční maticí chyb \mathbf{C}_{ε} a $h(\cdot)$ je kladná funkce, definovaná na intervalu $\langle 0, \infty \rangle$ s konečnými momenty do řádu $(n+1)$. Nejpoužívanějším rozdělením je vícerozměrné normální rozdělení $N(0, \mathbf{C}_{\varepsilon})$, které vyjde při volbě $h(x) = \exp(-0.5 x^2)$. Jeho sdružená hustota pravděpodobnosti má tvar

$$p_{\varepsilon}(\boldsymbol{\varepsilon}) = (2\pi)^{-n/2} (\det \mathbf{C}_{\varepsilon})^{-1/2} \exp(-0.5 \boldsymbol{\varepsilon}^T \mathbf{C}_{\varepsilon}^{-1} \boldsymbol{\varepsilon}) \quad (13)$$

Je možno také použít vícerozměrného Laplaceova rozdělení, Studentova rozdělení resp. dalších[1].

Tvar kovarianční matice chyb \mathbf{C}_ε souvisí s typem závislosti chyb. Jednoduchá je v případě heteroskedasticity, kdy jsou chyby vzájemně nezávislé, ale mají nekonstantní rozptyl

$E(\varepsilon_i^2) = \sigma_i^2$. Matice \mathbf{C}_ε je potom diagonální s prvky σ_i^2 na diagonále a sdružená hustota pravděpodobnosti (13) přechází na tvar (10). Při různých typech autokorelace není již matice \mathbf{C}_ε diagonální a její mimodiagonální prvky C_{ij} odpovídají kovariancím mezi ε_i a ε_j , tj.

$$C_{ij} = E(\varepsilon_i \varepsilon_j).$$

Při znalosti buď sdružené hustoty pravděpodobnosti chyb měření $p_\varepsilon(\boldsymbol{\varepsilon})$ nebo marginálních hustot $p_\varepsilon(\varepsilon_i)$ můžeme určit hustotu pravděpodobnosti $p(\mathbf{y})$ nebo $p(y_i)$ na základě vztahu pro hustotu pravděpodobnosti funkce náhodné veličiny.

Pro případ nezávislých náhodných chyb $\boldsymbol{\varepsilon}$ platí vztah

$$p(y_i) = p_\varepsilon \left[Z_i^{-1}(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\varepsilon}) \right] \left| \frac{\partial Z_i^{-1}(\cdot)}{\partial y_i} \right| \quad (14)$$

kde symbol $Z_i^{-1}(\cdot)$ označuje inverzi k funkci $Z(\cdot)$.

Pro *aditivní model měření* (6) je

$$Z_i^{-1}(\cdot) = y_i - f(\mathbf{x}_i, \mathbf{a}) \text{ a derivace } \left| \frac{\delta Z_i^{-1}(\cdot)}{\delta y_i} \right| = 1.$$

Dosazením do rov. (14) pak vyjde

$$p(y_i) = p_\varepsilon(y_i - f(\mathbf{x}_i, \mathbf{a})) \quad (14)$$

Z toho vyplývá, že aditivní model měření nezpůsobuje žádné deformace rozdělení měřených veličin vzhledem k rozdělení chyb.

Pro *multiplikativní model měření* (7) je

$$Z_i^{-1}(\cdot) = \ln y_i - \ln f(\mathbf{x}_i, \mathbf{a}) \text{ a derivace } \left| \frac{\delta Z_i^{-1}(\cdot)}{\delta y_i} \right| = \frac{1}{y_i}$$

Předpokládají se pouze kladné hodnoty měřených veličin. Po dosazení do (14) bude

$$p(y_i) = \frac{1}{y_i} p_\varepsilon(\ln y_i - \ln f(\mathbf{x}_i, \mathbf{a})) \quad (16)$$

Tato hustota pravděpodobnosti již neodpovídá hustotě pravděpodobnosti chyb $p_\varepsilon(\cdot)$.

Z přehledu modelů chyb vyplývá, že podle konkrétních podmínek experimentu a představ o působení různých druhů chyb lze odvodit rozdělení měřených hodnot veličiny \mathbf{y} . Ve většině případů se předpokládá normalita chyb měření a aditivní model jejich působení. Rozdíly jsou pouze v tom, zda kovarianční matice \mathbf{C}_ε , obsahuje jen diagonální a nebo i nediagonální prvky.

Sdružená hustota pravděpodobnosti pro vektor naměřených hodnot $\mathbf{y} = \{y_1, \dots, y_n\}^T$ se označuje jako *věrohodnostní funkce* $L(\boldsymbol{\alpha})$. Tato funkce je závislá na vektoru parametrů $\boldsymbol{\alpha}$, který obsahuje parametry modelu \mathbf{a} a distribuční parametr $\boldsymbol{\sigma}$. Maximálně věrohodné odhady parametrů se určují maximalizací logaritmu věrohodnostní funkce

$$\ln L(\boldsymbol{\alpha}) = \ln p(\mathbf{y}) = \sum_{i=1}^n \ln p(y_i) \quad (17)$$

Druhá rovnost v rovnici (17) platí pouze pro speciální případ nezávislých chyb měření. Maximálně věrohodné odhady mají asymptotický rozptyl, který je roven inverzi *očekávané Fisherovy informační matice*

$$\mathbf{D}(\boldsymbol{\alpha}) = \mathbf{I}^{-1}(\boldsymbol{\alpha}) \quad (18)$$

Prvky matice $\mathbf{I}(\boldsymbol{\alpha})$ jsou dány vztahem

$$\mathbf{I}_{ij} = -\mathbf{E} \left[\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_j} \right] \quad (19)$$

Pro praktické účely se očekávaná informační matice $\mathbf{I}(\mathbf{b})$ nahrazuje *odhadovanou informační maticí* $\hat{\mathbf{I}}(\cdot)$ s prvky

$$\hat{\mathbf{I}}_{ij} = - \left[\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \alpha_i \partial \alpha_j} \right]_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} \quad (20)$$

Odhadovaná informační matice je pro konstrukci intervalů spolehlivosti výhodnější než očekávaná informační matice.

Pro maximálně věrohodné odhady lze odvodit řadu důležitých výsledků:

1. Odhady $\hat{\boldsymbol{\alpha}}$ jsou pro $n \rightarrow \infty$ (tzn. asymptoticky) nevychýlené, tj. vychýlení

$$\mathbf{h} = \boldsymbol{\alpha} - \mathbf{E}(\hat{\boldsymbol{\alpha}}) = \mathbf{0} \quad (21)$$

je nulový vektor. Pro konečné počty měření n jsou však odhady vychýlené a velikost \mathbf{h} závisí na stupni nelinearity regresního modelu.

2. Odhady $\hat{\boldsymbol{\alpha}}$ jsou asymptoticky vydatné a rozptyly odhadů $\mathbf{D}(\hat{\boldsymbol{\alpha}}_j)$ jsou minimální ve třídě všech nevychýlených odhadů. Kovarianční matice $\mathbf{D}(\hat{\boldsymbol{\alpha}})$ leží na dolní mezi Cramerovy-Raovy nerovnosti [5]. Pro konečné výběry není ani tato vlastnost obecně zachována.
3. Náhodný vektor $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ má asymptoticky normální rozdělení $N(0, \mathbf{I}^{-1})$ s nulovou střední hodnotou a rozptylem rovným inverzi Fisherovy informační matice. Pokud je rozdělení chyb přibližně normální, platí normalita odhadů i pro konečné výběry.

Pro dostatečně velké rozsahy měření lze využít výše uvedených vlastností odhadů $\hat{\boldsymbol{\alpha}}$. Pro konečné počty měření činí značné potíže především vychýlení odhadů.

Při znalosti sdružené hustoty pravděpodobnosti $p(\mathbf{y})$ lze pak nalézt maximálně věrohodné odhady nebo kritérium pro jejich určení, tzv. *kritérium regrese*.

V praxi nejčastější je případ, kdy chyby měření ϵ jsou nezávislé, s nulovou střední hodnotou, konstantním rozptylem a normálním rozdělením $N(\mathbf{0}, \sigma^2 \mathbf{E})$. Platí také aditivní model měření definovaný vztahem (15). Rozdělení měřené veličiny y_i je pak také normální $N(f_i, \sigma^2)$. Platí že

$$p(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - f_i)^2}{2\sigma^2}\right),$$

kde $f_i = f(\mathbf{x}_i, \mathbf{a})$ Vektor hledaných parametrů necht' je $\boldsymbol{\alpha}^T = (\mathbf{a}^T, \sigma^2)$. Logaritmus věrohodnostní funkce $\ln L(\boldsymbol{\alpha})$ má pak tvar

$$\ln L(\boldsymbol{\alpha}) = \sum_{i=1}^n \ln p(y_i) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \mathbf{S}(\mathbf{a}) \quad (22)$$

Symbol $\mathbf{S}(\mathbf{a})$ značí součet čtverců odchylek

$$\mathbf{S}(\mathbf{a}) = \sum_{i=1}^n (y_i - f_i)^2 \quad (23)$$

Pro analytickou maximalizaci $\ln L(\boldsymbol{\alpha})$ podle σ^2 platí

$$\frac{\partial \ln L(\boldsymbol{\alpha})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbf{S}(\mathbf{a}) = 0 \quad (24)$$

Po úpravě vyjde

$$\sigma^2 = \frac{\mathbf{S}(\mathbf{a})}{n} \quad (25)$$

Odhad $\hat{\sigma}^2$ je pro malé rozsahy měření vychýlený. Jeho nevychýlená verze je

$$\hat{\sigma}^2 = \frac{\mathbf{S}(\mathbf{a})}{n - m} \quad (26)$$

Dosazením z rovnice (25) do (22) se určí tzv. *koncentrovaná věrohodnostní funkce* $\ln L(\mathbf{a})$, pro kterou platí

$$\ln(\mathbf{a}) = -\frac{n}{2} (1 + \ln(2\pi)) - 0.5 \ln \mathbf{S}(\mathbf{a}) \quad (27)$$

Maximum $\ln L(\mathbf{a})$ zřejmě odpovídá minimu kritéria $\mathbf{S}(\mathbf{a})$, což je vlastně kritérium metody nejmenších čtverců (MNČ). *Metoda maximální věrohodnosti je pro tento případ totožná s metodou nejmenších čtverců odchylek.*

Na základě rovnice (18) platí, že kovarianční matice odhadů $D(\hat{\boldsymbol{\alpha}})$ je rovna

$$D(\hat{\alpha}) = \begin{bmatrix} \sigma^2 (\mathbf{J}^T \mathbf{J})^{-1} & \mathbf{0} \\ \mathbf{0} & 2\sigma^4 / n \end{bmatrix} \quad (28)$$

kde symbol \mathbf{J} označuje Jakobián, tj. matici rozměru $(n \times m)$ prvních derivací modelu dle parametrů

$$J_{ij} = \frac{\partial f(\mathbf{x}_i, \mathbf{a})}{\partial a_j} \quad (29)$$

Z rovnice (28) plyne, že odhady σ^2 a $\bar{\mathbf{a}}$ jsou nezávislé a kovarianční matice

$$D(\bar{\mathbf{a}}) = \sigma^2 (\mathbf{J}^T \mathbf{J})^{-1}.$$

S využitím vlastností metody maximální věrohodnosti lze jednoduše konstruovat intervaly spolehlivosti a testovat statistické hypotézy [1].

Hledání lokálního extrému věrohodnostní funkce vede obecně na úlohu **nelineární optimalizace**. Pokud je známa kovarianční matice chyb \mathbf{C}_ε , lze pro případ, že platí aditivní model měření a normalita chyb $\boldsymbol{\varepsilon} \approx N(0, \mathbf{C}_\varepsilon)$, nalézt maximálně věrohodné odhady $\bar{\mathbf{a}}$ parametrů \mathbf{a} minimalizací kritéria zobecněných nejmenších čtverců

$$\mathbf{S}(\mathbf{a}) = (\mathbf{y} - \mathbf{f})^T \mathbf{C}_\varepsilon^{-1} (\mathbf{y} - \mathbf{f}) = \text{Tr}[\mathbf{C}_\varepsilon^{-1} \mathbf{e} \mathbf{e}^T] \quad (30)$$

kde $\mathbf{e} = \mathbf{y} - \mathbf{f}$ je vektor reziduí a symbol $\text{Tr}(\mathbf{A})$ značí stopu matice \mathbf{A} . Pokud je matice \mathbf{C}_ε diagonální, situace je značně jednodušší. Kriteriaální podmínka metody nejmenších čtverců (30) přechází na tvar

$$\mathbf{S}(\mathbf{a}) = \sum_{i=1}^n w_i (y_i - f(\mathbf{x}_i, \mathbf{a}))^2 \quad (31)$$

kde $w_i^2 = 1/C_{ii}$ jsou váhy úměrné reciprokým hodnotám diagonálních prvků kovarianční matice. Zavedením proměnných $y_i^* = w_i y_i$ a $f^* = w_i f(\mathbf{x}_i, \mathbf{a})$ přechází rov.(31) na kritérium klasické metody nejmenších čtverců odchylek

$$\mathbf{S}(\mathbf{a}) = \sum_{i=1}^n (y_i^* - f^*(\mathbf{x}_i, \mathbf{a}))^2 \quad (32)$$

Při znalosti vah w_i lze tedy úlohu vážené a zobecněné metody nejmenších čtverců převést na úlohu klasické metody nejmenších čtverců pro modifikované proměnné. Tento postup je vhodný i pro **neznámou diagonální matici** \mathbf{C}_ε , pokud se její prvky určují odděleně, např. na základě modelu heteroskedasticity. Pro případ heteroskedasticity vede klasická (nevážená) metoda nejmenších čtverců k nevychýleným odhadům $\bar{\mathbf{a}}$, odhady kovarianční matice jsou však vychýlené.

Při řešení kalibračních úloh se téměř výhradně využívá klasická nevážená MNC. To však může pro některé situace vést k nepoužitelným výsledkům. Pokud platí, že:

a) vyhovuje *multiplikativní model měření* je třeba použít kritéria nejmenších čtverců v logaritmech

b) měření probíhá za podmínek konstantní relativní chyby (variačního koeficientu) je třeba použít kritéria vážených nejmenších čtverců s vahami $w_i = 1/f(x, \mathbf{a}) \sim 1/y_i$

c) měření probíhá za podmínek úplné kumulace chyb je třeba použít kritéria nejmenších čtverců v prvních diferencích

d) měření probíhá za podmínek kde se mohou vyskytovat velké odchylky s vyšší pravděpodobností je vhodné použít kritéria nejmenších absolutních odchylek

e) mají být odchylky od kalibračního modelu co do velikosti přibližně stejné volí se kritérium minimalizace maximální odchylky

Tento neúplný výčet ukazuje, že kritérium MNČ není zdaleka univerzální a při kalibraci je třeba zkoumat i otázky typu poruch (chyb) a jejich rozdělení. Výhodou použití metody maximální věrohodnosti je to, že umožňuje také určení rozptylů jednotlivých odhadů další statistickou analýzou. Podrobnosti jsou uvedeny v práci [1].

3. INVERZNÍ KALIBRAČNÍ ÚLOHA

Inverzní kalibrační úloha se vyskytuje pouze u C-kalibrace. Ve fázi použití kalibračního modelu. Účelem je nalezení odhadu cílové veličiny \hat{y} a odpovídající nejistoty (intervalu spolehlivosti) ze známého signálu x^* . Jde tedy formálně o určení inverzní funkce

$$\hat{y} = f^{-1}(x^*, \mathbf{a}) = g(x^*, \mathbf{a}) \quad (33)$$

Tato úloha se dá matematicky převést na problém hledání kořene nelineární rovnice

$$f(y, \mathbf{a}) - x^* = 0 \quad (34)$$

vzhledem k proměnné y . Pokud nelze tuto úlohu řešit analyticky používá se celé řady postupů. Poměrně účinná je iterativní metoda Newtonova vyžadující znalost první derivace funkce $f(y)$.

Problémem je určení odpovídajícího rozptylu pokud je $f(y)$ nelineární. Na základě Taylorova rozvoje kalibrační funkce lze nalézt přibližný odhad rozptylu $D(y^*)$ ve tvaru

$$D(y^*) = \left(\frac{\partial f(y, \mathbf{a})}{\partial y} \right)^{-2} \left(D(x^*) + D(f(y^*, \mathbf{a})) \right) \quad (35)$$

Obyčejně je rozptyl signálu totožný s rozptylem chyb jeho měření σ_x^2 (viz. rov. (2)). Pro rozptyl predikce kalibrační funkce $f(y, \mathbf{a})$ v místě y^* přibližně platí

$$D(f(y^*, \mathbf{a})) \approx \mathbf{d}^T D(\hat{\mathbf{a}}) \mathbf{d} = \sigma^2 \mathbf{d}^T (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{d} \quad (36)$$

Zde \mathbf{d} je vektor prvních derivací kalibrační funkce s prvky

$$d_j = \frac{\partial f(y^*, \mathbf{a})}{\partial a_j} \quad (37)$$

a \mathbf{J} je Jakobián tj. matice prvních derivací kalibrační funkce ve všech bodech. Pro silně nelineární modely je rov. (34) velmi aproximativní a volí se jiné postupy (Bootstrap nebo odhady založené na věrohodnostním poměru [7])

Pro případ malých reziduí (to je u většiny kalibračních modelů splněno) a málo nelineárních modelů se konstruují intervaly spolehlivosti odhadované cílové veličiny y^* na základě předpokladu její normality. Pro přibližný 95%-ní interval spolehlivosti jsou dolní L resp. horní U meze ve tvaru

$$L = y^* - 1.96\sqrt{D(y^*)} \quad \text{resp.} \quad U = y^* + 1.96\sqrt{D(y^*)}$$

Přesnější odhady lze nalézt v práci [1]. Tam jsou také uvedeny odhady rozptylu predikce y^* pro případ inverzní kalibrace (odpovídá odhadu rozptylu predikce u nelineárních regresních modelů).

4. REGRESNÍ SPLINE

Pro nelineární kalibraci se často používá empirických modelů. Jako dostatečně flexibilní se uvažují zejména polynomické modely. Jejich výhodou je, že jde o lineární regresní modely, takže lze snadno odhadnout jejich parametry. Inverzní úloha vede na řešení polynomické rovnice, což je pro nízké stupně polynomu možno realizovat analyticky. Nevýhodou polynomů zejména vyšších stupňů je jejich tendence oscilovat mimo experimentální body. Navíc je předem definován počet extrémů a inflexních bodů, což je pro neasociativní fyzikální modely nepřijatelné.

Odstranění těchto problémů při zachování jednoduchosti manipulace a odhadu parametrů zajišťují tzv. *lokální funkce (polynomy)*. Kalibrační model je zde tvořen soustavou lokálních funkcí (polynomů) spojených v celém definičním intervalu v zadaném počtu derivací. Pro tyto modely je třeba kromě experimentálních bodů definovat také posloupnost tzv. uzlových bodů t_j $j=1, \dots, k$. Uzlové body tvoří hranice intervalů, kde jsou jednotlivé lokální funkce definované. V každém intervalu I_j ohraničeném uzly t_{j-1}, t_j je kalibrační model vyjádřen jako $g_j(x)$. Kvalitu kalibrace zde závisí na počtu a polohách jednotlivých uzlových bodů t_j , tvaru funkcí $g_j(x)$ a třídě C^m , do které kalibrační model $g(x)$ patří. Uveďme, že funkce $g(x)$ ze třídy C^m je spojitá v prvních m derivacích.

Speciálním typem *lokální funkce* jsou lokální polynomy jsou *regresní spline*. Spline $S_{m+1}(x)$ jsou funkce třídy C^m , které jsou definovány jako lokální polynomy maximálního stupně $(m+1)$. Vlastnosti spline funkcí jsou definovány v práci [2]. Pro účely kalibrace vyhovují dobře kvadratické spline $S_1(x)$, které jsou spojitě a hladké (tj. spojitě v první derivaci).

Spline $S_1(x)$ je možno jednoduše definovat pomocí uřezaných polynomů

$$S_1(x) = b_1 + b_2x + b_3x + \sum_{j=1}^k (x - t_j)_+^2 \quad (38)$$

Platí, že

$$(x)_+^2 = x^2 \quad \text{pro } x > 0$$

$$(x)_+^2 = 0 \quad \text{pro } x \leq 0$$

Pro známé hodnoty t_j představuje $S_1(x)$ lineární regresní model vzhledem k parametrům \mathbf{b} . Jestliže platí aditivní model měření a chyby ε_j jsou vzájemně nezávislé náhodné veličiny s konstantním rozptylem lze odhadnout parametry b_j $j=1, \dots, k+3$ pomocí metody lineárních nejmenších čtverců.

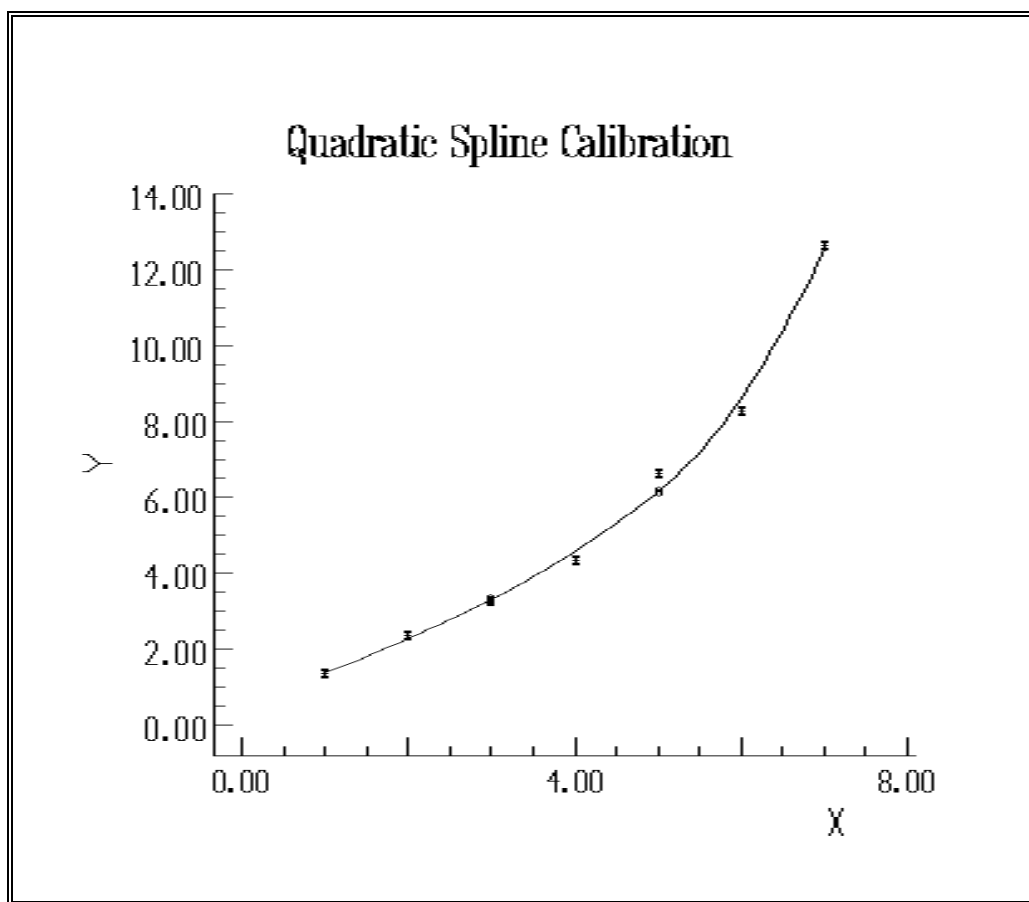
S ohledem na speciální tvar kubického spline vyjádřeného přes useknuté polynomy lze převést úlohu lineárních nejmenších čtverců na řešení soustavy lineárních (normálních) rovnic [1].

Z numerického hlediska je tato soustava rovnic špatně podmíněná, a proto je vhodné použít speciálních algoritmů (např. SVD). Ke zlepšení numerické stability je vhodné provést lineární transformaci x-ové souřadnice do intervalu [1, 2].

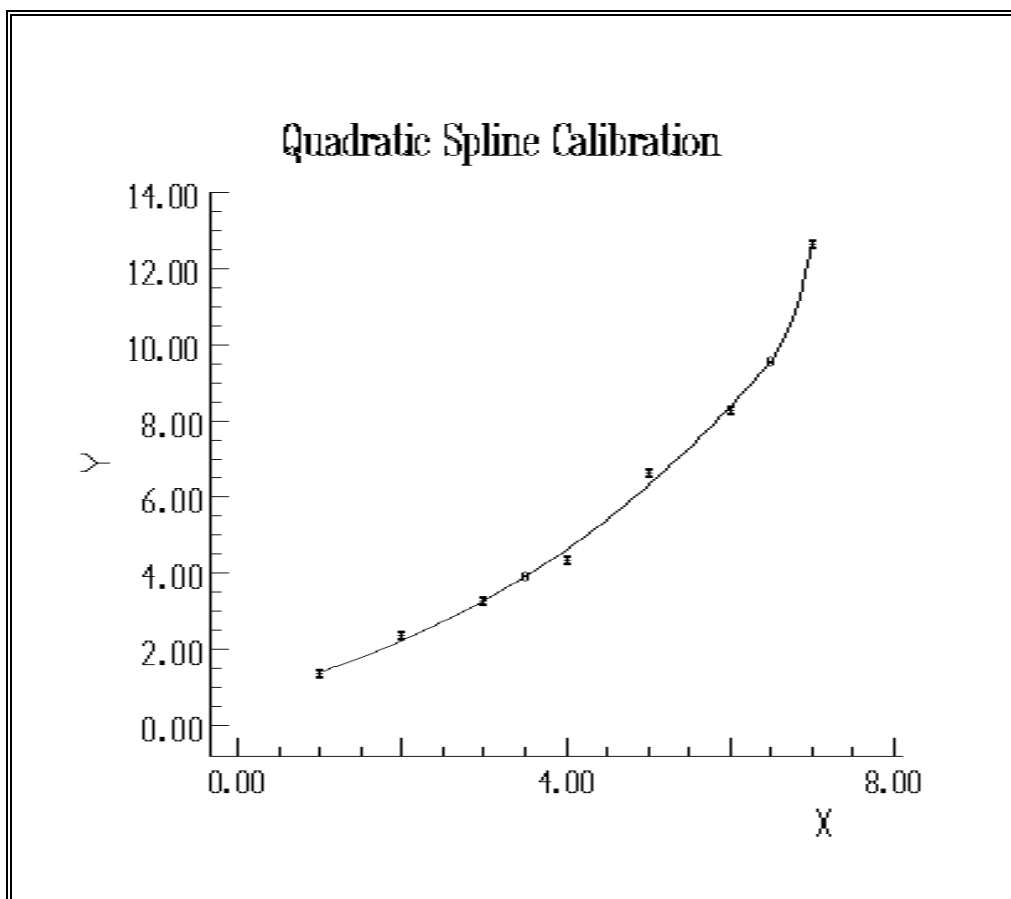
Flexibilitu regresních spline zajišťuje především vhodná selekce uzlových bodů t_i . V programu KALIBRACE systému ADSTAT jsou k dispozici tři alternativy:

- a) konstantní délka intervalu mezi uzlovými body,
- b) polohy uzlů takové, že v každém intervalu I_j je stejný počet dat,
- c) uživatelem volené polohy a počet uzlových bodů

Příklad vlivu různé strategie výběru uzlových bodů na výsledek spline kalibrace je na obr.3 a obr.4.



Obr 3. Kvadratická spline regrese (varianta a) - dva uzly)



Obr 4. Kvadratická spline regrese (varianta b) - tři body

5. KALIBRAČNÍ PŘÍMKA

Kalibrační model ve tvaru přímky patří mezi často používané v praktických aplikacích. Lze použít jak klasické C tak i inverzní I kalibrace resp. celé řady modifikací. Je tedy třeba zvolit vhodná kritéria pro výběr typu kalibrace resp. porovnání jejich vhodnosti. Problémem je, že existuje celá řada přístupů k výběru typu kalibrace, které vedou k různým typům závěrů. V této kapitole je učiněn pokus o porovnání různých typů kalibrace s ohledem na praktické použití.

Výchozí data je n- tice bodů (y_i, x_i) $i = 1, \dots, n$ Pro výpočty v obou fázích tvorby kalibračních modelů postačuje znalost průměrů, rozptylů a kovariance

$$\bar{x}, \bar{y}, s_y^2, s_x^2, C(x, y) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})$$

V dalším jsou ukázány jak postupy odhadu cílové hodnoty, tak i způsoby porovnání jednotlivých typů kalibrace

C - kalibrace

Vychází se z předpokladu platnosti modelu měření

$$x = a_0 + a_1 y + \varepsilon_x, \quad \varepsilon_x = N(0, \sigma_x^2) \quad (39)$$

Pro MNČ odhad kalibračního modelu platí

$$\hat{x} = \bar{x} + \frac{C(x, y)}{s_y^2} (y - \bar{y})$$

Odhady směrnice a úseku lze vyjádřit ve tvaru

$$\hat{a}_1 = C(x, y) / s_y^2, \quad \hat{a}_0 = \bar{x} - \hat{a}_1 \bar{y}$$

Odhad *inverzní funkce* a predikce cílové veličiny pro velikost signálu x jsou ve tvaru

$$\hat{y} = \frac{1}{a_1} x - a_0 / a_1, \quad \hat{y}_{ce} = \bar{y} + \frac{s_y^2}{C(x, y)} (x - \bar{x}) \quad (40)$$

Odhad \hat{y}_{cl} je vychýlený, protože obecně

$$E(\hat{y}_{cl}) \neq \frac{E(x - \hat{a}_0)}{E(\hat{a}_1)} = y_{cl}$$

Na druhou stranu jde o maximálně věrohodný odhad., pro který lze jednoduše sestavit interval spolehlivosti (viz [1]). Tento interval je konečný jen pokud platí, že \hat{a}_1 je dostatečně veliké. Na základě předpokladu, že chyby ϵ_x jsou dostatečně malé byly odvozeny asymptotické aproximace pro vychýlení B a střední kvadratickou chybu MSE klasického odhadu \hat{y}_{cl} (viz. [9])

I - kalibrace:

Tato technika byla navržena Krutchoffem pro případ, že je x deterministické. Vychází se z předpokladu platnosti modelu měření

$$y = b_0 + b_1 x + \epsilon_y, \quad \epsilon_y \approx N(0, \sigma_y^2)$$

MNČ odhad kalibrační přímky je přímo predikce \hat{y}_{in} cílové veličiny pro velikost signálu x

$$\hat{y}_{in} = \bar{y} + \frac{C(x, y)}{s_x^2} (x - \bar{x}), \quad \frac{C(x, y)}{s_x^2} = \hat{b}_1 \quad (41)$$

Na základě předpokladu, že chyby ϵ_y jsou dostatečně malé byly odvozeny asymptotické aproximace pro vychýlení B a střední kvadratickou chybu MSE inverzního odhadu \hat{y}_{in} (viz. [9])

U obou modelů je možné odhadnout příslušný rozptyl chyb (σ_x^2 resp. σ_y^2) z reziduálního rozptylu

$$\hat{\sigma}_c^2 = \frac{RSC}{n - 2}$$

Existují práce, které ukazují na přednosti klasického odhadu. Na druhé straně je inverzní odhad Bayesovým odhadem a platí pro něj, že

$$\hat{y}_{in} = \left[\frac{F}{F + (n - 2)} \right] \hat{y}_{cl} \quad (42)$$

kde

$$F = \frac{\hat{a}_1^2 \sum_{i=1}^n (y_i - \bar{y})}{\hat{\sigma}_c^2}$$

Pokud je F dostatečně větší než kvantil F rozdělení $F_\alpha(1, n-2)$ je klasický odhad y_{cl} dostatečně přesný. Pokud to neplatí, a F je o málo větší než $F_\alpha(1, n-2)$, je klasický odhad nepřesný a lépe je použít inverzní odhad y_{in} . Pokud je $\hat{\sigma}_c^2$ malé (tj. data leží téměř na kalibrační přímce) a F je veliké (kalibrační přímka má významnou směrnici) nejsou rozdíly mezi oběma typy kalibrace prakticky významné.

Je tedy užitečné porovnat odchylky mezi oběma odhady. Snadno lze ukázat, že

$$\frac{\hat{y}_{in} - \bar{y}}{\hat{y}_{cl} - \bar{y}} = \frac{C^2(x, y)}{s_x^2 s_y^2} = R^2 \quad (43)$$

Z tohoto výrazu je patrné, že:

- pokud je R^2 rovno jedné, jsou oba odhady stejné,
- protože musí platit, že R^2 je menší nebo rovno jedné, je inverzní odhad vždy blíže k průměru \bar{y} než odhad klasický

V práci [8] byl odvozen vztah pro určení pravděpodobnosti, že R^2 se bude lišit od jedné o malou předepsanou hodnotu q . Tato pravděpodobnost P_q je vyjádřena vztahem

$$P_q = P(F_N < (n - 2)(1 - q) / q)$$

Zde F_N označuje necentrální F rozdělení s 1 a $n - 2$ stupni volnosti a parametrem necentrality $s_y^2 a_1^2 / \sigma_y^2$. Při volbě $P_q=0:05$ a $q=1\%$ platí, že oba odhady se prakticky neliší, pokud

$$A = \frac{s_y^2 a_1^2}{n \sigma_y^2} \leq 152 \quad (44)$$

Při praktických výpočtech je možno jednotlivé parametry nahradit jejich odhady. Přesnější postup je popsán v práci [8]. Je zajímavé, že asymptotické vychýlení inverzního odhadu je $(n - 3)$ krát vyšší než asymptotické vychýlení klasického odhadu.

Pro extrémně malé výběry $2 < n \leq 8$ je z hlediska MSE odhad \hat{y}_{in} vždy lepší než odhad \hat{y}_{cl} .

Porovnání C a I kalibrace

S využitím rov (43) je možno vyjádřit vztah mezi oběma odhady ve tvaru

$$|\hat{y}_{in} - \bar{y}| \leq |\hat{y}_{cl} - \bar{y}|$$

Z této nerovnosti plyne, že:

a) \hat{y}_{in} je blíže k centru než \hat{y}_{ce} ,

b) pro $\hat{\sigma}_c^2 \rightarrow 0$ je $\hat{y}_{in} = \hat{y}_{ce}$,

c) I - kalibrace lépe vystihuje chování dat v oblasti centra (\bar{x}, \bar{y}) a C - kalibrace na krajích ,

S ohledem na minimální střední kvadratickou chybu $MSE = E(y - \hat{y})$ je I - kalibrace lepší v intervalu

$$\left\langle \bar{y} - \sqrt{2 s_y^2 + \sigma_x^2 / a_1}, \bar{y} + \sqrt{2 s_y^2 + \sigma_x^2 / a_1} \right\rangle$$

než C - kalibrace. Při praktických výpočtech se opět nahrazují parametry a_1 a σ_x^2 odhady \hat{a}_1 a $\hat{\sigma}_c^2$ určenými podle výše uvedených vztahů.

Kombinovaná kalibrace

Využívá se vhodné vážené kombinace \hat{y}_{cl} a \hat{y}_{in} resp. \bar{y} . Jeden z prvních kombinovaných odhadů se snaží „posunout“ klasický odhad \hat{y}_{cl} směrem k těžišti

$$\hat{y}_p = c \hat{y}_{cl} + (1 - c) \bar{y}$$

Dá se ukázat, že existuje hodnota c , pro kterou je tento odhad lepší než odhad \hat{y}_{cl} . V práci [9] je navržen jiný kombinovaný odhad

$$\hat{y}_{op} = v \hat{y}_{cl} + (1 - v) \hat{y}_{in}$$

Při volbě

$$v = \frac{1}{n - 2}$$

vyjde asymptoticky nevychýlený odhad s malým MSE. Je vždy lepší než klasický odhad \hat{y}_{cl} a také lepší než inverzní odhad \hat{y}_{in} pokud jsou skutečné hodnoty vzdálené od průměru. V okolí průměru jsou nejlepší inverzní odhady \hat{y}_{in} .

O - kalibrace

Pro známý poměr rozptylů $P = C_y^2 / \sigma_x^2$ lze určit predikci cílové veličiny pro velikost signálu x přímo ze vztahu

$$\hat{y}_0 = \bar{y} - \left[\Theta + s_y (C(x, y)) \sqrt{\Theta^2 + P} \right] (x - \bar{x}), \quad \Theta = \frac{s_y^2 - P s_x^2}{2 C(x, y)}$$

Podrobnosti o této kalibraci a konstrukci intervalů spolehlivosti pro cílovou veličinu lze nalézt v práci [1].

Příklad

K ilustraci byl použit jednoduchý příklad z práce [3]. Data jsou uvedena v tab.1

Tabulka 1 Data pro lineární kalibraci

číslo měření	Target y	Measurement x
1	2	5
2	4	6,5
3	4,5	6,25
4	5	6,5
5	5	8
6	6	6,5
7	6,5	8
8	6,5	9
9	7,75	10,5
10	8	11
11	9	12,5
12	9,5	10,5
13	10,5	12

Základní statistické informace jsou

$$\bar{x} = 8,6346 \quad s_x^2 = 72,327 \quad \bar{y} = 6,4808 \quad s_y^2 = 70,308 \quad C(x, y) = 66,284$$

C - kalibrace

Použitím lineární MNČ vyšlo

$$x = k_0 + k_1 y + \varepsilon_x, \quad x = 2.524 + 0.9427 y, \quad \hat{y}_{ce} = -2.68 + 1.06 x$$

Korelační koeficient vyšel 0,9051 a reziduální rozptyl $\hat{\sigma}_c^2 = 0,89427$.

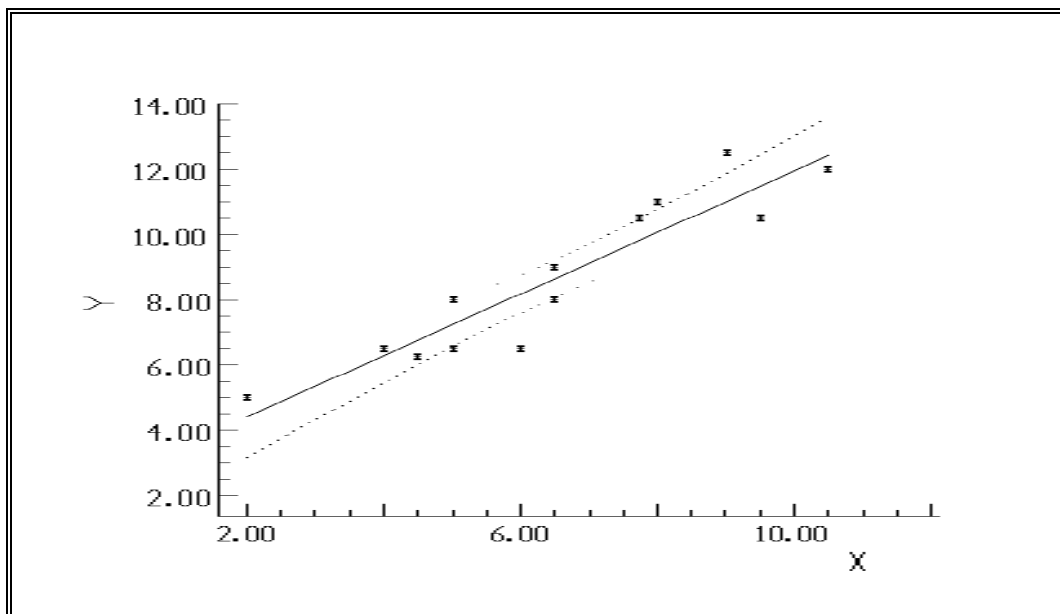
Byly vypočteny také kalibrační limity (viz. [1]):

Mez detekce: 5.718

Mez kritická: 2.859

Mez stanovení: 10.320

Na obr. 5 je znázorněna kalibrační přímka spolu s 95 % intervalem spolehlivosti.



Obr. 5 C-kalibrační přímka

I - kalibrace

Použitím lineární MNČ vyšlo

$$\hat{y}_{in} = -1.432 + 0.916x$$

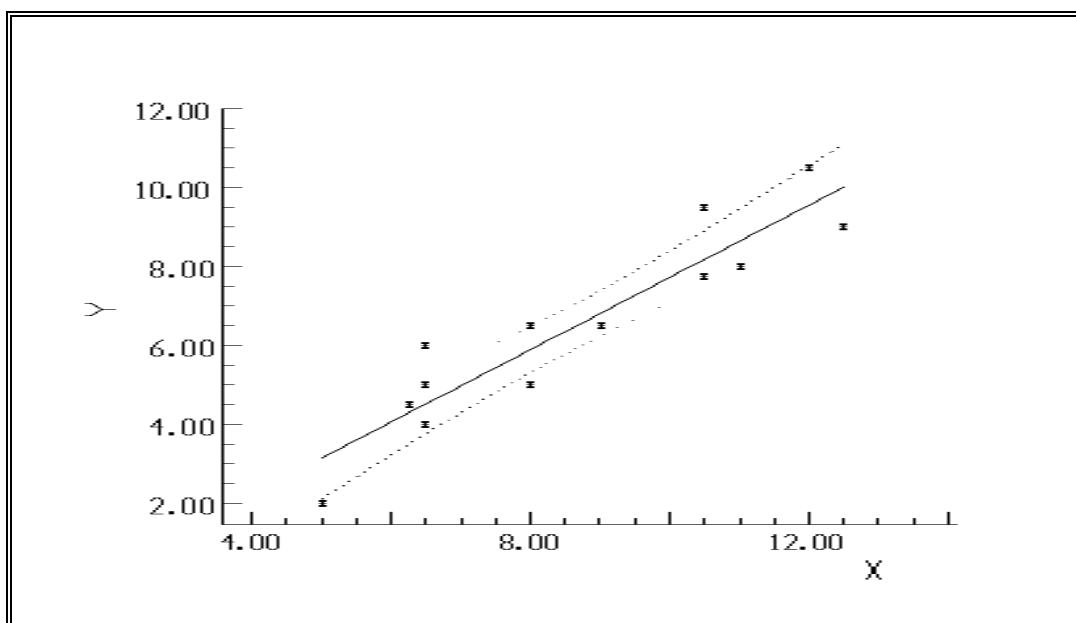
Byly vypočteny kalibrační limity (viz. [1]):

Mez detekce: 4.53

Mez kritická: 1.54

Mez stanovení: 13.53

Na obr. 6 je znázorněna kalibrační přímka spolu s 95 % intervalem spolehlivosti.



Obr. 6 I-kalibrační přímka

Z rov. (43) byl určen parametr $R^2 = 0,864$ a z rov (44) parametr $A = 5,375$. Klasický odhad se tedy významně neliší od inverzního. Interval, ve kterém je \hat{y}_{in} lepší než \hat{y}_{cl} z hlediska MSE vyšel $\langle -5.4173, 18.378 \rangle$. Lze tedy shrnout, že \hat{y}_{in} je sice lepší než \hat{y}_{cl} v širokém intervalu, ale rozdíl mezi oběma odhady je prakticky nevýznamný.

6. ZÁVĚR

Jak je patrné z výše uvedeného záleží **výběr typu kalibrace** na charakteru proměnných a jejich variabilitě.

Proměnná x (M) :

- obvykle dosti přesně stanovená (elektrická veličina, absorbance),
- ε_x zahrnují především neuvažované proměnné (teplota, . . .)

σ_x^2 . . . může být nekonstatní

Proměnná y (T):

- určená z externích informací (jiné přístroje, etalony, . . .),
- ε_y zahrnuje chyby měření,
- σ_y^2 je obvykle rostoucí funkcí y.

Poměr rozptylů $P = \sigma_y^2 / \sigma_x^2$ se může měnit v mezích $(0, \infty)$. Výběr mezi I - nebo C - kalibrací se dá provést buď na základě poměru rozptylů nebo použitím jiných kritérií. Také při lineární kalibraci je obecně třeba řešit problém výběru modelu měření.

Poděkování: *Tato práce vznikla s podporou grantu MŠMT č. VS 97084.*

7. LITERATURA

- [1] Meloun M., Militký J.: Zpracování experimentálních dat, EAST Publishing, Praha 1998
- [2] De Boor C.: A Practical Guide to Splines, Springer Verlag, New York 1998
- [3] Martens H. Naes T.: Multivariate Calibration,, Wiley, Chichester, 1993
- [4] Singh M., Kanji G. K., El-Bizri K. S.: A note on inverse estimation in non-linear models, J. Appl. Statistics. **19**, 473, 1992
- [5] Cheng Ch.-L., van Ness J. W.: On estimating linear relationship when both variables are subject to errors, J. R. Stat. Soc., **56**, 167-183, 1994
- [6] Militký J.: Non-linear calibration in practice ,Proc. First Mansoura Engineering Conference, Mansoura, March 1995
- [7] Huet S. a kol.: Statistical tools for nonlinear regression, Springer New York 1996
- [8] Chow S., Shao J.: On the difference between classical and inverse methods of calibration, J. Roy. Stat. Soc. **C39**, 219, 1990
- [9] Srivastava V. K., Singh N.: Small disturbance asymptotic theory for linear calibration Technometrics **31**, 373, 1989

Název souboru: kalibra
Adresář: E:\Pom
Šablona: D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název: KALIBRACE
Předmět:
Autor: Militky
Klíčová slova:
Komentáře:
Datum vytvoření: 14.09.00 13:39
Číslo revize: 2
Poslední uložení: 14.09.00 13:39
Uložil: Milan Meloun
Celková doba úprav: 1 minuta
Poslední tisk: 14.09.00 13:43
Jako poslední úplný tisk
Počet stránek: 18
Počet slov: 4 822 (přibližně)
Počet znaků: 27 488 (přibližně)