

Různé přístupy při vyhodnocování kalibračních údajů

Lenka Křazovická, Radek Strnad

V článku jsou stručně představeny a posouzeny různé základní způsoby výpočtu hodnot parametrů kalibrační rovnice s použitím sady výchozích údajů pocházejících z kalibrace termoelektrického článku typu S ve vybraných pevných bodech teplotní stupnice ITS-90. Uvedené metody a závěry jejich posouzení mají při vyhodnocování souborů reálných údajů širší využití.

In the article, some basic methods used for calibration model parameter calculations are briefly described and evaluated. A set of data obtained by the calibration of type S thermocouple at various ITS-90 temperature scale fix points is used for demonstration. The discussed methods as well as results of their acceptability assessment could be widely exploited at real laboratory data processing.

1. Úvod

K určování hodnot parametrů kalibračních rovnic je v laboratorních používáno několik postupů, které jsou založeny na evropských normách, např. EN 60584. V normách jsou stanoveny základní polynomy, které by měly popisovat chování jednotlivých typů snímačů (čidel) teploty. Cílem tohoto článku je ukázat, jaký význam mají při výpočtu hodnot parametrů kalibrační rovnice chybějící nebo odlehle kalibrační údaje.

Různé způsoby výpočtu hodnot parametrů kalibrační rovnice a příklady výsledků výpočtu jsou ukázány při použití jedné a téže sady výchozích údajů pocházejících z kalibrace termoelektrického článku typu S ve vybraných pevných bodech teplotní stupnice ITS-90.

Výpočty byly provedeny s použitím programu Matlab[®], verze 2011a.

2. Matematický aparát

2.1 Kalibrační rovnice versus sada údajů z kalibrace

Při kalibraci snímače teploty se stanovuje velikost jeho výstupního signálu jen v několika vybraných, přesně daných bodech teplotní stupnice. Snímač je ovšem v praxi používán k měření teploty v celém jeho měřicím (pracovním) rozsahu. Aby toto bylo možné, je třeba vhodně stanovit závislost výstupního signálu snímače na teplotě, jíž je snímač vystaven (kterou měří), a *vice versa* – tedy statickou převodní charakteristiku snímače, v praxi i dále v tomto článku označovanou jako *kalibrační rovnice*. Vhodná kalibrační rovnice je taková, která dostatečně vystihuje chování toho kterého snímače. V praxi obecně není dostatečně vhodná kalibrační rovnice ve tvaru přímky (výstupní signál se nemění lineárně s teplotou), ale polynomu. Doporučuje se popisovat chování snímačů kalibračními rovnicemi ve tvaru polynomu druhého

řádu. Při použití řádů vyšších často vznikají výrazné chyby a fyzikální nesmysly.

Hodnoty parametrů kalibrační rovnice bývají i nyní určovány s použitím údajů z použitých tří namísto ze všech teplotních bodů, v nichž byl snímač kalibrován (tj. namísto využití úplné sady kalibračních údajů). Tento způsob výpočtu může vést k nevhodnému stanovení kalibrační rovnice s ohledem na úplnou sadu kalibračních údajů v důsledku nesprávného výběru bodů použitých pro výpočet.

K demonstraci možných chyb je v článku zkoumán vztah mezi kalibrační rovnicí určenou při použití úplné sady kalibračních údajů a kalibračními rovnicemi určenými jinými metodami výpočtu. K proložení údajů získaných při kalibraci vhodnou kalibrační závislostí je použita jednak dobře známá metoda nejmenších čtverců, jednak její modifikace. Kalibrační rovnice má ve všech případech tvar polynomu druhého řádu.

2.2 Metoda nejmenších čtverců a její vážená forma

Metoda nejmenších čtverců (MNC) představuje způsob, jak jednoduchým výpočtem při použití lineární algebry určit regresní polynom (kalibrační rovnici) co nejlépe prokládající soubor bodů získaných měřením. Poloha každého z těchto bodů v rovinném kartézském souřadném systému je určena dvojicí hodnot: hodnotou nezávisle proměnné x (v daném případě příslušná referenční teplota t_{90}) a závisle proměnné y (naměřená hodnota výstupního signálu z kalibrovaného snímače). Hledaná regresní funkce $f(x, \theta)$ vykazuje v každém změřeném bodě odchylku e , tedy pro i -tý změřený bod platí

$$e_i = y_i - f(x_i, \theta) \quad (1)$$

kde θ je vektor hodnot hledaných koeficientů kalibrační rovnice.

Podle metody nejmenších čtverců představuje nejlepší proložení daných údajů regresní

funkce minimalizující součet druhých mocnin (čtverců) odchylek (odtud název metody) přes všechny body vstupující do výpočtu, tedy

$$\hat{\theta} = \arg \min e^T e = \arg \min \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \quad (2)$$

kde

$\hat{\theta}$ je odhad vektoru θ ,

e je vektor hodnot odchylek e_i , $i = 1, \dots, n$, n – počet bodů ve výpočtu, přičemž parciální derivace odchylky podle hledaných koeficientů je rovna nule

$$\frac{\partial e}{\partial \theta} = 0 \quad (3)$$

Hledané hodnoty koeficientů kalibrační rovnice se poté určí řešením soustavy lineárních rovnic. Důležitá je skutečnost, že při použití metody nejmenších čtverců vstupující do výpočtu všechny naměřené údaje se stejnou váhou.

Modifikací uvedené metody nejmenších čtverců je *vážená metoda nejmenších čtverců* (VMNČ), také popsána v mnoha zdrojích, např. [3], [4]. Přestože jde o metodu vcelku jednoduchou, při aproximaci průběhu kalibračních závislostí představuje mocný nástroj. Do výpočtu hledaných hodnot koeficientů regresní funkce vstupuje další člen – vektor, popř. matice vah w (popř. W). Hodnoty vah w_i přiřazují každému jednotlivému páru údajů určitou důležitost (váhu), a tím určují vliv konkrétního bodu ve finálním výpočtu regresního polynomu – čím větší w_i , tím větší váhu má daný bod ve výpočtu. Rovnici (2) lze v tomto případě přepsat do tvaru

$$\hat{\theta} = \arg \min e^T W e \quad (4)$$

Váhy w_i lze stanovit mnoha různými způsoby; v daném případě je vhodné jejich vyjádření při použití nejistot u_i způsobem

$$w_i = \frac{1}{u_i} \quad (5)$$

2.3 Robustní přístup

Nedostatkem již uvedených metod MNC a VMNČ je jejich velká citlivost na odlehle hodnoty (extrémní hodnoty, které jsou platnou částí sady údajů). Obecně se předpokládá, že chyby při kalibraci mají normální (gaussovské) rozdělení pravděpodobnosti a že odlehle hodnoty se vyskytují zřídka. Druhá mocnina jim příslušných odchylek

ovšem výrazně zvyšuje efekt odlehlých hodnot a tyto mají na výslednou regresi velký vliv.

K minimalizaci vlivu odlehlých hodnot je možné použít robustní metody regrese, např. metodu LAR (*Least Absolute Residuals*), spočívající v minimalizaci součtu absolutních hodnot odchylek e_i místo jejich druhých mocnin podle vztahu

$$\hat{\theta} = \arg \min \sum_{i=1}^n [|y_i - f(x_i, \theta)|]^p \quad (6)$$

kde p je volitelný parametr splňující podmínku $p \geq 1$, v tomto článku je $p = 1$.

Výsledkem je, že odlehlé hodnoty mají ve výsledné regresi menší vliv na výsledek.

2.4 Statistické vyhodnocení

Za účelem objektivního porovnání výsledků regrese dosažených při použití uvedených metod MNČ, VMNČ a LAR jsou dále sledována následující tři kritéria:

1. Prostá hodnota součtu druhých mocnin odchylek S_{SE} (*Sum of Squares Error*), s níž pracují MNČ a VMNČ (1), když

$$S_{SE} = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \quad (7)$$

2. Hodnota koeficientu determinance R_{square} jako statistiky informující o tom, jak velkou část variability proměnné y vysvětluje regresní polynom (model) a jak velká část zůstala neobjasněna; definice je následující

$$R_{square} = \frac{\sum_{i=1}^n w_i (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n w_i (y_i - \bar{y})^2} \quad (8)$$

kde

\hat{y}_i je hodnota y_i předpovídaná regresním modelem (kalibrační rovnicí),

\bar{y} průměrná hodnota ze všech (měřených, dostupných) dat.

Statistika R_{square} může nabývat hodnot mezi 0 a 1. Čím je hodnota blíže jedné, tím větší část variability je vysvětlena modelem.

3. Hodnota ukazatele je R_{MSE} (*Root Mean Square Error*), opět založeného na metodě nejmenších čtverců odchylek

$$R_{square} = \frac{\sum_{i=1}^n w_i (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n w_i (y_i - \bar{y})^2} \quad (9)$$

4. Výsledky

4.1 Výchozí kalibrační údaje

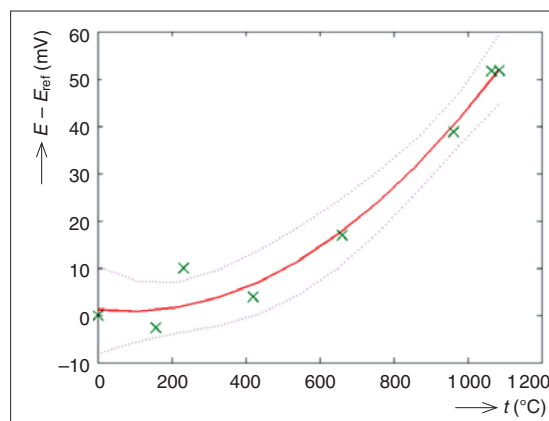
Regresní metody uvedené v kap. 2.2 a 2.3 byly porovnány při použití údajů (naměřeného elektrického napětí, resp. elektromotorické síly E) získaných při kalibraci termoelektrického článku typu S. Kalibrace byla pro-

vedena v pevných bodech In, Zn, Al, Ag, Au a Cu teplotní stupnice ITS-90. K regresi byl použit polynom druhého řádu.

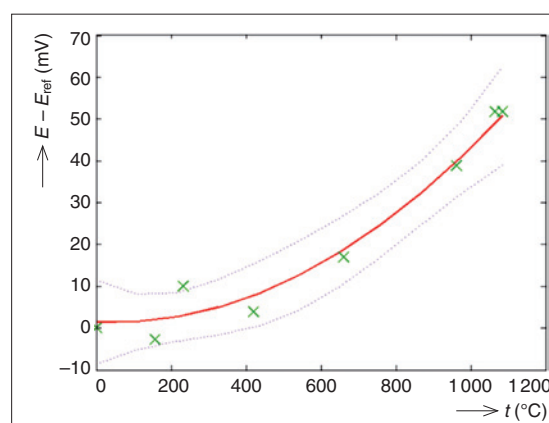
Údaje vstupující do regresních výpočtů jsou v odchylkovém tvaru uvedeny v tab. 1, kde hodnoty ve sloupci Nejistota U_c jsou použity k výpočtu vah podle (5). Zpracovávaná sada dat obsahuje jeden odlehlý údaj E naměřený v pevném bodě Sn (231,9 °C). Vliv od-

Tab. 1. Naměřené kalibrační údaje (odchylkové vyjádření)

Teplota t_{90} (°C)	Odchylka od reference ($E - E_{ref}$) podle ČSN EN 60584-1 (μV)	Nejistota U_c (μV)
156,598 5	-2,7	1,4
231,928	10,0	1,0
419,527	3,9	1,8
660,323	17,0	1,6
961,78	38,8	2,3
1 064,18	51,7	8,0
1 084,62	51,7	2,4



Obr. 1. Regrese kalibračních údajů při použití metody nejmenších čtverců odchylek (nevážená, nerobustní)



Obr. 2. Regrese kalibračních údajů při použití vážené metody nejmenších čtverců odchylek (vážená, nerobustní)

lehlého údaje na výslednou kalibrační rovnici je v dalším textu vysvětlen jednotlivě pro každou použitou regresní metodu. Pojem kalibrační rovnice je v této kapitole použit ve významu kalibrační rovnice v odchylkovém vyjádření.

4.2 Výsledky regrese při použití různých metod

Nejprve byla kalibrační rovnice určena při použití prosté metody nejmenších čtverců (MNČ, obr. 1). Tato metoda využívá k určení kalibrační rovnice všechny dostupné údaje, přičemž všechny body mají tutéž váhu. Hledání nevhodnějšího regresního polynomu probíhá tak, že je hledán polynom, který proloží kalibrační údaje tak, že hodnota S_{SE} (7) je nejmenší. Jelikož MNČ představuje nerobustní regresní metodu, není vhodná ke zpracování souborů údajů obsahujících odlehlé hodnoty.

Následně byly kalibrační údaje zpracovány při použití vážené metody nejmenších čtverců (VMNČ, obr. 2), umožňující vzít při výpočtu v úvahu i nejistoty, s jakými byla uskutečněna jednotlivá měření. Čím větší je nejistota, tím menší je vliv konkrétního bodu ve výsledném výpočtu hodnot parametrů kalibrační rovnice. Výpočet nejistot je ovšem závislý na mnoha okolnostech, takže nepředstavuje spolehlivý způsob detekce odlehlých hodnot.

Robustní metoda regrese umožňuje detekovat odlehlé údaje, které sice zůstávají platnou součástí sady kalibračních dat, ale jejich vliv je při výpočtu hodnot parametrů kalibrační rovnice minimalizován. Výsledky dosažené při použití robustních metod regrese jsou ukázány na obr. 3 a obr. 4.

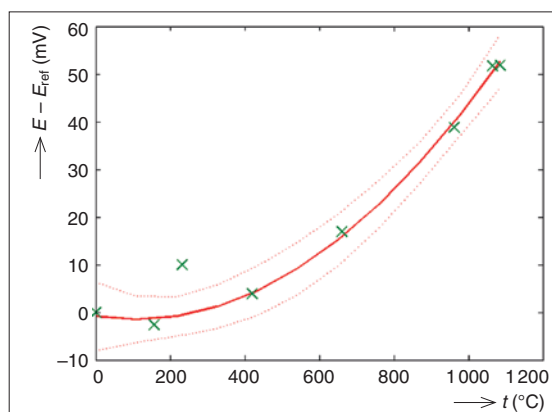
Na obr. 3 je zobrazen výsledek získaný při použití MNČ v kombinaci s robustní metodou regrese. Je dobře patrné, že určená kalibrační rovnice vhodně prokládá reálné údaje a že vliv odlehlého údaje je jen minimální. Obdobně lze hodnotit i výsledek použití VMNČ v kombinaci s robustní metodou (obr. 4).

Ve všech čtyřech uvedených obrázcích jsou současně zobrazeny meze konfidenčních intervalů určující pásmo, v němž se kalibrační závislost nachází s pravděpodobností 95 %.

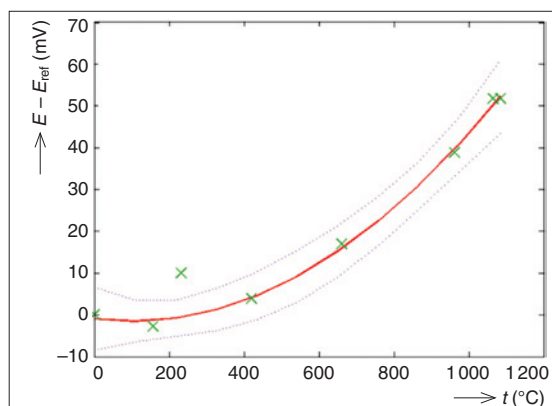
4.3 Statistické porovnání výsledků regrese

Po nalezení kalibračních závislostí (rovnice) znázorněných na obr. 1 až obr. 4 byly dále pro porovnání stanoveny hodnoty objektivních statistických kritérií uvedených v kap. 2.4. Výsledky jsou v tab. 2.

Důvodem významně větších hodnot S_{SE} u vážených metod regrese je skutečnost, že údaj naměřený s větší nejistotou má menší váhu, a tedy také menší vliv na výsledné proložení než údaj naměřený s menší nejistotou, jehož váha je větší.



Obr. 3. Regrese kalibračních údajů při použití metody nejmenších čtverců odchylek v kombinaci s robustní metodou (nevážená, robustní)



Obr. 4. Regrese kalibračních údajů při použití vážené metody nejmenších čtverců odchylek v kombinaci s robustní metodou (vážená, robustní)

Tab. 2. Statistické porovnání výsledků regrese (získaných kalibračních rovnic)

Kritérium	Regresní metoda			
	nevážená & nerobustní	nevážená & robustní	vážená & nerobustní	vážená & robustní
S_{SE}	152,3	173,4	1 987	2 398
R_{square}	0,956	0,949	0,925	0,909
R_{MSE}	5,5	5,9	19,9	21,9

Robustní metody regrese dokážou identifikovat odlehlé údaje a znalost vah není nezbytná. Protože odlehlé údaje mají při těchto výpočtech velmi malý vliv, je hodnota S_{SE} větší než u nerobustních metod. To ale nezna-

mená, že by nerobustní metody regrese popisovaly chování snímače lépe. Z obdržených výsledků je patrné, že robustní metody regrese díky schopnosti identifikovat odlehlé údaje jsou mocným nástrojem při vyhodnocování reálných údajů naměřených v kalibrační laboratoři. Vykazují také užší konfidenční interval. Statistiky podle kap. 2.4 zjišťují u souborů reálných údajů velmi velké hodnoty S_{SE} a R_{MSE} a velmi malé hodnoty R_{square} . Důvodem je, že tyto statistiky jsou z definice vypočítávány ze všech naměřených údajů, bez jakýchkoliv vah nebo detekce údajů odlehlých hodnot.

5. Závěr

Ke zpracování reálných údajů naměřených v laboratoři na termočláncích i jiných typech snímačů se pro jejich výhodné vlastnosti doporučuje používat vážené robustní metody regrese.

V článku je probírána vhodnost použití čtyř různých přístupů ke zpracování údajů pocházejících z kalibrací snímačů teploty. V souborech údajů tohoto druhu se často vyskytují údaje odlehlých hodnot, s nimiž je třeba se korektně vypořádat.

ní rovnice jsou použity všechny naměřené údaje bez rozdílu a s touž vahou. Použití této metody regrese je vhodné za účelem rychlého odhadu hodnot parametrů kalibrační rovnice. Jelikož ovšem nedokáže odhalit údaje odlehlých (extrémních) hodnot, není pro zpracování souborů reálných údajů příliš vhodná.

Vážená metoda nejmenších čtverců umožňuje zahrnout do výpočtu nejistoty měření, podle nichž je každému páru údajů přiřazena určitá váha, která pak ovlivňuje výsledné stanovení kalibrační rovnice. S čím větší nejistotou byl bod změřen, tím méně ovlivní výsledek regrese. Jde tudíž o metodu, kterou lze doporučit ke zpracování souborů reálných údajů.

Robustní přístup umožňuje objektivně zjistit údaje odlehlých hodnot, které pak mají na výsledné stanovení kalibrační rovnice jen minimální vliv. Při použití tohoto přístupu výsledná kalibrační závislost respektuje především všechny „správné“ údaje a nedochází k posunu výsledku v důsledku přítomnosti odlehlých údajů.

Na základě předložených výsledků je vhodné používat pro soubory reálných údajů kombinaci robustní a vážené metody regrese, která spojuje přednosti plynoucí z respektování nejistot měření i detekce odlehlých údajů.

Literatura:

- [1] NICHOLAS, J. V. – WHITE, D. R.: *Traceable Temperatures*. 1994, ISBN 0471938033.
- [2] EN 60584-1 *Thermocouples Part 1: Reference tables* (1995). Ref. No. EN 60584-1:1995.
- [3] http://www.efunda.com/math/least_squares/least_squares.cfm. October 2011.
- [4] RAO, R. – TOUTENBURG, H. – FIEGER, A. – HEUMANN, C. – NITTNER, T. – SCHEID, S.: *Linear Models: Least Squares and Alternatives*. Springer Series in Statistics, 1999, ISBN 0387988483.

Lenka Kňazovická,
Vysoká škola chemicko-technologická
v Praze, ústav počítačové a řídicí techniky
(lknazovicka@cmi.cz),
Radek Strnad,
Český metrologický institut, oddělení primární
metrologie tepelně-technických veličin
(rstrnad@cmi.cz)

SNÍMAČE A MĚŘIDLA TLAKU

rozsahy 100 Pa - 40 MPa
důlní průmysl - I M1 Ex ia I Ma
plynárenství - II1G Ex ia IIB T4 Ga

CRESSTO s.r.o.
PRESSURE MEASUREMENTS

www.cressto.cz