

VALIDÁCIA ANALYTICKÝCH METÓD V MATLABE

Miloslav Foltin^a, Martin Foltin^b

^aKatedra analytickej chémie, Prírodovedecká fakulta Univerzity Komenského ,
Mlynská dolina, 842 15 Bratislava

^bKatedra ASR, Fakulta elektrotechniky a informatiky, Slovenská technická univerzita,
Ilkovičova 3, Bratislava,

1. Úvod

Validácia analytických metód je v súčasnosti nevyhnutným predpokladom pre zavedenie systému zabezpečenia kvality (QA/QC) podľa noriem ISO 17 025 v chemických laboratóriách. Validačný proces je rozdelený do niekoľkých samostatných fáz. V prvej fáze sa určuje tvar kalibračného modelu, testuje homoskedasticita dát a prítomnosť vplyvných bodov a odhadujú sa predikčné schopnosti kalibračného modelu na nízkej koncentračnej úrovni (odhad medze dôkazu-LOD a medze stanoviteľnosti-LOQ). Všetky pokusy sa v tejto fáze vykonávajú s kalibračnými roztokmi analytu bez prítomnosti matrice.

Pri spracovaní validačných údajov sa používajú rôzne štatistické postupy a nástroje, ktoré sú súčasťou rôznych štatistických softverových produktov. Väčšinou sú to programy, ktoré sú uzavreté, bez možnosti zásahu či úpravy zo strany užívateľa.

Predložený program napísaný v jazyku MATLAB slúži na získanie základných informácií o tvare kalibračného modelu (lineárneho alebo kvadratického), overuje homoskedasticitu dát a testuje prítomnosť odľahlých výsledkov (outliers) a odhaduje hodnotu LOD a LOQ. Jeho výhodou je, že sa dá pružne meniť a dopĺňať podľa požiadaviek užívateľa tak, aby bolo možné získať komplexnejšie informácie o validovanej analytickej metóde.

2. Určenie matematického tvaru kalibračného modelu

Pri spracovaní validačných údajov je prvou úlohou určiť matematický tvar kalibračného modelu. Väčšine analytických metód vyhovuje lineárny kalibračný model:

$$y = \beta_0 + \beta_1 x \quad (1)$$

S ohľadom na základné fyzikálno-chemické princípy analytických metód (Lambertov-Beerov zákon v spektrofotometrických metódach, Ilkovičova rovnica v polarografii a pod.) sa v mnohých prípadoch dá tento model ešte zjednodušiť na tvar:

$$y = \beta x \quad (2)$$

V niektorých prípadoch, ako napr. v emisnej spektrálnej analýze, sa uplatnia skôr nelineárne kalibračné modely, najčastejšie kvadratické:

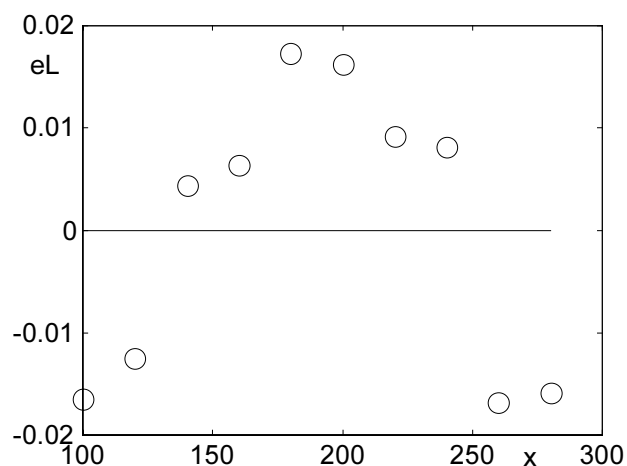
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \quad (3)$$

Podľa algoritmu na spracovanie validačných dát, prezentovaného v literatúre [1], je potrebné najskôr rozhodnúť o tvare kalibračného modelu. Na tento účel sa doporučuje použiť numerický Mandelov test [1,2], ktorý je založený na porovnaní reziduálnych rozptylov lineárneho a kvadratického modelu:

$$F_{calc} = \frac{(n-2)s_{e,L}^2 - (n-3)s_{e,Q}^2}{s_{e,Q}^2} \quad (4)$$

Nulová hypotéza H_0 o platnosti lineárneho modelu sa zamietá ak $F_{calc} > F_{krit, 1, n-3, 0.01}$, kde F_{krit} je kvantil Fisher-Snedecorovho rozdelenia.

Vhodnosť navrhnutého modelu môžeme posúdiť aj vizuálne z grafu rezíduí. Ak obraz závislosti rezíduí od predikovanej hodnoty analytického signálu \hat{y} , alebo od nezávisle premennej x vytvára obrazec nelineárneho pásu navrhnutý model môže byť nesprávny (obr.1).

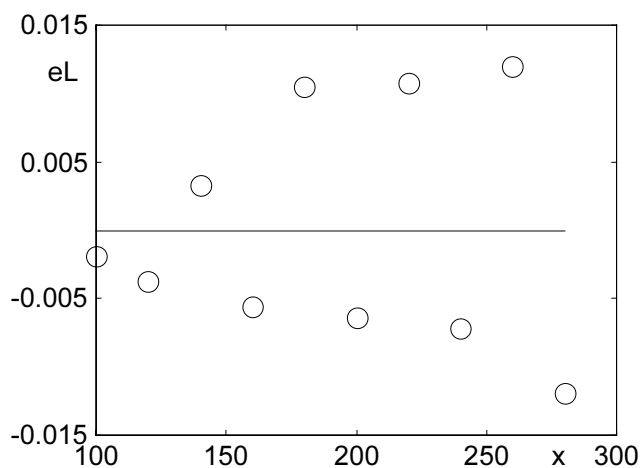


Obr.1. Závislosť reziduálov lineárneho modelu od nezávisle premennej x .

3. Testovanie homoskedasticity

Po určení správneho kalibračného modelu nasleduje test homogenity rozptylu merania analytického signálu (test homoskedasticity).

Na predbežnú identifikáciu heteroskedsticity môžeme opäť použiť graf reziduálov. Homoskedastické dáta vytvárajú obrazec približne eliptického tvaru rovnomerne rozložený okolo strednej línie s hodnotou $e = 0$. Heteroskedastické dáta vytvoria obrazec výseče (obr.2).



Obr. 2. Graf reziduálov heteroskedastických dát

Spôľahlivejšie určenie heteroskedasticity umožňuje neparametrický test trendu rezíduí [2]:

$$D = \sum_{i=1}^n (R|\hat{e}_i| - i)^2 \quad (5)$$

$$\rho_S = 1 - \frac{6 \cdot D}{n^3 - n} \quad (6)$$

Testovanie heteroskedasticity je potom založené na významnosti Spearmanovho korelačného koeficientu. Ak je nulová hypotéza $H_0: \rho_S = 0$ zamietnutá znamená to, že dáta môžu byť heteroskedastické. $R|e_i|$ je poradie absolútnej hodnoty i -teho rezídua.

4. Testovanie prítomnosti vplyvných bodov

Súčasťou validácie kalibračného modelu je testovanie prítomnosti vplyvných bodov – odľahlých bodov (outliers) a extrémnych bodov (extrems).

Za odľahlé body sa považujú tie, ktorých y -ová súradnica sa výrazne odlišuje od ostatných kalibračných bodov. prítomnosť takýchto bodov významne zvyšuje hodnotu reziduálneho rozptylu, čím významne zhoršuje presnosť predikčných vlastností kalibračného modelu. Z týchto dôvodov je potrebné tieto body identifikovať, zistiť príčiny odľahlosti a znovu premerať podozrivý bod. V krajnom prípade sa takýto bod z regresie vylúči.

Extrémy sa v dátach vyskytujú obyčajne ako dôsledok nevhodného plánovania kalibračných koncentrácií. Ich pôsobenie je dvojsečné. Ak vyhovujú teoretickému kalibračnému modelu a sú v súlade so zvyšnými dátami zvyšujú presnosť predikčných vlastností kalibračného modelu. V opačnom prípade ich pôsobenie je negatívne a preto by sme sa mali radšej ich prítomnosti v kalibračných dátach vyhnúť.

Na identifikáciu prítomnosti odľahlých resp. extrémnych bodov sa používajú rôzne testy založené na analýze rôznych typov reziduálov (normované, štandardizované, Jackknife), diagonálnych prvkov projekčnej matice H_{ii} [2], resp. na vyhodnotení vplyvových funkcií

(Cookova vzdialenosť, Atkinsonova vzdialenosť, Cook-Weisbergova vierohodnostná vzdialenosť, Andrews-Pregibonova diagnostika) [2,3], alebo rôzne neparametrické testy [3,4]. Uvedené testy sa používajú buď ako grafické alebo numerické.

V predloženom programe sú k dispozícii grafické testy Wiliamsov, Pregibonov, predikovaných rezíduí, a z numerických testov je použitá Cookova vzdialenosť CD_i , veľkosť normalizovaných rezíduí e_N a veľkosť diagonálnych prvkov projekčnej matice H_{ii} . Vzhľadom na to, že citlivosť jednotlivých testov je rôzna mali by sme podozrivý bod vylúčiť po dôkladnej analýze príčin odľahlosti. V niektorých prípadoch je vhodné urobiť testy vplyvných bodov urobiť pred testovaním homoskedasticity a testovaním kalibračného modelu, pretože tieto parametre môžu byť prítomnosťou vplyvných bodov významne skreslené.

Užívateľ si môže doplniť ľubovoľný ďalší test bez ovplyvnenia funkčnosti zvyšného programu.

5. Testovanie predikčných schopností kalibračného modelu na nízkej koncentračnej úrovni (odhad medze detekcie a medze stanoviteľnosti)

Posledným krokom 1.fázy validačného procesu je odhad detekčných vlastností validovanej analytickej metódy. V literatúre je opísaných viacero rôznych definícií medze dôkazu (detekčných limitov LOD) a medze stanoviteľnosti (LOQ) a postupov ich odhadu. [1,5,6,7,8,9,10]. Veľké množstvo koncepcií a spôsobov odhadu LOD a LOQ naznačuje, že v analytickej chémii dodnes nie je ustálená jednotnotná definícia a postup odhadu týchto základných charakteristík analytickej metódy. Hodnoty LOD a LOQ získané rôznymi postupmi sa líšia niekedy aj veľmi výrazne. Z týchto dôvodov je potrebné k získaným hodnotám uvádzať vždy aj postup, ktorým boli odhadnuté. V predloženom programe je použitý postup, ktorý je založený na vlastnostiach kalibračného modelu v oblasti veľmi

nízkych koncentrácií, ktorý je doporučený v literatúre [1, 10]. Doplnenie ďalších postupov podľa potrieb užívateľa resp. vývoja legislatívy a doporučení organizácií, ktoré určujú pravidlá pre QA/QC v analytických laboratóriách (IUPAC, AOAC, ILAC, Eurachem, ISO a iné) je možné.

Literatúra

1. W.Funk, V.Dammann, G.Donnenevert: Quality Assurance in Analytical Chemistry. VCH Weinheim, 1995.
2. M.Meloun, J.Militký: Statistické zpracování experimentálních dat. PLUS, Praha 1994.
3. D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke: Handbook of Chemometrics and Qualimetrics. Part A. Elsevier, Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo, 1997.
4. J.N. Miller, J.C. Miller: Statistics and Chemometrics for Analytical Chemistry. 4th edition. Pearson Education, Harlow, 2000.
5. H.Kaiser: Fresenius Z.Anal.Chem. **209**,1 (1965).
6. L.A.Currie: Anal.Chem. **40**, 586 (1968).
7. L.M. Schwartz: Anal.Chem. **55**, 1424 (1984).
8. Analytical Methods Committee. Recommendations for definition, estimation and use of the detection limit. Analyst, **112**, 199 (1987).
9. J.Mocák, A.M.Bond, S.Mitchell, G.Scollary: Pure and Appl. Chem. **69**, 297 (1997).
10. ISO-CD 11843-1: Decision Limit, Detection Limit, Capability of Detection; Terms and Definitions

foltin@fns.uniba.sk, foltin@kasr.elf.stuba.sk