

## 4.8 Vícerozměrná kalibrace

Podobně, jako jednorozměrná kalibrace, i vícerozměrná kalibrace se používá především v analytické chemii. Bude vysvětlena na příkladu spektroskopie: cílem je popis závislosti mezi blokem naměřených dat, např. spektrem  $X$ , a veličinami kalibračního modelu, např. koncentracemi  $C$  látek ve směsi. Aplikace totiž umožňuje stanovit hodnoty koncentrací jednotlivých složek v neznámém vzorku ( $c^*$ ) z jeho naměřeného spektra ( $x^*$ ). Existují dvě alternativy vícerozměrné kalibrace stejné jako u jednorozměrné, a to tzv. *klasická kalibrace* a *kalibrace inverzní*.

### 4.8.1 Klasická vícerozměrná kalibrace

Uplatňuje se v případě, kdy všechny sledované látky směsi jsou známy stejně jako látky, které s nimi mohou interferovat. Navíc jsou látky i interferenty fyzicky k dispozici. Pak jsou připraveny umělé standardy a proměřeny při uvažovaných  $m$  vlnových délkách. Matice absorpčních  $X$  je výsledkem přítomnosti  $p$  látek o koncentracích  $C$  s absorpčními koeficienty  $\beta$  dle vztahu

$$\begin{matrix} X & = & C & \beta & E_x \\ (n, m) & & (n, p) & (p, m) & (n, m) \end{matrix}$$

$$\begin{vmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{vmatrix} = \begin{vmatrix} c_{11} & \dots & c_{1p} \\ c_{21} & \dots & c_{2p} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{np} \end{vmatrix} \times \begin{vmatrix} \beta_{11} & \dots & \beta_{1m} \\ \beta_{21} & \dots & \beta_{2m} \\ \dots & \dots & \dots \\ \beta_{p1} & \dots & \beta_{pm} \end{vmatrix} + \begin{vmatrix} e_{11} & \dots & e_{1m} \\ e_{21} & \dots & e_{2m} \\ \dots & \dots & \dots \\ e_{n1} & \dots & e_{nm} \end{vmatrix}$$

kde  $x_{ij}$  je absorbance vzorku  $i$  při vlnové délce  $j$ ,  $c_{ir}$  je koncentrace látky  $r$  ve vzorku  $i$ ,  $\beta_{ij}$  je molární absorpční koeficient látky  $r$  při vlnové délce  $j$  a  $e_{ij}$  je chyba absorbance pro vzorek  $i$  při vlnové délce  $j$ .

Rovnice je řešitelná metodou nejmenších čtverců MNČ, přičemž se minimalizuje součet čtverců reziduí v matici  $E_x$ . Výsledkem je odhad neznámých parametrů  $\beta$  klasického kalibračního modelu

$$\beta = (C^T C)^{-1} C^T X$$

$$(p, m) \quad (p, n) \quad (n, p) \quad (p, n) \quad (n, m)$$

Každý  $r$ -tý řádek matice  $\beta$  reprezentuje absorpční spektrum látky  $r$  při jednotkové koncentraci a jednotkové délce kvyety čili molární absorpční koeficient.

Znalost  $\beta$  dovoluje stanovit koncentraci všech  $p$  látek přítomných v neznámém vzorku, tj.  $c^*$ , na základě jeho naměřeného spektra  $x^*$  dle vztahu

$$c^* = (\beta^{-1} x^*)$$

$$(p, 1) \quad (p, m) \quad (m, p) \quad (p, m) \quad (m, 1)$$

Ve srovnání s jednorozměrnou kalibrací má klasická vícerozměrná kalibrace následující výhody: (1) absorbanční měření  $X$ , použitá ke kalibraci, nemusí být selektivní vzhledem k  $C$ , čistá spektra stanovovaných látek se mohou totiž libovolně překrývat, (2) přesnost stanovení koncentrací neznámých vzorků je lepší, (3) dovoluje automatickou korekci na změny průběhu spektrální základní linie a (4) dovoluje odhadnout *čistá a reziduální spektra* stanovovaných látek. Znalost obou těchto spekter může vést k lepšímu pochopení studovaného kalibrovaného systému.

Hlavní nevýhodou klasické vícerozměrné kalibrace je omezení na případy, kdy všechny látky směsi musí být známy včetně interferentů a musí být fyzicky k dispozici. Vzhledem ke vzácnému splnění těchto předpokladů se klasická vícerozměrná kalibrace v praxi příliš nepoužívá. Naopak, značně se v praxi rozšířila tzv. *inverzní vícerozměrná kalibrace*.

### 4.8.2 Inverzní vícerozměrná kalibrace

Na rozdíl od klasické vícerozměrné kalibrace jsou požadavky pro aplikaci inverzní vícerozměrné kalibrace minimální. Vyžaduje znalost koncentrace sledované látky v kalibračních vzorcích. Kalibrační vzorky mohou být komplexní směsi se složitou maticí a neznámými interferenty, např. přírodní, potravinářské nebo petrochemické materiály. Pro vzorky jsou zaznamenána spektra absorbance  $X$  a sestaven kalibrační model, popisující vztah mezi koncentrací sledované látky  $c$  a absorpční  $X$  rovnicí

$$\begin{array}{c}
 \mathbf{c} \\
 (n, 1)
 \end{array}
 =
 \begin{array}{c}
 \mathbf{X} \\
 (n, m)
 \end{array}
 \begin{array}{c}
 \boldsymbol{\beta} \\
 (m, 1)
 \end{array}
 +
 \begin{array}{c}
 \mathbf{e}_c \\
 (n, 1)
 \end{array}$$

$c_1$	$x_{11}$	$\dots$	$x_{1m}$	$\beta_1$	$e_1$
$c_2$	$x_{21}$	$\dots$	$x_{2m}$	$\beta_2$	$e_2$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$c_n$	$x_{n1}$	$\dots$	$x_{nm}$	$\beta_m$	$e_n$

kde  $c_i$  je koncentrace sledované látky v kalibračním vzorku  $i$ ,  $x_{ij}$  je absorbance tohoto vzorku při  $j$ -té vlnové délce,  $\beta_j$  je regresní koeficient čili molární absorpční koeficient pro vlnovou délku  $j$  a  $e_i$  je koncentrační reziduum vzorku  $i$ , čili rozdíl mezi koncentrací stanovenou referenční metodou a koncentrací vypočtenou proložením kalibračních dat modelem.

Místo maticového zápisu lze kalibrační model prezentovat i vektorově

$$\mathbf{c} = \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \dots + \mathbf{x}_m \beta_m + \mathbf{e}_c$$

Odhad parametrů modelu metodou nejmenších čtverců MNC je dán vztahem

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{c}$$

Řešení existuje pouze v případě, když počet kalibračních vzorků  $n$  je větší než počet vlnových délek  $m$  spektra,  $m > n$ . Jinak nelze invertovat součin  $\mathbf{X}^T \mathbf{X}$  a tím ani odhadnout parametry  $\boldsymbol{\beta}$ . Moderní analytické přístroje umožňují naměřit spektrum při stovkách vlnových délek a zmíněný předpoklad je pak zásadním omezením použitelnosti metody. Byly proto navrženy dvě strategie jak postupovat, pokud uvedená podmínka není splněna:

1) Z původního spektra vhodnou metodou vybrat omezený počet vlnových délek tak, aby podmínka byla splněna.

2) Zkomprimovat původní spektra získaná při stovkách vlnových délek do formy latentních proměnných, a ty pak použít ke kalibraci namísto původních absorbančních proměnných.

### 1. Selekcce originálních proměnných

Klasickou metodou k selekci proměnných z původního výběru všech nezávislých proměnných spektra je postupná *kroková vícenásobná lineární regrese*. Alternativních metod je několik, přičemž nejnámější metodou je *genetický algoritmus*.

*Postupná kroková vícenásobná lineární regrese* má následující postup:

(i) V prvním kroku se využitím korelačního koeficientu v absorbanční matici  $\mathbf{X}$  nalezne proměnná  $x_j$ , která je nejvíce korelována s kalibrovanou koncentrací  $c$ . Pak se sestrojí jednorozměrný kalibrační model

$$\mathbf{c} = \beta_0 + \beta_j x_j + \mathbf{e}_c$$

Testuje se statistická významnost odhadnuté směrnice tohoto modelu  $\beta_j$ . Pokud je parametr  $\beta_j$  shledán statisticky významným, selekcce proměnných v modelu pokračuje.

(ii) Ve druhém kroku je vybrána ta proměnná  $x_j$ , která má nejvyšší parciální korelační koeficient s koncentrací  $c$ , a to je ta, která má nejvyšší korelační koeficient s vektorem reziduí  $\mathbf{e}_c$ , získaným z první regrese. Postup se nazývá *dopředná selekcce parametrů lineárního regresního modelu*. Pak se sestrojí lineární regresní model

$$\mathbf{c} = \beta_0 + \beta_j x_j + \beta_l x_l + \mathbf{e}_c$$

a testuje se statistická významnost obou odhadnutých regresních parametrů  $\beta_j$  a  $\beta_l$ . Nevýznamné členy jsou z rovnice eliminovány. Postup se pak nazývá *zpětná eliminace parametrů v lineárním regresním modelu*.

(iii) Dopředná selekcce a zpětná eliminace parametrů regresního modelu jsou opakovány tak dlouho, až model již nelze zlepšit přidáním nějaké proměnné a současně všechny proměnné zahrnuté v modelu jsou statisticky významné. Takové řešení je pak považováno za optimální.

*Genetický algoritmus* je obecný optimalizační nástroj, simulující proces přírodní evoluce, kde přežívá ten jedinec, který se nejlépe hodí do okolního prostředí. Tento jedinec se pak dále rozvíjí a adaptuje. Analogicky ve vícerozměrné kalibraci hledají genetické algoritmy optimální kombinaci původních proměnných. Začíná se od náhodně zvolených kombinací těchto proměnných. U počátečních kombinací se testuje schopnost popsat variabilitu

závisle proměnné koncentrace  $c$ . S pravdě-podobností odvozenou od této schopnosti jsou pak nejlepší řešení podrobena dvěma operacím: míšení a mutaci. *Míšení* je řízené promíchání nejlepších řešení z posledního cyklu. *Mutate* simuluje proces náhodné změny, která se v přírodě vyskytuje s nižší frekvencí. Celá procedura má iterativní charakter a je založena na srovnání vícenásobných lineárních modelů pro všechny testované kombinace.

## 2. Redukce originálních proměnných metodami s latentními proměnnými

Latentní proměnné jsou lineárními kombinacemi všech původních proměnných (viz kap. 4.5.1). Sloupcový vektor  $i$ -té latentní proměnné  $y_i$  je matematicky definován jako

$$y_i = w_{i1} x_1 + w_{i2} x_2 + \dots + w_{im} x_m + \dots$$

kde  $i$  nabývá hodnot od 1 do  $\min(n, m)$ . Koeficienty  $w_{ij}$  odpovídají zátěži  $j$ -té původní proměnné do  $i$ -té latentní proměnné a  $x_j$  je sloupcový vektor absorbancí, měřených při vlnové délce  $j$ . Soubor latentních proměnných dává matici latentních proměnných  $Y$ . Smyslem konstrukce latentních proměnných je zkomprimovat informaci obsaženou ve velkém počtu původních proměnných v absorbanční matici do několika málo nových proměnných. Důležitou vlastností latentních proměnných je jejich vzájemná ortogonalita čili kolmost. Tato vlastnost zaručuje, že metoda nejmenších čtverců MNČ, aplikovaná na soubor latentních proměnných, poskytne matematicky stabilní řešení a přítomnost či absence určité latentní proměnné neovlivní hodnoty regresních parametrů ostatních členů. Nejpoužívanější metodou konstrukce latentních proměnných je *metoda hlavních komponent* PCA a *metoda částečných nejmenších čtverců* PLS (anglicky Partial Least Squares):

**A. Využití metody hlavních komponent.** Nejběžnější metodou, umožňující konverzi původních proměnných na latentní, je *metoda hlavních komponent* PCA a vzniklé proměnné se nazývají *hlavní komponenty*. Kalibrace závisle proměnné  $c$  pomocí metody nejmenších čtverců MNČ, aplikovaná na malý, správně zvolený podsoubor hlavních komponent, se nazývá *regrese na hlavních komponentách* (Principal Component Regression PCR). Metodou hlavních komponent lze zdrojovou absorbanční matici měření  $X$  rozložit dle vztahu

$$X = V \Gamma W^T$$

(n, m)   (n, n)   (n, n)   (n, m)

kde  $V$  je matice *normovaných skóre*, diagonální matici vlastních čísel  $\Gamma$  nazýváme *maticí zátěží*, matice  $W$  pak zajišťuje transformaci původních proměnných na latentní. Vlastní čísla obsažená v  $\Gamma$  udávají důležitost jednotlivých vektorů matice  $V$  vzhledem ke zdrojové matici  $X$ , tj. příspěvek jednotlivých latentních proměnných  $y_i$  do celkové variability  $X$ :  $y_1$  popisuje nejvíce variability,  $y_2$  méně atd. Součin  $V$  a  $\Gamma$  dává matici nenormovaných skóre  $Y$

$$Y = V \Gamma$$

a  $i$ -tý sloupec matice  $Y$  obsahuje hodnoty skóre pro všechny kalibrační vzorky na  $i$ -té latentní proměnné. Jelikož matice  $Y$  má rozměr  $(n, n)$ , lze z ní nebo z podsouboru  $k$  sloupcových vektorů aplikací metody nejmenších čtverců sestrojít model pro koncentraci  $c$  dle vztahu:

$$c = Y q + e_c$$

(n, 1)   (n, k)   (k, 1)   (n, 1)

Tato rovnice má řešení

$$q = (Y^T Y)^{-1} Y^T c$$

(k, 1)   (k, n)   (n, k)   (k, n)   (n, 1)

Vedle metody hlavních komponent PCA je zde regrese metodou nejmenších čtverců MNČ druhým základním krokem v metodě regrese na hlavních komponentách PCR.

Protože několik málo  $k$  prvních hlavních komponent vysvětluje většinu variability v matici  $X$ , vyšší komponenty  $k+1$  atd. není třeba při konstrukci modelu vůbec uvažovat. Ty totiž popisují *šum v datech*.

Hledání a optimalizace počtu  $k$  latentních proměnných je jedním z nejdůležitějších úkolů ve vícerozměrné inverzní kalibraci. Jelikož model regrese na hlavních komponentách PCR není sestrojen na základě původních, ale latentních proměnných, je před stanovením koncentrace neznámého vzorku  $c^*$  z jeho spektra třeba vypočítat skóre vzorku na jednotlivých latentních proměnných. Výpočet se provede pomocí zátěží  $W$

$$y^{(T)} = x^{(C)} W$$

(1, k)   (1, m)   (m, k)

Získaná skóre se nakonec použijí k odhadu koncentrace stanovovaného vzorku

$$c^{(k)} = y^{(T)} q^{(k, 1)}$$

**B. Částečné nejmenší čtverce (PLS).** Hlavní nevýhodou metody regrese na hlavních komponentách PCR je fakt, že používá ke konstrukci modelu všech  $k$  prvních hlavních komponent. Jelikož některé z nich mohou popisovat zdroj variability v datech, který však není svázán s koncentrací  $c$ , může při použití takového modelu dojít k nesprávné predikci. Tento problém lze potlačit selekcí latentních proměnných na základě korelace s koncentrací  $c$  nebo aplikací metody částečných nejmenších čtverců PLS. Tato metoda se při konstrukci latentních proměnných zaměřuje na tu část variability ve spektrech  $X$ , která je korelována s koncentrací  $c$ . Popis činnosti vhodného iterativního algoritmu částečných nejmenších čtverců NIPALS obsahuje následující kroky:

- (1) Centrování zdrojové absorbanční matice  $X$  a kalibrované proměnné  $c$ .
- (2) Rozklad absorbanční matice  $X$  za použití vah  $w$  maximalizujících kovarianci  $X^T c$

$$X = c w_{iter}^T + E_X$$

(n, m) (n, 1) (1, m) (n, m)

kde  $iter$  značí číslo iterace. Rovnici lze řešit metodou nejmenších čtverců s výsledkem

$$w_{iter} = (c^T c)^{-1} X^T c$$

(m, 1) (1, n) (n, 1) (m, n) (n, 1)

Vektor  $w_{iter}$  se skládá z vah pro každou vlnovou délku  $j$ . V první iteraci je tento krok identický s klasickou vícerozměrnou kalibrací za předpokladu, že je známá koncentrace pouze jedné látky. Veličinu  $w_{iter}$  lze interpretovat jako spektrum dotyčné látky.

(3) Normalizace vah  $w_{iter}$ : první normalizovaný vektor vah  $w_1$  je úměrný váženému průměru centrovaných spekter  $X$ .

(4) Rozklad absorbanční matice  $X$  s cílem získat  $X$  skóre, označená  $Y$ . Jde o krok ekvivalentní prvnímu kroku v metodě regrese na hlavních komponentách

$$X = y_m w_{iter}^T + E_X$$

(n, m) (n, 1) (1, m) (n, m)

s řešením metodou nejmenších čtverců MNČ

$$y_{iter} = (w_{iter}^T w_{iter})^{-1} X w_{iter}$$

(n, 1) (1, p) (p, 1) (n, p) (p, 1)

Tento krok je srovnatelný s predikcí v klasické kalibraci. Protože vektor vah  $w_{iter}$  je funkcí  $c$ ,  $Y$  v metodě částečných nejmenších čtverců PLS, reprezentuje společnou část  $X$  a  $c$ .

(5) Stanovení vnitřního vztahu metody částečných nejmenších čtverců mezi  $y_{iter}$  a  $c$

$$c^T = q_{iter} y_{iter}^T + e_c$$

(1, n) (1, 1) (1, n) (1, n)

Rovnice je řešitelná metodou nejmenších čtverců MNČ a je ekvivalentní regresnímu kroku v regresi na hlavních komponentách PCR s výsledkem

$$q_{iter} = (y_{iter}^T y_{iter})^{-1} c^T y_{iter}$$

(1, 1) (1, n) (n, 1) (1, n) (n, 1)

Na rozdíl od metody regrese na hlavních komponentách PCR,  $q_{iter}$  se určí iterativní metodou.

(6) Stanovení  $X$  zátěží  $p_{iter}$ :

$$X^T y_{iter} p_{iter}^T \approx E_x$$

$$(n, m) \quad (n, 1) \quad (1, m) \quad (n, m)$$

a rovnice je řešitelná metodou nejmenších čtverců MNC

$$p_{iter} = (y_{iter}^T y_{iter})^{-1} X^T y_{iter}$$

$$(m, 1) \quad (1, n) \quad (n, 1) \quad (m, n) \quad (n, 1)$$

zátěže  $p_{iter}$  mají stejnou funkci jako váhy  $w_{iter}$  a jsou zapotřebí pouze ke korekci faktu, že  $y_{iter}$  nejsou známy a musely být nahrazeny odhadem.

(7) iter-tá aproximace spekter částečnými nejmenšími čtverci PLS a koncentrace je dána vztahy

$$E_x = X & y_{iter} p_{iter}^T$$

$$(n, m) \quad (n, m) \quad (n, 1) \quad (1, m)$$

$$e_c = c & q_{iter} y_{iter}$$

$$(n, 1) \quad (n, 1) \quad (1, 1) \quad (n, 1)$$

rezidua  $E_x$  a  $e_c$  přebírají roli původního  $X$  a  $c$  a celá procedura se iterativně opakuje od kroku (ii). To se provádí  $k$ -krát, resp. tak dlouho, dokud nedojde ke konvergenci. Konvergence nastává, když veškerá kovariance mezi  $E_x$  a  $e_c$  je popsána modelem.

Postup výpočtu koncentrací neznámých vzorků se provede následovně:

(a) Zkonstruuje se výsledný model částečnými nejmenšími čtverci PLS, tj. odhadnou se parametry  $\beta$ , popisující přímo vztah mezi  $c$  a  $X$

$$\hat{\beta} = \hat{W}^T (\hat{P} \hat{W}^T)^{-1} \hat{q}$$

$$(m, 1) \quad (m, k) \quad (k, m) \quad (m, k) \quad (k, 1)$$

kde  $\hat{W}$  obsahuje  $k$ -tý řádek vah  $w_{iter}$  a  $\hat{P}$  obsahuje  $k$  řádků zátěží  $p_{iter}$ .

(b) Bude se centrovat spektrum neznámého vzorku  $x^*$  použitím sloupcového průměru kalibračních dat.

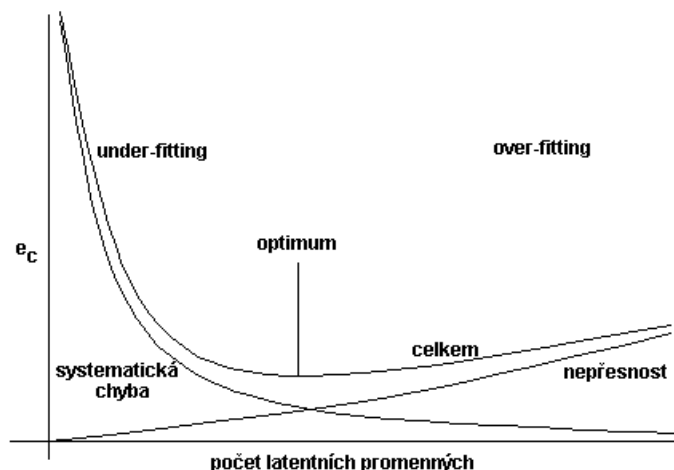
(c) Stanoví se koncentrace  $c^*$  tohoto vzorku dle vzorce

$$c^* = x^{*T} b \approx \bar{c}$$

$$(1, 1) \quad (1, m) \quad (m, 1) \quad (1, 1)$$

kde  $\bar{c}$  je průměrná koncentrace kalibračních vzorků.

**C. Validace optimálního modelu a určení jeho predikční schopnosti.** Ve více-rozměrné inverzní kalibraci se pod pojmem validace chápe jednak stanovení optimálního počtu proměnných  $k$ , jež mají být zahrnuty v modelu, a jednak odhad predikční schopnosti modelu. Cílem optimalizace počtu proměnných je nalézt vhodný kompromis mezi příliš těsným (over-fitting) a nedostatečným (under-fitting) proložením kalibračních dat. Nedostatečným proložením je míněna situace, kdy významná část koncentrace  $c$  není popsána modelem, protože počet latentních proměnných uvažovaný v modelu je příliš malý. Na druhé straně, příliš těsné proložení znamená, že je modelován i nesystematický rozptyl v datech (tj. šum), a to z důvodu zahrnutí příliš mnoha komponent do modelu. Následkem toho je zhoršená věrohodnost predikované koncentrace. Obě uvedené situace i hledaný kompromis, označený *optimum*, jsou na obr. 4.24.



Obr. 4.24 Predikční chyba versus počet latentních proměnných v modelu.

V praxi lze použít dva postupy stanovení optimálního počtu latentních proměnných:

(i) metodu příčné validace (cross-validation, CV), nebo

(ii) metodu predikčního testování externí validací, EV. Externí validace je postavena na použití tzv. tréninkového a monitorovacího souboru, které vznikají rozdělením původních kalibračních dat.

Nevýhodou druhého postupu je právě zmíněné dělení kalibračního souboru. To musí být provedeno tak, aby vznikly dva reprezentativní podsoubory. Dělení navíc vyžaduje, aby celkový počet kalibračních vzorků byl velký.

Cílem stanovení predikční schopnosti modelu je kvantifikace chyby, kterou je odhad koncentrace průměrně zatížen. Ke stanovení predikční schopnosti lze použít jak metodu příčné validace, tak predikčního testování externí validací. Jelikož náklady spojené s měřením monitorovacího souboru pro predikční testování bývají vysoké, častěji se užívá metoda příčné validace. Typickou charakteristikou sloužící ke kvantifikaci chyby predikce je *střední kvadratická chyba predikce*, kterou lze vyčíslit dvojitou technikou:

**(a) Příčnou validaci (CV).** Metoda je založena na simulaci predikční fáze při skutečné kalibraci. Kalibrační data jsou nejprve rozdělena na určitý počet skupin ( $gr$ ). V extrémním případě je  $gr = n$  a metoda se nazývá "odlož-jeden-mimo" příčná validace (leave-one-out validation). V prvním kroku je sestaven kalibrační model na základě všech dat kromě první skupiny. Tento model je použit na stanovení koncentrace vzorků z první skupiny. Konstrukce modelu a predikce se poté opakuje  $gr$ -krát, takže koncentrace každého vzorku je stanovena právě jednou.

Srovnání stanovených ( $c^*$ ) a referenčních ( $c_{ref}^*$ ) koncentrací kalibračních vzorků poskytuje odhad očekávané chyby koncentrace pro neznámé vzorky v budoucnu, střední kvadratickou chybu predikce RMSECV:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (c_i^* - c_{i,ref}^*)^2}{n}}$$

kde  $i$  je index kalibračního vzorku od 1 do  $n$  nebo do  $gr$ .

**(b) Predikčním testováním externí validací (EV).** Predikční testování lze použít jak k optimalizaci počtu proměnných v modelu, tak k odhadu průměrné chyby predikce. Kalibrační soubor je rozdělen do dvou skupin, na tzv. tréninkový a monitorovací pod-soubor. Kalibrační model je sestaven na základě tréninkových dat. Je stanovena koncentrace pro vzorky v monitorovacím souboru a vypočtena střední kvadratická chyba predikce RMSEP:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_t} (c_i^* - c_{i,ref}^*)^2}{n_t}}$$

kde  $i$  je index monitorovacího vzorku od 1 do  $n_t$ .

**D. Aplikace vícerozměrné kalibrace v analytické chemii.** Klasická vícerozměrná kalibrace se používá především v UV/VIS spektroskopii, a to v případech, kdy je cílem kvantifikovat množství látek ve směsi, obsahující známé složky. Na druhé straně, inverzní vícerozměrná kalibrace se využívá hlavně v oblasti NIR, UV/VIS, ICP a Ramanovy spektroskopie, a to při analýzách směsí obsahujících neznámé látky v neznámém množství.

Technika, která je na možnostech vícerozměrné kalibrace vysloveně závislá, je blízká infračervená spektroskopie NIR, pokrývající spektrální oblast 700 - 2500 nm. Na rozdíl od IR jsou totiž NIR spektra naprosto neselektivní a neumožňují aplikaci jednorozměrné kalibrace.

S rozvojem počítačů a vícerozměrné kalibrace vystoupila metoda NIR do popředí zájmu analytických chemiků a dnes se už běžně používá k analýzám v petrochemii, potravinářství i farmacii a chemii. V petrochemii se konkrétně využívá jako náhrada kalibrovaného motoru při stanovení oktanových čísel benzinů, dále při stanovení složení benzinů, destilačních charakteristik, hydroxylového čísla polyeter-polyolů apod. V potravinářství se aplikuje zejména při stanovení vody, proteinů a uhlohydrátů v potravinách a nápojích. Existuje však množství dalších specifických aplikací. Ve farmacii a chemii lze NIR využít k identifikaci surovin, produktů a obalových materiálů, dále ke stanovení obsahu hlavní látky, vody a polymorfismu, ale i ke stanovení množství mikroorganismů.