

## OPTIMALIZACE ANALYTICKÝCH POSTUPŮ POMOCÍ PLACKETTOVA-BURMANOVA PLÁNU

MIROSLAV HOLÍK

*Katedra teoretické a fyzikální chemie, Přírodovědecká fakulta Masarykovy univerzity v Brně, Kotlářská 2, 611 37 Brno*  
holik@chemi.muni.cz

Došlo 8.11.02, přepracováno 22.5.03, přijato 5.6.03.

Klíčová slova: úplný faktorový plán, Plackettův-Burmanův plán, confounding, víceúrovňový plán

### Obsah

1. Úvod
2. Úplný faktorový plán
3. Neúplný faktorový plán
4. Plackettův-Burmanův plán
5. Fiktivní (dummy) proměnné
6. Confounding
7. Víceúrovňové plány

### 1. Úvod

V roce 1975 napsal S. N. Deming v *Science*<sup>1</sup>: „Recent awareness of the finite character of both material and energy resources has stimulated a renewed interest in the optimization of reaction yields“. A opravdu, v sedmdesátých a osmdesátých letech se objevila řada publikací a článků popisujících nejen optimalizaci syntéz, ale také nastavení měřicích zařízení. Také u nás se optimalizaci, hlavně v průmyslu, věnovala značná pozornost. Tak již v roce 1968 vydalo Státní nakladatelství technické literatury knihu Jiřího Likeše *Navrhování průmyslových experimentů*. Časopis *Chemický průmysl* přinášel občas příklady matematického modelování technologických podmínek průmyslové syntézy, např.<sup>2-4</sup>. Do laboratoří se u nás optimalizace dostala hlavně knihou, kterou spolu s K. Doerfflelem vydal vynikající český chemometrik Karel Eckschlagger<sup>5</sup>. Tato kniha je překladem práce obou autorů, která vyšla<sup>6</sup> v němčině v roce 1981. Již před tím se ale v *Chemických listech* objevil krátký článek „Optimalizace analytických postupů“ věnovaný prof. Čútovi k 80. narozeninám<sup>7</sup>. V současné době kromě úspory „materiálu a energie“ hraje důležitou roli při optimalizacích také čas. Proto se příliš nepoužívají klasické postupy<sup>8</sup>, ale důležitou úlohu převzaly tzv. neúplné faktorové plány, z nichž nejčastěji používaný je Plackettův-Burmanův plán<sup>9</sup>. Protože není originální literatura<sup>9</sup> snadno dostupná a v řadě publikací je tento plán uváděn jen velmi stručně, případně s nepřesnostmi a chy-

bami<sup>5,6,10-13</sup>, dovoluji si nabídnout českému čtenáři tento článek. Než se ale dostanu k Plackettovu-Burmanovu plánu, je třeba stručně vysvětlit, co je to úplný faktorový plán.

### 2. Úplný faktorový plán

Pro jednoduchost uvažujme, že výsledek pokusu mohou ovlivnit dvě proměnné veličiny. Při syntéze to mohou být např. teplota a polarita rozpouštědla, při nastavení přístroje rychlost průtoku plynu a otevření nebo zavření přidavného zařízení. Proměnná může být buď plynule měnitelná (teplota, průtok plynu) nebo binární (otevřeno – zavřeno, polární – nepolární). Při plánování takového pokusu převádíme všechny proměnné na binární,  $-1$  a  $+1$ : u plynule měnitelných proměnných volíme tedy jen dva stavy: nominální a extrémní. Počet úrovní proměnné veličiny je tedy  $L = 2$ . Pro teplotu může být nominálním stavem třeba teplota místnosti a stavem extrémním teplota varu použitého rozpouštědla, při průtoku plynu zvolíme podle své zkušenosti dvě vhodné hodnoty: nízký průtok – vysoký průtok. Při popisování plánu pak nižší hodnotě přiřadíme číslo  $-1$  a vyšší hodnotě číslo  $+1$ . Přiřazení čísel veličinám binárním je libovolné – polární rozpouštědlo může mít  $+1$  a nepolární  $-1$  nebo naopak. Pro uvedený příklad vyplývá, že provedeme čtyři pokusy, a to pro nižší a vyšší hodnotu jedné proměnné, zatímco druhá proměnná je na své nižší nebo vyšší hodnotě. Číselně lze tento plán zapsat takto:

1. proměnná	2. proměnná	výsledek
$-1$	$-1$	$y_1$
$-1$	$+1$	$y_2$
$+1$	$-1$	$y_3$
$+1$	$+1$	$y_4$

U každého pokusu zaznamenáme výsledek ( $y_1$  až  $y_4$ ); podle povahy pokusu to může být výtěžek produktu, jeho čistota, nebo účinnost přístroje, jeho citlivost nebo dělicí (rozlišovací) schopnost. K tomu, abychom odhadli, která proměnná ovlivní výsledek a jak mnoho, použijeme lineární regresi: proměnná  $Y$  (výsledek) závisí na nezávisle proměnné  $X$  a citlivost této závislosti udávají hodnoty směrnice (parametrů)  $A$ . Matici  $X$  sestavíme ze sloupců plánu pro první a druhou proměnnou (tj. základní matice  $Z$ ) a předřadíme jim sloupec jedniček pro výpočet tzv. lokačního parametru (úsek na ose  $y$ ). Naměřené výsledky jsou  $Y$ , výsledky vypočtené z lineární regrese se označují  $\hat{Y}$ :

$$Y = X * A + E \quad \hat{Y} = X * A \quad (1)$$

kde

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

a  $\mathbf{E}$  je vektor náhodných chyb. Žádané parametry  $\mathbf{A}$  získáme převedením matice  $\mathbf{X}$  z jedné strany rovnice  $I$  na druhou tzv. pseudoinverzí.

$$\mathbf{A} = (\mathbf{X}' * \mathbf{X})^{-1} * \mathbf{X}' * \mathbf{Y} \quad (2)$$

Protože jsou sloupce matice  $\mathbf{X}$  ortogonální, je výsledkem násobení matic  $\mathbf{X}'$  a  $\mathbf{X}$  matice diagonální.

$$\mathbf{X}' * \mathbf{X} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad (\mathbf{X}' * \mathbf{X})^{-1} = \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix}$$

Ortogonalita vede k výraznému zjednodušení výpočtu parametrů  $\mathbf{A}$ . Rovnici (2) můžeme nahradit rovnicí (3).

$$\mathbf{A} = (1/m) * \mathbf{X}' * \mathbf{Y} \quad (3)$$

kde  $m$  je počet řádků matice  $\mathbf{X}$ , tj. počet experimentů. Výpočet je pak velmi jednoduchý – do transponované matice  $\mathbf{X}$  dosadíme místo jedniček odpovídající hodnoty výsledků  $y$ ; je to vlastně násobení matice  $\mathbf{X}'$  sloupcovým vektorem  $\mathbf{Y}$ . Potom čísla v řádcích sečteme a výsledek vydělíme počtem experimentů  $m = 4$ .

$$\mathbf{X}' * \mathbf{Y} = \begin{bmatrix} +y_1 & +y_2 & +y_3 & +y_4 \\ -y_1 & -y_2 & +y_3 & +y_4 \\ -y_1 & +y_2 & -y_3 & +y_4 \end{bmatrix} \begin{matrix} \text{suma} / 4 = a_0 \\ \text{suma} / 4 = a_1 \\ \text{suma} / 4 = a_2 \end{matrix}$$

Jak je patrné,  $a_0$  je součet všech výsledků  $y$ ,  $a_1$  a  $a_2$  pak představují směrnice (citlivosti) závislosti výsledků na proměnných 1 a 2.

Pozor! V literatuře<sup>5,6,9-13</sup> se často setkáváme s odlišným postupem v tom, že se parametry  $a_1$  a  $a_2$  vypočítávají jako rozdíly průměrů kladných a záporných hodnot  $y$ . Tedy pro výpočet např.  $a_1$  se bere  $(y_3+y_4)/2 - (y_1+y_2)/2$ . Je zřejmé, že takto vypočítané hodnoty  $a_1$  a  $a_2$  jsou dvojnásobné vzhledem k těm, které získáme lineární regresí a nejsou konzistentní s výpočtem parametru  $a_0 = (y_1 + y_2 + y_3 + y_4)/4$ .

Někdy se používají místo binárních proměnných  $-1$  a  $+1$  proměnné  $0$  a  $1$ . Pochopitelně je to možné a lineární regrese poskytne stejné směrnice jako při použití  $-1$  a  $+1$ . Protože však v tomto případě není matice  $\mathbf{X}' * \mathbf{X}$  diagonální, nelze použít zjednodušenou rovnici (3), ale je třeba počítat s pseudoinverzí (rovnice 2). Lokační člen není v tomto uspořádání průměrem všech výsledků, ale odpovídá výsledku měření při nastavení všech proměnných na nominální hodnoty.

Počet experimentů  $m$ , a tedy i velikost matice  $\mathbf{X}$ , roste s počtem proměnných  $n$  podle  $2^n$ .

Pro jednu proměnnou máme jen dva stavy – nižší a vyšší. Při dvou proměnných se dva stavy druhé proměnné přidávají ke dvěma stavům proměnné první, tak jak bylo ukázáno výše. Tři proměnné ( $A$ ,  $B$ ,  $C$ ) pak vyžadují  $2^3$ , tj. 8 experimentů. Tzv. úplný faktorový plán pro tento případ je v následující tabulce I.

Tabulka I

Úplný faktorový plán pro tři proměnné

	1	A	B	C	AB	AC	BC	ABC
	1	-1	-1	-1	1	1	1	-1
	1	1	-1	-1	-1	-1	1	1
	1	-1	1	-1	-1	1	-1	1
	1	1	1	-1	1	-1	-1	-1
	1	-1	-1	1	1	-1	-1	1
	1	1	-1	1	-1	1	-1	-1
	1	-1	1	1	-1	-1	1	-1
	1	1	1	1	1	1	1	1

Základní matici  $\mathbf{Z}$  zde představují sloupce  $A$ ,  $B$  a  $C$ . Hodnoty v řádcích nám říkají, jak se mají provést jednotlivé experimenty, např. v prvním experimentu se použijí u všech proměnných nominální hodnoty. V každém sloupci základní matice je stejný počet kladných a záporných znamének. To může sloužit jako kontrola správnosti sestavení plánu.

Další sloupce v tabulce dostaneme vynásobením sloupců základní matice tak, jak je uvedeno v záhlaví tabulky I. Tyto čtyři poslední sloupce představují vzájemné ovlivnění jednotlivých proměnných. Řešení podle tohoto plánu si ukážeme na příkladě z literatury<sup>8</sup>. Matici  $\mathbf{X}$  tvoří čísla ( $-1$  a  $+1$ ) z uvedené tabulky I. Sloupcový vektor  $\mathbf{Y}$  jsou výsledky experimentů (observations):

	11,8	$a(I)$	=	13,71
	9,9	$a(A)$	=	-0,64
	8,5	$a(B)$	=	-1,51
	20,9	$a(C)$	=	4,14
$\mathbf{X}' * \mathbf{Y} = \mathbf{X}' * \mathbf{Y}$	8,1	$a(AB)$	=	0,49
	18,3	$a(AC)$	=	-0,06
	16,2	$a(BC)$	=	-0,24
	16,0	$a(ABC)$	=	0,11

Kontrolu správnosti výpočtu lze provést takto: výsledky experimentů se umocní na druhou a zprůměrují:  $\Sigma (y^2) / 8 = 208,1562$ . Stejně číslo se musí dostat součtem druhých mocnin vypočítaných parametrů:  $\Sigma (a^2) = 208,1562$ .

Hodnoty parametrů u součinných proměnných jsou relativně malé, což napovídá, že se proměnné  $A$ ,  $B$  a  $C$  vzájemně neovlivňují. V takových případech je vhodné celý plán zjednodušit a snížit tak počet potřebných experimentů.

### 3. Neúplný faktorový plán

Při úplném faktorovém plánu se počet experimentů s každou další proměnnou zdvojnásobí. To může někdy vést až k ekonomicky a časově neuskutečnitelnému případu. Například při optimalizaci postupu pro silanizaci silikagelu pro kapalinovou chromatografii museli autoři posoudit významnost 23 proměnných<sup>10</sup>. V případě, že by jeden experiment trval jen 10 minut, bylo by třeba k provedení úplného faktorového plánu asi 160 roků nepřetržité práce. V takových náročných případech zanedbáváme možnost vzájem-

ného ovlivňování proměnných a uchylujeme se k neúplnému faktorovému plánu. Takových plánů existuje celá řada a lze je najít v literatuře<sup>8</sup>. Mezi nimi významné místo zaujímá tzv. Plackettův-Burmanův plán<sup>9</sup> a tomuto plánu je věnován další výklad.

#### 4. Plackettův-Burmanův plán

Vraťme se k tabulce I pro úplný faktorový plán se třemi proměnnými. Omezme ji jen na čtyři první sloupce a vyberme řádky 4, 6, 7 a 1. Dostaneme tabulku II.

Tabulka II  
Plackettův-Burmanův plán pro 3 proměnné

Řádek	I	A	B	C
4	1	1	1	-1
6	1	1	-1	1
7	1	-1	1	1
1	1	-1	-1	-1

Jestliže u prvního řádku základní matice  $Z$  (tj. sloupce A, B a C v tabulce II) provedeme cyklickou záměnu, tj. první prvek vyjmeme a vložíme za poslední, dostaneme řádek 2. Při další cyklické záměně dostaneme řádek 3; tím cyklické záměny končí (při další bychom dostali zase řádek 1). Poslední řádek vytvoříme ze samých -1. To je podstata Plackettova-Burmanova plánu pro počet úrovní  $L = 2$ , u kterého platí:

1. pro  $n$  proměnných budeme mít jen  $n+1 = m$  experimentů,
2. přitom musí být  $n + 1$  dělitelné 4, pokud není, přidá se k reálným proměnným potřebný počet fiktivních (dummy) proměnných,
3. pomocí cyklické záměny prvního řádku definovaného autory<sup>9</sup> (viz tab. III) se vytvoří základní matice  $Z$ , kde poslední řádek je tvořen samými -1.

Tabulka III  
Základní vektory Plackettova-Burmanova plánu

$m$	První řádek Plackettova-Burmanova plánu										
8	1	1	1	-1	1	-1	-1				
12	1	1	-1	1	1	1	-1	-1	-1	1	-1
16	1	1	1	1	-1	1	-1	1	1	-1	-1
20	1	1	-1	-1	1	1	1	1	-1	1	-1
24	1	1	1	1	1	-1	1	-1	1	1	-1

4. provede se  $m$  experimentů, při nichž se použijí proměnné v jejich nominální nebo extrémní hodnotě podle znamének u jedniček v jednotlivých řádcích,
5. před výpočtem citlivostí výsledků na jednotlivých proměnných se k základní matici  $Z$  předřadí vektor jedni-

ček a citlivosti ( $A$ ) se vypočítají jako směrnice z lineární regrese pomocí rovnice (2),

6. protože jsou sloupce základní matice  $Z$  ortogonální, lze použít i v této regresi zjednodušenou rovnici (3).

Příklad se třemi proměnnými, který byl vyřešen pomocí úplného plánu, bude nyní vypadat takto:

$$Z = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & -1 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$$

$$\begin{array}{r} +8,1 \\ +8,1 \\ +8,1 \\ -8,1 \end{array} \begin{array}{r} +18,3 \\ +18,3 \\ -18,3 \\ +18,3 \end{array} \begin{array}{r} +16,2 \\ -16,2 \\ +16,2 \\ +16,2 \end{array} \begin{array}{r} +11,8 \\ -11,8 \\ -11,8 \\ -11,8 \end{array} \begin{array}{r} +54,4/4 = +13,60 \\ -1,6/4 = -0,40 \\ -5,8/4 = -1,45 \\ +14,6/4 = +3,65 \end{array}$$

Parametry vypočtené podle úplného plánu (13,71; -0,64; -1,51 a 4,14) se mírně liší od hodnot získaných pomocí neúplného faktorového plánu Plackettova-Burmanova. V každém jednotlivém případě je proto třeba se rozhodnout, zda dát přednost úspoře času a materiálu na úkor přesnosti v získaných parametrech (citlivostech na proměnné).

#### 5. Fiktivní (dummy) proměnné

Zatím jsme se zabývali výpočty parametrů, ale nezjistili jsme nic o jejich přesnosti a statistické významnosti. Plackettův-Burmanův plán umožňuje i tato zjištění. Dosáhne se toho zavedením tzv. fiktivních proměnných (známých pod anglickým označením dummy). Tyto fiktivní proměnné umožňují prove-

dení více experimentů s různou kombinací nastavení reálných proměnných a doplňují reálné proměnné na počet potřebný pro sestavení Plackettova-Burmanova plánu. Tím, že fiktivní proměnné neovlivňují nijak průběh experimentu, měly by mít odpovídající parametry nulové. Odchylka vypočtených parametrů

od nuly umožňuje vytvořit si představu o přesnosti parametrů pro reálné proměnné.

Vraťme se k našemu příkladu s proměnnými  $A$ ,  $B$  a  $C$ . Místo plánu s  $m = 4$ , použijeme plán s  $m = 8$  a čtyři chybějící proměnné ( $d_1 - d_4$ ) budeme považovat za fiktivní. Matice  $\mathbf{X}$ , sestavená podle Plackettova-Burmanova plánu pro 8 experimentů, je v levé části tabulky IV.

Tabulka IV  
Plackettův-Burmanův návrh matice  $\mathbf{X}$  spolu s vektorem  $\mathbf{y}$

l	A	B	C	$d_1$	$d_2$	$d_3$	$d_4$	y
1	1	1	1	-1	1	-1	1	16,0
1	1	1	-1	1	-1	1	1	8,1
1	1	-1	1	-1	-1	1	1	18,3
1	1	1	1	1	1	1	1	8,5
1	1	-1	-1	1	1	1	-1	9,9
1	-1	-1	1	1	1	-1	1	20,9
1	-1	1	1	1	1	1	-1	16,2
1	-1	-1	-1	-1	-1	-1	-1	11,8

Experimentální uspořádání se řídí opět jen sloupci reálných proměnných  $A$ ,  $B$  a  $C$ , tak jako u úplného faktorového plánu. V prvním experimentu budou všechny tři proměnné na své extrémní úrovni, v posledním, osmém experimentu budou proměnné v nominálním stavu.

Pokud se soustředíme na nastavení těchto proměnných  $A$ ,  $B$  a  $C$ , můžeme jednotlivým experimentům přiřadit výsledky y uvedené výše u úplného plánu. Následující tabulka V uvádí srovnání výsledků výpočtu parametrů podle Plackettova-Burmanova plánu a podle úplného faktorového plánu.

Tabulka V  
Parametry vypočtené podle neúplného Plackettova-Burmanova a podle úplného plánu

Neúplný plán			Úplný plán	
parametr	a	$ a /s_a$	parametr	a
$a(l)$	13,7125	49,21	$a(l)$	13,7125
$a(A)$	-0,6375	2,29	$a(A)$	-0,6375
$a(B)$	-1,5125	5,43	$a(B)$	-1,5125
$a(C)$	4,1375	14,85	$a(C)$	4,1375
$a(d_1)$	0,0625	0,22	$a(AB)$	0,4875
$a(d_2)$	0,1125	0,40	$a(AC)$	-0,0625
$a(d_3)$	-0,4875	1,75	$a(BC)$	-0,2375
$a(d_4)$	0,2375	0,85	$a(ABC)$	0,1125

První čtyři parametry (tj. lokační a pro proměnné  $A$ ,  $B$  a  $C$ ) jsou v obou případech stejné; to je pochopitelné, pro-

tože pro tyto parametry jsou v obou plánech stejné kombinace nominálních a extrémních hodnot proměnných.

Všimněme si však parametrů, odpovídajících fiktivním proměnným. Pokud by měření v experimentech byla bez chyb, byly by tyto parametry nulové. Skutečně naměřené hodnoty můžeme tedy použít k odhadu rozptylu  $s_a^2$  podle rovnice (4).

$$s_a^2 = \sum (a(d) - 0)^2 / p \quad (4)$$

kde  $p$  je počet fiktivních proměnných (v našem případě = 4) a 0 představuje očekávanou hodnotu parametru  $a(d)$ .

Odmocninou rozptylu  $s_a^2$  je odhad standardní odchylky  $s_a$ . Když touto hodnotou vydělíme jednotlivé parametry (v absolutní hodnotě), dostaneme hodnoty testu  $t$ , které srovnáme s kritickou hodnotou  $t(\text{krit})$  Studentova rozdělení pro  $p$  stupňů volnosti na 95% nebo 90% hladině spolehlivosti. Výběr z kritických hodnot je v tabulce VI.

Tabulka VI  
Výběr z kritických hodnot Studentova rozdělení

p	1	2	3	4	5	6	7
$t(\text{krit})$ 95%	12,706	4,303	3,182	2,776	2,571	2,447	2,365
$t(\text{krit})$ 90%	6,314	2,920	2,353	2,132	2,015	1,943	1,895

V našem případě je počet fiktivních proměnných a tedy i počet stupňů volnosti 4. To znamená, že všechny proměnné, u nichž je  $|a|/s_a$  větší než 2,776, jsou statisticky významné z 95 % a ty, u nichž je tento poměr větší než 2,132, jsou významné na 90% úrovni.

Pozor! Někdy se v literatuře<sup>6</sup> setkáme s jinou rovnicí pro výpočet rozptylu, podobnou rovnici (4). Místo nuly je v čitateli průměrná hodnota  $a(d)$  parametrů a ve jmenovateli je  $p - 1$ . Tento způsob není správný, protože očekávaná hodnota parametrů u fiktivních proměnných je nula a pokud se nepracuje s průměrem, je správné použít ve jmenovateli počet všech použitých parametrů  $p$ .

V našem případě je

$$s_a^2 = \sum ((0,0625)^2 + (0,1125)^2 + (-0,4875)^2 + (0,2375)^2) / 4$$

a standardní odchylka  $s_a = 0,279$ . Z toho plyne, že všechny tři proměnné  $A$ ,  $B$  a  $C$  jsou statisticky významné;  $B$  a  $C$  na 95% a  $A$  na 90% hladině spolehlivosti. Všechny čtyři fiktivní proměnné ( $d_1 - d_4$ ) jsou statisticky nevýznamné. Tento druhý fakt je důležité zjištění. Všimněme si, že stejné hodnoty, které jsme použili pro výpočet standardní odchylky, představují v úplném plánu vzájemné ovlivnění jednotlivých proměnných. Kdyby k takovému ovlivnění docházelo, vyšel by pro odpovídající parametr test  $t$  jako významný.

Pro posouzení kvality uvedené regrese nelze použít korelační koeficient. Když dosadíme  $\mathbf{A}$  z rovnice (2) do rovnice (1), dostaneme rovnici pro přepočtené výsledky  $\mathbf{Y}$  na výsledky vypočítané z regrese  $\hat{\mathbf{Y}}$  pomocí tzv. „hat“ matice  $\mathbf{H}$ , rovnice (5).

$$\hat{Y} = X*(X'*X)^{-1}*X'*Y = H*Y \quad (5)$$

Protože je matice  $H$  v tomto případě jednotková, je  $\hat{Y} = Y$  a korelační koeficient, který charakterizuje podobnost obou vektorů  $\hat{Y}$  a  $Y$  je vždy roven jedné.

## 6. Confounding

Confounding je termín, který je snad lepší nepřekládat. Žádný český ekvivalent totiž nepopisuje odpovídající jev srozumitelně (confound = smíchat, poplést, zahanbiti).

Nejblíže skutečnému významu je anglický výklad: confound = mistake for another, čili omylem považovat za něco jiného. Co se tedy může omylem považovat za něco jiného?

Všimněme si sloupců pět až osm v Plackettově-Burmanově matici  $X$  (tabulka IV). Označíme-li každý ze sloupců této matice jako vektor s číslem, tedy  $v_1$  pro první sloupec, atd., můžeme psát:

$$v_5 = (-1) \cdot v_2 \cdot v_4, \quad v_7 = (-1) \cdot v_2 \cdot v_3, \quad v_8 = (-1) \cdot v_3 \cdot v_4$$

ale také

$$v_6 = v_2 \cdot v_3 \cdot v_4$$

To znamená, že významná hodnota parametrů pro vektory  $v_5$ – $v_8$  může být způsobena interakcí proměnných tvořících základní matici (vektory  $v_2$ – $v_4$ ). Pozor! Není to ale přesně to, co pozorujeme v úplném faktorovém plánu. Totiž, vznikne-li confounding vynásobením dvou vektorů, násobí se tento součin ještě  $-1$ . V případě confoundingu ze tří vektorů je tento násobitel  $+1$ . I když vzájemné ovlivnění čtyř a více proměnných je většinou málo pravděpodobné, dá se vypočítat, že střídání násobitelů  $-1$  a  $+1$  je pravidelné:  $-1$  pro sudý počet vektorů,  $+1$  pro lichý. U úplného faktorového plánu s  $m = 8$  platilo:

$$v_5 = v_2 \cdot v_3, \quad v_6 = v_2 \cdot v_4, \quad v_7 = v_3 \cdot v_4 \quad \text{a} \quad v_8 = v_2 \cdot v_3 \cdot v_4$$

a šlo o výsledek vzájemného ovlivnění proměnných. Tady nemohlo jít o mylné přiřazení, protože takto byly sloupce matice  $X$  konstruovány úmyslně. Při sestavování Plackettovy-Burmanovy matice se vzájemným ovlivňováním proměnných nepočítáme, ale pokud tam je, projeví se jako confounding. Toto srovnání úplného plánu a plánu Plackettova-Burmanova napovídá, že confounding bude existovat pro takový počet experimentů  $m$ , pro který lze sestavit úplný faktorový plán, tj. obecně pro  $m = 2^k$ , kde  $k = 2, 3, 4, 5, \dots$ . Protože se Plackettův-Burmanův plán dá sestavit pro  $m = 4k$ , kde  $k = 1, 2, 3, \dots$ , snadno zjistíme, že existují Plackettovy-Burmanovy plány, u nichž není confounding přítomen. Jsou to plány pro  $m = 12, 20, 24$  atd. Proto se doporučuje, máme-li 7 a méně proměnných, nepoužívat nejblíže možný plán, tj. pro  $m = 8$ , ale raději přidat více fiktivních proměnných a pracovat podle plánu pro  $m = 12$ .

## 7. Víceúrovňové plány

Plackett a Burman popsali<sup>9</sup> velmi stručně i plány pro případ, kdy nechceme mít experimenty jen na dvou úrovních

( $L = 2$ ), ale na třech, pěti a sedmi. V praxi se tento přístup prakticky nepoužívá. Je to proto, že se velmi zvyšuje počet potřebných experimentů; ten musí být totiž dělitelný  $L^2$ . Přitom počet proměnných je omezen počtem možných cyklických záměn základních sloupců (nikoliv řádků jako u  $L = 2$ ), který je  $(L^2-1)/(L-1)-1$ . Pro  $L = 3$  je tento počet 3, tj. celkový počet sloupců základní matice a tedy i proměnných je 4. V následující tabulce VII je základní Plackettův-Burmanův sloupec<sup>9</sup> uveden tučně; v posledním experimentu jsou opět všechny proměnné v nominální hodnotě (tj. 0).

Tabulka VII  
Plackettův-Burmanův plán pro  $L = 3$

Experiment	Proměnná			
	<i>m</i>	<i>A</i>	<i>B</i>	<i>C</i>
1	<b>0</b>	1	2	2
2	<b>1</b>	2	2	0
3	<b>2</b>	2	0	2
4	<b>2</b>	0	2	1
5	<b>0</b>	2	1	1
6	<b>2</b>	1	1	0
7	<b>1</b>	1	0	1
8	<b>1</b>	0	1	2
9	<b>0</b>	0	0	0

Z hlediska správnosti výpočtu směrnice pro jednotlivé proměnné  $A, B, C$  a  $D$  je opět jedno, zda se jako úrovně proměnných použijí 0, 1 a 2 nebo  $-1, 0$  a  $+1$ . Pouze druhý způsob však poskytne též správnou hodnotu lokačního parametru jako průměr všech hodnot výsledků  $y$ . V osmdesátých letech se snažil K. Jones<sup>14,15</sup> využít tento plán k zpřesnění výsledků při optimalizaci silanizace chromatografických materiálů. Bohužel nepostřehl, že tentokrát jde v Plackettově-Burmanově plánu o cyklickou záměnu sloupců a ne řádků. Pro matici  $9 \times 8$ , kterou dostal analogicky jako pro dvouúrovňový plán, se pak potýkal s vysvětlením opakovaných hodnot směrnice, navíc poněkud divně počítaných. Vyvážený tříúrovňový plán publikovali Massart a spol.<sup>16</sup>

Domnívám se, že nemá smysl snažit se zpřesňovat směrnice získané pomocí dvouúrovňového plánu. Ten slouží především k tomu, abychom si z velkého množství podezřelých proměnných vybrali ty, které skutečně významně ovlivňují studovaný proces. Zpřesnění, případně dokončení optimalizace navrhovaného postupu, je pak třeba provést jinou metodou, jako je třeba vícerozměrná regrese nebo během optimalizace simplexovou metodou<sup>17</sup>.

## LITERATURA

- Dean W. K., Heald J., Deming S. N.: *Science* 189, 805 (1975).

2. Vanko I., Komora L., Sakáloš Š.: Chem. Prum. 39/64, 245 (1989).
3. Vanko I., Komora L., Sakáloš Š.: Chem. Prum. 40/65, 175 (1990).
4. Polakovič M., Štefuca V., Bálež V., Michalková E., Welward L.: Chem. Prum. 40/65, 184 (1990).
5. Doerffel K., Eckschlager K.: *Optimální postup chemické analýzy*. SNTL, Praha 1985.
6. Doerffel K., Eckschlager K.: *Optimale Strategien in der Analytik*. VEB Deutscher Verlag fuer Grundstoffindustrie, Leipzig 1981.
7. Suchánek M., Šůcha L., Urner Z.: Chem. Listy 72, 1037 (1978).
8. Bennett C. A., Franklin H. L.: *Statistical Analysis in Chemistry and the Chemical Industry*. Wiley, New York 1954.
9. Plackett R. L., Burman J. P.: Biometrika 33, 305 (1946).
10. Jones K.: J. Chromatogr. 392, 1 (1987).
11. Stowe R. A., Mayer R. P.: Ind. Eng. Chem. 58, 36 (1966).
12. Abel M.: Trends Anal. Chem. 3, VII (1984).
13. Vindevogel J., Sandra P.: Anal. Chem. 63, 1530 (1991).
14. Jones K.: Int. Lab., 16, 32 (1986).
15. Jones K.: J. Chromatogr. 392, 11 (1987).
16. Van der Hayden Y., Khots M.S., Massart D.L.: Anal. Chim. Acta 276, 189 (1993).
17. Routh M. W., Swartz P. A., Denton M. B.: Anal. Chem. 49, 1422 (1977).

**M. Holík** (*Department of Theoretical and Physical Chemistry, Faculty of Science, Masaryk University, Brno*):  
**Optimization of Analytical Procedures with Plackett-Burman Design**

Reduced factorial design invented by Plackett and Burman in 1946, which reduces the number of experiments necessary for the determination of important variables in regression, is compared with the corresponding full factorial design. The calculation procedure is explained in detail using published data; some errors in literature are pointed out. Problems with confounding and how to avoid them are presented.