

## **ПРИЛОЖИМОСТ НА ХЕМОМЕТРИЧНИЯ ПОДХОД В МЕХАНИКАТА И МАТЕРИАЛОЗНАНИЕТО**

**Детелин Василев<sup>1</sup>, Мартин Иванов<sup>2</sup>, Николай Тончев<sup>1</sup>**  
[dvasilev@vtu.bg](mailto:dvasilev@vtu.bg), [martinivanov@abv.bg](mailto:martinivanov@abv.bg), [tontchev@vtu.bg](mailto:tontchev@vtu.bg)

<sup>1</sup> *Висше Транспортно Училище „Тодор Каблешков“ - София*

<sup>2</sup> *Нов Български Университет  
БЪЛГАРИЯ*

**Ключови думи:** *Моделиране на свойства, хемометриката, титанови сплави.*

**Резюме:** *Хемометриката се определя като съвременна научно-приложна дисциплина, обединяваща и систематизираща различни математически модели и формални техники за обработка на експериментални данни в задачите на химията и сродните ѝ науки. Тя включва методи на математическата статистика, многомерния анализ на данни, машинното обучение, изкуствения интелект и др. Доколкото хемометричният подход се е доказал като ефективен и целесъобразен при разкриването на типични вътрешни структурни връзки, той намира приложение и в редица други области на научните изследвания, дори и в такива, които привидно стоят твърде далеч от първоначалните източници на идеята. Хемометриката предлага разширени методи за анализ на данни и за оптимизиране на лабораторните изследвания. В този смисъл тя е елемент от развитието на тенденцията за създаване на интелигентни лабораторни системи. Представената публикация има за цел да обоснове приложимостта на хемометричния подход в задачите на материалознанието и механиката. Изложението се придружава от илюстративен пример.*

### **1. УВОД**

Разширяващият се интерес към хемометриката като методология засяга редица области, оказващи се сродни с първоначалния ѝ източник – химическите изследвания, но също и забележимо отдалечени от нея. Като подход хемометриката е ефективна във всички случаи, когато обектът на изследването е сложна система със скрити (ненаблюдаеми директно) вътрешни връзки. В този дух идеите на хемометриката биват възприети и в изследванията в областта на биологията, геологията, медицината ([2],[6]), както дори и в социалните и икономически науки, в които наблюдаемите данни могат да бъдат само косвени и да изискват продължителен исторически период (напр. психология, социология, финанси и т.н.). Съществено място хемометричният подход има и в осигуряването на качеството на технологични процеси.

Задачите на хемометриката се класифицират в две големи групи – качествени и количествени изследвания ([1],[2],[3]). В първата група попадат основно класификационни задачи – разпознаване на шаблони, идентификация), които се отбелязват като по-ранни в нейното развитие. Към втората група се отнасят

предиктивни задачи. Те се основават на методи на математическата статистика, представени преди всичко чрез различни регресионни модели и са насочени към изучаването на свойствата на изследваните обекти.

Една съществена особеност на хеометричните задачи е високата размерност на данните на изследването и съществената сложност на връзките между величините, описващи изследваните обекти.

Хеометриката използва концепцията на модела. Хеометричният модел най-общо описва връзката между експериментално наблюдавани независими променливи и зависими променливи, придружени от наличието на случаен шум:

$$Y = f(X) + E, \quad (1)$$

където  $X$  е обобщена форма на представяне на съвкупността от стойности на независимите променливи,  $Y$  представя съвкупността от наблюдавани стойности на зависимите променливи, а  $E$  съдържа адитивната компонента на грешката (шума).

Типична задача за хеометриката е задачата за т.нар. "калибровка" ([3], [4], [8], [9]). Това е приложна форма на класическата линейна многомерна регресия, в която се конструира математически модел, обвързващ показанията на инструмент със свойствата на изследваните обекти. Поради обобщаващия си характер и тенденцията в този модел да се установят и непряко наблюдаеми структурни особености на данните, калибровката като подход и термин е получила по-широка интерпретация и се използва и в случаи, когато не са налице технически средства на измерването. Най-съществената разлика между задачата за калибровка и линейната регресия от статистиката е нейната прагматична насоченост и изчислителната процедура, която допуска известни опростявания, изискващи се при оценяването на статистическата значимост на параметрите на модела.

Често калибровката се представя като процес, включващ няколко етапа, сред които са началната подготовка на данните, самата изчислителна процедура и проверката и обработката на резултата от моделирането.

Основните инструменти, които се използват от калибровката са три типа регресионни модели:

- Класическа многомерна линейна регресия (MLR);
- Регресия по главни компоненти (PCR);
- PLS-регресия.

Последователното прилагане на тези техники позволява да бъдат изяснени освен основните статистически значими връзки, така и някои латентни структурни характеристики на изследвания обект.

Методологическата схема на калибровката включва обикновено и два особено полезни (но незадължителни) етапа:

- Валидация – при който качествата на моделите се оценяват върху независим набор данни от типа на тези, с които е съставен модела (обучаващи данни).

- Диагностика – техника, при която се установява отсъствието на съществени предиктори в модела (непълнота на модела) или пък наличието на такива, чието участие в модела е несъществено („излишни” предиктори).

## **2. РЕГРЕСИОННИ И ПРОЕКТИВНИ МЕТОДИ ЗА АНАЛИЗ НА ДАННИ В ХЕМОМЕТРИКАТА**

**Класическият многомерен регресионен анализ (MLP)** ([3], [4], [8], [9]) (инструмент на множествената калибровка) решава задачата за калибровка посредством линейния модел:

$$Y = XB + E, \quad (2)$$

където  $B$  е матрицата на регресионните коефициенти, свързваща наблюдаваните стойности на предикторите  $X$  с резултата  $Y$ .

Този инструмент оценява количествено връзките между входните величини (предиктори) от  $X$  и резултативния параметър  $Y$ , но не позволява разкриването на вътрешната структура на изследвания обект. Основава се на стандартното приложение на метода на „най-малките квадрати“ (МНК). Методът се отличава с голяма популярност, но основните проблеми при неговото използване възникват в случаите на висока колинеарност на векторите, съдържащи наблюдаваните стойности на променливите, високо равнище на шума в наблюденията, недостатъчен брой на наблюденията спрямо независимите променливи и др.

**Регресия по главни компоненти (PCR, Principal Components Regression)** ([3], [4], [8], [9]). Това е мощно средство за преодоляване на проблемите, възникващи поради наличието на колинеарност между независимите променливи, съдържащи се в матрицата  $X$ . Методът реализира проекция на пространството, в което са представени многомерните данни в пространство със значително по-ниска размерност. Оригиналните данни, съдържащи се в матрицата  $X$  се трансформират в нови, представени чрез ортогонални дименсии. Това води до създаването на нови не корелирани латентни променливи (фактори), които са линейна комбинация на изходните предиктори:

$$t_a = p_{a1}x_1 + \dots + p_{aJ}x_J, \quad a = 1, \dots, A, \quad (3)$$

където  $t_a$  са новите латентни променливи, които представят главните не корелирани компоненти в модела,  $A$  е броят на включените в модела латентни променливи.

Представянето на независимите променливи, съдържащи се в матрицата  $X$  получава вида:

$$X = TP^t + E = \sum_{a=1}^A t_a p_a + E, \quad (4)$$

където:  $T$  – е матрицата на факторните стойности. Тя обединява векторите, които са взаимно ортогонални и представят проекциите на данните в новото пространство.

$P$  е матрицата на факторните тегла;

$E$  е матрицата на грешките в модела (шум, несъдържащ значима за модела информация).

В метода на главните компоненти проекцията се изпълнява само по стойностите на независимите „входни“ величини, представен в матрицата  $X$ . Стойностите на зависимите променливи от  $Y$  не се използват при съставянето на модела, което означава, че величината на ковариацията между предикторите и резултатните променливи не се взема под внимание. Съществено за решаваните задачи е, че този тип регресия дава възможност да се установят неяви, но обективно действащи фактори, които влияят върху съвместно изменящите се стойности на непосредствено наблюдаваните величини в матрицата  $X$ . При това броят на включените в модела фактори  $A$  може да бъде редуциран според степента на фактическото им влияние върху поведението на резултативните променливи, съдържащи се в матрицата  $Y$ . Това на практика може да доведе до преосмислянето и на резултата от моделирането, и на фактическата структура на изследваното явление, което е от значение в технологичните задачи.

Съществено развитие и обобщение на метода на главните компоненти представлява **методът на проекция по латентни структури (PLS-регресия)** ([3], [4], [7] [8], [9], [11]), който е един от най-популярните методи за калибровка в хемометриката.

Методът разчита на множество компоненти  $t$  (наречени „латентни“ вектори), които извършват едновременна декомпозиция на матриците  $X$  и  $Y$ , съдържащи предикторите и резултантните величини така, че да се максимизира ковариацията между съответните факторни променливи по  $X$  и  $Y$ .

Формалното представяне на съгласуваните отношения между групите фактори в този подход има следните особености:

- Декомпозицията на предикторите има вида  $X=TP^T$ , където  $T$  е матрицата на факторните стойности, а  $P$  –матрицата на факторните тегла.
- Матрицата  $Y$  се оценява чрез  $\hat{Y} = TBC^T$ , където  $B$  е диагонална матрица на регресионните тегла, а  $C$  е матрица на теглата за зависимите променливи от  $Y$ .

Както и в PCR-регресията колоните на  $T$  съдържат латентните вектори  $t$ . За да се постигне линейна комбинация на колоните на  $X$  и  $Y$ , при която ковариацията е максимална, PLS-регресията използва две множества от тегловни оценки  $w$  и  $c$ , участващи в съотношенията  $t = Xw$  и  $u = Yc$ , където  $w^T w = I$  и  $c^T c = I$  и  $t^T u$  е максимално (процедурата на изчисляване на латентните вектори е итеративна и е описана в ([7], [11]).

PLS-декомпозицията описва сложните вътрешни връзки на изследваното явление значително по-добре от другите регресионни модели, като при това използва по-малък брой главни компоненти. Това може съществено да допринесе за разкриването на неяви вътрешни структурни зависимости на изследвания технологичен обект (респ.материал).

В случая, когато се наблюдават едновременно няколко резултиращи променливи в матрицата  $Y$ , се използват две разновидности на PLS метода – PLS1 и PLS2В първата от тях за всяка от зависимите променливи в  $Y$  се съставя собствено проекционно подпространство. Във втората разновидност проекционното подпространство е едно за всички зависими променливи, включени в  $Y$ , т.е. PLS2 позволява съвместното моделиране на всички зависими променливи във всякаква възможна комбинация.

За проекционните методи PCR и PLS качеството и сложността на модела се определят в крайна сметка от броя на главните компоненти  $A$ , което в много случаи на практика се оказва трудна и нееднозначна задача. Пътищата за решаването ѝ са различни, но един от ефективите подходи е изследването на изменението на оценките на корена на средноквадратичната грешка RMSE. Подходът е преди всичко прагматичен. В тези случаи се търси или минимума на оценката RMSE, или точката, в която темпа на промяната ѝ рязко намалява. Параметърът RMSE, наред с коефициента на регресия  $R$  между наблюдавани и моделирани стойности за  $Y$  (респ.коефициента на детерминация  $R^2$ ) е сред най-използваните критерии, определящи качеството на регресионния модел ([8], [9]):

$$RMSE_k = \sqrt{\frac{1}{I} \sum_{i=1}^I e_{ki}^2}, \quad (5)$$

където  $e_{ki}$  е установената грешка между наблюдаваната и моделираната стойност на  $k$ -тата зависима променлива (отклик) за  $i$ -то наблюдение.  $I_p$  е общият брой на наблюденията в експеримента.

### 3. СОФТУЕРНИ ИНСТРУМЕНТИ, ИЗПОЛЗВАНИ В ПРАКТИКАТА НА ХЕМОМЕТРИКАТА

Популярността на хемометричния подход е довело до разработването на разнообразни специализирани софтуерни приложения, които облекчават и правят значително по-ефективен процеса на прилагането му. Сред най-известните и

използваните приложения следва да бъдат отбелязани The Unscrambler ([12]), Simica-p, Polytex ([13]), Chemometrics Toolbox за Matlab и др. The Unscrambler е мощен пакет за многомерен анализ и планиране на експеримента с над 30-годишна история. Притежава удобен интерфейс, множество функционалности, позволяващи подготовка на данни, изпълнение на разнообразни аналитични задачи (PCA/PCR/PLSR) и за визуализиране на резултатите. Широко разпространени и масово използвани са също и допълнения (add-ins) към MS Excel, предназначени също за многомерен анализ и визуализация на експериментални данни – Multivariate Analysis Add-In на Хемометричния център на университета в Бристол ([14]), XLStat. ([10]). Почти всички продукти са комерсиални и изискват притежаването на лиценз от потребителя.

#### 4. ИЛЮСТРАТИВЕН ПРИМЕР ОТ ОБЛАСТТА НА ИЗСЛЕДВАНЕ НА МЕХАНИЧЕСКИ СВОЙСТВА НА СПЛАВИ

Приведените по-горе разсъждения, свързани с приложението на регресионни и проективни методи в хемометриката ще бъдат илюстрирани с кратък пример. Данните за примера включват 55 наблюдения на връзка между химически състави и свойства на титанови сплави и са получени като ограничена извадка от съществуваща база от данни за проведени лабораторни експерименти. Предложеното тук решение демонстрира резултати от прилагането на трите метода MLR, MCR и PLS, и е извършено експериментално с използването на специализирано за целта софтуерно разширение на MS Excel – XLSTAT (Ver.10.4). Следва да се вземе под внимание, че примерът е преди всичко онагледяващ и за по-сериозно обобщаване на резултатите от анализа са необходими разширени и задълбочени изследвания.

Експерименталните данни включват стойности на девет предиктора, представляващи процентния състав на легиращите елементи на титанова сплав: Al, Mo, Sn, Zr, Cr, Fe, V, Nb и O в сплавта, както и на две зависими от този състав механични характеристики границата на провлачване  $R_{p02}$  и относителното удължение  $E/elongation/$ . Сред всички механични показатели, това са двете най-важни за проектирането характеристики от гледна точка на съпротивлението на материалите и материалознанието [15].

##### Приложение на метода на многомерната линейна регресия (MLR)

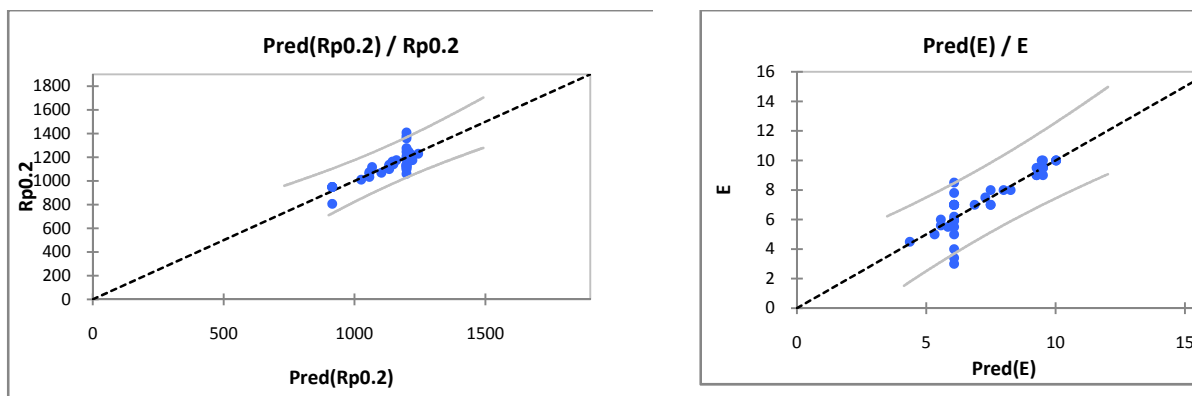
Корелационният анализ на наличните от експеримента данни показва като цяло не висока степен на корелация между стойностите на повечето предиктори. По-високи стойности на корелация са установени между елементите Mo-Nb ( $r(Mo,Nb) = 0,676$ ), Sn-Nb ( $r(Sn,Nb) = -0,565$ ), Zr-Nb ( $r(Zr,Nb) = -0,547$ ), Cr-Nb ( $r(Cr,Nb) = -0,609$ ), Mo-V ( $r(Mo,V) = -0,628$ ), Mo-V ( $r(Mo,V) = -0,749$ ), Cr-Sn ( $r(Cr,Sn) = 0,596$ ), Al-Zr ( $r(Al,Zr) = -0,534$ ). Корелацията между зависимите променливи е значително по-висока: за  $R_m-R_{p0}$  ( $r(R_m,R_{p0}) = 0,984$ ),  $R_m-E$  ( $r(R_m,E) = -0,806$ ),  $E-R_{p0}$  ( $r(E,R_{p0}) = -0,819$ ).

Регресионният анализ е изпълнен по стандартна схема. Най-съществените показатели за качеството на моделите по зависимите променливи са показани в таблица 1.

Таблица 1: Оценки на модела MVR за зависимите променливи,  $R_{p02}$  и  $E$ .

	$R_{p02}$	$E$
$R^2$	0,633	0,729
Adjusted $R^2$	0,535	0,657
RMSE	81,434	1,136

На фиг.1 са изобразени диаграмите „предсказано/наблюдавано” за четирите оценявани механични параметъра за модела MLR.



Фиг. 1: Диаграми „предсказано/наблюдавано” за зависимите променливи  $R_{p0.2}$ ,  $E$  за модела MLR.

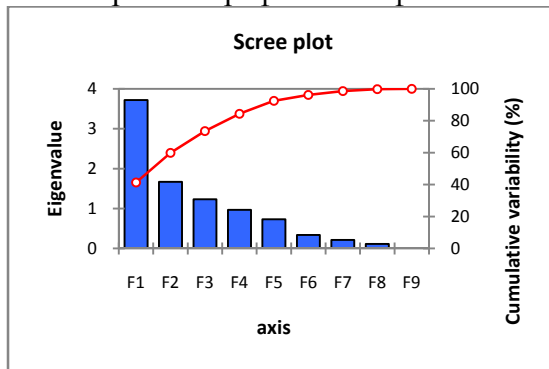
### Резултати от моделирането с PCR - регресията.

Табл. 2 илюстрира основните резултати от анализа по главни компоненти на данните (PCA), съдържащи се в матрицата  $X$ . От таблицата се вижда какъв е процентът на обяснената дисперсия по  $X$  при въвеждането на съответния брой главни компоненти в модела. При използване на първите шест или седем компонента (F6, F7) се постига високо качество на апроксимацията

Таблица 2: Резултати от PCA-анализа на данните за предикторите

	F1	F2	F3	F4	F5	F6	F7	F8	F9
Eigenvalue	3,719	1,672	1,234	0,970	0,731	0,338	0,214	0,113	0,010
Variability (%)	41,322	18,574	13,712	10,781	8,118	3,753	2,375	1,254	0,111
Cumulative %	41,32	59,897	73,608	84,389	92,507	96,260	98,635	99,889	100,00

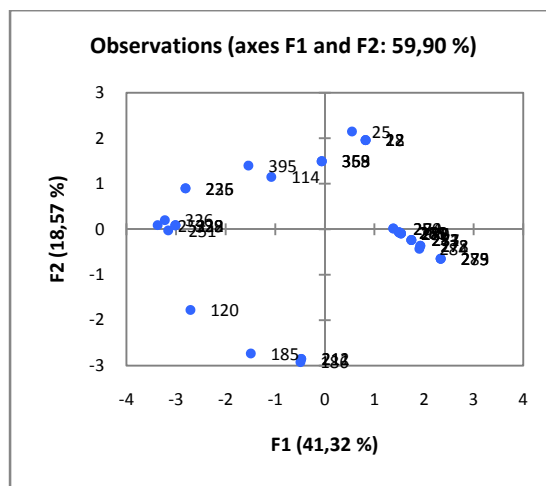
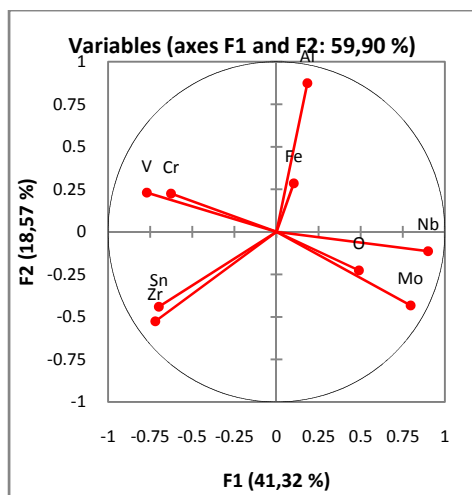
Резултатите от табл. 2 са изобразени графично на фиг. 2.



Фиг. 2: Собствени стойности и кумулативна обяснена дисперсия по факторите F1 до F9 в MLR-модела.

От показаните в таблица 2 и от фиг. 2 резултати се вижда, че добра апроксимация на входните данни се постига чрез въвеждането на значителен брой факторни променливи (7 или 8), което не дава възможност да се търсят изразени структурни връзки в данните за предикторите  $X$ . Очевидно PCR-моделът няма потенциал да доведе до съществено съкращаване на броя на предикторите.

На фиг. 3(a) е показана диаграмата на корелационните коефициенти между предикторите  $X$  и първите два фактора. Проекцията на изходните стойности на предикторите във факторно пространство, определено от първите две главни компоненти F1 и F2 е представена на диаграмата от фиг. 3(b).



(a)

(b)

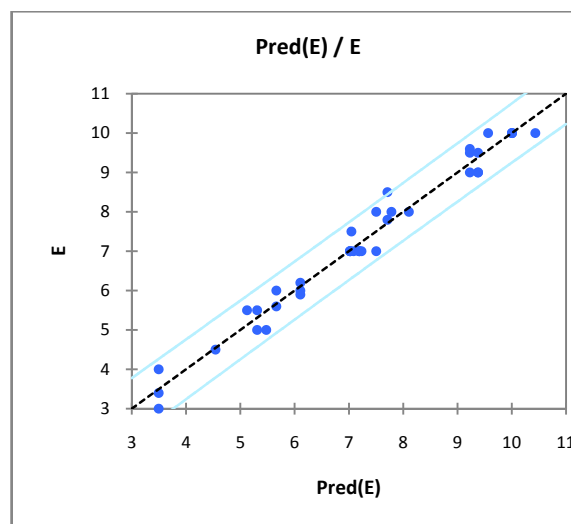
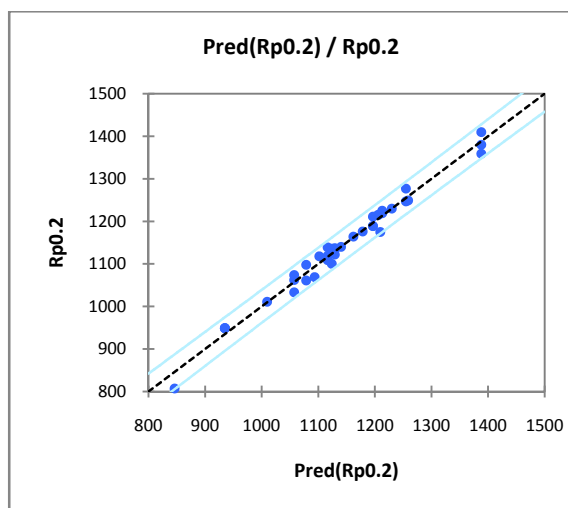
Фиг. 3: Диаграми на корелацията на предикторите спрямо факторите F1 и F2 (a) и на проекцията на наблюденията върху осите F1 и F2 (b).

Очевидно е оформянето на няколко групи от наблюденията, като групирането е твърде вероятно да е свързано с технологията на получаване на състава на сплавта. Забележима е положителната корелация между наличието на елементите Mo и Nb в сплавта и отрицателната им корелация с елементите V и Cr.

Таблица 3: Оценки на качеството на PCR-моделите по зависимите променливи  $R_{p0.2}$  и  $E$ .

	$R_{p0.2}$	$E$
$R^2$	0.982	0.974
Adjusted $R^2$	0.976	0.966
RMSE	18.505	0.358

Диаграмите "предсказано/наблюдавано" за четирите моделирани зависимы променливи са показани на фиг. 4. Очевидно е разсейването в наблюдаваните експериментални стойности на параметрите  $E$  и  $RA$ , което може да се обясни с несъвършенства на лабораторното изследване.



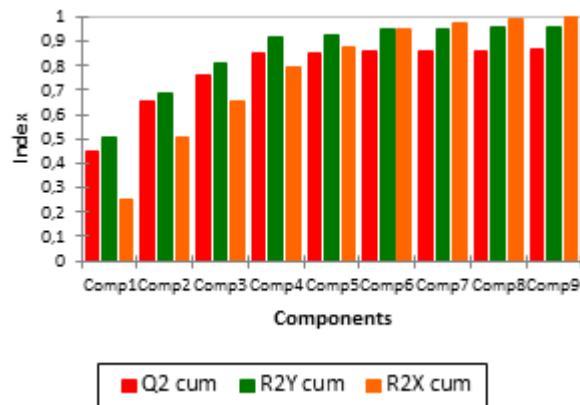
Фиг. 4: Диаграми „предсказано/предсказано“ за зависимите променливи  $R_{p0.2}$  и  $E$  за модела PCR.

### Резултати от моделирането с PLS-метода.

Основните резултати, оценяващи качеството на моделирането на връзката между предикторите и зависимите променливи от модела PLS са показани в таблица 4.

Таблица 4: Качествени показатели на PLS-модела според броя на компонентите

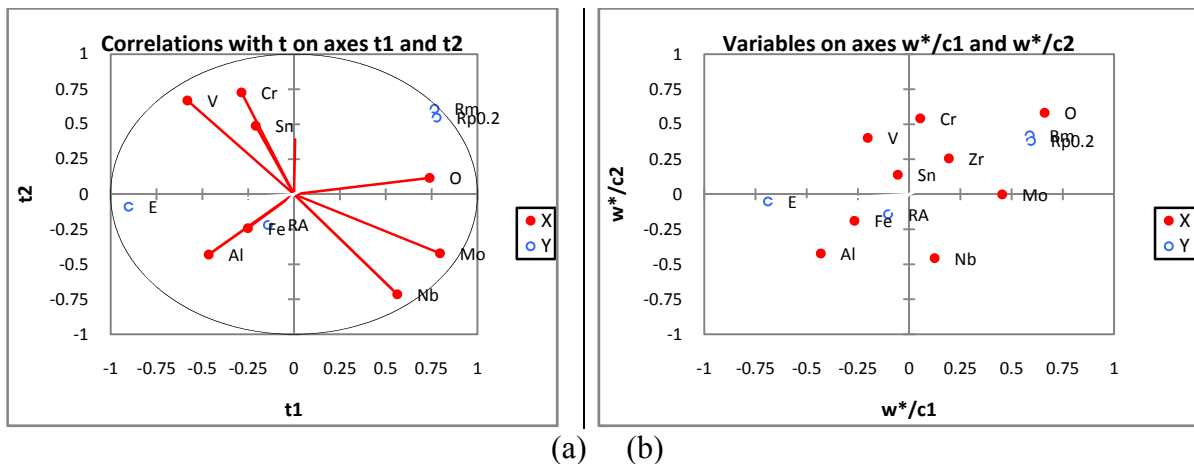
Index	Comp1	Comp2	Comp3	Comp4	Comp5	Comp6	Comp7	Comp8	Comp9
Q2 cum	0,452	0,655	0,757	0,847	0,855	0,857	0,862	0,863	0,867
R2Y cum	0,506	0,688	0,814	0,915	0,927	0,947	0,953	0,955	0,960
R2X cum	0,249	0,506	0,653	0,792	0,877	0,949	0,976	0,989	1,000



Фиг. 5: Качество на PLS-модела според броя на компонентите - визуализация.

Кумулативният индекс  $Q^2$  оценява с общото качествена модела, в който са включени първите компоненти от анализа, Параметрите  $R^2Y$  и  $R^2X$  също са кумулативни и оценяват степента на корелация на предикторите (X) и (Y) с генерираните от метода компоненти. В конкретния случай се наблюдава едно съществено намаляване на темпа на нарастване на тези показатели след шестия и седмия компонент, което насочва вниманието към един целесъобразен компромис с размерността на модела при нищожна загуба на предиктивните му качества.

На фиг. 6(a) е показана корелацията на предикторите с първите два компонента  $t_1$  и  $t_2$ . На фиг. 6(b) е изобразена графика (карта), която представя общите отношения между променливите в модела. Тя изобразява предикторите в системата на векторите на теглата  $w^*$  и зависимите променливи – в системата на векторите  $c1$  и  $c2$ .



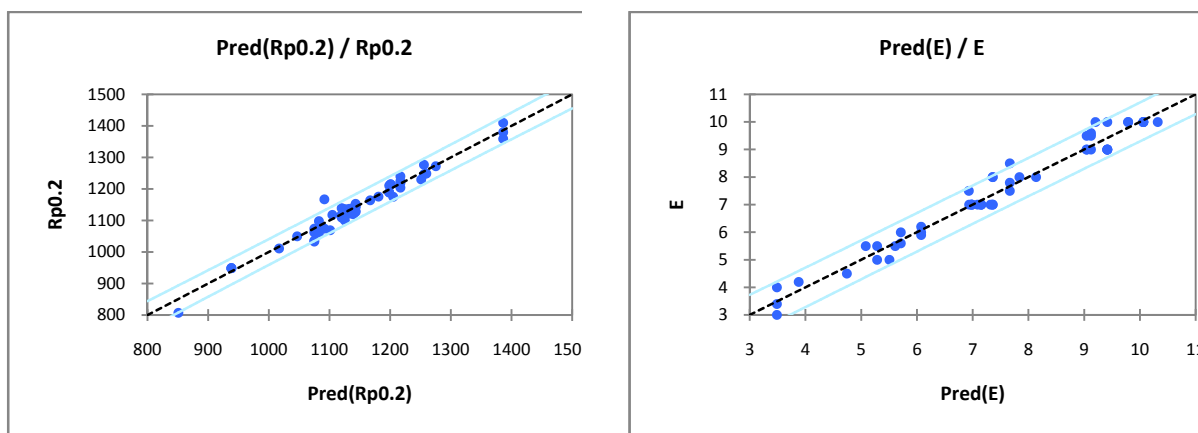
Фиг. 6: Корелация на предикторите с компонентите  $t_1$  и  $t_2$  (a). Проекции на предикторите X и зависимите променливи Y в координатната система на теглата  $w^*$  и  $c$  (b).



Фиг. 7 показва диаграми „наблюдавано/предсказано” за четирите зависими променливи (механични параметъра) в модела. Вижда се добрата корелация на моделираните и наблюдаваните стойности за  $R_m$  и за  $R_{p0}$  и по-слабата за  $E$  и  $RA$ , причина за което може да бъде начинът на провеждане на лабораторния експеримент.

Таблица 5 : Оценки на качеството на PLS-апроксимацията за зависимите променливи.

	$R_{p02}$	$E$
$R^2$	0.973	0.966
Std.deviat.	21.769	0.378
RMSE	19.679	0.340



Фиг.7: Диаграми „наблюдавано/предсказано” за зависимите променливи  $R_{p02}$  и  $E$  за модела PLS.

Проведеният изчислителен експеримент показва, че използването на PCR и PLS-регресията дава съществени предимства в точността на апроксимацията, както и по отношение на разкриването на вътрешната структура и връзки между изходните предиктори от матрицата  $X$ . Последователното изпълнение на методите и сравняването на резултатите им има смисъл в процеса на анализ на физическото явление и изясняване на характеристиките му. Следва да се обърне внимание на факта, че оценките за качеството на PCR и на PLS-модела са близки, но вторият има предимства по отношение на коректното отчитане на ковариацията между предикторите и резултативните променливи, както и по отношение на понижаването на размерността на пространството, в което се представят некорелираните латентни променливи, описващи изследваните свойства и връзки.

## 5. ЗАКЛЮЧЕНИЕ.

Хемометричният подход очевидно излиза извън границите на първоначалната си приложна област. Той представлява интерес с прагматичната си насоченост и с възможността да се определи целесъобразна и ефективна технология за обработка на лабораторни и експериментални резултати. В този смисъл отбелязаният подход има своето място в провеждането на експериментални изследвания в областта на техниката механиката и материалознанието.

## ЛИТЕРАТУРА:

- [1] Matthias Otto, Chemometrics. Statistics and Computer Application in Analytical Chemistry, John Wiley & Sons, Apr 9, 2007.
- [2] Tauler R., Beata Walczak, Steven Brown, Comprehensive Chemometrics, 1st Edition, Chemical and Biochemical Data Analysis, 2009.
- [3] Richard G. Brereton, Applied Chemometrics for Scientists, Willey, 2007.

- [4] Practical Guide To Chemometrics, Second Edition, edited by Paul Gemperline, CRC Press, Taylor& Francis Group, Elsevier Inc., 2006.
- [5] Wehrens, Ron, Chemometrics with R, Multivariate Data Analysis in the Natural Sciences and Life Sciences, Springer, 2011.
- [6] Wold Svante, Chemometrics; what do we mean with it, and what do we want from it?, Chemometrics and Intelligent Laboratory Systems 30 (1995) 109-115
- [7] Wold Svante, Michael Sjöström, Lennart Eriksson, PLS-regression: a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems, Volume 58, Issue 2, 28 October 2001, Pages 109–130
- [8] <http://rcs.chemometrics.ru/english.htm>
- [9] <http://www.chemometrics.ru/materials/articles/>
- [10] <https://help.xlstat.com>
- [11] <https://www.utdallas.edu/~herve/Abdi-PLSR2007-pretty.pdf>
- [12] <http://www.camo.com/rt/Products/Unscrambler/unscrambler.html>
- [13] <http://www.polytec.com/>
- [14] <http://www.chm.bris.ac.uk/org/chemometrics/addins/>
- [15] Malinov, S., Sha, W., Software products for modeling and simulation in material science, Computational Material Science, vol.28, 2003, pp.179-198.

## APPLICABILITY OF CHEMOMETRIC APPROACH IN MECHANICS AND MATERIALS SCIENCE

Detelin Vasilev<sup>1</sup> Martin Ivanov<sup>2</sup>, Nikolai Tonchev<sup>1</sup>  
 dvasilev@vtu.bg, martinivanov@abv.bg, tontchev@vtu.bg

<sup>1</sup>*Todor Kableshkov University of Transport - Sofia*  
<sup>2</sup>*New Bulgarian University*  
**BULGARIA**

**Key words:** Modeling properties chemometrics, titanium alloys.

**Abstract:** Chemometrics is defined as a modern scientific and applied discipline integrating and systematizing various mathematical models and formal techniques for experimental-data processing in problems of chemistry and its related sciences. It includes methods of mathematical statistics, multidimensional data analysis, machine learning, artificial intelligence, etc. Since the chemometric approach has proven to be effective and expedient in revealing typical internal structural connections, it is applied in many other areas of research and even in such that seem to be too far away from the initial sources of the idea. Chemometrics offers expanded methods for data analysis and for optimization of laboratory research. In this sense it is a part of the development of the trend to create intelligent laboratory systems. The presented paper aims to justify the applicability of the chemometric approach to problems of material sciences and mechanics. The representation is accompanied by an illustrative example.