

Feature Extraction Methods for Electronic Nose Responses

A.A. Kalinichenko^{*†}, L.U. Arseniyeva[†], U.P. Butsenko[‡]

[†]National University of Food Technologies, 68 Volodymyrska str., Kyiv, 01601, Ukraine;

*e-mail: asya.kalini4enko@gmail.com

[‡]NTUU "Igor Sikorsky Kyiv Polytechnic Institute", 37 Pobedy ave., Kyiv, 03056, Ukraine;

Received: September 12, 2017; Accepted: December 06, 2017

DOI: 10.17721/moca.2017.112-122

This paper reports about feature extraction methods from the original dynamic responses of electronic nose sensors. The aim of feature extraction methods in this research is to extract the information about quantitative changes of oils oxygenated volatile compounds. This paper presents a novel and well-known approaches and algorithms of feature extraction, such as extraction from the original responses, log-normal and polynomial curve fitting parameters, recognition of odor fingerprints based on selected times of sensor array responses as representatives. Optimized mathematical models and new algorithms for data extraction and compression into a structured part is shown to significantly enhance the performance of subsequent the partial least squares regression. The most accurate calibration models pretended for prediction of quality (peroxide value) of new vegetable oils samples were obtained with log-normal parameters such as asymmetry and width of response curves and with new geometric parameters for recognition the odor fingerprints such as star coefficient, the coefficient of non-smoothness and the perimeter of a polygon.

The proposed approach of evaluation the peroxide value of vegetable oils using algorithm of recognition the odor fingerprints in combination with partial least squares regression allows for fast analysis of samples quality with a relative error of about 8%.

Keywords: response curve, feature extraction, electronic nose, curve fitting, quartz-microbalance sensors, partial least squares regression

Підходи до вилучення параметрів з відгуків електронного носа

A.O. Kalinichenko^{*†}, L.Yu. Arseniyeva[†], Yu.P. Butsenko[‡]

[†]Національний університет харчових технологій, вул. Володимирська 68, Київ, Україна, 01601;

*e-mail: asya.kalini4enko@gmail.com

[‡]НТУУ "Київський політехнічний інститут ім. Ігоря Сікорського", просп. Перемоги 37, Київ, Україна, 03056

Надійшла: 12 вересня 2017 р; Прийнята: 06 грудня 2017 р

DOI: 10.17721/moca.2017.112-122

В статті розглядаються підходи до вилучення параметрів з вихідних динамічних відгуків сенсорів електронного носа з метою отримання інформації щодо кількісної зміни летких продуктів окиснення ліпідів. Запропоновано нові та оптимізовані відомі підходи та алгоритми вилучення інформативних параметрів: з вихідних цифрових сигналів одичних сенсорів, з моделей сигналів, апроксимованих лог-нормальною та поліноміальною функціями, з візуального образу запаху, побудованого з використанням скорочених вихідних відгуків масиву сенсорів. Оптимізовані математичні моделі та запропоновані алгоритми вилучення та стиснення змістовних даних демонструють значне підвищення ефективності наступного використання регресійного методу проєкції на латентні структури. Найбільш точні калібрувальні моделі та придатні для прогнозування показника якості олій (пероксидного числа) за новими зразками побудовані із використанням параметрів, вилучених з лог-нормальної моделі (коефіцієнта асиметрії та ширини розподілу даних кожного сенсора масиву) та геометричних параметрів оцінки зміни візуальних образів запахів зразків (коефіцієнта зірчастості, коефіцієнта негладкості та периметра фігури). Запропонований спосіб визначення пероксидного числа олій із використанням алгоритму оцінки геометричної форми візуального профілю запаху та методу проєкції на латентні структури дасть змогу провести експрес-аналіз якості зразків з відносною похибкою вимірювання не більше 8%.

Ключові слова: крива відгуку, вилучення параметрів, електронний ніс, апроксимація кривої, п'єзокварцові сенсори, проєкція на латентні структури

Використання багатоканальних сенсорних масивів з перехресною чутливістю типу електронний ніс дає змогу вилучити великий об'єм інформації для оцінки складу об'єкта аналізу. Кожен сенсор генерує вихідний відгук, який містить декілька сотень значень та являє

собою інформацію щодо різних властивостей об'єкта та особливостей вимірювальної системи. Тому інтелектуальний аналіз даних (або виявлення знань в базах даних) електронного носа починається з вилучення похідних ознак або набору параметрів (feature extraction) з

відгуків сенсорів. Отримані параметри повинні мати фізико-хімічний та кількісний взаємозв'язок з необхідними властивостями об'єкта. Обробка матриці інформативних та статистично надійних параметрів сенсорів дасть змогу побудувати точну та стабільну класифікаційну чи калібрувальну модель.

Форма відгуків п'єзокварцових сенсорів залежить від особливостей взаємодії досліджуваних парів з сорбційними покриттями, умов та режиму аналізу летких сполук, тобто тривалості стимулу у випадку дискретної інжекції аналіту в закриту комірку детектування, а також швидкості газоносія під час безперервної подачі аналіту в проточний детектор.

Динамічна екстракція дає змогу вилучити більшу кількість летких сполук в широкому діапазоні летючості, що є особливо важливим під час дослідження олій з метою оцінки ступеня їх окиснення. Як показали наші дослідження, під час окиснення олій зменшується масова частка води та легколетких сполук, яка в свіжих зразках становить 0.1–0.2 мас.%, та збільшується вміст ліпофільних летких сполук.

Зазвичай відгук сенсора електронного носа має форму піка, який рідко є ідеальним аналітичним піком Гауса, а здебільшого має асиметричну форму (змінюється ширина зростаючої або спадної частини піка, гостровершинність). Сигнал сенсора, отриманий під час інжекції аналіту в закриту систему детектування, можна поділити на три періоди: адсорбція (зростаюча частина), фаза рівноваги (плато) та десорбція (спадна частина). Під час динамічної екстракції та прямої подачі аналіту в проточну комірку детектування електронного носа (фронтальне введення проби) послідовно вилучаються групи летких сполук відповідно до збільшення їх коефіцієнтів розподілу та швидкості масопереносу [1], стадії адсорбції та десорбції на поверхні сенсора чергуються, а форма динамічного відгуку відображає поточну концентрацію летких сполук в потоці осушеного повітря, яка проходить через детектор за одиницю часу, сорбційну спорідненість покриття сенсора до вилучених летких сполук або класів речовин, ступінь вилучення летких сполук олій в приймаючу фазу. Таким чином, кінетика зміни відгуків сенсорів надає інформацію про ступінчасту зміну концентрації різних фракцій запаху відповідно до їх летючості.

Мета роботи – розробити підходи до вилучення похідних ознак з вихідних відгуків сенсорів; запропонувати параметри, які характеризують кількісні зміни різних фракцій запаху з динамічного сигналу в формі піка, отриманого під час безперервного введення аналіту в проточну комірку детектування електронного носа; оцінити ефективність запропонованих параметрів для побудови калібрувальних моделей з метою прогнозування

якості олій.

Матеріали та методи

Дослідження ступеня окиснення олій за зміною складу летких сполук проводили на лабораторному аналізаторі газів (МАГ-8, виробник ТОВ «СенТех») на базі восьми п'єзокварцових резонаторів РК-169 з основною частотою коливання 11.0 МГц на срібних електродах діаметром 5 мм, які модифікували сорбційними покриттями масою 10–20 мкг [2]. Використовували сорбенти різної полярності та сорбційної активності відносно певних груп летких сполук: поліетиленгліколь сукцинат (PEG suc – сенсор 1), поліетиленгліколь адипінат (PEG ad – сенсор 2), поліетиленгліколь себацинат (PEG seb – сенсор 3), Твін 80, або поліоксиетилен (20) сорбітанмоноолеат (Tween 80 – сенсор 4), Triton X-100, або поліетиленгліколь моно(тетраметилбутил) феніловий ефір (TX-100 – сенсор 5), дициклогексано-18-краун-6 (DCG18k6 – сенсор 6), триоктилфосфін оксид (ТОРО – сенсор 7), бджолиний віск (BW – сенсор 8).

Детектування летких сполук масивом сенсорів в проточній комірці електронного носа відбувалося після попереднього регенерування системи потоком осушеного повітря протягом 10 хв. до повної десорбції летких сполук з комірки детектування та плівки сорбентів, стабілізації частоти коливання сенсорів в потоці осушеного повітря на рівні дрейфу ± 2 Гц.

Досліджували склад летких сполук середньої проби олії масою 10 ± 0.01 г в склянці Дрекселя об'ємом 100 мл. Збільшення чутливості аналізу за рахунок зниження коефіцієнтів розподілу летких сполук досягали термостатуванням проби за температури $30 \pm 0.1^\circ\text{C}$ протягом 30 хв. Час вимірювання складав 10 хв.: 1 хв. – перевірка дрейфу базової лінії в потоці, 9 хв. – динамічна екстракція летких сполук барботуванням зразка олії зі швидкістю газу-носія $150 \text{ мл} \cdot \text{хв}^{-1}$. Швидкість газу-носія підбиралася такою, щоб кінетика процесу адсорбції на плівках визначалася переважно концентрацією летких речовин в потоці газу.

Вологість осушеного повітря перед початком вимірювання контролювали за холостою пробую в аналогічних умовах. За необхідності фільтр з силікагелем регенерували, оскільки покриття сенсорів проявляють високу чутливість до парів води, які також суттєво знижують сорбційну ємність сорбенту відносно інших складових газової суміші.

Для оброблення вихідних відгуків та вилучення параметрів нами були розроблені програмні забезпечення «eNose Curve Parameters» та «Odor pattern recognition». За допомогою програмного забезпечення The Unscrambler 10.4 (CAMO Software Oslo, Norway) будували калібрувальні моделі для прогнозування показника якості олії (пероксидного числа) з використанням регресійного методу проєкції на латентні структури (partial least squares

regression, PLSR) [3]. Показник пероксидного числа зразків олії визначали за арбітражною методикою [4].

Результати та обговорення

Вилучення параметрів

Запропоновано чотири підходи та алгоритми вилучення інформативних параметрів: з вихідних цифрових сигналів одиничних сенсорів, з моделей сигналів, апроксимованих лог-нормальною та поліноміальною функціями, з візуального образу запаху, побудованого з використанням скорочених вихідних відгуків масиву сенсорів.

Перша група складається з параметрів, розрахованих з вихідних сигналів, таких як: максимальна зміна частоти коливання j -го сенсора (ΔF_j^{\max} , Гц); час досягнення максимального значення сигналу (τ_{\max} , с) та час досягнення половини максимальної зміни частоти коливання сенсора на спадній частині відгуку ($\tau_{1/2} \Delta F_j^{\max}$, с); площа під кривою сигналу i -го сенсора (S_i , Гц·с) та сумарна площа під кривими масиву сенсорів (ΣS , Гц·с), розраховані методами інтегрування; коефіцієнт асиметрії (K_{skew}) та ексцесу (K_{kurt}) піка, кривина кривої на заданому інтервалі часу вимірювання сигналу (k) та радіус кривини кривої (R), розраховані методами чисельного диференціювання. Ефективність використання перелічених параметрів для вирішення різних задач аналізу летких сполук описано в літературі [5–12].

Нами запропоновано такі параметри: швидкість зміни сигналів на заданому інтервалі часу вимірювання (v , Гц/с) та відхилення від лінійності (L , Гц).

Більшість перелічених параметрів характеризують лише частину сигналу на заданому інтервалі часу вимірювання, що є актуальним під час аналізу фракцій запаху легколетких сполук з невеликими значеннями коефіцієнта розподілу за зростаючою частиною та вершиною піка кривої сенсора з полярним покриттями, а також визначення вмісту груп сполук з середніми та великими значеннями коефіцієнтів розподілу, які характеризують процеси глибокого окиснення ліпідів олій за спадною частиною сигналу після точки перегину кривої.

Таким чином, головним завданням аналізу є вибір оптимального набору параметрів, який містить інформацію про необхідну властивість об'єкта для побудови калібрувальної моделі, а також урахування їх впливу та кореляції між собою.

Друга група параметрів складається з коефіцієнтів рівнянь моделей сигналів, апроксимованих поліноміальною та лог-нормальною функціями. Відповідні коефіцієнти, на відміну від параметрів, вилучених з вихідного відгуку, характеризують повний сигнал.

Результати підгонки математичних моделей до вихідних відгуків (рис. 1 та 2) наведено для двох сенсорів різних за полярністю, селективністю та кінетикою адсорбції покриттів. ТОРО – екстрагент, що характеризується високою гідрофобністю, як неіоногенна поверхнево-активна речовина має довгий часовий ефект адсорбції [13]. У зв'язку з низькою чутливістю покриття ТОРО до параметрів навколишнього середовища, зокрема, до температури та вологості, сенсор проявляє високу стабільність в динамічних умовах аналізу, також характеризується довшим часом адсорбційно-десорбційних процесів порівняно з іншими покриттями ($BW > \text{ТОРО} > \text{PEGad} \approx \text{PEGsuc} \approx \dots \text{TX-100}$).

PEGsuc як хроматографічна фаза характеризується найбільшою полярністю з покриттів масиву сенсорів (значення сумарної полярності за коефіцієнтами Мак-Рейнольдса – 3300 [14]), проявляє високу сорбційну ємність та масову чутливість, але потребує більшого часу для стабілізації дрейфу базової лінії в потоці, характеризується швидкою реакцією в часі (акти адсорбції та десорбції).

Модель сигналу, апроксимованого лог-нормальною функцією визначається рівнянням (1) [15].

$$y(\tau) = \begin{cases} h \exp \left\{ -\ln 2 \left(\frac{\ln [1 + 2b(\tau - \tau_c)/\Delta]}{b} \right)^2 \right\}, & \frac{2b(\tau - \tau_c)}{\Delta} > -1 \\ 0, & \text{інакше} \end{cases} \quad (1)$$

де h – амплітуда сигналу, Гц; b – натуральний логарифм коефіцієнта асиметрії; τ_c – центр, положення якого відповідає амплітуді сигналу, с; Δ – ширина розподілу даних.

Апроксимацію вихідного сигналу здійснено симплекс-методом Нелдера – Міда з мінімізацією за ваговими сумами квадратів різниць (2) [16].

$$\chi^2(a) = \sum_{i=0}^{N-1} w_i [y_i - y(x_i | a)]^2 \quad (2)$$

де a – масив апроксимуючих параметрів, y_i – вихідний сигнал в точці x_i ;

$y(x_i | a)$ – апроксимуюча функція;

$w_i = 1/|y_i|$ – ваговий коефіцієнт, який нормує відхилення до значення сигналу в кожній точці. Оскільки вагові коефіцієнти виявляються найменшими за пікових значень сигналу та максимальними на «хвостах», то найбільш точно апроксимується сигнал на часових проміжках його зростання та спадання.

Отже, було вирішено зафіксувати амплітуду та центр сигналу, які розраховували за дискретним сигналом (відповідно ΔF_j^{\max} та τ_{\max}), та здійснювати процедуру підгонки моделі за двома параметрами (b та Δ), що сприяло підвищенню точності апроксимації всієї кривої та відповідно зменшенню середньоквадратичної помилки (RMSE) (рис. 1).

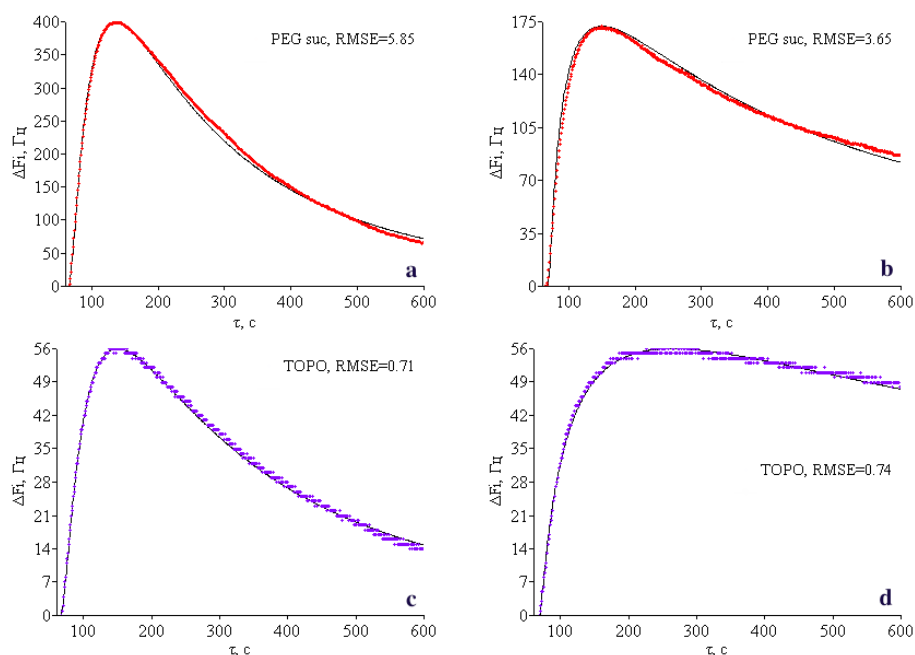


Рис. 1. Дискретні вихідні відгуки сенсорів (+ + +) та лог-нормальні моделі сигналів (—) для свіжих а, с та окиснених b, d зразків гарбузової олії.

Лог-нормальна модель відповідає простому фізичному опису процесу вимірювання, використовуються чотири параметри з точним фізичним змістом. Модель відтворює характерні геометричні риси відгуків, але не так точно проходить через дискретні точки експериментальної кривої.

Під час оброблення експериментальних даних встановлено, що точне відображення характеру сигналу забезпечується використанням апроксимуючого полінома дев'ятого степеня (рис. 2).

$$y = p_1 x^9 + p_2 x^8 + p_3 x^7 + p_4 x^6 + p_5 x^5 + p_6 x^4 + p_7 x^3 + p_8 x^2 + p_9 x + p_{10} \quad (3)$$

де p_i – коефіцієнти полінома.

Значення коефіцієнтів апроксимуючого полінома $p_1 \dots p_{10}$ знаходили, розв'язуючи систему рівнянь з використанням матриці Вандермонда.

Поліпшення результату апроксимації (помітне зменшення суми вагових квадратів залишків WSSE) досягається з використанням робастного методу найменших квадратів (LAR-метод, або мінімізація за абсолютними помилками), де матрицю Вандермонда та вектор експериментальних значень помножують на матрицю вагових коефіцієнтів.

$$w_i = \frac{1}{|y_i - y(x_i | a)|} \quad (4)$$

де w_i – вагові коефіцієнти.

Після кожної ітерації вагові коефіцієнти перераховуються, процес оптимізації зупиняється, коли різниця зміни параметрів полінома становить не більше 0.0001%.

Поліноміальна модель є суто емпіричною, апроксимує динамічні сигнали різної форми з високою точністю, але за рахунок використання

десяти параметрів не має фізичної інтерпретації. Математична модель, окрім набору коефіцієнтів для побудови калібрувальних моделей, може також використовуватися для інтерполяції і згладжування сигналу та наступного вилучення параметрів, які неможливо розрахувати за дискретним сигналом.

Нами запропоновано розраховувати за апроксимованими сигналами поліноміальною та лог-нормальною функціями кут піка на 0.9–0.95 % його висоти ($\alpha_{0.95}$) та точки перегину кривої (τ_{IP} , с).

Математичні моделі відтворюють вихідні криві різних за селективністю сенсорів електронного носа та описують характерні зміни форми відгуку під час зміни складу летких сполук протягом псування олій. Отримані коефіцієнти є набором параметрів для подальшого аналізу методами обробки багатовимірних даних.

Четвертий підхід вилучення інформативних ознак з відгуків сенсорів полягає в побудові візуального образу запаху зразка та його розпізнаванні за новими геометричними параметрами. Алгоритм розпізнавання образів запахів дає змогу візуалізувати та одночасно аналізувати структуру взаємодії даних всіх сенсорів масиву (дослідити інформативність перехресної чутливості сенсорів та визначити вибіркочку селективність певних сенсорів до груп та класів летких сполук), оптимізувати структуру даних та поліпшити дискримінацію візуального образу (профілю) запаху. Даний підхід дає змогу суттєво знизити розмірність даних та дискримінувати зразки за простими критеріями, підвищити ефективність будування моделей класифікації чи прогнозування, збільшивши їх точність за рахунок відсутності малоінформативних даних та викидів.

Образ запаху являє собою полярний графік, де головні осі – це зріз за часом вимірювання (τ_i , вилучені інформативні секунди відгуку), а допоміжні осі ($\Delta F_i(\tau_i)$) – це сигнали сенсорів (на рис. 3 вилучені відгуки шести сенсорів масиву) у відповідний час

вимірювання τ_i . Інформативність полярного графіка можна змінювати під час побудови профілю запаху зразка, а також проводити попередню обробку даних.

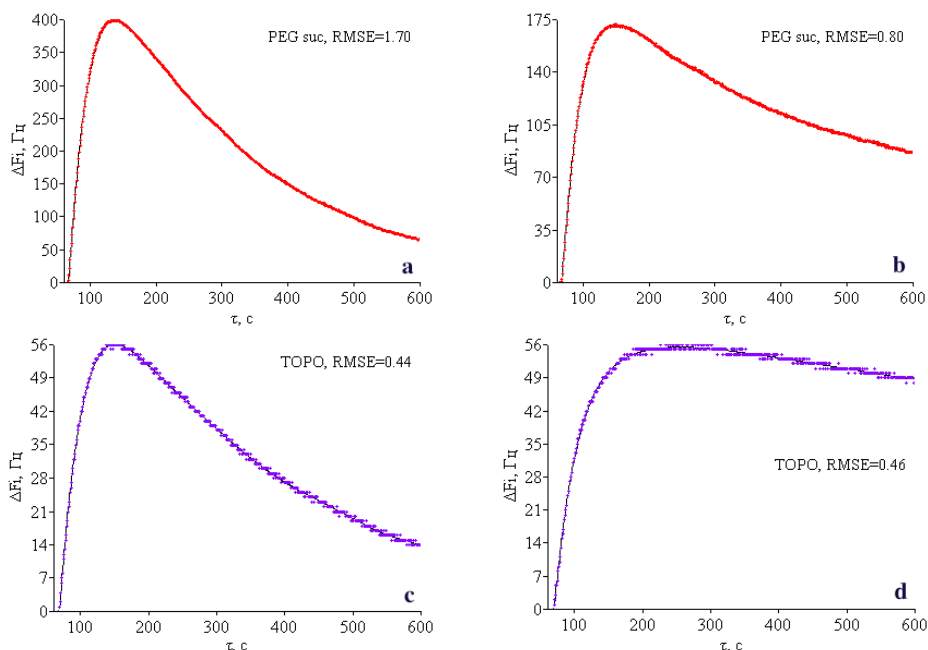


Рис. 2. Дискретні вихідні відгуки сенсорів (+ + +) та поліноміальні моделі сигналів (—) для свіжих а, с та окиснених b, d зразків гарбузової олії.

Попередня обробка даних може покращити точність та статистичну надійність моделі, стабілізувати обчислювальні алгоритми. Для оцінки якісних змін профілів запахів під час порівняння зразків буває корисним урівняти внесок в модель сигналів від різних сенсорів. Нормалізація за мінімальним та максимальним сигналами, яку застосовали під час аналізу зміни форми профілів (рис.3 та табл.1), змінює не структуру даних, а вагу різних частин даних під час їх аналізу (5). Даний вид попередньої обробки даних дає змогу отримати розподіл значень ознаки в межах широкого інтервалу [0, 100].

$$F'_i = (F_i - F_{min}) * 100 / (F_{max} - F_{min}) \quad (5)$$

де F'_i – сигнал після попередньої обробки протягом реакції в часі τ_i ; F_i – вихідний сигнал протягом реакції в часі τ_i ; F_{min} та F_{max} – мінімальне та максимальне значення сигналу за час вимірювання.

Нами запропоновано кількісно характеризувати візуальні образи (профілі) запахів за допомогою коефіцієнта зірчастості (G), що розраховується за полярними координатами графіка; коефіцієнта негладкості фігури (Sm), а також периметра (P , у.о.) та площі візуального образу запаху (A , кв.у.о.), які розраховуються після перетворення полярних координат в декартові координати зображення.

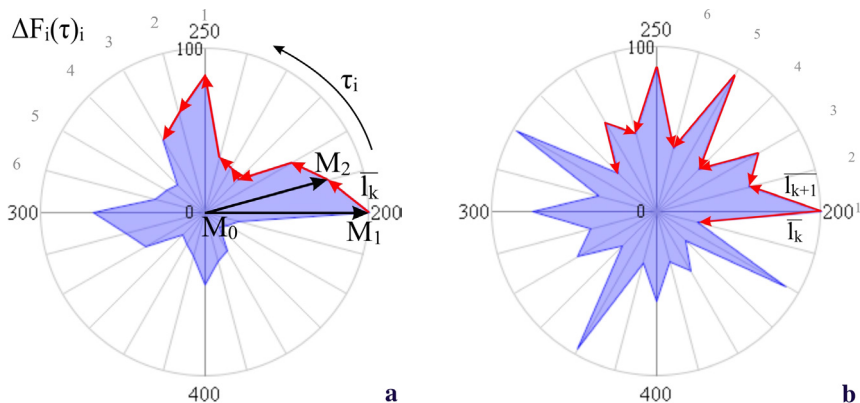


Рис. 3. Візуалізація етапів розрахунку параметра «негладкості» фігури.

Параметр негладкості візуального образу запаху розраховується в декілька етапів (рис.3): знаходимо вектори, що характеризують контур фігури, маючи сукупність точок $M_1(x_1, y_1) \dots M_n(x_n, y_n)$ відносно центру декартових координат $M_0(x_0, y_0)$ (рис.3а); наступним кроком – величини їх скалярних добутоків (рис.3б), що слугують мірою близькості векторів, характеризують довжини векторів-співмножників та кут між ними.

Для визначення коефіцієнта негладкості фігури (S_m) розраховуємо сумарні зміни її контуру:

$$S_m = \sum_{k=1}^n \overline{l_k} \cdot \overline{l_{k+1}} \quad (6)$$

Таким чином, у випадку аналізу окиснених профілів зразків олій переважно утворюються гострі кути суміжними векторами $\overline{l_k}$ та $\overline{l_{k+1}}$, скалярні добутки яких близькі до добутоків довжин цих векторів. Під час аналізу свіжих зразків олій у багатьох випадках утворюються тупі кути між суміжними векторами, що дає від'ємні значення

скалярних добутоків та істотно менше значення їх суми, ніж відповідне значення для окиснених зразків.

Коефіцієнт зірчастості G характеризує зміну геометричної форми фігури візуального образу запаху (рис. 4) і розраховується як сума кількостей перетинів фігури з круговою сіткою, заданою з певним кроком (Γ ц). Таким чином враховуються зміни сорбції летких сполук на плівках сенсорів з перехресною чутливістю під час накопичення інших фракцій запаху та прояву вибіркової селективності до летких продуктів окиснення ліпідів, за рахунок чого збільшується кількість піків (вершин багатокутника).

Результати розрахунків геометричних параметрів візуальних образів під час зміни якості гарбузової олії (рис.5) свідчать, що в процесі окиснення олії змінюється кількісний та якісний склад летких сполук, оскільки суттєво змінюється профіль зразків.

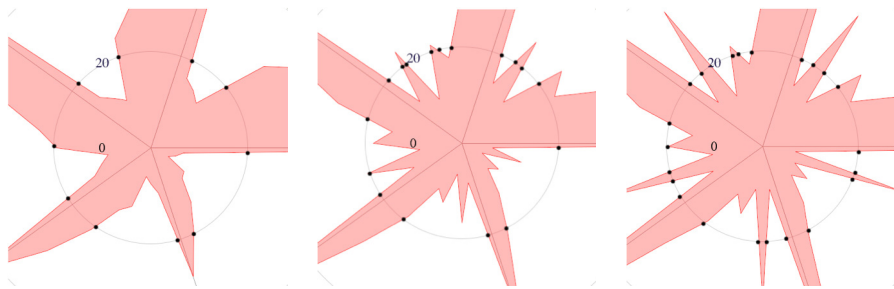


Рис. 4. Візуалізація параметра G для профілів олій різної якості в процесі окиснення (зліва направо).

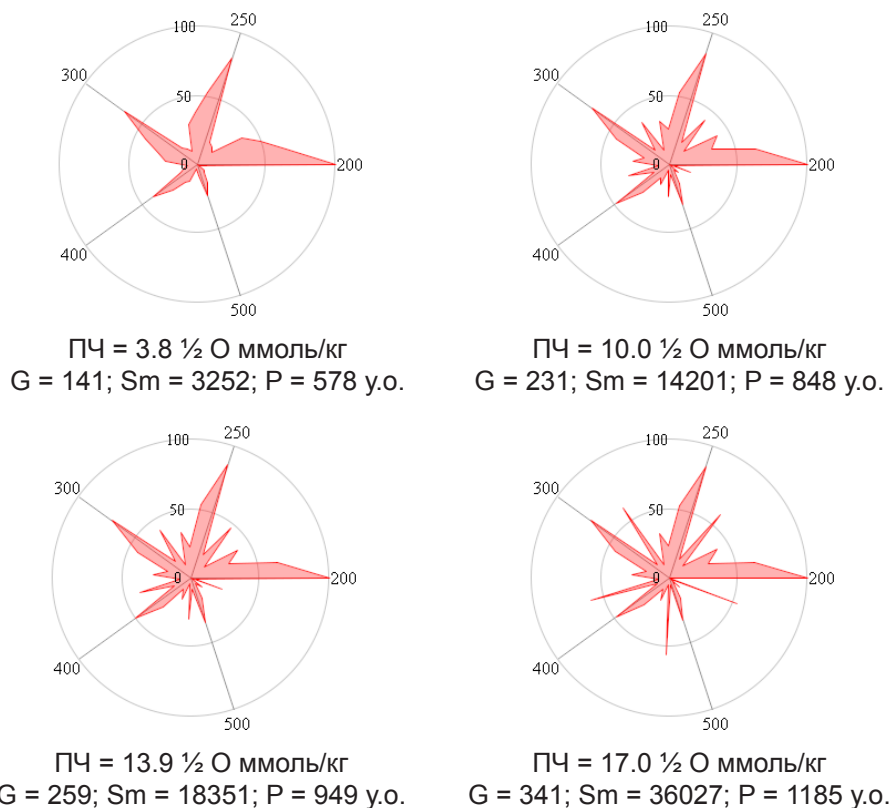


Рис. 5. Зразки гарбузової олії різної якості та параметри оцінки зміни геометричної форми профілів.

Побудова регресійних моделей

Найбільш популярним методом багатовимірного калібрування в хемометриці є метод проєкції на латентні структури PLS, який здійснює одночасну декомпозицію матриць X (матриці предикторів, параметрів сенсорів) та Y (матриці відгуків прогнозуючого показника, пероксидного числа, ПЧ). Проєкції будуються таким чином, щоб

максимізувати кореляцію X та Y рахунків, що дає змогу описувати складні зв'язки, використовуючи меншу кількість головних компонент [17]. Однією з умов успішного використання PLS є лінійність даних, тому оцінювали кореляцію запропонованих параметрів з прогнозуючим показником ПЧ, критичне значення коефіцієнта кореляції Пірсона для набору даних $r_{47, 0.05} = 0.29$ (табл. 1).

Таблиця 1. Значення коефіцієнтів кореляції параметрів електронного носа з пероксидним числом.

Параметри відгуків сенсорів	r	Параметри відгуків сенсорів	r	Параметри відгуків сенсорів	r
<i>I матриця параметрів, вилучених з вихідних цифрових сигналів одиничних сенсорів</i>					
T 1/2 ΔF_j^{\max} 1	0.88	K _{kurt} 4	0.51	v 8	0.86
T 1/2 ΔF_j^{\max} 2	0.72	K _{kurt} 5	0.39	L 1	-0.84
T 1/2 ΔF_j^{\max} 3	0.65	K _{kurt} 6	0.36	L 2	-0.82
T 1/2 ΔF_j^{\max} 4	0.76	K _{kurt} 7	-0.36	L 3	-0.80
T 1/2 ΔF_j^{\max} 5	0.78	K _{kurt} 8	0.32	L 4	-0.81
T 1/2 ΔF_j^{\max} 6	0.85	$\alpha_{0,95}$ 1	0.89	L 5	-0.83
T 1/2 ΔF_j^{\max} 7	0.90	$\alpha_{0,95}$ 2	0.91	L 6	-0.80
T 1/2 ΔF_j^{\max} 8	0.38	$\alpha_{0,95}$ 3	0.89	L 7	-0.80
K _{skew} 1	0.83	$\alpha_{0,95}$ 4	0.86	L 8	-0.41
K _{skew} 2	0.34	$\alpha_{0,95}$ 6	0.86	R 1	0.89
K _{skew} 3	0.35	$\alpha_{0,95}$ 7	0.73	R 2	0.81
K _{skew} 4	0.58	$\alpha_{0,95}$ 8	0.86	R 3	0.86
K _{skew} 5	0.40	v 1	0.89	R 4	0.89
K _{skew} 6	0.70	v 2	0.87	R 5	0.82
K _{skew} 7	-0.02	v 3	0.88	R 6	0.89
K _{skew} 8	0.33	v 4	0.87	R 7	-0.23
K _{kurt} 1	0.78	v 5	0.86	R 8	0.41
K _{kurt} 2	0.47	v 6	0.88	-	-
K _{kurt} 3	0.26	v 7	0.76	-	-
<i>II матриця параметрів, вилучених з моделей сигналів, апроксимованих лог-нормальною функцією</i>					
Δ 1	0.80	Δ 4	0.60	Δ 7	0.76
b 1	0.84	b 4	0.69	b 7	0.80
Δ 2	0.61	Δ 5	0.69	Δ 8	0.33
b 2	0.44	b 5	0.52	b 8	0.29
Δ 3	0.54	Δ 6	0.76	-	-
b 3	0.46	b 6	0.71	-	-
<i>III матриця параметрів, вилучених з моделей сигналів, апроксимованих поліноміальною функцією</i>					
1p1	-0.01	3p8	0.09	6p5	-0.07
1p2	0.02	3p9	-0.22	6p6	0.11
1p3	-0.04	3p10	0.34	6p7	-0.17
1p4	0.06	4p1	0.24	6p8	0.24
1p5	-0.09	4p2	-0.23	6p9	-0.34
1p6	0.12	4p3	0.21	6p10	0.43
1p7	-0.17	4p4	-0.18	7p1	-0.11
1p8	0.23	4p5	0.15	7p2	0.12
1p9	-0.32	4p6	-0.10	7p3	-0.12
1p10	0.41	4p7	0.03	7p4	0.13
2p1	0.35	4p8	0.06	7p5	-0.14
2p2	-0.34	4p9	-0.21	7p6	0.16
2p3	0.31	4p10	0.36	7p7	-0.18

Параметри відгуків сенсорів	r	Параметри відгуків сенсорів	r	Параметри відгуків сенсорів	r
<i>III матриця параметрів, вилучених з моделей сигналів, апроксимованих поліноміальною функцією</i>					
2p4	-0.29	5p1	0.06	7p8	0.21
2p5	0.25	5p2	-0.04	7p9	-0.25
2p6	-0.20	5p3	0.02	7p10	0.29
2p7	0.14	5p4	0.01	8p1	0.41
2p8	-0.05	5p5	-0.04	8p2	-0.40
2p9	-0.09	5p6	0.08	8p3	0.38
2p10	0.23	5p7	-0.13	8p4	-0.36
3p1	0.20	5p8	0.21	8p5	0.33
3p2	-0.18	5p9	-0.32	8p6	-0.29
3p3	0.16	5p10	0.42	8p7	0.24
3p4	-0.13	6p1	0.02	8p8	-0.16
3p5	0.10	6p2	0.00	8p9	0.03
3p6	-0.06	6p3	-0.02	8p10	0.11
3p7	-0.01	6p4	0.04	-	-
<i>IV матриця параметрів, вилучених з візуального образу запаху</i>					
G	0.98	Sm	0.97	P	0.96

Окрім коефіцієнтів полінома, що описують повний сигнал, всі інші параметри характеризують частину відгуку після максимального значення сигналу, що відповідає кількісним змінам груп летких сполук з середніми та великими значеннями коефіцієнтів розподілу, які характеризують процеси глибокого окиснення ліпідів олій. Значення коефіцієнтів кореляції запропонованих параметрів з ПЧ (табл.2) вказують на їх високу інформативність для прогнозування даного показника якості. Малоінформативні параметри не виключали з матриці даних під час побудови регресійних моделей, так як головним завданням було запропонувати універсальний підхід вилучення матриці інформативних параметрів з метою отримання робастних моделей.

На етапі підготовки змінних (параметрів сенсорів, пероксидного числа) використовували автошкалування, тобто поєднання центрування та стандартизації даних ($1/Sdev$) [3], щоб отримати одиничну дисперсію змінних та нульове середнє значення та уникнути превалювання одних змінних над іншими, що можуть мати домінуючий вплив на модель через порядок величини. Об'єм навчальної вибірки для побудови PLSR моделей становив 30 зразків, всі калібрувальні моделі перевіряли за допомогою перехресної валідація з числом сегментів 15.

Точність калібрування та крос-валідації прийнято характеризувати значеннями середньоквадратичних помилок (RMSE cal та RMSE val). Відповідно, чим менше RMSE cal та RMSE val, тим точніше описуються дані. Крім того, якість калібрування та валідації характеризується квадратами коефіцієнтів кореляції R^2 cal та R^2 val,

які розраховуються в масштабі від нуля до одиниці, де нуль вказує на відсутність кореляції, а одиниця відображає відмінну лінійну кореляцію.

Придатність моделей для прогнозування оцінювали за тестовим набором із 17 нових зразків. RMSEP є середньоквадратичною помилкою прогнозування тестових зразків (7). Значення пероксидного числа в процесі зберігання олії змінювалося в межах від 3.8 до 17.0 $\frac{1}{2}$ O ммоль/кг, таким чином відносна похибка RMSEP розраховується як відношення максимального та мінімального значення ПЧ до RMSEP моделі.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (7)$$

де y_i та \hat{y}_i – вихідні та прогнозовані значення ПЧ, відповідно; N – кількість тестових зразків.

Значення RMSE cal, RMSE val та RMSEP в програмі виражається в вихідних одиницях Y, тобто в одиницях ПЧ.

Для порівняння метрологічних характеристик методу з арбітражним [4] розраховували значення середньої відносної похибки прогнозування зразків тестової вибірки. Відповідно до стандарту [4] значення відносної похибки вимірювання для зразків з ПЧ $\leq 3.0 \frac{1}{2}$ O ммоль/кг становить 8%, для зразків з ПЧ $\geq 3.0 \frac{1}{2}$ O ммоль/кг – 4%.

Встановлено, що найбільш точна регресійна модель побудована (табл.2) в результаті оброблення навчального набору параметрів, вилучених з візуального образу запаху. IV модель також характеризується кращою лінійною кореляцією порівняно з іншими моделями.

Таблиця 2. Результати побудови регресійних моделей та прогнозування пероксидного числа олії за тестовим набором.

Статистичні параметри та метрологічні характеристики методу	Матриці параметрів			
	I	II	III	IV
Точність регресійних моделей				
RMSE cal	1.10	1.03	1.36	0.77
RMSE val	1.28	1.23	1.57	0.85
R ² cal	0.90	0.91	0.84	0.95
R ² val	0.88	0.88	0.80	0.94
Точність прогнозування нових зразків				
RMSEP	1.16	1.16	1.38	0.79
Відносна похибка RMSEP, %	6.82...30.53	6.82...30.53	8.12...36.32	4.65...20.79
Відносна похибка, %	14.34	12.08	14.36	7.72

Слід зазначити, що під час дослідження пояснювальної дисперсії для кожної моделі обрано оптимальну кількість головних компонент (ГК) або латентних змінних: I модель – 2 ГК (77 %, 90 %), II модель – 3 ГК (93 %, 92 %), III модель – 3 ГК (99 %, 84 %), IV модель – 1 ГК (99 %, 95 %), де в дужках зазначені значення залишкових X та Y дисперсій. В більшості випадків перші компоненти описують найбільші зміни значень в Y, тому є найбільш інформативними.

IV матриця даних добре апроксимується з використанням лише однієї головної компоненти, що пояснюється сильною кореляцією всіх змінних, що призводить до зниження розмірності близької до одиниці, так як залежність майже лінійна. Це пов'язано з тим, що параметри вилучаються з візуального образу запаху, який будується під час одночасної оптимізації структуру даних та оцінювання дискримінації зразків. Таким чином, матриця параметрів представляє змістовну структуру даних з майже відсутніми залишками (відсутнім шумом та надлишковості в даних).

В матрицях великої розмірності I, II та особливо III присутні шум та надмірність в даних, що обов'язково проявляють себе через кореляційні зв'язки між змінними (мультиколінеарністю) та потребують більш детального аналізу вкладу кожної змінної в ГК.

Точна підгонка моделі не завжди вказує на те, що вона описує системні зміни даних, а не випадкові коливання (шум). Тому особливо важливо перевіряти робастність та ефективність прогнозування моделей за новими даними. На рис. 6 наведено прогнозовані значення тестових зразків з відхиленнями для всіх моделей, де прогнозоване значення – червона горизонтальна лінія (з червоною вертикальною лінією RMSEP), а синій прямокутник – відхилення кожного зразка. Відхилення вказує наскільки надійними є прогнозовані результати, оскільки враховує залишок та коефіцієнт впливу спостереження відносно калібрувальних даних.

Як видно, найбільш точною та стабільною є IV модель, оскільки прогнозовані зразки мають невеликі відхилення відносно калібрувального набору. Помилка прогнозування становить $\pm 0.79 \frac{1}{2}$ O ммоль/кг.

Зважаючи на похибку RMSEP та відхилення спостережень, придатність до роботи з новими даними та надійність прогнозування моделей знижується в ряду: IV > II > I > III.

Отримане значення відносної похибки для зразків майже вдвічі більше допустимої похибки вимірювання стандартним методом визначення пероксидного числа [4]. Однак, межа значень відносної похибки 8% є достатньою для тест-методу та вирішення задач аналізу якості та скринінгу об'єкта.

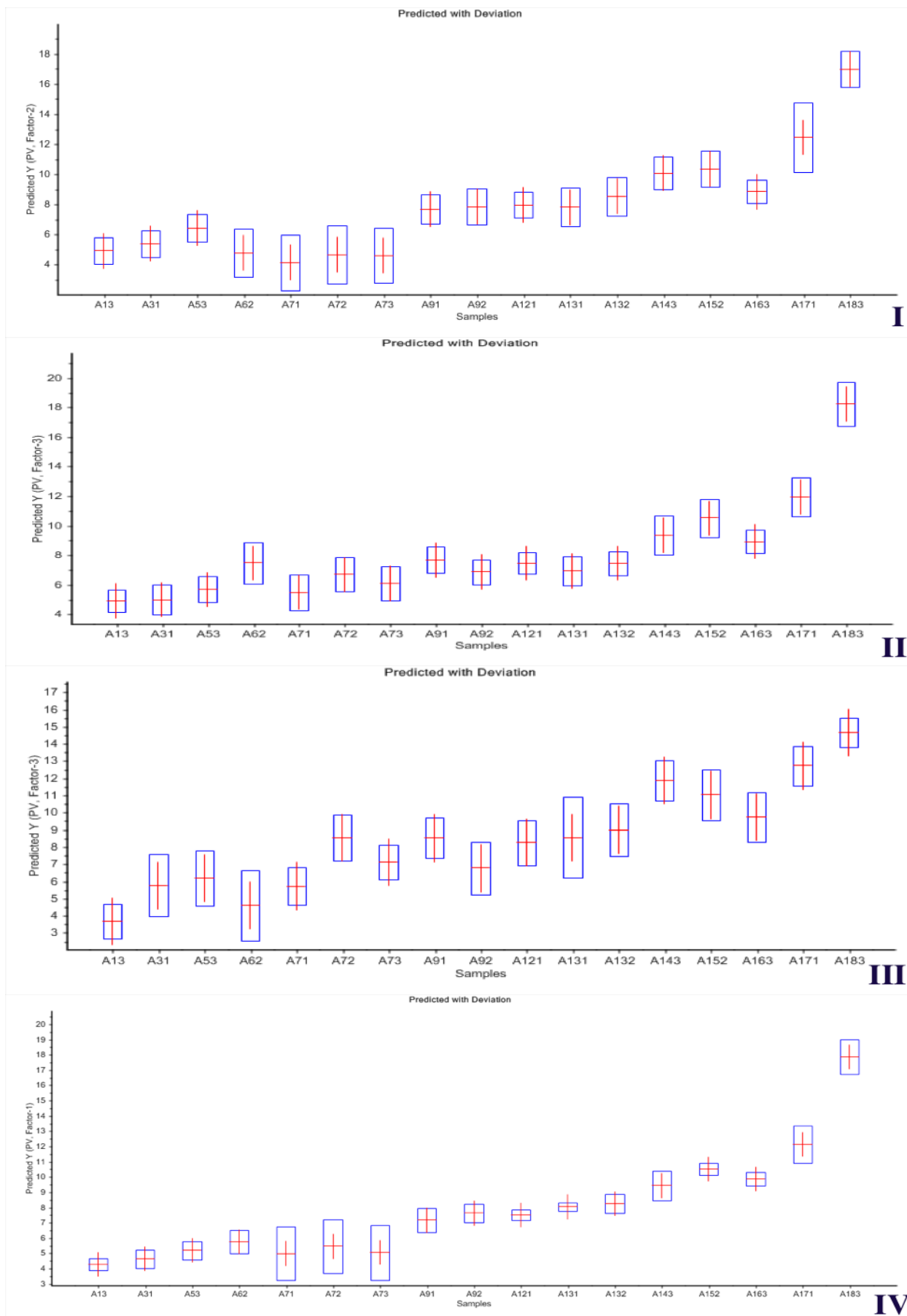


Рис. 6. Результати прогнозування ПЧ тестових зразків гарбузової олії.

Висновки

Запропоновано чотири підходи до вилучення похідних ознак з вихідних динамічних відгуків сенсорів електронного носа, які характеризують кількісні зміни летких продуктів окиснення ліпідів: з вихідних цифрових сигналів одиничних сенсорів, з моделей сигналів, апроксимованих лог-нормальною та поліноміальною функціями, з візуального образу запаху, побудованого з використанням скорочених вихідних відгуків масиву сенсорів. Формування матриць параметрів дає змогу досягти різної інформативності для оцінки як загальних показників об'єкта, так і вилучити інформацію про необхідну властивість.

Запропоновано матриці інформативних параметрів для побудови калібрувальних моделей з метою прогнозування показника пероксидного числа олій. Нові параметри оцінки геометричної форми візуальних образів (профілів) запахів зразків, а саме коефіцієнт зірчастості та коефіцієнт негладкості фігури, дозволили знизити розмірність даних та під час декомпозиції матриці перетворити багатовимірну

калібрувальну модель в одновимірну лінійну.

Проведено порівняльне оцінювання точності чотирьох регресійних моделей із використанням методу проекції на латентні структури. Установлено, що найбільш точні моделі отримано під час обробки навчальних наборів параметрів, вилучених з візуального образу запаху, та параметрів, вилучених з моделей апроксимованих сигналів лог-нормальною функцією, з похибкою RMSE 0.77...0.85 $\frac{1}{2}$ O ммоль/кг та 1.03...1.23 $\frac{1}{2}$ O ммоль/кг відповідно.

Оцінено придатність моделей для прогнозування пероксидного числа за новими даними. Відносна помилка прогнозування для моделі, побудованої з використанням параметрів, вилучених з візуальних профілів зразків, становить $\pm 0.79 \frac{1}{2}$ O ммоль/кг.

Запропонований спосіб визначення пероксидного числа олій дасть змогу з високою точністю провести експрес-аналіз якості зразків, знизивши економічні затрати на використання хімічних реактивів.

Література

1. Витенберг А. Г., Иоффе Б. В. Газовая экстракция в хроматографическом анализе: Парофазный анализ и родственные методы. Л.: Химия, 1982. С. 56–64.
2. Пат. 111253 Україна, МПК G01N 33/02, A23L 13/60. Спосіб виявлення фальсифікації варених ковбасних виробів соєвими продуктами/Арсеньєва Л. Ю., Калініченко А. О., Іванов С. В., Кучменко Т. А.; заявник патентовласник Нац. ун-т харчових технологій. № а201407714; заявл. 09.07.2014; опубл. 11.04.2016, Бюл. № 7.
3. Esbensen Kim H. Multivariate Data Analysis – in practice. An Introduction to Multivariate Data Analysis and Experimental Design, 5th Edition. Aalborg University, Esbjerg. Oslo: Camo Process AS, 2004. pp. 75–78.
4. Жири рослинні та олії. Метод визначення пероксидного числа: ДСТУ 4570:2006. Чинний від 2008–01–01. К.: Держспоживстандарт України, 2007. 6 с. (Національні стандарти України).
5. Jia Yan, Xiuzhen Guo, Shukai Duan, Pengfei Jia, Lidan Wang, Chao Peng, Songlin Zhang. Electronic Nose Feature Extraction Methods: A Review. *Sensors*. 2015, 15 (11), 27804–27831.
6. Carmel L., Levy S., Lancet D., Harel D. A feature extraction method for chemical sensors in electronic noses. *Sensors and Actuators B*. 2003, 93, 67–76.
7. Yaqi Jing, Qinghao Meng, Peifeng Qi, Ming Zeng, Wei Li, Shugen Ma. Electronic nose with a new feature reduction method and a multi-linear classifier for Chinese liquor classification. *Review of scientific instruments*. 2014, 85, 055004-1–055004-10.
8. Eklov T., Mirtensson P., Lundstrom I. Enhanced selectivity of MOSFET gas sensors by systematical analysis of transient parameters. *Analytica Chimica*

Acta. 1997, 353, 291–300.

9. Brudzewski K., Ulaczyk J. An effective method for analysis of dynamic electronic nose responses. *Sensors and Actuators B*. 2009, 140, 43–50.

10. Botre B. A., Gharpure D. C., Shaligram A. D. Embedded electronic nose and supporting software tool for its parameter optimization. *Sensors and Actuators B*. 2010, 146, 453–459.

11. Eugenio Martinelli, Christian Falconi, Arnaldo D'Amico, Corrado Di Natale. Feature Extraction of chemical sensors in phase space. *Sensors and Actuators B*. 2003, 95, 132–139.

12. Wei Z., Wang J., Zhang W. Detecting internal quality of peanuts during storage using electronic nose responses combined with physicochemical methods. *Food Chemistry*. 2015, 177, 89–96.

13. Prochaska K., Walczak M., Staszak K. Estimation of trioctylphosphine oxide (TOPO) diffusion coefficients by dynamic adsorption measurements in model extraction systems. *Journal of Colloid and Interface Science*. 2002, 248, 143–148.

14. Harald R. Stationary phases in gas chromatography. *Journal of chromatography library*. 1991, 48. pp. 93.

15. Bingemann D., Ernsting N. P. Femtosecond solvation dynamics determining the band shape of stimulated emission from a polar styryl dye. *The Journal of Chemical Physics*. 1995, 102 (7), pp. 2691–2700.

16. Press W., Teukolsky S., Vetterling W., Flannery B. Numerical recipes. *The Art of Scientific Computing*. Cambridge University Press, Third Edition, 2007. pp. 778.

17. Родионова О.Е. Хемометрический подход к исследованию больших массивов химических данных. *Рос. хим. ж. (Ж. Рос. хим. об-ва им. Д. И. Менделеева)*. 2006, 2, С. 128–144.