Journal of Plant Protection Research eISSN 1899-007X

ORIGINAL ARTICLE

Application of chemometric analysis using physicochemical and chromatographic data to differentiate the origin of plant protection products containing trinexapac-ethyl

Patrycja Marczewska*, Joanna Rolnik, Tomasz Stobiecki

Pesticide Quality Testing Laboratory, Institute of Plant Protection – National Research Institute, Sosnicowice Branch, Sosnicowice, Poland

Vol. 65, No. 2: 200–210, 2025

DOI: 10.24425/jppr.2025.155059

Received: August 08, 2024
Accepted: October 21, 2024
Online publication: July 08, 2025

*Corresponding address:
p.marczewska@iorpib.poznan.pl

Responsible Editor:
Piotr Kaczyński

Abstract

The European market for plant protection products (PPPs) faces significant challenges related to counterfeit and substandard PPPs, posing threats to sustainable agriculture and food safety. This study explored the application of chemometric methods based on physical, chemical, and technical parameters, as well as data obtained by high-performance liquid chromatography with a diode array detector (HPLC-DAD) and headspace gas chromatography coupled with mass spectrometry (HS-GC/MS), to verify the authenticity of PPPs containing trinexapac-ethyl. A total of 44 formulations were analyzed, including authentic samples and substandard PPPs obtained from various retail points and manufacturers. The developed analytical methods demonstrated robustness in determining physicochemical parameters and generating chromatographic profiles distinguishing between genuine and non-genuine products. Chemometric tools such as principal component analysis (PCA), hierarchical clustering analysis (HCA), and Soft Independent Modeling of Class Analogy (SIMCA) facilitated data interpretation, revealing distinct clusters of samples based on their chemical fingerprints. SIMCA models exhibited their potential for routine quality control assessments. Overall, integrating advanced analytical techniques and chemometrics offers a promising strategy to safeguard the integrity of PPPs, enhance regulatory compliance, and mitigate the risks associated with counterfeit products in the European agricultural market. This approach supports sustainable agricultural practices by ensuring product authenticity and safety, thereby fostering consumer trust and regulatory adherence. In the context of increasing global demand for agricultural products, effective verification of PPPs authenticity becomes a crucial element in ensuring food security, human health, and environmental protection.

Keywords: counterfeit plant protection products, HCA, PCA, quality control, SIMCA, trinexapac-ethyl

Introduction

The market for plant protection products (PPPs) in the European Union is influenced by various factors affecting demand and future trends. National action plans by member states aim to reduce risks related to the use of PPPs, promote sustainable pesticide use, integrated pest management (IPM), and encourage organic farming. However, the global population is projected to exceed 9.1 billion by 2050, necessitating a significant increase in food production. This may result in intensified agricultural production and increased use of agrochemicals (Biondi *et al.* 2012). Europe's share in the global agrochemical market from 2010 to 2018 is estimated at 11.8% (FAOSTAT 2018). The market for plant protection products is particularly vulnerable to irregularities. Data published in articles and reports indicate that the main cause of these irregularities is the circulation of counterfeit or illicitly sourced plant protection products (Frezal and Garsous 2020).

A 2011 report by the Food and Agriculture Organization of the United Nations (FAO) and the World Health Organization (WHO) states that in developing countries, the rate of non-compliance with quality standards for plant protection products reaches 20-30% (FAO 2010). In Europe, the European Union Intellectual Property Office (EUIPO) estimates that counterfeit pesticides make up approximately 14 percent (around EUR 1.3 billion) of the EU market (EUIPO 2017).

Due to legal loopholes, enforcement issues, accessible technology for producing counterfeit plant protection products (PPPs), and the profitability of this process, the number of counterfeit and substandard products on the market is rising (Europol 2011). Illegal products may contain unknown chemical mixtures with unpredictable effects. Using PPPs from unauthorized sources risks exposure to banned or restricted substances. The unexplored phytotoxicity of these PPPs threatens farmer and consumer health and can cause crop losses or destruction (Jankowska *et al.* 2022). Additionally, counterfeit and illegal PPPs result in financial and reputational damage for producers, jeopardizing sustainable agriculture. Their production, transport, and trade bypass legal regulations, depriving states of revenue from customs duties and taxes (Streloke 2018).

According to the 2016 IUPAC recommendations, chemometrics employs mathematical and statistical methods to extract relevant information from multidimensional data. Modern measurement techniques generate extensive datasets of physicochemical properties and process parameters, which are challenging to interpret. Chemometrics provides robust techniques for assessing product originality and has been increasingly applied in recent years for quality control, using advanced chromatographic techniques to create "chemical fingerprints" of analyzed substances. Data obtained by HPLC was applied to similarity analysis for the quality control of polyphenols extracted from pomegranate peel (Li *et al.* 2015). "Chemical fingerprints" obtained by HS-GC/MS of authentic and counterfeit samples of Viagra® and Cialis® were used for PCA, projection analysis, classification and regression trees, and SIMCA modeling of individual groups (Custers *et al.* 2014). Chromatograms obtained by HPLC with a photodiode array detector (HPLC-PDA) and HPLC with mass spectrometry (HPLC-MS) were also utilized to analyze impurities in Cialis®. The obtained data were analyzed and modeled using PCA, partial least squares (PLS), and classifiers SIMCA and k-nearest neighbors (k-NN) to detect cases of Cialis® counterfeiting (Custers *et al.* 2016). A method using gas chromatography (GC) for obtaining similarity analysis data for the quality control of gasoline was developed, optimized, and validated (Flumignan *et al.* 2008). "Chemical fingerprints" of the phenolic fraction of extra virgin olive oil were obtained using HPLC with DAD and FLD detectors. After preliminary PCA analysis, partial least squares discriminant analysis (PLS-DA), SIMCA, and k-nearest neighbors (k-NN) algorithms were applied (Bajoub *et al.* 2017). HS-GC/MS was utilized to classify various commercially available rums, with data analyzed using unsupervised learning methods like HCA and PCA, as well as supervised methods, including LDA (Belmonte-Sánchez *et al.* 2018). The impact of environmental and post-harvest processing factors on thyme's metabolic composition was investigated using UHPLC-QTOF-HRMS fingerprinting, which traced thyme origins and evaluated processing effects, identifying key differentiation compounds for product traceability and quality assessment (Rivera-Pérez *et al.* 2023). A portable spectrophotometer and chemometrics method was introduced to authenticate and differentiate Amazon stingray meats, supporting conservation efforts by combating illegal trade (Craveiro de Andrade *et al.* 2024). Non-targeted HPLC-FLD fingerprinting was employed to classify and authenticate paprika using methods like PLS-DA, PCA-LDA, RF for first-order fingerprints, and NPLS-DA for second-order fingerprints, achieving accurate classification and effective detection of origin blends (Sun *et al.* 2023).

These studies collectively demonstrate the broad applicability and effectiveness of chemometric analysis in various domains of quality control, ensuring product authenticity, safety, and regulatory compliance. While chemometrics is widely used for quality control across many fields, its application in agrochemical sciences remains limited. The Pesticide Quality Testing Laboratory published a study using cluster analysis (CA) and PCA based on the results obtained by the Headspace Gas Chromatography-Mass Spectrometry method and some selected physicochemical properties of examined pesticides including pH, density, stability, active ingredient and water content to verify the authenticity of plant protection products containing chlorpyrifos as the active substance (Miszczyk *et al.* 2015). Another study verified the authenticity of plant protection product formulations containing azoxystrobin as the active substance using chromatographic methods and the SIMCA method. The purpose of this work was to illustrate the use of chemometric analysis for the modeling of chemical fingerprints obtained by HPLC-DAD and HS-GC-MS to detect illegal and counterfeit PPPs without determining the individual ingredients of the formulation (Marczewska *et al.* 2019). The aim of study was to explore the application of chemometric methods based on physical, chemical, and technical

parameters, as well as data obtained by high-performance liquid chromatography with a diode array detector (HPLC-DAD) and headspace gas chromatography coupled with mass spectrometry (HS-GC/MS), to verify the authenticity of PPPs containing trinexapac-ethyl. For this purpose, two chromatographic methods using HPLC-DAD and HS-GC/MS were developed to obtain specific chromatographic profiles, and determinations of physical, chemical, and technical parameters characteristic of the formulation of the tested plant protection product were performed. The parameters were determined according to the methodologies included in CIPAC collections: pH of 1% product solution (MT 75.3), foam persistence (MT 47.2), acidity/alkalinity (MT 191), emulsion stability (MT 36.6), water content (MT 30.6), and guidelines from the Organization for Economic Co-operation and Development (OECD) for density (OECD 109), dynamic viscosity (OECD 114), according to registration documentation, recommendations published by FAO in the "Manual on Development and Use of FAO and WHO Specifications for Pesticides," and guidelines from the European Commission in the document "Reference Document Illustrating Best Practices on Analytical Strategies and Interpretation of Results for the Formulation Analysis of Plant Protection Products Obtained During Official Market Control".

Materials and Methods

Materials

To develop and verify the applicability of the analytical method in practice, 44 formulations of emulsifiable concentrate containing the trinexapac-ethyl active substance were utilized. Trinexapac-ethyl [IUPAC: ethyl 4-[cyclopropyl(hydroxy)methylidene]-3,5-dioxocyclohexane-1-carboxylate] is a compound from the cyclohexanedione group acting as a gibberellin inhibitor. It shortens and stiffens cereal stems, preventing lodging. As a growth regulator, it is widely used in crops such as winter wheat, winter barley, winter rye, spring barley, and oats. Fourteen samples of PPPs were sourced directly from different production batches provided by their manufacturers, accompanied by certificates confirming their origin and adequate quality (original/reference samples numbered from 100 to 113). The remaining 30 samples were supplied to the Laboratory for the Quality Assessment of Plant Protection Products by the State Plant Protection and Seed Inspection, as well as by contractors, and were purchased from various retail outlets (tested samples numbered from 1 to 44, including two illegal samples numbered 16 and 17). In 2015, the European Commission (EC) defined illegal PPPs as any plant protection product not considered legal. This category contains the two sub-categories of counterfeit and substandard. These sub categories are defined as follows: substandard PPPs - products which contain substances not approved under EU legislation (or which contain no active substances) and falsified PPPs (e.g., falsified content, falsified country of origin) and counterfeit PPPs – illegal copies of legitimate, branded products. Whenever feasible, a minimum of three containers of each product with identical size, batch number, and production date were acquired to assess batch variability and its implications on data interpretation.

Methods

To determine the active substance content and obtain product-specific chromatographic profiles, a method utilizing HPLC-DAD was developed and validated. The procedure was conducted according to guidelines outlined in SANCO/3030/99 rev. 5, as well as the guidelines provided by the Collaborative International Pesticides Analytical Council (CIPAC): "Guidelines on method validation to be performed in support of analytical methods for agrochemical formulations" and "Guideline for analytical methods for the determination of relevant impurities referred to in FAO and/or WHO specifications for pesticide technical grade active ingredients and formulations".

The development and validation of the analytical method involved optimizing chromatographic conditions, assessing method specificity, determining critical parameters for method linearity, and establishing method precision and accuracy. Chromatographic analysis was performed using an HPLC-DAD system (Agilent Technologies, Santa Clara, CA, USA) equipped with an Altima C18 column (250 mm × 4.6 mm, 5 µm, W.R. Grace & Co-Conn.). The solvent used in the analysis was a mixture of acetonitrile HPLC grade acetonitrile (JT Baker, USA) and deionized water (deionized water was obtained from a Milli-Q SP reagent water system, Millipore, Bedford, USA) (60:40, v/v). The mobile phase consisted of acetonitrile and 0.1% H_3PO_4 (orto-Phosphoric acid solution 85% pure p.a., Merck, Germany). Samples were injected into the column in 20 µl quantities. The mobile phase flow rate was 1.0 ml·min^{-1}, with the column temperature set at 20°C and the injector temperature at 5°C. Analyte detection was carried out at a wavelength of 280 nm over a 15-minute analysis period. Chromatographic data were collected and integrated using ChemStation software rev. B.03.02. Based on the obtained validation parameter results, it was concluded that the methods met the requirements specified in the SANCO/3030/99 rev. 5 document for quality control of plant protection products (Tab. 1).

Table 1. Linearity parameters and precision obtained by HPLC-DAD

Substance	Calibration levels	Slope $a \pm SD_a$*	Intercept $b \pm SD_b$**	Correlation coefficient	Residual st. dev. $S_{y/x}$***	Precision (H_r) (%RSD/%RSD_r)****	Mean recovery % *****
Trinexapac--ethyl	5	1071505.5 ± 21447.2	142.2 ± 172.7	0.99	195.8	0.50	101.5

* – a – slope of the calibration curve, SD_a – standard deviation of the slope
** – b – intercept of the calibration curve, SD_b – standard deviation of the intercept
*** – $S_{y/x}$ – residual standard deviation
**** – $Hr \leq 1$, acceptable
***** – for the active substance content above 10% according to SANCO/3030/99 rev.5 the acceptable range is 97–103%

The method was further optimized to enable comparative studies of the examined plant protection products using HS-GC/MS (Agilent Technologies, Santa Clara, CA, USA). Various equilibration times before analysis, sample heating temperatures, and different sample masses were optimized. As a result, a method was developed that allows for comparative studies of the composition of the tested samples. A non-polar capillary column (HP-5MS, 30 m × 0.25 mm × 0.25 µm, J&W GC Columns) was used as the chromatographic column. Samples of 1 ml were incubated in 20 ml headspace vials and gently agitated at 50°C for 15 minutes. The headspace loop and transfer line temperatures were set at 80°C and 110°C, respectively. After equilibration, 1 ml of the vapor phase was injected into the GC-MS system in split injection mode (split ratio 1:10), using a standard split/splitless injector set at 250°C. Helium (5.0, Messer), at a flow rate of 1.8 ml · min^{-1}, was employed as the carrier gas. The chromatographic oven temperature was held at 40°C for 5 minutes, then increased to 100°C at a rate of 5°C · min^{-1}, followed by a ramp to 280°C at 25°C · min^{-1}, and maintained for 5 minutes. Ionization was performed in electron impact mode, and mass spectrometry detection was set to full scan mode to identify the solvents present in the samples. The ion source temperature was maintained at 230°C, while the interface temperature was held at 300°C. Chromatographic data were collected and integrated using the Agilent MSD ChemStation software. Subsequently, all acquired samples were analyzed using both methods in triplicate. The obtained chromatograms, which serve as fingerprints, were subjected to chemometric analysis.

Results

The samples of the formulations were characterized by determining the values of physical, chemical, and technical parameters. Instrumental signals were also recorded for all samples using HPLC-DAD and HS-GC/MS. The averaged results from three repetitions obtained from physicochemical determinations are presented in Table 2.

Data preparation for chemometric analysis

Preliminary data preparation is an obligatory step before starting the process of constructing chemometric models and drawing conclusions from the obtained results. For the dataset containing physicochemical parameters measured across different ranges of variability, averaging and autoscaling operations were performed to assign each variable unit variance.

For the chromatographic signals obtained using the HPLC-DAD technique on the studied material, the baseline presence and signal-to-noise ratio parameters were evaluated and found to be acceptably low. Peak shift corrections were made using the correlation optimized warping (COW) method relative to a reference signal (target) that showed the highest correlation with all signals in the analyzed dataset. The characteristic parameter values used for the COW method were 200 and 3 for window width and degree of fit, respectively. For chromatographic signals obtained using HS-GC/MS, the presence of baseline and signal-to--noise ratio parameters was evaluated and found to be acceptably low. The obtained data showed consistent retention times for analytes, making peak shift correction unnecessary. Centering operations were performed on data matrices created from instrumental signals. Data processing was conducted using Matlab (version R2021b).

The application of unsupervised chemometric methods

A PCA model was built for the set of autoscaled physicochemical data. The distribution of points representing individual samples in the space of successive pairs of principal components was analyzed. The degree of data compression in the table of physicochemical properties, subjected to autoscaling and analyzed using the PCA method for the studied preparation, made it possible to infer that the first principal component describes 48.8% of the data variance, the second component 68.6%, and the third component 77.2%. In the next step of data interpretation, the distribution of samples on planes defined by selected pairs of principal components was analyzed, and the distribution of

Table 2. Summary of averaged results of physicochemical parameter determinations

	Sample number	Density [g/mL]=	Water content [g/kg]	Acidity [%]	pH after 1 min	pH after 2 min	pH after 10 min	Foam after 10 sec	Foam after 1 min	Foam after 3min	Foam after 12 min	Viscosity [mPa·s]	Emulsion [ml]	Content of a.s.[g/L]
Tested samples	1	0.9814	7.338	4.953	3.710	3.711	3.710	32.7	24.7	20.0	12.7	0.9814	0	244.13
	2	0.9812	5.751	4.920	3.702	3.702	3.702	26.7	22.7	20.0	12.0	0.9812	0	242.89
	3	0.9812	5.184	4.920	3.702	3.702	3.703	30.0	25.3	18.7	13.3	0.9812	0	254.99
	4	0.9819	2.390	4.937	3.692	3.695	3.698	28.7	26.7	23.3	18.7	0.9819	0	246.53
	5	0.9818	2.152	4.953	3.699	3.699	3.697	31.3	28.7	26.7	20.0	0.9818	0	247.88
	6	0.9819	2.793	4.920	3.651	3.665	3.665	26.0	20.0	16.0	12.0	0.9819	0	246.06
	7	0.9813	2.788	4.937	3.668	3.670	3.672	24.7	18.7	12.7	10.0	0.9813	0	238.99
	8	0.9813	3.333	4.888	3.675	3.678	3.677	24.0	20.7	17.3	13.3	0.9813	0	242.16
	9	0.9813	2.342	4.888	3.683	3.685	3.684	25.3	23.3	20.0	14.7	0.9813	0	241.23
	10	0.9816	1.657	4.888	3.675	3.682	3.684	26.0	22.0	20.0	12.0	0.9816	0	245.12
	11	0.9816	1.650	4.888	3.524	3.535	3.544	27.3	22.0	16.7	12.0	0.9816	0	245.24
	12	0.9817	1.067	4.937	3.542	3.551	3.555	27.3	24.0	19.3	12.0	0.9817	0	244.43
	13	0.9845	1.804	5.035	3.585	3.591	3.588	24.0	20.7	18.0	12.0	0.9845	0	250.50
	14	0.9840	2.150	4.937	3.584	3.589	3.590	24.0	21.3	18.7	14.0	0.9840	0	251.78
	15	0.9834	2.293	4.986	3.592	3.597	3.599	23.3	21.3	17.3	12.0	0.9834	0	252.23
	16	0.9876	2.372	5.018	3.617	3.614	3.612	15.3	11.3	8.7	10.7	0.9876	0	233.28
	17	1.0225	3.138	4.920	3.560	3.560	3.554	12.7	8.0	6.7	5.3	1.0225	0	224.17
	18	0.9821	0.490	4.920	3.662	3.660	3.657	24.7	22,7	19.3	16.0	0.9821	0	237.08
	19	0.9819	0.779	4.920	3,671	3.667	3.662	25.3	21,3	18.0	13.3	0.9819	0	243.05
	20	0.9832	0.659	4.904	3.660	3.659	3.653	27.3	24,7	20.0	15.3	0.9832	0	250.61
	21	0.9839	1.141	4.888	3.662	3.662	3.659	24.0	20.0	16.7	10.0	0.9839	0	238.33
	22	0.9825	1.753	4.855	3.683	3.681	3.677	22.0	12.0	10.0	4.7	0.9825	0	247.31
	23	0.9826	1.351	4.888	3.683	3.682	3.677	26.7	22.0	18.0	12.7	0.9826	0	241.60
	24	0.9826	1.356	4.855	3.678	3.677	3.674	27.3	22.0	18.7	14.7	0.9826	0	241.20
	25	0.9806	1.496	4.773	3.691	3.691	3.688	26.0	20.7	19.3	15.3	0.9806	0	242.13
	26	0.9850	1.345	4.937	3.663	3.664	3.659	24.0	20.7	16.0	10.0	0.9850	0	238.66
	27	0.9848	1.653	5.002	3.663	3.663	3.659	26.0	22.7	19.3	14.7	0.9848	0	235.08
	28	0.9837	1.422	4.986	3.661	3.662	3.660	24.7	22.7	18.7	12.7	0.9837	0	240.24
	29	0.9816	0.794	4.904	3.682	3.688	3.685	25.3	22.0	18.0	13.3	0.9816	0	236.08
	30	0.9815	1.362	4.904	3.681	3.685	3.682	26.7	20.7	19.3	14.0	0.9815	0	241.31
Reference samples	100	0.9811	3.375	4.855	3.686	3.685	3.682	28.0	22.0	18.7	14.7	0.9811	0	248.03
	101	0.9815	1.161	4.904	3.686	3.683	3.676	24.0	22.0	18.7	14.7	0.9815	0	251.33
	102	0.9815	1.019	4.953	3.677	3.676	3.670	26.0	22.7	18.0	12.7	0.9815	0	261.37
	103	0.9815	0.887	4.953	3.662	3.665	3.668	26.7	22.0	19.3	14.0	0.9815	0	252.48
	104	0.9815	0.959	4.953	3.670	3.666	3.667	26.7	23.3	18.7	12.0	0.9815	0	251.66
	105	0.9822	0.639	4.937	3.647	3.640	3.644	27.3	25.3	22.0	17.3	0.9822	0	246.58
	106	0.9815	0.747	4.969	3.665	3.660	3.666	28.7	23.3	20.0	15.3	0.9815	0	255.88
	107	0.9815	0.765	5.035	3.664	3.659	3.662	25.3	22.0	20.0	15.3	0.9815	0	256.25
	108	0.9815	0.673	4.953	3.657	3.647	3.654	26.7	21.3	18.7	16.7	0.9815	0	256.78
	109	0.9815	0.533	4.986	3.660	3.655	3.660	25.3	22.7	20.0	15.3	0.9815	0	251.00
	110	0.9814	0.752	4.986	3.670	3.664	3.668	28.0	23.3	18.0	14.0	0.9814	0	252.53
	111	0.9814	0.718	4.986	3.669	3.665	3.669	25.3	21.3	19.3	18.0	0.9814	0	255.62
	112	0.9814	0.698	4.969	3.668	3.667	3.669	26.7	24.0	20.0	16.7	0.9814	0	263.49
	113	0.9822	0.671	4.969	3.667	3.665	3.667	25.3	20.7	18.0	12.0	0.9822	0	253.40

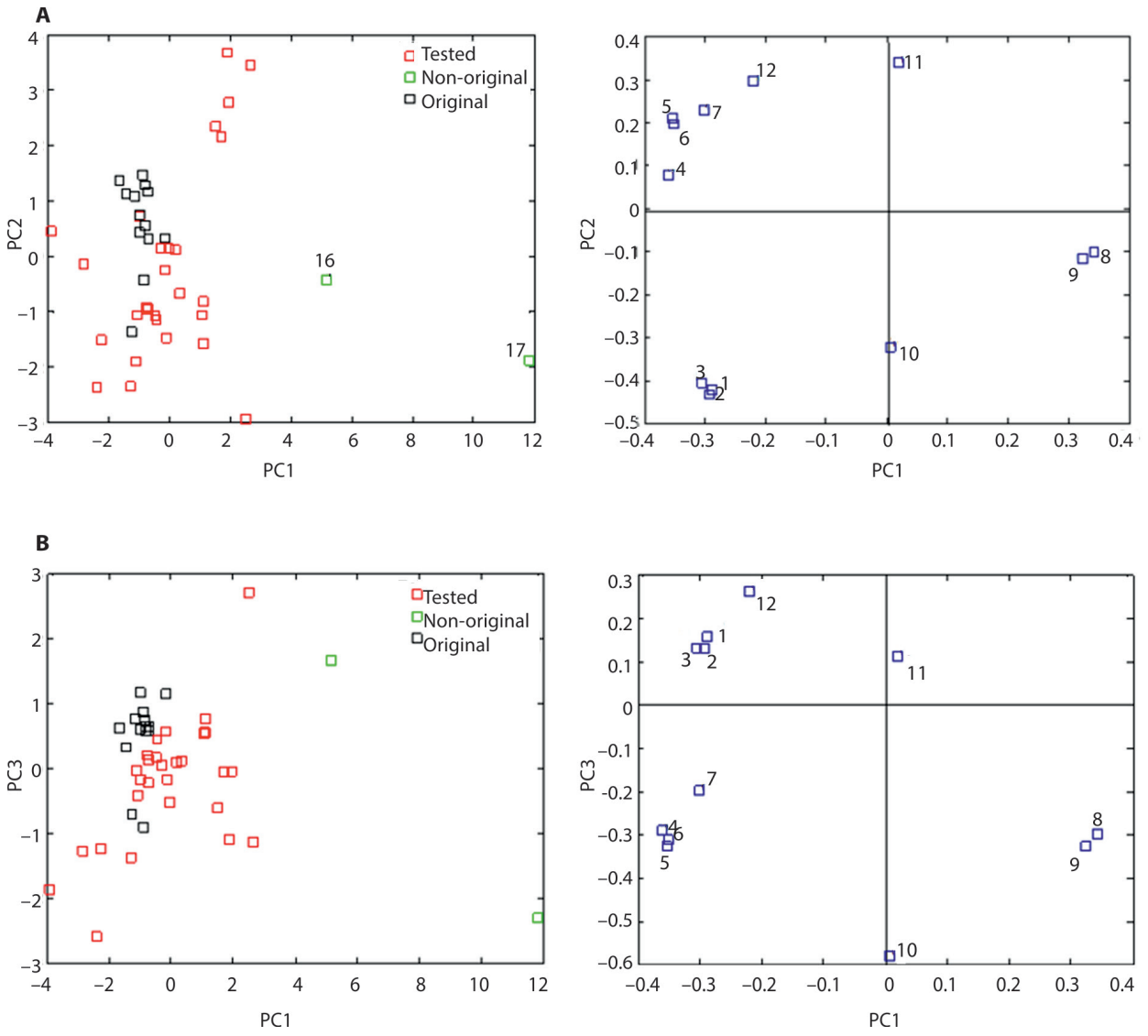

Fig. 1. Projection of samples on the plane defined by factors: A – PC 1 and PC 2; B – PC 1 and PC 3 (1 – pH after 1 min; 2 – pH after 2 min; 3 – pH after 10 min; 4 – foam stability after 10 s; 5 – foam stability after 1 min; 6 – foam stability after 3 min; 7 – foam stability after 12 min; 8 – density; 9 – viscosity; 10 – water content; 11 - acidity; 12 – active substance content)

weights for the components was interpreted. Figure 1 shows the distribution of samples on the plane defined by PC 1 and PC 2, PC 1 and PC 3, PC 2 and PC 3, and the contribution of parameters contained in the table of physicochemical properties (excluding the emulsion designation values, as the emulsion value was zero for all studied preparations).

The distribution of samples on the plane defined by the first (PC 1) and second (PC 2) component was analyzed. Non-original samples were moderately separated from the other samples concerning PC 1, with the highest contribution to this component being variable 8 (density), and variable 9 (viscosity). The distribution of samples on the plane defined by PC 1 and PC 3 showed a slight separation of non-original samples from the other samples concerning PC 1. The highest contribution to this component was made by variable 8 (density) and variable 9 (viscosity). Analyzing the distribution of samples on the plane defined by the second (PC 2) and third (PC 3) component did not reveal any substantial trend in grouping.

Exploratory analysis was conducted on the entire data matrix containing 396 chromatographic signals obtained using the HPLC-DAD technique, consisting of 2250 data points. In subsequent steps of the exploratory analysis, the degree of data compression and the distribution of samples on planes defined by selected pairs of principal components were assessed. A PCA model was built, achieving a data variance compression of 99.9% for the first three components, with the first component explaining 77.3% of the variance and the second component explaining 99.6%. In the next step of data exploration, the distribution of

samples on planes defined by selected pairs of principal components PC 1 and PC 2, PC 1 and PC 3, PC 2 and PC 3 for unaveraged centered data is presented in Figure 2. In the presented figures, reference samples (original) are marked in red, samples classified by the laboratory as non-original are marked in green, and samples belonging to the studied group are marked in black.

In the case of the chromatographic data analysis obtained by HPLC using the PCA method, no grouping of samples was observed. Non-original samples marked in green are mingled with reference samples marked in red.

Exploratory analysis was conducted on the entire data matrix containing 132 chromatographic signals obtained using the HS-GC/MS technique, consisting of 4061 data points. In subsequent steps of the exploratory analysis, the degree of data compression and the distribution of samples on planes defined by selected pairs of principal components were assessed. A PCA model was built, achieving a data variance compression of 92.3% for the first three components, with the first principal component explaining 58.0% of the variance and the second component explaining 85.8%. The first 19 principal components describe nearly 100% of the data variance. In the next step of data exploration, the distribution of samples on planes defined by selected pairs of principal components PC 1 and PC 2, PC 1 and PC 3, PC 2 and PC 3 for unaveraged centered data is presented in Figure 3.

In the case of the distribution of samples on the plane defined by the first (PC 1) and the second and third factors (PC 2, PC 3), one sample considered non-original (sample no. 17) was observed to be separated. Sample 17 separates from the rest concerning factor PC 1. The observed trend towards grouping required the analysis of weights for the first factor.

Figure 4 shows the weights indicating which peaks contribute the most to its construction. Analyzing the distribution of samples on the plane defined by the second (PC 2) and third (PC 3) factors did not reveal any interesting trend in grouping.

The analysis of weights for the first factor showed that the greatest contribution to its construction came from variables located in the chromatographic signals around data points 1166, 2335, and 2921, corresponding to analysis times of 9.0, 15.9, and 19.4 minutes.

In the next stage of the analysis, the hierarchical clustering method was used. Dendrograms were constructed for the set of raw and autoscaled data using various combinations of distance evaluation between objects (Euclidean and Mahalanobis) and methods for clustering them (single linkage, complete linkage, Ward's method, average linkage). Different combinations of

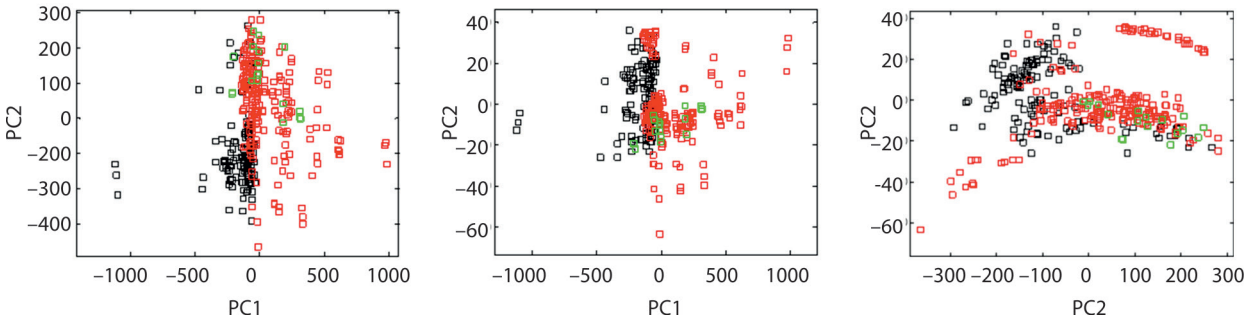

Fig. 2. Projection of samples on the plane defined by factors (unaveraged data) (red markers – tested; green markers – non-original; black markers - original)

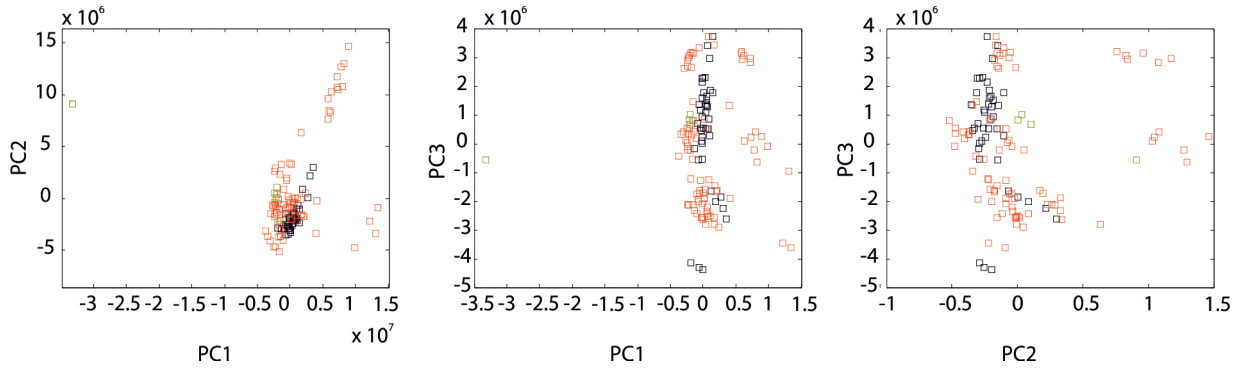

Fig. 3. Projection of samples on the plane defined by factors (unaveraged data) (red markers - tested; green markers – non-original; black markers - original)

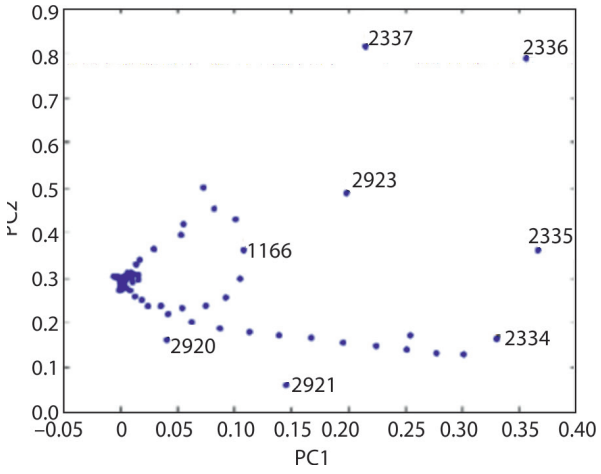

Fig. 4. Distribution of weights for the first and second factors

distance calculations and object clustering yielded the same results. Example results are presented in Figure 5.

The analysis of dendrograms obtained using various combinations of distance evaluation between objects and clustering methods yielded the same results. Samples 16 and 17, initially classified by the laboratory as non-original, form a separate cluster distant from the rest of the objects in the dataset.

The application of supervised chemometric methods

Centering was applied to the data matrices derived from the instrumental signals to eliminate constant elements that do not affect data variability. The prepared

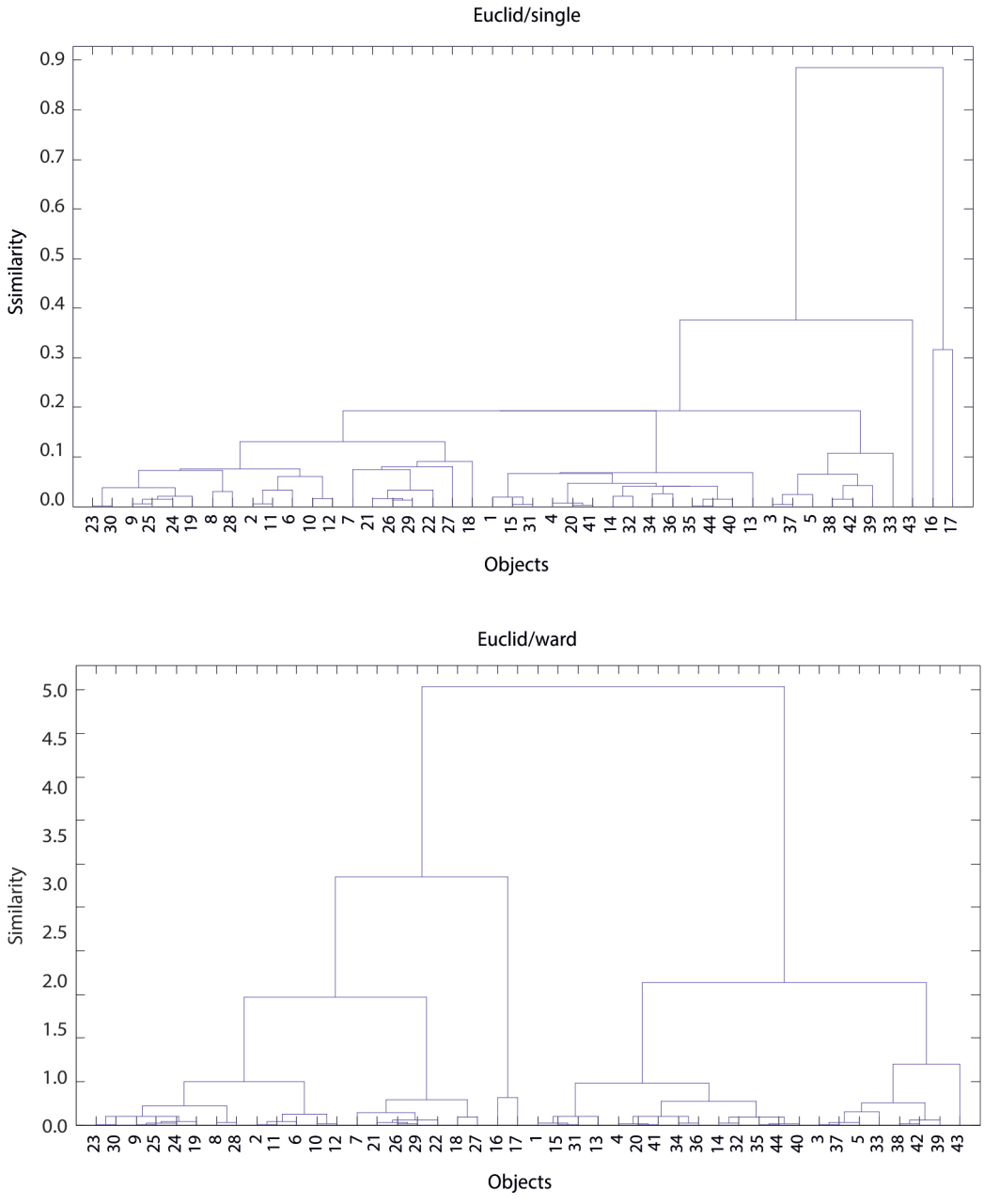

Fig. 5. Example dendrograms constructed using Euclidean distance and different linkage methods (single linkage, Ward's method)

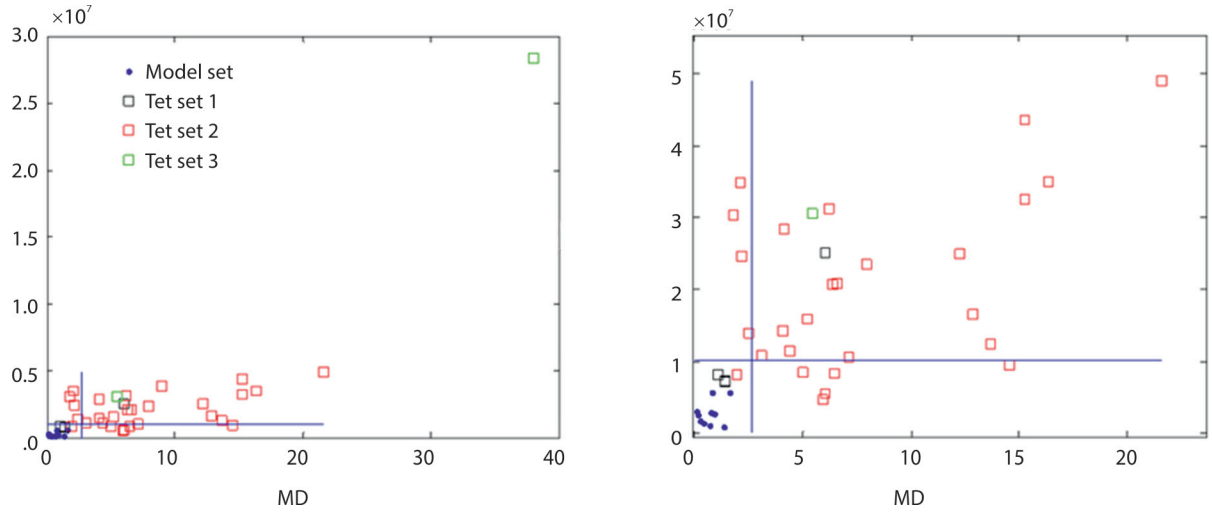

Fig. 6. Performance results of the SIMCA model built for chromatographic data obtained using HS-GC/MS, with a complexity of five factors, for the model and test sets

instrumental signals were then analyzed using the SIMCA method. In SIMCA, classification rules were constructed based on individual PCA models for the studied group of authentic products. To determine the optimal complexity of the PCA model, leave-one-out cross-validation was applied. After completing the cross-validation procedure, the root mean square error of cross-validation (RMSECV) was calculated from the resulting matrix. From the plot of RMSECV values against the number of principal components, the point where the curve stabilizes indicates the optimal number of components. Based on the obtained plots, nine principal components were selected, describing 98.3% of the studied data.

The instrumental signals obtained using both HS-GC/MS and HPLC techniques for the studied formulation were initially divided into three subsets: Subset I consisted of signals recorded for reference samples (model and Test 1); Subset II included signals recorded for samples preliminarily identified as "original" (Test 2); and Subset III comprised signals recorded for samples preliminarily identified as "non-original" (Test 3). From Subset I, 75% of the samples were selected using the Kennard-Stone algorithm to form the model set, while the remaining signals were included in test set 1. These constructed sets were used to build and evaluate the performance of SIMCA models.

For the data obtained using HS-GC/MS, five factors were selected for constructing the SIMCA model for the class of "original" samples using the scree plot method, as including more principal components did not yield a significant increase in information.

Analyzing the model performance shown in Figure 6, it was determined that all objects from test set 3 had MD and OD values exceeding the threshold values determined for the modeled class. Two objects from test set 1 (samples no. 105, 106) and one object from test set 2 (sample no. 23) were recognized as belonging to the "original" class.

For the data obtained using HPLC-DAD, the scree plot method was used to select two factors for constructing the SIMCA model for the class of "original" samples, as adding more principal components did not provide a substantial increase in information.

Analyzing the performance quality of the model shown in Figure 7, it was found that all objects from test sets 2 and 3 have MD and OD values exceeding the threshold values determined for the modeled class. Two objects from test set 1 (sample no. 105 and 106) were identified as belonging to the "original" class.

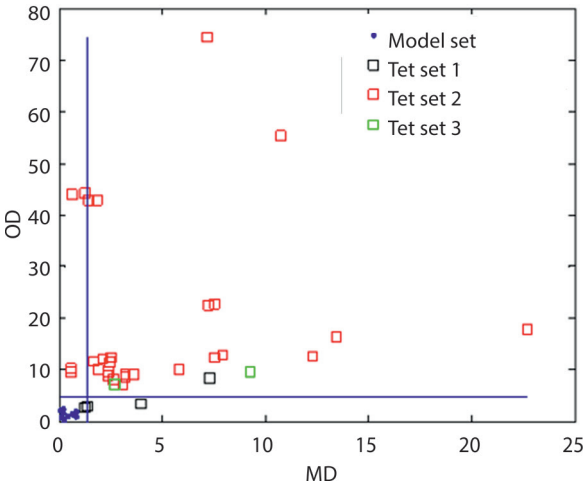

Fig. 7. Performance results of the SIMCA model, constructed with a complexity of two factors for chromatographic data obtained using HPLC-DAD, for both the model and test sets

Discussion

The developed method for determining the active substance content, in accordance with document Sanco 3030 rev.5, is characterized by appropriate linearity (correlation coefficient higher than the required minimum value > 0.99), acceptable precision based on the Horrat value meeting the required criterion (Hr ≤ 1), and accurate recovery values confirming the reliability of the developed methods (the recovery condition for active substances with analyte concentration ≥ 10% is met within the range of 97–103%).

The determined values of physical, chemical, and technical parameters and the obtained instrumental signals from the laboratory analyses of the tested formulations were subjected to exploratory analysis using methods such as PCA and HCA. The exploratory data analysis showed a tendency for the samples to cluster according to the vector containing preliminary information about the originality and non-originality of the formulations. A clear separation of samples 16 and 17, which are apart from the rest of the objects in the set, can be observed. The physicochemical parameters having the most significant impact on the observed trend are the amount of stable foam after 10 seconds, density, and viscosity. These parameters should be thoroughly analyzed during planned laboratory analyses.

For the analysis of chromatographic data obtained by HPLC using PCA, no grouping of the samples was observed. In the case of data obtained by HS-GC/MS and the analysis of the distribution of samples on the plane defined by the first (PC 1) and second and third factors (PC 2, PC 3), a separation of one sample deemed non-original (sample no. 17), which separates from the rest relative to factor PC 1, was observed. The observed tendency for grouping required an analysis of the weights for the first factor, which showed that the variables at 9.0, 15.9, and 19.4 minutes of analysis had the most significant contribution to its construction - during future analyses, this time should be closely monitored.

The quality of the SIMCA model was assessed based on its performance for two independent test sets. The first one – test set 1 – included only samples from the same class for which the model was constructed. Based on the number of correctly recognized samples, a parameter called sensitivity was calculated. The second test set – test set 2 – contained only objects outside the modeled class. Based on the number of samples from this set, which were recognized as not belonging to the modeled class, a parameter called specificity was calculated. The number of all samples contained in the model, test set 1, and test set 2, which were correctly assigned to previously defined classes, is the basis for calculating the quality parameter known as the correct classification rate. In the case of building models for instrumental data obtained by HPLC for the tested formulation, the model sensitivity is at the level of 50% for data obtained by HS-GC/MS at the level of 75%. However, low specificity does not confirm the possibility of using this model to verify the originality of the formulation, even though samples preliminarily classified by the laboratory as non-original were classified by the model in the same way. Therefore, to obtain reliable results of chemometric analysis, it is necessary to use samples of the original product supplied directly by the manufacturer and samples whose quality has been unequivocally questioned. Given the obtained results, consideration should be given to the possibility of using SIMCA models, built according to appropriately prepared chromatographic signals, in routine analyses. This will be possible after verifying the models' performance for data with an increased number of "non-original" samples. Therefore, when planning further laboratory studies, special attention should be paid to balancing the number of samples classified as "original" and "non-original." Chemometric techniques verify product authenticity and help eliminate counterfeit or low-quality products from the market. This proactive approach maintains the integrity of agricultural practices, protects plant health, and minimizes potential environmental and human health hazards. This analytical method strengthens overall quality control and regulation efforts for plant protection products.

Acknowledgements

This work was supported by the National Centre for Research and Development under Grant GOSPOSTRATEG 1/385957/5/NCBR/2018.

References

Bajoub A., Medina-Rodríguez S., Gómez-Romero M., Ajal E.A, Bagur-González M.G., Fernández-Gutiérrez A., Carrasco-Pancorbo A. 2017. Assessing the varietal origin of extra-virgin olive oil using liquid chromatography fingerprints of phenolic compound, data fusion and chemometrics. Food Chemistry 215: 245–255. DOI: https://doi.org/10.1016/j.foodchem.2016.07.140

Belmonte-Sánchez J.R., Gherghel S., Arrebola-Liébanas J., González R.R., Martínez Vidal J.L., Parkin I., Garrido-Frenich A. 2018. Rum classification using fingerprinting analysis of volatile fraction by headspace solid phase microextraction coupled to gas chromatography-mass spectrometry. Talanta 187: 348–356. DOI: https://doi.org/10.1016/j.talanta.2018.05.025

Biondi A., Desneux N., Siscaro G., Zappalà L. 2012. Using organic-certified rather than synthetic pesticides may not be safer for biological control agents: Selectivity and side ef-

fects of 14 pesticides on the predator *Orius laevigatus*. Chemosphere 87 (7): 803–812. DOI: https://doi.org/10.1016/j.chemosphere.2011.12.082

Craveiro de Andrade J., Teixeira de Oliveira A., Glauciney Fernandes Macedo Amazonas M., Galvan D., Tessaro L., Conte-Junior C.A. 2024. Fingerprinting based on spectral reflectance and chemometrics – An analytical approach aimed at combating the illegal trade of stingray meat in the Amazon. Food Chemistry 436: 137637. DOI: https://doi.org/10.1016/j.foodchem.2023.137637

Custers D., Canfyn M., Courselle P., De Beer J. O., Apers S., Deconinck E. 2014. Headspace-gas chromatographic fingerprints to discriminate and classify counterfeit medicines. Talanta 123: 78–88. DOI: https://doi.org/10.1016/j.talanta.2014.01.020

Custers D., Krakowska B., De Beer J.O., Courselle P., Daszykowski M., Apers S., Deconinck E. 2016. Chromatographic impurity fingerprinting of genuine and counterfeit Cialis® as a means to compare the discriminating ability of PDA and MS detection. Talanta 146: 540–548. DOI: https://doi.org/10.1016/j.talanta.2015.09.029

EUIPO. 2017. The economic cost of IPR infringement in the pesticides sector. [Available from: https://europa.eu/pesticides_sector_en.pdf][Accessed: January 2024].

Europol. 2011. Growth in the trade in counterfeit and other illegal pesticides across. [Available from: https://croplife.org/wp-content/uploads/pdf_files/Europol-AC-Policy-Brief.pdf] [Accessed: January 2024].

FAO. 2010. International Code of Conduct on the distribution and use of pesticides. Guidance on Pest and Pesticide Management Policy Development No. 978-92-5-106827-4. Rome: Inter-Organization Programme for the Sound Management of Chemicals (IOMC).

FAOSTAT. 2018. [Online] Available from: https://www.fao.org/faostat/en/#data [Accessed January 2024].

Flumignan D., Ferreira F., Tininis A., Lopes M., Oliveira J. 2008. Development, optimization and validation of gas chromatographic fingerprinting of Brazilian commercial gasoline for quality control. Journal of Chromatography A 1202 (2): 181–188. DOI: https://doi.org/10.1016/j.chroma.2008.06.040

Frezal C., Garsous G. 2020. New digital technologies to tackle trade in illegal pesticides. OECD Trade and Environment Working Papers, No. 2020/02. OECD Publishing. DOI: https://doi.org/10.1787/9383b310-en

Jankowska M., Hrynko I., Łozowicka B. 2022. Human health risk assessment of pesticide residues in fruit, vegetable and cereal samples from Poland – a 5-year survey. Journal of Plant Protection Research 62 (4): 385–392. DOI: https://doi.org/10.24425/jppr.2022.143231

Li J., He X., Li M., Zhao W., Liu L., Kong X. 2015. Chemical fingerprint and quantitative analysis for quality control of polyphenols extracted from pomegranate peel by HPLC. Food Chemistry 176: 7–11. DOI: https://doi.org/10.1016/j.foodchem.2014.12.040

Marczewska P., Miszczyk M., Płonka M., Kronenbach-Dylong D., Szeremeta D., Sajewicz M. 2019. Application of different chromatographic techniques and chemometric analysis in authenticity testing of plant protection products containing azoxystrobin as an active substance. Journal of Environmental Science and Health, Part B 54 (7): 590–597. DOI: https://doi.org/10.1080/03601234.2019.1610298

Miszczyk M., Płonka M., Bober K., Dołowy M., Pyka A., Pszczolińska K. 2015. Application of chemometric analysis based on physicochemical and chromatographic data for the differentiation origin of plant protection products containing chlorpyrifos. Journal of Environmental Science and Health, Part B 50 (10): 744–751. DOI: https://doi.org/10.1080/03601234.2015.1048108

Rivera-Pérez A., García-Pérez P., Romero-González R., Garrido Frenich A., Lucini L. 2023. UHPLC-QTOF-HRMS metabolomics insight on the origin and processing authentication of thyme by comprehensive fingerprinting and chemometrics. Food Chemistry 407: 135123. DOI: https://doi.org/10.1016/j.foodchem.2022.135123

Streloke M. 2018. Illegal trade of plant protection products: a highly profitable way to smuggle chemicals. Journal of Consumer Protection and Food Safety 13: 255–256. DOI: https://doi.org/10.1007/s00003-018-1177-6

Sun X-D., Zhang M., Zhang S., Wang P-J., Chen J-H., Gao X-L. 2023. Non-targeted HPLC–FLD fingerprinting for the classification, authentication, and fraud quantitation of Guizhou paprika by chemometrics. Journal of Food Composition and Analysis 120: 105346. DOI: https://doi.org/10.1016/j.jfca.2023.105346