

Influence of Raman spectra measurement conditions on the dating results of writing compositions

KSENIA ERSHOVA

ORCID: 0000-0002-6815-8467

Nanotechnology Research and Education Centre of the Russian Academy of Sciences,
Saint Petersburg Academic University, Russia

SVETLANA KOCHEMIROVSKAYA

ORCID: 0000-0001-7592-3860

Peter the Great St. Petersburg Polytechnic University, Russia

NATALIA P. KIRILLOVA

ORCID: 0000-0002-2692-4346

St. Petersburg State University, Russia

VLADIMIR KOCHEMIROVSKY

ORCID: 0000-0002-0352-3793

Peter the Great St. Petersburg Polytechnic University, Russia

MIKHAIL RYAZANTSEV

ORCID: 0000-0003-3413-1706

Nanotechnology Research and Education Centre of the Russian Academy of Sciences,
Saint Petersburg Academic University, Russia

Abstract

When establishing the age of a document using spectral methods, the result is influenced not only by the storage conditions, but also by the methods of measuring the spectra and mathematical processing of the results. This work analyzes the influence the modes of Nowa Kodyfikacja Prawa Karnego 60, 2021

© for this edition by CNS

measuring Raman spectra have on the result of measuring the writing composition spectra. The analysis area includes the correlation coefficients of the full spectral matrices as a criterion for the reliability of the determination date. The influence of focusing accuracy, correctness of subtraction of the baseline, sample irradiation time, and varying the number of repetitions is investigated. In the study, chemometric tools for the analysis of Raman spectra were also applied. The influence of mathematical spectra processing shows that the components PC4, P5, and PC6 are most closely related to the shelf life, although they do not make the maximum contribution to the variance in the system. Taking into account all the requirements for measuring spectra, the PCA method applied to the Raman spectra of writing compositions allows clustering samples by manufacturer's brands and by the sample shelf life.

Keywords: dating documents, questioned document examination, Raman spectroscopy, chemometric analysis, PCA

Despite the large number of chromatographic methods used to determine the dating of documents, no one has yet achieved breakthrough results. The accuracy of determinations varies around 70–80% of hits on control samples with a known initial composition and storage conditions. Interlaboratory tests show insufficient convergence of results; therefore, chromatographic studies are currently developing in three main areas.¹

Attempts were made to calculate the regression parameters of the volatile component content that are independent or weakly dependent on the initial compositions and storage conditions of the handwritten inscription. In the terminology of Díaz-Santana et al.² – these are called “dynamic dating methods,” the study of changes in the concentration of one or more ink components over time.³ It also includes searching for time markers that correspond to a particular life period of a handwritten item. Sharma

¹ O.A. Díaz-Santana, F.B. Conde-Hardisson, D.C. Vega-Moreno, “Comparison of the main methods for six ballpoint pen inks”, *Microchemical Journal* 138, 2018, pp. 550–561.

² Ibid.

³ V. Sharma, R. Kumar, “Dating of ballpoint pen writing inks via spectroscopic and multiple linear regression analysis: A novel approach”, *Microchemical Journal* 134, 2017, pp. 104–113; V.N. Aginsky, “Measuring ink extractibility as a function of age. Why the relative aging approach is unreliable and why it is more correct to measure ink volatile components than dyes”, *International Journal of Forensic Document Examiners* 4, 1998, pp. 214–230.

and Kumar call these static dating methods⁴ – relative dating methods based on comparing two independent ink samples available in the same document. This is mainly used for analyzing documents with some changes (additions, corrections).⁵

By default, all researchers assume that there are two main reasons for the span of results on determining the inscription age: the uncertainty of the initial composition and of the storage conditions, primarily temperature. However, there are many factors that have not yet been studied in detail, so the list can still be expanded.

The problem of normalizing the chromatographic signal to the inscription mass has long been discussed in the literature. A large number of authors suggest solving it by normalizing the chromatographic signal intensity to the optical density of the dye extract according to spectrophotometric analysis.⁶

This approach seems indisputable, given that the optical density of the extract is additive for each component, provided that the content of all other components in the mixture is constant. A real mixture undergoes temporary changes and can change its composition and non-additively change the optical density simultaneously for several components. The proposal to use the inscription width as a mass characteristic looks entirely frivolous.⁷

However, this approach to spectroscopic data gives correct results, since the lines of the same spectrum are compared as a result of a single measurement rather than two independent ones performed by two completely different physical and chemical methods on two heterogeneous objects (dye and volatile components).

⁴ C. Weyermann, D. Kirsch, B. Spengler, “A GC/MS study of the drying of ballpoint pen ink on paper”, *Forensic Science International* 168, 2007, pp. 119–127; R.L. Brunelle, “Ink dating: The state of the art”, *Journal of Forensic Sciences*, 37, 1992, pp. 113–124.

⁵ A. Cantú, R.S. Prough, “On the relative aging of ink: The solvent extraction technique”, *Journal of Forensic Sciences* 32, 1987, pp. 1151–1174.

⁶ Y. Xu, J. Wang, L. Yao, “Dating the writing age of black roller and gel inks by gas chromatography and UV-vis spectrophotometer”, *Forensic Science International* 162, 2006, pp. 140–143.

⁷ E.A. Trosman et al., “Methodology: Age-dating of document attribute entries by the relative content of volatile solvents in strokes”, *Теория и практика судебной экспертизы (Theory and Practice of Forensic Science)* 30, 2013, no. 2, pp. 80–88.

The second direction of chromatographic research – the search for time markers corresponding to a particular life period, is based on a similar principle that has long been used in art history, archeology, and historical analysis.⁸ The method is to find the concentration ratios of various ink components that have been stored for a long time and subjected to artificial aging. The results obtained are not presented in an analytical form, but rather compared with different existence periods of the inscription. The most prominent representatives of this direction are the works by Aginsky.⁹ The ratio of the amounts of the inscription component recovered by the extraction method compared to standard samples is selected here as a temporary marker.

Given the new data, researchers have a growing understanding that the possibilities of the monomethodic approach are limited and combinations of heterogeneous methods are the future. There is a growing interest in research using spectral methods. This is because, despite the multiplicity of chromatographic methods, no one has yet been able to achieve ideal results. At the same time, the trend is such that it is not a question of abandoning chromatography; the possibilities of spectroscopy complement its capabilities.¹⁰

⁸ L. Giorgi et al., “In-situ technical study of modern paintings. Part 2: Imaging and spectroscopic analysis of zinc white in paintings from 1889 to 1940 by Alessandro Milesi (1856–1945)”, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 219, 2019, pp. 504–508; J. Lee et al., “Scientific investigation into the water sensitivity of twentieth century oil paints”, *Microchemical Journal* 138, 2018, pp. 282–295; A. Burnstock, K.J. van den Berg. “Twentieth century oil paint. The interface between science and conservation and the challenges for modern oil paint research”, *Issues in contemporary oil paint: Proceedings of the Issues in Contemporary Oil Paint (ICOP) symposium*, 28–29 March 2013, Amersfoort, Netherlands.

⁹ V.N. Aginsky, “Forensic examination of ‘slightly soluble’ ink pigments using thin-layer chromatography”, *Journal of Forensic Sciences* 38, 1993, no. 5, pp. 1131–1133; V.N. Aginsky, “Determination of the age of ballpoint pen ink by gas and densitometric thin-layer chromatography”, *Journal of Chromatography A* 678, 1994, no. 1, pp. 119–125; V.N. Aginsky, “Measuring ink extractibility as a function of age. Why the relative aging approach is unreliable and why it is more correct to measure ink volatile components than dyes”, *International Journal of Forensic Document Examiners* 4, 1998, pp. 214–230.

¹⁰ N. Grechukha et al, “Analysis of the aging processes of writing ink: Raman spectroscopy versus gas chromatography aspects”, *Applied Sciences* 7, 2017, no. 10, p. 991.

The electronic spectrum of the writing compositions (a complex organic compound) is a non-selective and non-additive parameter. Therefore, temporary chemical changes can affect the spectrum in different directions. An example of this is the dynamics of the Raman scattering spectra of writing compositions studied by Gorshkova et al.¹¹ In this work, physicochemical mechanisms of degradation of triarylmethane dyes were proposed and spectral lines were determined that change over time as a result of dye oxidation by atomic oxygen.

Raman dye spectroscopy has many advantages over other spectral and chromatographic methods. This is primarily a non-invasive nature of the impact on the document. The spot of “exposure” of the inscription is only 5 microns in diameter and is not noticeable to the human eye. Dyes are the most stable components of the writing composition. All transformations with them are slow. This allows going beyond the two years that chromatographic studies are limited to. Finally, the main advantage is selectivity. The Raman spectrum of an ink dye consists of a set of narrow lines, each of which can be correlated with the oscillation of a specific bond in the dye molecule. UV and IR spectra consist of wide absorption bands, in which it is often difficult to distinguish the contribution of a particular ink component. Therefore, interest in the method is growing, although so far it is used mainly for tasks of identifying ink, toners, and other writing materials.¹²

Therefore, the primary current trend in developing methods for dating inscriptions of writing compositions is the simultaneous use of several heterogeneous physical and chemical methods and mathematically pro-

¹¹ K.O. Gorshkova et al., “The investigation of dye aging dynamics in writing inks using Raman spectroscopy”, *Dyes and Pigments* 131, 2016, pp. 239–245.

¹² I. Geiman, M. Leona, J.R. Lombardi, “Application of Raman spectroscopy and surface-enhanced Raman scattering to the analysis of synthetic dyes found in ballpoint pen inks”, *Journal of Forensic Science* 54, 2009, no. 4, pp. 947–951; M. Kunicki, E. Fabiańska, A. Parczewski, “Raman spectroscopy supported by optical methods of examination for the purpose of differentiating blue gel pen inks”, *Problems of Forensic Sciences* 95, 2013, pp. 627–641; P. Buzzini, C. Polston, M. Schackmuth, “On the criteria for the discrimination of inkjet printer inks using micro-Raman spectroscopy”, *Journal of Raman Spectroscopy* 49, 2018, no. 11, pp. 1791–1801.

cessing a complete data array.¹³ The use of spectral and chromatographic analysis methods to determine the dating of handwritten inscriptions leads to the need to use new methods of mathematical signal processing. First of all, these are methods for processing multidimensional data arrays.

Currently, principal component analysis (PCA) is widely used for chemometric data processing. This method reduces the dimension of the original data set by calculating a new set of variables called principal components. The capabilities of chemometrics allow obtaining more objective results than visual assessment. The chemometric approach allows determining the number of main factors R equal to the number of ink component forms, whose the mutual conversion is due to the dispersion of the analytical signal.¹⁴ This method provides a unique solution to the equation that describes the maximum variance using a fixed number of variables. However, for reliable application of the method, it is necessary to evaluate the influence of measurement modes in order to exclude all possible systematic errors that are not associated with temporary dye degradation.

This paper describes experiments related to obtaining reliable spectral data suitable for applying chemometric data processing procedures for Raman spectroscopy of writing compositions to the problem of dating handwritten inscriptions on documents.

Methods and materials

Writing compositions of various brands that are most often found in the market were used for the research.

The following substances were used in the study: dyes/pigments in the form of powders: fat-soluble violet K, alcohol-soluble blue phthalocyanine, cationic turquoise, straight violet, basic violet K, acid violet C, acid bright blue Z, blue phthalocyanine pigment, Victoria blue, crystal violet, methyl violet, rhodamine Y, varnish basic blue K.

¹³ K.O. Gorshkova et al., “Investigation of the new possibility of mathematical processing of Raman spectra for dating documents”, *Science & Justice* 60, 2020, no. 5, pp. 451–465.

¹⁴ E.R. Malinowsky, D.G. Howery, *Factor analysis in chemistry*, New York 1980, p. 251; R. Tauler, A. de Juan, “Multivariate curve resolution”, [in:] *Practical guide to chemometrics*, ed. P. Gemperline, London 2006, p. 421.

Raman spectra were obtained using a Bruker SENTERRA Raman spectrometer equipped with three lasers with wavelengths of 488, 532, and 785 nm. The authors also used an Olympus confocal microscope module and a fiber-optic sensor for recording the spectra of samples in an external cuvette compartment. The spectra were measured in the range of 280–1735 cm^{-1} .

The method of mathematical signal processing by PCA was used. For this purpose, a database was formed, which is a set of Raman spectra obtained for samples of writing compositions no. 1–5 (Table 1) with a different period of application on a paper base. The selected writing compositions include a different set of coloring agents.

Table 1. The main characteristics of the studied writing compositions

No. (DB)	Writing composition brand	Type ¹⁵
1 (17)	Paper Mate Korea pen	B*
2 (23)	Easy Clicker BEIFA pen	B
3 (25)	Erich Krause refill	B
4 (4)	ICO Silver pen (Parker type)	B
5 (24)	Erich Krause U18 Pen	B

*B – ballpoint pen

Raman spectra were obtained from the same inscription from different points with an interval of about 5 months, for storage periods: 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90 months; 100 spectra for each date. The database used for calculations has more than 7 000 spectra.

The PCA of spectra was implemented in the R Studio program in the R programming language. All the spectra were loaded into the program in the form of tables without preprocessing, after which the “prcomp” function was applied, which uses singular value decomposition rather than spectral (unlike the “princomp” function). The variables were scaled before the analysis using the “scale” function – it accepts a numeric matrix

¹⁵ According to K.O. Gorshkova et al., “The investigation of dye aging dynamics...”; K.O. Gorshkova et al., “Investigation of the new possibility...”.

as input and performs column scaling, where scaling is performed using the following formula:

$$\frac{x_i - \text{mean}(x)}{\text{sd}(x)}$$

where $\text{sd}(x)$ – standard deviation, $\text{mean}(x)$ – the average value of x . After that, the main components were calculated.

The Raman spectra were obtained using the same parameters. Python was chosen as the programming language as it is one of the most widely used languages for data analysis and has a large number of packages for all modern methods of data processing and visualization.

Three collections of samples were formed to solve this problem:

1. provided by arbitration courts in different cities of Russia,
2. provided by a third-party commercial organization,
3. provided by the State Expert Institution of the Russian Federation.

Samples are extracts of inscriptions of writing compositions on a paper basis with different storage periods under natural conditions (pressure 1 bar = 105 Pa = 750.06 mm Hg; temperature 298.15 K = 25°C) without exposure to extreme factors such as high temperature, direct light, etc.

For each series of images, the possible terms for applying the corresponding writing compositions to the paper base are determined. At this stage, the authors focused on the study of writing compositions assigned to type B,¹⁶ since the majority of samples of type B are randomly selected. Fifteen B-type samples were obtained from each collection. All deadlines were set by the blind method to avoid subjectivity in the study.

Results

The Raman spectra for the writing composition assigned to type B, obtained using a wavelength of 532 nm, are resonant, since the maximum absorption for blue dyes of the triarylmethane group is in the area of 540–600 nm. These stripes are related to the π - π^* transitions in the aromatic

¹⁶ K.O. Gorshkova et al., “The investigation of dye aging dynamics...”; K.O. Gorshkova et al., “Investigation of the new possibility...”.

ring of the dye, due to which the symmetric vibrations of the chromogen, in this case – fuchsonimine, are more active in the Raman spectrum.¹⁷

Table 2. Frequencies and waveforms in the Raman spectrum obtained for writing compositions containing triarylmethane group dyes (assigned to type B) and literature data for the crystal violet dye

Raman shift, cm ⁻¹		Oscillation [75, 80–81]
Exp.	Lit. data [75, 80–81]	
526	524	δ (CNC)
563	558	γ (CCC) / δ (CNC) / δ (CC _{center} C)
729	724	ν (CN)
765	767	ν_s (CC _{center} C) / ν_s (CN)
805	805	δ (CH) _{ring}
915	915	ν (CC) _{ring}
974	973	δ (CC _{center} C)
1176	1178	ν_s (CC _{center} C) / δ (CCC) _{breathing} / δ_r (CH ₃)
1308	1300	ν_{as} (CC _{center} C) / δ (CCC) _{ring} / δ (CH)
1369	1369	ν (C _{center} C)
1448	1447	δ_{as} (CH ₃)
1488	1480	δ_{as} (CH ₃)
1536	1536	ν (C _{ring} N) / δ_s (CH ₃)
1587	1591	ν (C – C) _{ring}
1620	1622	ν (C – C) _{ring}

ν : valence vibrations

δ : plane deformation vibrations

γ : out-of-plane deformation vibrations

s: symmetric

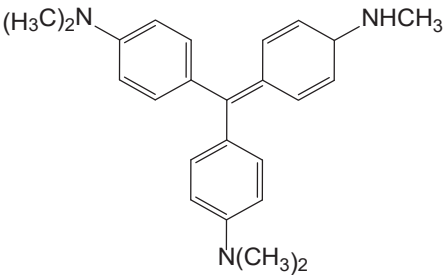
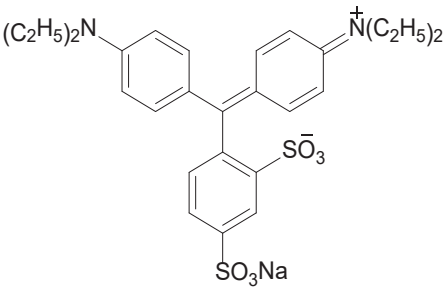
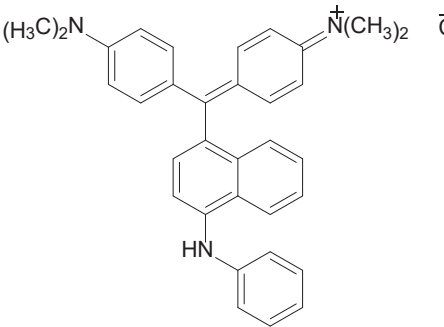
as: asymmetric

It is established that the basis of writing compositions assigned to type B may include the following triarylmethane group dyes: basic violet K, crystalline violet, methyl violet, fat-soluble violet K, Brilliant Acid Blue G, Victoria blue BO; the structural formulas of the latter are presented in Table 3.

¹⁷ B.I. Stepanov, *Vvedenie v himiju i tehnologiju organicheskikh krasitelej* (Introduction to chemistry and technology of organic dyes), Moscow 1984.

Table 3. Structural formulas of coloring substances that make up the basis of writing compositions assigned to type B

Name	Structural formula
Basic purple K	<chem>CNc1ccc(cc1)C(=C2C=CC(=[NH+](C)C)C=C2)c3ccc(N(C)C)cc3.[Cl-]</chem>
Crystal purple	<chem>CN(C)c1ccc(cc1)C(=C2C=CC(=[NH+](C)C)C=C2)c3ccc(N(C)C)cc3.[Cl-]</chem>
Methyl violet	<chem>CN(C)c1ccc(cc1)C(=C2C=CC(=[NH+](C)C)C=C2)c3ccc(N(C)C)cc3.[Cl-]</chem>

Fat-soluble purple R	
Brilliant Acid Blue G	
Victoria blue BO	

The interpretation of the obtained Raman spectrum (Figure 1), based on the literature data, is presented in Table 2.

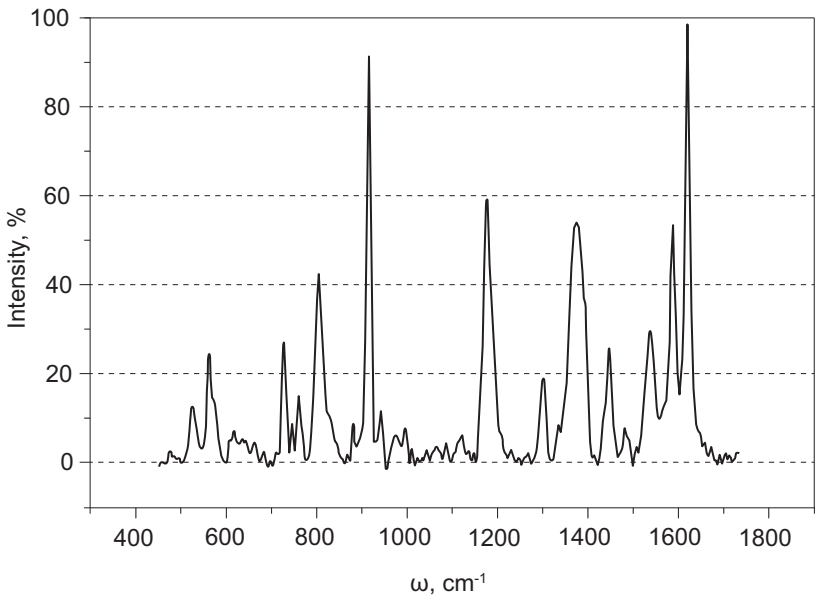


Figure 1. Raman spectrum of the writing composition from the database (Erich Krause Fiore), the coloring base of which includes dyes from the triarylmethane group (assigned to type B). The spectrum was obtained using a laser radiation wavelength of 532 nm

Errors are calculated for each series of writing compositions. Table 4 shows the calculated confidence intervals for determining the timing (in months) for samples from different collections.

Table 4. Confidence intervals for the average deviation from real terms for confidence probabilities of 99 and 95%

Confidence probability	Confidence interval for the average deviation from real terms, months				
No. 1 collection (AC)					
95%	0.9	<	1.6	<	2.3
99%	0.7	<	1.6	<	2.6
No. 2 collection (CO)					
95%	1.6	<	3.1	<	4.6
99%	1.0	<	3.1	<	5.2

No. 3 collection (EC MD)					
95%	6.5	<	10.8	<	15.1
99%	4.9	<	10.8	<	16.7

Since 15 samples from each collection were obtained, student quantile values of 2.145 with a confidence probability of 0.95 and 2.977 with a confidence probability of 0.99 were selected from the tables. A normal distribution is constructed for the relations of experimentally obtained terms from real ones (Figure 2).

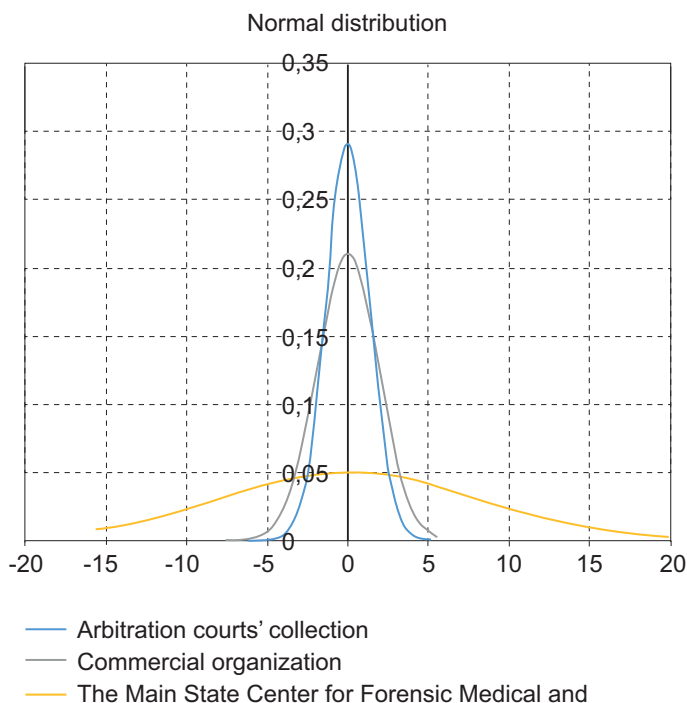


Figure 2. Normal distribution for the analyzed collections

It was found that the results obtained for collection 1 were the best: most of the values were based on the model dependence¹⁸ with the smallest

¹⁸ K.O. Gorshkova et al., "The investigation of dye aging dynamics...".

spread – about 1.6 months. The most unsatisfactory results were obtained for collection 3: the deviation was about 10.8 months.

Full correlation matrices were calculated for each sample to understand what might be the cause of deviations for different collections. For the convenience of performing and presenting calculations, a script for implementing a simplified version of the database interface was written. The elements of the correlation matrix are the correlation ratios between all variables of this population; in this case, ones between the intensities of all peaks in 5 spectra obtained for 1 sample.

The correlation ratios were calculated for the 13 most intense peaks in the sample spectrum (Figure 3) using the formula:

$$r_{xy} = \frac{\sum_{i=1}^m (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} = \frac{\text{cov}(x, y)}{\sqrt{s_x^2 s_y^2}}$$

where \bar{x}, \bar{y} – sample means x^m and y^m ; s_x^2, s_y^2 – sample variances $r \in [-1.1]$; $|r_{xy}| = 1$ – x, y are linearly dependent; $r_{xy} = 0$ – x, y are linearly independent.

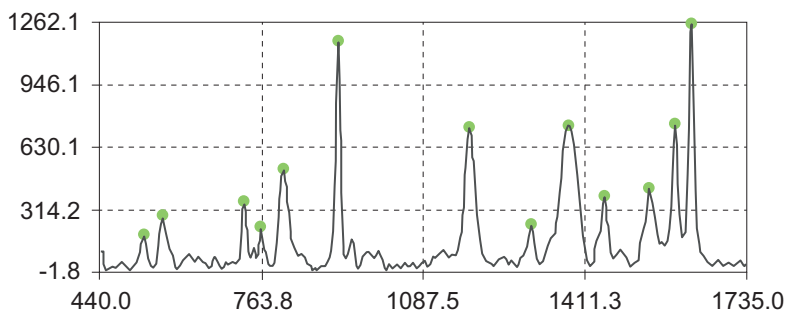
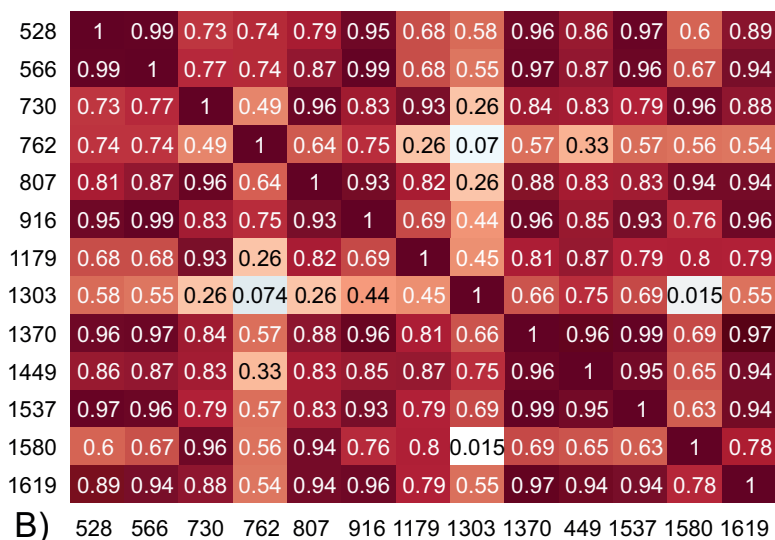
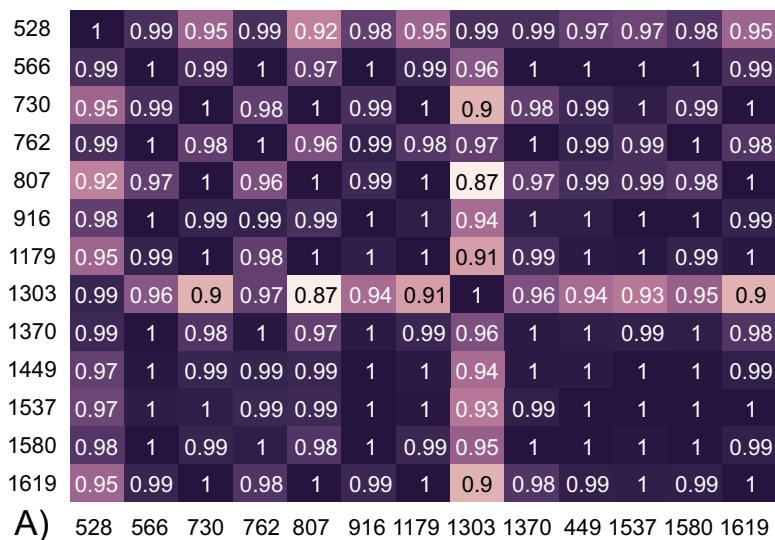


Figure 3. Peaks used for the calculation of correlation matrices

The “complete correlations matrices” are calculated (elements of the main diagonal are equal to 1, denoted by R_1 ; Figure 4). By placing 1 on the main diagonal or correlations of each variable with itself, the authors take into account the full variance of each variable represented in the matrix, thus taking into account the influence of general and specific factors:

It was expected that samples with uncorrelated matrices would have the largest deviation of the experimentally calculated term from the real one.



A) correlated: the sum of the correlation ratios for 13 peaks was 165.85 (1.9% less than the maximum);

B) uncorrelated: the sum of the correlation ratios is 129.74 (23.2% less than the maximum).

Figure 4. Examples of calculated correlation matrices

Table 5. Deviation of the experimentally established term from the real one

Δ (months)	AC	CO	EC MD
≤ 5	100%	54%	31%
6–10	0	38%	19%
≥ 11	0	8%	50%

Table 6. Sums of correlation ratios for 13×13 matrices

Σr_{xy} (%)	AC	CO	EC MD
100–95	71%	43%	63%
94–89	18%	21%	25%
≤ 88	11%	36%	12%

Table 7. Sums of correlation ratios for 8×8 matrices (with deduction of uncorrelated peaks)

Σr_{xy} (%)	AC	CO	EC MD
100–95	88%	72%	81%
94–89	6%	7%	7%
≤ 88	6%	21%	12%

Tables 5–7 show the percentage of samples with different deviations of the experimentally established period from the real one (for deviations less than 5 months, from 6 to 10 and more than 11 months); with different amounts of correlation ratios for the 13×13 and 8×8 matrices (for 100–95%, 94–89%, and less than 88%) for each of the studied collections.

The greatest deviations from the actual experimental terms were found for the collection provided by the State Expert Institution of the Russian Federation. The developed method is only suitable for analyzing the collection provided by arbitration courts since the deviations for all the samples under study were less than 5 months. This may indicate that the method is suitable for the study of samples stored under the same conditions.

From the results, the assumed relationship of the correlation ratios sum with the actual term of applying the writing composition to paper is

absent, and the sum is not a determining factor for the sample to fall into the model dependence.¹⁹

It is also important to note that the correlation ratios for peak intensities at 729 and 1580 cm^{-1} used for calculations range from 0.983 for the collection provided by a commercial organization to 0.997 for the collection provided by arbitration courts (Table 8), which indicates a close relationship between these peaks.

Based on the data obtained, it can be concluded that to develop the most accurate and effective method for determining the timing of writing compositions on paper, it is necessary to further study possible ways of external factors influencing the process of writing compositions.

Table 8. Average values of the correlation ratio between the peaks of 729 and 1580 cm^{-1}

Peak position (cm^{-1})	AC	CO	EC MD
729/1580	0.997	0.983	0.991

Influence of conditions on obtaining Raman spectra

The authors can distinguish sample degradation during spectrum measurement, focusing accuracy, and the correctness of subtraction of the baseline among the factors that can have a significant impact on the characteristics of the spectra of writing compositions and dyes/pigments. Usually, these parameters are monitored visually during the measurement process. A number of experiments were conducted, aimed at studying the conditions for obtaining Raman spectra and their relationship with the obtained measurement results to improve the measurement accuracy.

For the experiments, the authors used the writing composition inscriptions of the ballpoint pen no. 21 from the database – Paper Mate 2 in 1 (Table A, no. 1). The basis of the writing composition no. 21 includes dyes from the triarylmethane group, presumably crystal violet and Victoria blue BO.

¹⁹ Ibid.

Varying number of repetitions. Influence of mathematical processing of spectra

Ten consecutive measurements of Raman spectra were performed from the same point for the sample under study. The following instrumental parameters are used to measure the spectra: $50 \times$ lens, aperture $25 \times 1000 \mu\text{m}$, resolution $3\text{--}5 \text{ cm}^{-1}$, the measured range $440\text{--}1735 \text{ cm}^{-1}$, wavelength 532 nm , laser radiation power of 0.2 mW , integration time – 5 s , the number of repetitions – 2. The 11th spectrum was measured 5 minutes after the end of a series of 10 measurements. A series of 11 spectra (hereinafter referred to as “the series”) was measured for each test sample three times with a difference of 24 hours. This number of measurements was carried out to unambiguously exclude random errors in the process of measuring the spectra for a single sample.

Three different approaches were used to process the results to account for the impact of mathematical processing of spectra (such as baseline subtraction, smoothing).

1. The resulting spectrum was processed in the OPUS 7.5 program: baseline correction – 2 iterations, smoothing – 25 iterations; then the absolute peak intensities were calculated.

2. Two sections were cut from the resulting spectrum: $681\text{--}783 \text{ cm}^{-1}$ and $1490\text{--}1700 \text{ cm}^{-1}$, after which each section was processed in the software: baseline correction – 2 iterations, smoothing – 25 iterations; then the absolute peak intensities were calculated.

3. Two sections were cut from the resulting spectrum: $681\text{--}783 \text{ cm}^{-1}$ and $1490\text{--}1700 \text{ cm}^{-1}$, after which each of the sections was processed in the software: baseline correction – 2 iterations, smoothing – 25 iterations. In the OriginPro 9 program, each of the sections was decomposed into Gaussians, after which the absolute peak intensities were calculated.

According to the experiment results, dependency graphs of peak intensities in the area of 729 cm^{-1} to the peak in the area of 1580 cm^{-1} to the time of sample irradiation (during spectrum acquisition) were obtained using three methods of spectrum processing (Figure 5). The lowest confidence interval with a confidence probability of 0.95 was obtained using the third processing method and is ± 0.023 .

Even after irradiating the sample for 100 seconds (after measuring the tenth spectrum from one point), no visible changes occurred with the sam-

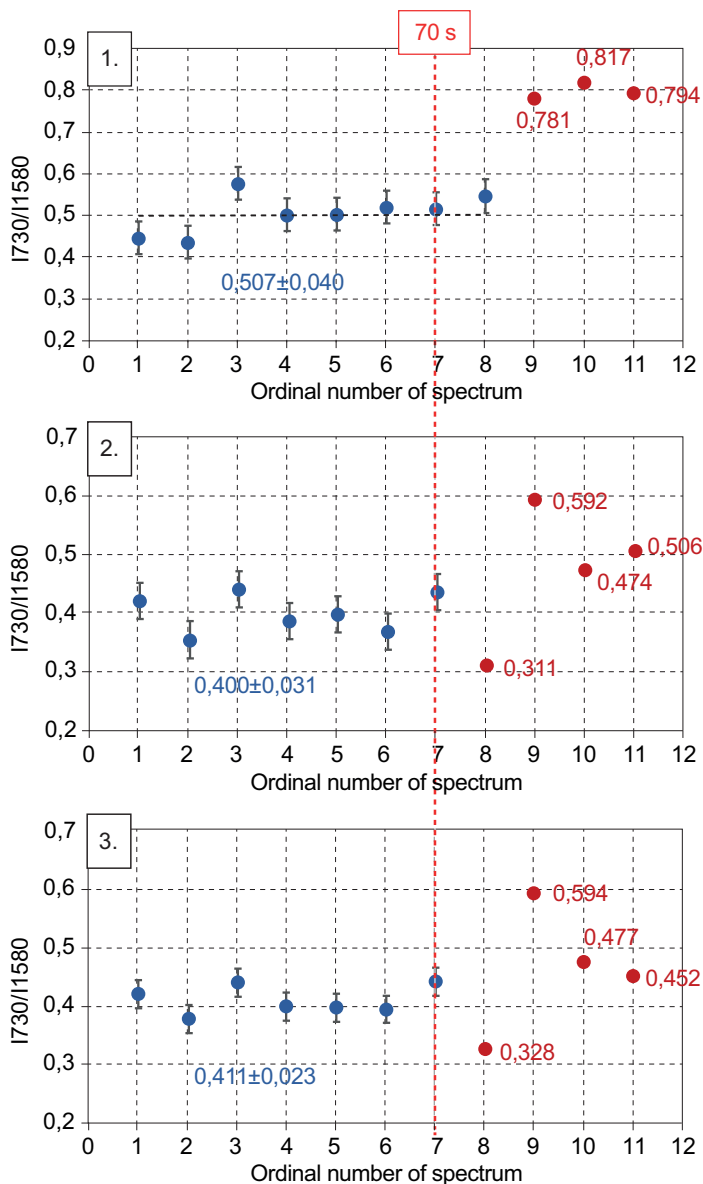


Figure 5. Dependency graphs of the peak intensities ratio in the 730 cm^{-1} area to the peak in the 1580 cm^{-1} area when using three methods for processing the Raman spectra of the writing composition sample no. 59 from the database (experiment no. 1)

ple (Figure 6), but after irradiating the sample for 70 seconds (after the 7th spectrum from one point), there was a sharp change in the value of the peak intensity ratio. After holding a pause of 10 minutes and measuring the 11th spectrum, the value of the peak intensity ratio became significantly greater than in the range from the 1st to the 7th spectrum (the difference was more than 0.1 relative units), which indicates that irreversible changes occur with the sample due to degradation under the influence of laser radiation. It can be concluded that the maximum sample irradiation time should be no more than 70 seconds at the power 0.2 mW.

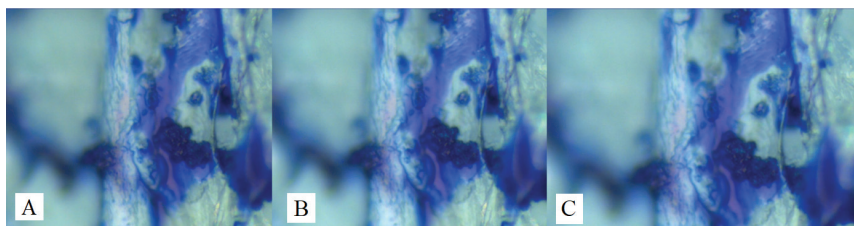


Figure 6. Optical micrographs of the area from which the Raman spectra were obtained after measuring the 1st (A), 5th (B), and 10th (C) spectra (50 × magnification)

Varying the integration time

The sample was measured 5 times in a row in modes with the same number of repetitions (2 repetitions) with increasing integration time: 5×2 , 10×2 , 20×2 , 40×2 , 80×2 from the same point. The measurements were taken on a single day. Three approaches were used to process the results, as in the previous experiment (cl. 4.8.1).

The experiment results show that the luminescence signal level increases linearly with increasing integration time. The intensity of the Raman signal increases linearly with increasing sample irradiation time (Figure 7).

The dependence of the ratio of peak intensities 730/1580 on the sample irradiation time (Figure 8) obtained in experiment 2 is similar to the dependence obtained in experiment 1 (Figure 5). A sharp change in the peak intensity ratio occurs after irradiation of the sample for more than 70 seconds (after the 3rd spectrum in experiment 2; Figure 8).

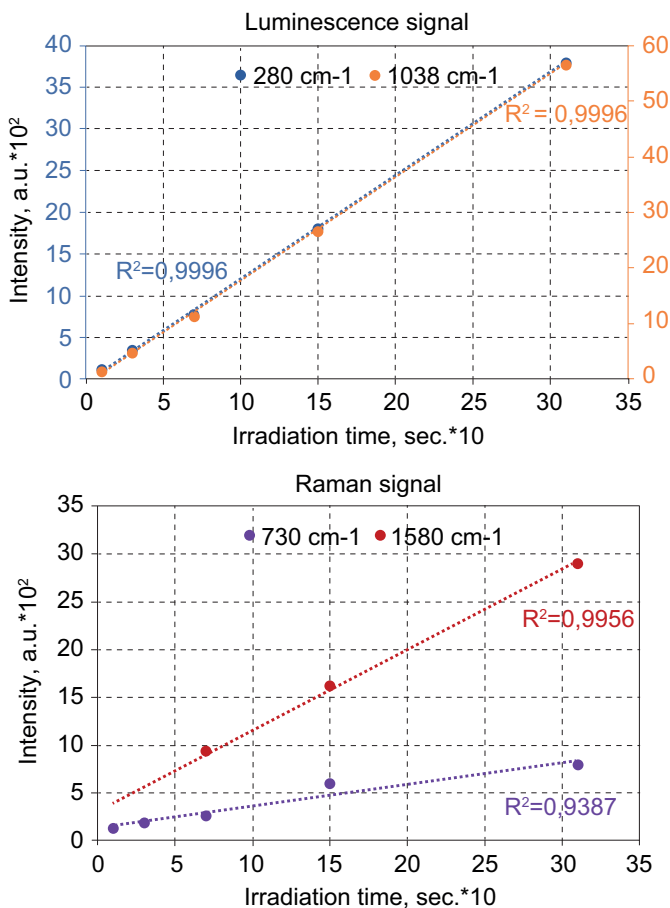


Figure 7. Dependency graphs of the signal level (on the left – the luminescence signal, on the right – the Raman signal) on the irradiation time of the writing composition sample no. 59 from the database

Even after irradiating the sample for 310 seconds (after measuring the 5th spectrum from one point), no visible changes occurred with the sample (Figure 9), but after irradiating the sample for 70 seconds (after the 3rd spectrum from one point), there is a sharp change in the ratio of peak intensities, as in the previous experiment (Figure 5). After holding a pause

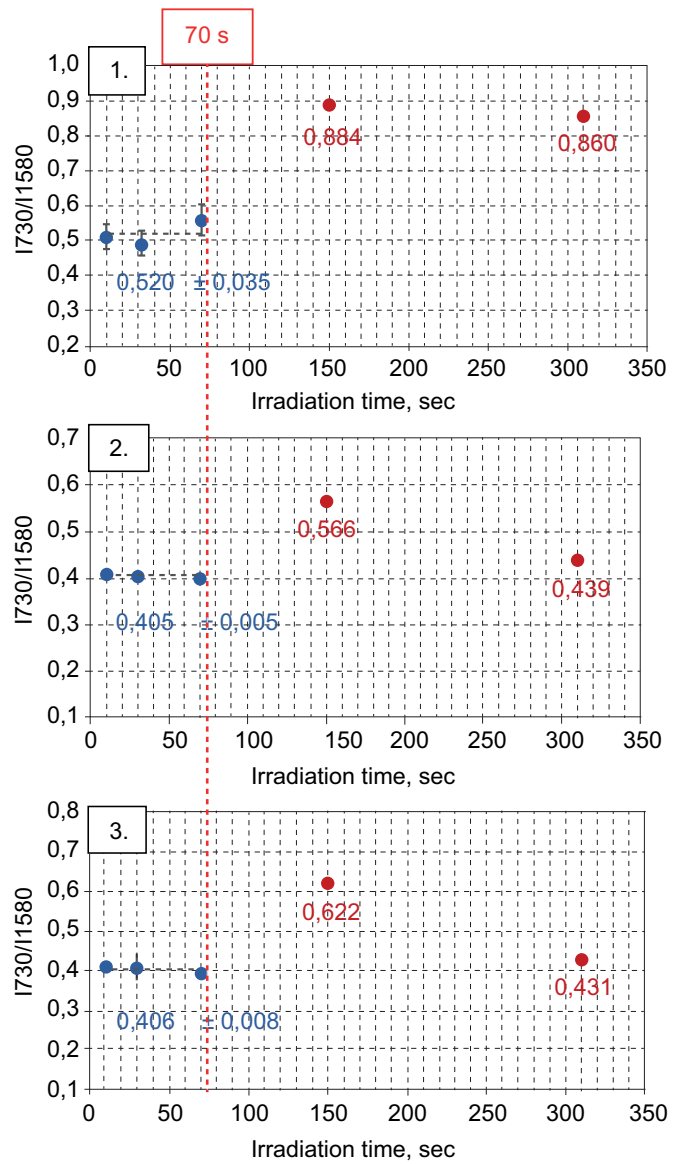


Figure 8. Dependency graphs of the ratio of peak intensities in the area of 730 cm^{-1} to the peak in the area of 1580 cm^{-1} to the irradiation time of the sample of writing composition no. 59 from the database using three methods of spectrum processing (experiment no. 2)

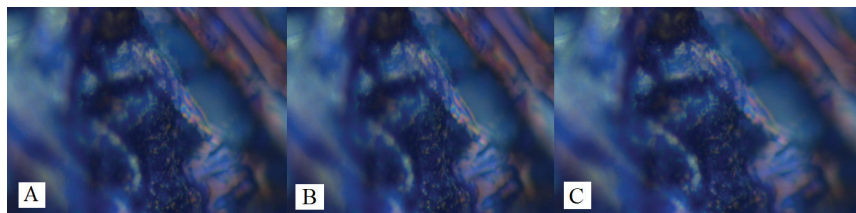


Figure 9. Optical micrographs of the area from which the Raman spectra were obtained after measuring the 1st (A), 3rd (B), and 5th (C) spectra (50 × magnification)

of 10 minutes and measuring the 5th spectrum, the value of the peak intensity ratio remains greater than in the range from the 1st to the 3rd spectrum. It can be concluded that to prevent the introduction of errors related to the degradation of the sample according to the results of experiments no. 1 and no. 2, the recommended irradiation time of the sample should not exceed 70 seconds.

The values of the peak intensity ratios obtained when processing the Raman spectra by various methods (Figures 5, 8) differ significantly. When using the 1st method of processing, the spectrum is considered as a whole; the baseline is subtracted immediately in the entire range of 440–1735 cm⁻¹. Since the baseline is complex, automatically subtracting the last one immediately over the entire range can lead to additional systematic error, which affects the results obtained.

Figure 10 shows the spectra obtained using the 1st processing method (used for the first model). As previously stated, the value of luminescence signal increases with more prolonged exposure to laser radiation on the sample, while there is an increase in “breaking” the baseline, which can be interpreted in two ways: either this effect is associated with the presence of multiple maxima of the luminescence and can be explained directly by the composition of the sample or by the feature associated with the processes occurring in the sample with increasing time of exposure to laser radiation. As a result, it is not possible to correctly subtract the baseline in software (for example, using Opus) immediately in the entire range. When trying to subtract the baseline using the first method of spectrum processing, there is an area of “residual luminescent background” in the area of the peak of 730 cm⁻¹. The intensity of this “residual luminescent background” area also increases proportionally to the integration time and its position shifts to the short-wave part of the spectrum (Figure 10–11).

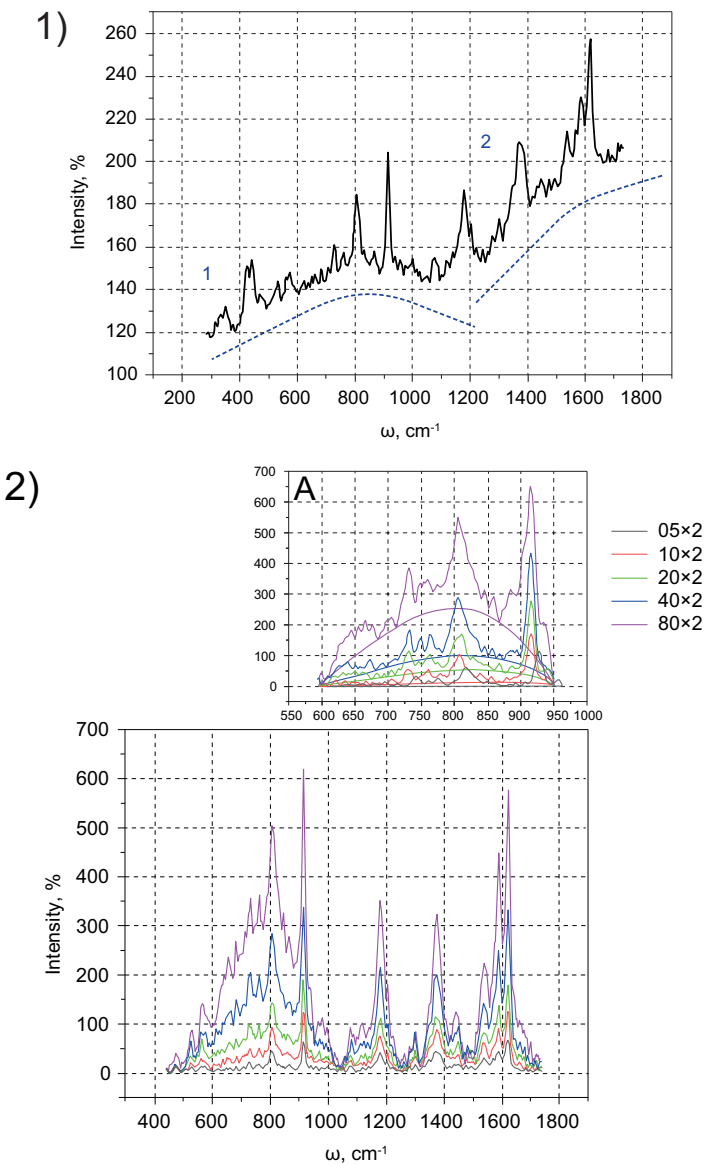


Figure 10. Raman spectra of the writing composition sample no. 59 from the database: 1) before processing by the 1st method; 2) after processing by the 1st method; A – enlarged spectral area in the area of 400–950 cm^{-1} with an approximation of the “residual luminescent background” area for better visualization

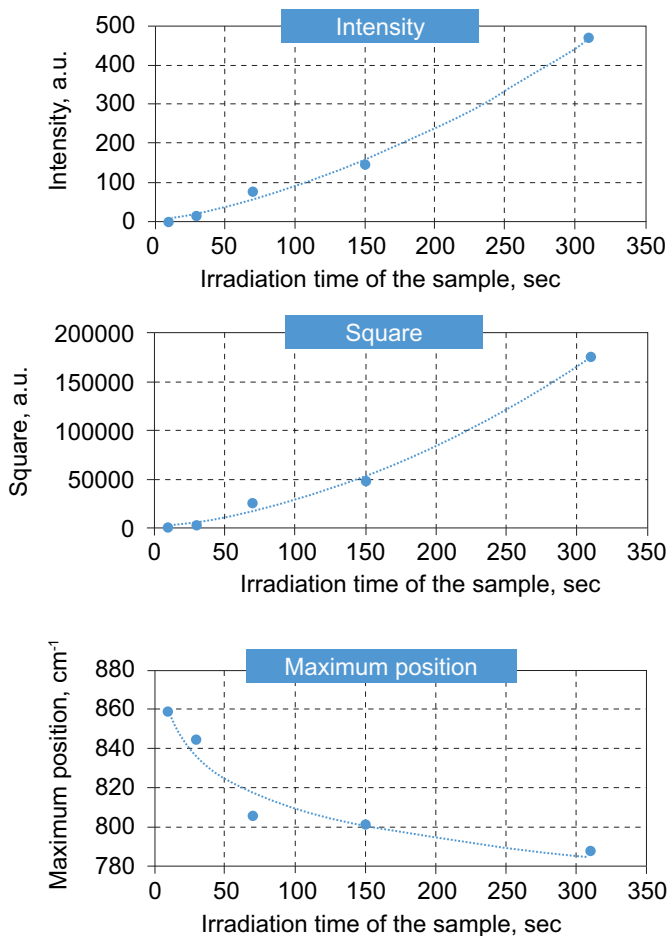


Figure 11. Characteristics of the “residual luminescent background” area obtained by subtracting the baseline using the 1st method from the irradiation time of the sample for the area, the maximum value, and the maximum position

The results indicate that the calculations must take into account changes in the luminescence signal from the sample.

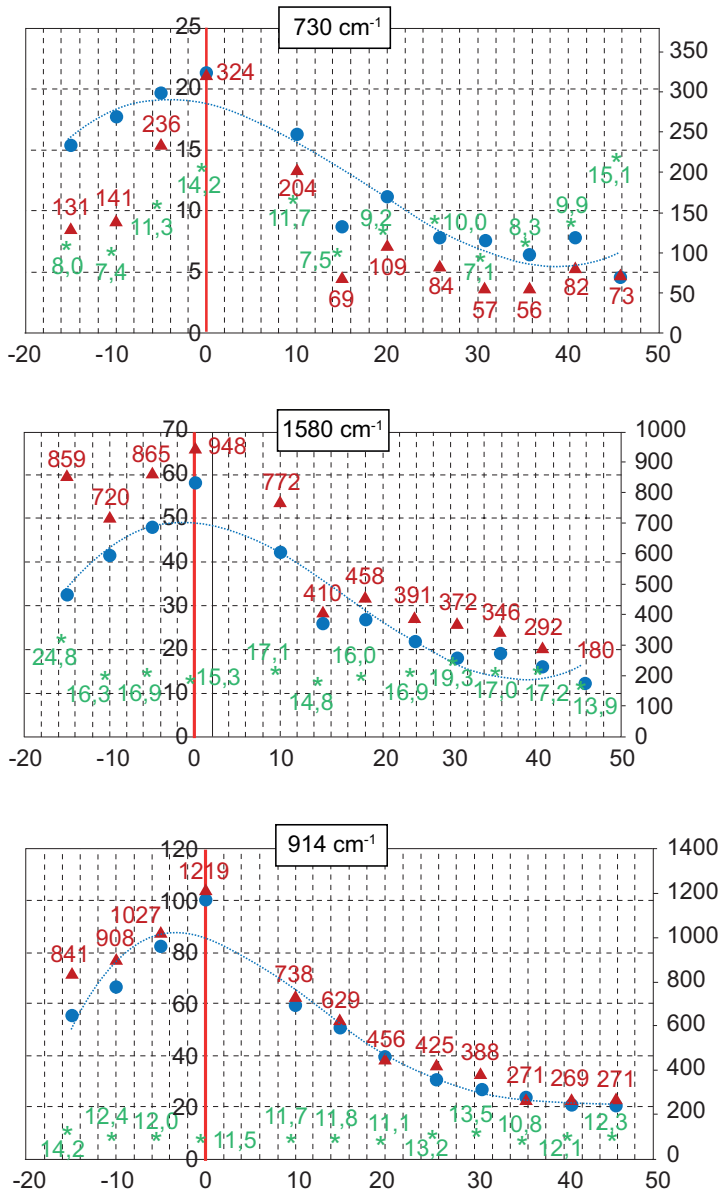


Figure 12. Dependency graphs of intensity (points), areas (triangles), and half-width at half-height (asterisks) for the peaks 730, 1580, and 914 cm⁻¹ on the z-axis coordinate, the visual focal point is assumed to be 0

Evaluating the contribution of focus accuracy to measurement results

The maximum achievable focusing on the sample was performed (focusing accuracy was determined visually) to study the effect of focusing accuracy on the measurement results, after which a coordinate 40 units higher was set for the motorized table along the z-axis. Then a series of spectra was obtained from one point in increments of 5 units along the z-axis. The spectra were processed using the 3rd processing method (cl. 4.8.1), since the minimum value of the standard deviation for the average is achieved wthrough it. The spectra were obtained using the following parameters: the laser radiation power – 0.2 mW, the integration time – 3 seconds, and the number of repetitions – 2. For the study, the main characteristics were calculated for the peaks of 730 and 1580 cm^{-1} , since they were used to obtain time dependencies.²⁰ The “model” peak was used at 914 cm^{-1} , as it is an intense, free-standing peak.

Based on the experiment results, the authors can conclude that when approaching the focus point, the values of the intensity and area of the peaks increase and fall after passing the focus point. The half-width at half-height remains constant and does not depend on the focus (Figure 12). From the presented graphs, it is clear that the visually defined focus point coincides with the actual one.

For the peaks of 730 cm^{-1} and 1580 cm^{-1} , the largest spread of values is observed, due to the error that occurs when approximating these peaks by Gaussian (for example, the spread of half-width values at half-height can reach 30%). Next to the 730 cm^{-1} peak, there are a number of closely spaced peaks (for example, the peak at 726 cm^{-1}), as a result of which the latter overlap and the Gaussian approximation of the peak at 730 cm^{-1} becomes much more complicated. In addition, the intensity of the 730 cm^{-1} peak is significantly less than the 914 cm^{-1} peak, which increases the approximation error of this peak. The optimal pattern is observed for a peak of 914 cm^{-1} . The half-width value at half-height averages 12.2, with a standard deviation of 1, which is 10% of the average. There are no other nearby peaks near the 914 cm^{-1} peak, i.e., there is no overlap with other peaks. The behavior of this peak corresponds to the theoretical assumptions as much as possible.

²⁰ Ibid.; K.O. Gorshkova et al., “Investigation of the new possibility...”.

Similar dependencies were obtained for samples with storage periods of 25, 39, 58, and 88 months. Dependency graphs of peak intensity on the z-axis position can be divided into three linear sections (Figure 13). For further calculations, the authors selected three angles between linear sections and the positive direction of the x-axis as well as the fourth angle between straight lines (Figure 13).

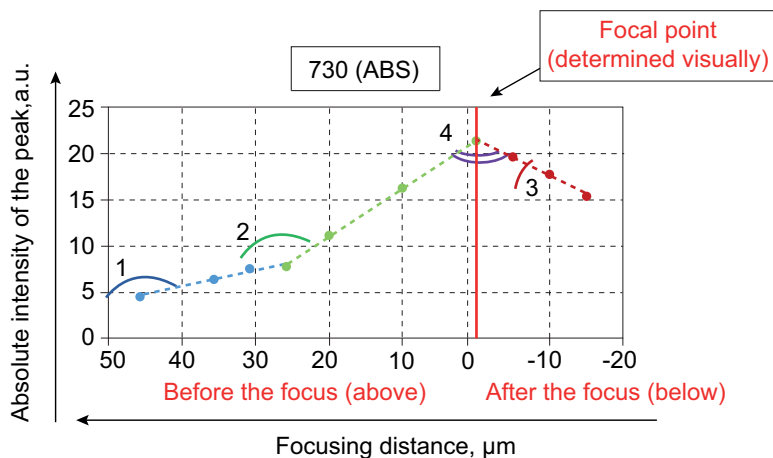


Figure 13. Example of a dependency graph of peak intensities on the z-axis position indicating the angles used for calculations, the visual focal point is taken as 0

Figure 14 shows the dependences of the tangents of angles 1–4 on the sample shelf life for peaks 730, 1580, and 914 cm^{-1} . The standard deviation for the studied angles 1–3 does not exceed 0.4.

A different character of the dependence is observed for angle 4. If the dependence for the angle tangent within 25–58 months has a form similar to the parabolic dependence with a minimum of 39 months, then after 58 months there is a sharp increase in the value of the angle tangent, which indicates a decrease in the sensitivity of peak intensities to focusing. The peak intensity is most sensitive to focusing at the age of 39 months, and this sensitivity decreases with increasing time. It can be assumed that after reaching 39 months, the heterogeneity of the sample is maximal, and after 39 months, the heterogeneity of the sample decreases. When con-

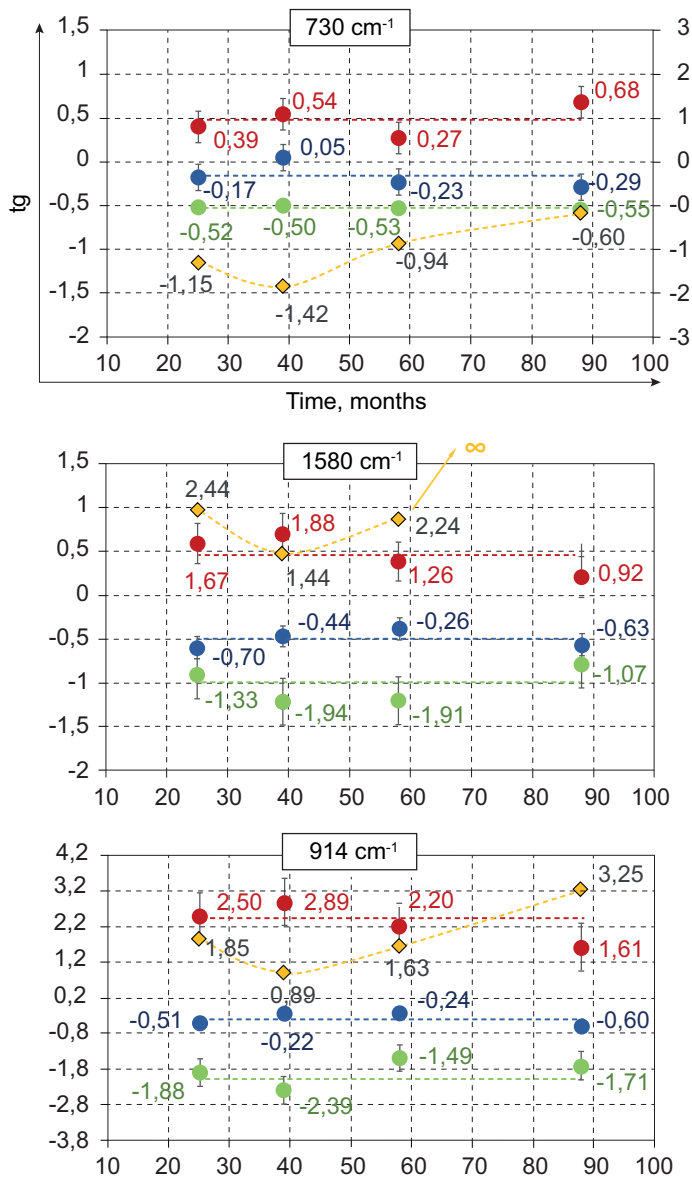


Figure 14. The dependence of the angle tangent 1–4 on the sample shelf life for peaks 730, 1580, and 914 cm⁻¹, the visual focal point is taken as 0

structing the first model of methodological development,²¹ the minimum intensity ratio was observed for about 40 months. It can be assumed that 40 months is the period of maximum fluctuation of the sample properties, after which the processes are stabilized.

Dependencies of the ratio of peak intensities 730/1580 on the z-axis position were obtained (Figure 15). The intensities ratio has a large spread when obtaining spectra further than 30 μm from the focus point. Within 30 μm before the focus and 10 μm after the focus, the intensity ratio varies within the standard deviation ~ 0.03 . When determining the focus visually, the error can be no more than 5 μm , and the standard deviation of the ratio within these limits is less than 0.01.

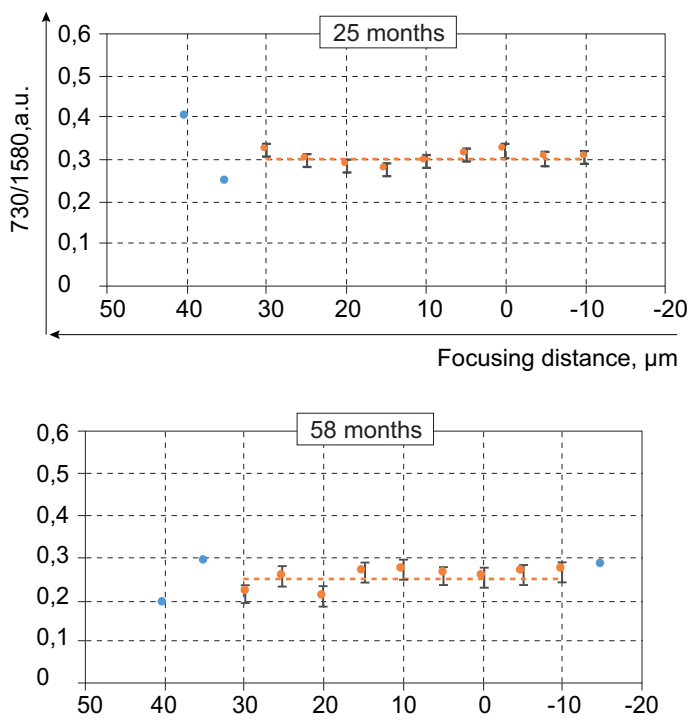


Figure 15. Dependence of the ratio of peak intensities 730/1580 on the z-axis coordinate, the visual focal point is taken as 0

²¹ K.O. Gorshkova et al., “The investigation of dye aging dynamics...”.

Based on the data obtained, it can be concluded that the visual definition of the focus point is sufficient for the study and does not significantly distort the peak intensity and characteristics of the writing compositions' Raman spectrum.

Application of chemometric tools for the analysis of Raman spectra

An attempt was made to use new methods of mathematical signal processing, namely the PCA method, to eliminate the contradictions identified at the previous stages.

The retention period was used to calculate the Pearson correlation ratio to identify the relationship between the components. Before the calculation, the nature of the data distribution was checked using histograms and quantile graphs, since both analyzed variables must be distributed normally to calculate the Pearson ratio.

The eigenvalue greater than 1 indicates that the main component preserves more variance than was accounted for by one of the original variables in the standardized data. This is usually used as the cut-off point for which the principal components are preserved.²² The system under study has 4 main components. The main contribution is made by PC1, which explains 97% of the data variance. However, for further analysis, the authors used all the components to take into account the maximum number of factors that affect the system.

The correlation ratios of the components and the sample shelf life are calculated (Table 23). The components PC4 (0.286), P5 (0.252), and PC6 (0.259) are most closely related to the shelf life, although they do not make the maximum contribution to the variance in the system. This explains the need to consider not only the main components for analysis, but also those that have a connection with the sample shelf life.²³

For better visualization, a 3D graph was constructed for averaged component values for samples with the same shelf life (Figure 16). Indeed,

²² G.B. Anderson, *Principal component analysis in R. An examination of the different functions and methods to perform PCA*, <https://www.ime.usp.br/~pavan/pdf/PCA-R-2013>.

²³ K.O. Gorshkova et al., "Investigation of the new possibility...".

samples with a shelf life of up to 50 months and after 50 months of storage are separated in space, while samples with a shelf life of 40 and 65 months actually lie at the interface. The obtained result indicates the potential possibility of separating samples depending on the shelf life by the PCA method by improving the mathematical processing algorithm.

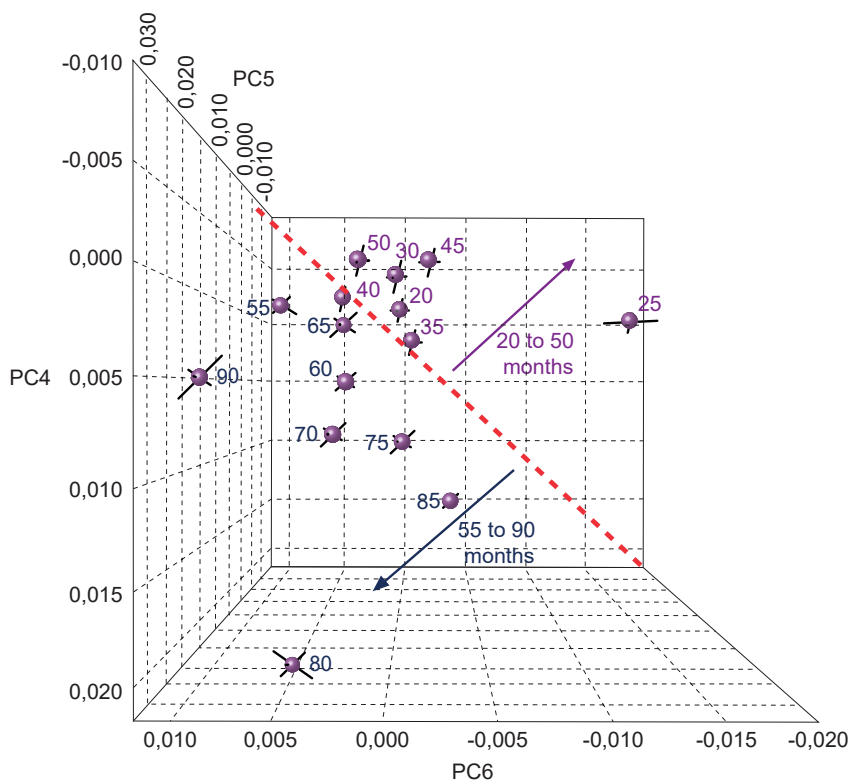


Figure 16. 3D graph in coordinates PC4/PC5/PC6 for the system under study, points – averaged component values for all samples with the same shelf life

The authors also managed to achieve clustering depending on the brand (Figure 17). For different writing compositions, the value of PC2 differs, which allows separating the studied writing compositions. Figure 17

shows a graph of the PC2 values for the number of writing compositions. Table 9 illustrates the calculated values of the average value of the second components and the standard deviation of the latter. It should be noted that the values of PC2 for writing compositions no. 3 and 5 are very close since both writing compositions belong to the Erich Krause brand (Table 1).

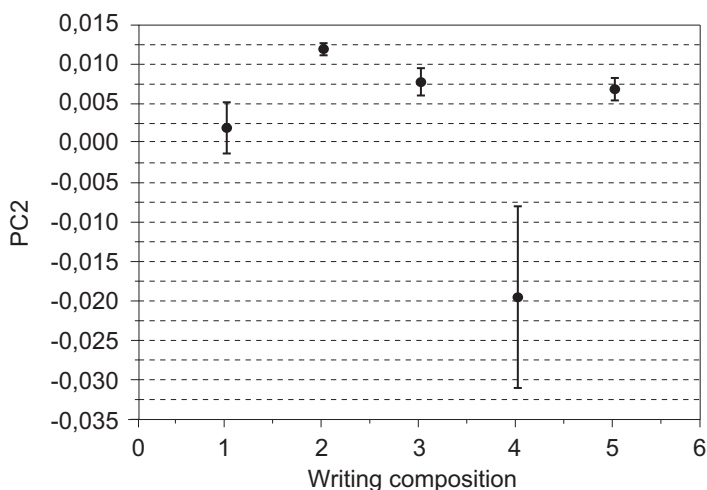


Figure 17. Values of PC2 for each studied writing composition

Table 9. Calculated value of PC2 for each studied writing composition

No. of pen	PC2 value	Standard deviation
1	0.0020	0.0032
2	0.0119	0.0008
3	0.0078	0.0017
4	-0.0195	0.0115
5	0.0069	0.0014

At this stage, the entire sample spectrum was used for PCA analysis. When analyzing such data, too many fluctuations contribute to each of the components, so it is difficult to interpret the resulting picture unambiguously.

Discussion

The following can be concluded based on the results of the work performed:

- the dynamics of the Raman spectra of writing compositions is determined by the spectral dynamics of the dye that is the basis of the latter;
- the main parameters for measuring the Raman spectra of writing compositions are determined;
- the laser power – 0.2 mW, the integration time – 10 seconds, and the number of repetitions – 4;
- in the process of obtaining the spectrum, the irradiation time of the sample should not exceed 70 seconds;
- it is necessary to consider individual sections of the spectrum containing characteristic peaks instead of the entire spectrum to calculate the physical characteristics of the spectrum. The peaks must be approximated using the Gaussian distribution.

Visual determination of the focus point is sufficient for conducting research and does not significantly distort the peak intensity and characteristics of the Raman spectrum of writing compositions. Calculations must take into account changes in the intensity of Raman peaks and the luminescence signal from the sample.

It can also be noted that 40 months is the period of maximum fluctuation of the writing composition properties, after which the transformation processes of the writing material components are stabilized. A similar result was obtained earlier in Gorshkova et al.²⁴

Taking into account all the requirements for measuring spectra, the PCA method applied to the Raman spectra of writing compositions allows clustering samples by manufacturer brands and by the sample shelf life. The dependency graph of the components confirms this with the highest values of the correlation ratio with the shelf life – PC3/PC5 (Figure 18). For Figure 18, 5 areas correspond to different sample shelf life: 20–25, 30–50, 55–70, 80–85, and 90 months.²⁵

²⁴ K.O. Gorshkova et al., “The investigation of dye aging dynamics...”; K.O. Gorshkova et al., “Investigation of the new possibility...”.

²⁵ K.O. Gorshkova et al., “Investigation of the new possibility...”.

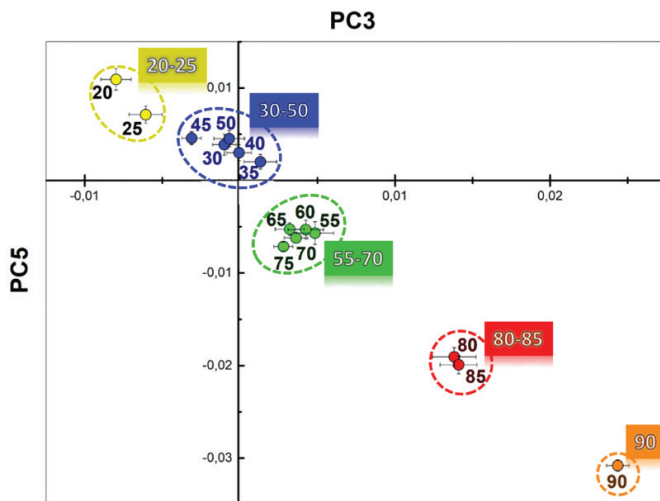


Figure 18. Dependency graph of PC5 values on PC3 values

The developed clustering method is the most objective since its implementation uses all points in the Raman spectra, and the “human factor” error is minimized.

A similar approach to examining the questioned document was also described by Łydźba-Kopczyńska et al. and Cieśla.²⁶ The developed clustering method is the most objective since its implementation uses all points in the Raman spectra, and the “human factor” error is minimized.

Acknowledgments

The work was supported by the Ministry of Education and Science of the Russian Federation (Project FSRM-2020-0006).

²⁶ B. Łydźba-Kopczyńska et al., “Application of chemometric methods for the determination of fading and age determination of blue ballpoint inks”, *Journal of Raman Spectroscopy* 52, 2021, no. 1, pp. 159–169; R. Cieśla, “Questioned document examination with the use of alternative and complementary non-destructive methods”, *Nowa Kodyfikacja Prawa Karnego* 57, 2020, pp. 149–167.

The work was supported by Federal State Budgetary Institution “Assistance Fund development of small enterprises in the scientific and technical sphere,” Agreement no. 4101ГC1/65634.

The research was partially funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program “Priority 2030” (Agreement 075-15-2021-1333, dated 30.09.2021).

References

- Aginsky V.N., “Determination of the age of ballpoint pen ink by gas and densitometric thin-layer chromatography”, *Journal of Chromatography A* 678, 1994, no. 1.
- Aginsky V.N., “Forensic examination of ‘slightly soluble’ ink pigments using thin-layer chromatography”, *Journal of Forensic Sciences* 38, 1993, no. 5.
- Aginsky V.N., “Measuring ink extractibility as a function of age. Why the relative aging approach is unreliable and why it is more correct to measure ink volatile components than dyes”, *International Journal of Forensic Document Examiners* 4, 1998.
- Anderson G.B., *Principal component analysis in R. An examination of the different functions and methods to perform PCA*, <https://www.ime.usp.br/~pavan/pdf/PCA-R-2013>.
- Brunelle R.L., “Ink dating: The state of the art”, *Journal of Forensic Sciences*, 37, 1992.
- Burnstock A., Berg K.J. van den. “Twentieth century oil paint. The interface between science and conservation and the challenges for modern oil paint research”, *Issues in Contemporary Oil Paint: Proceedings of the Issues in Contemporary Oil Paint (ICOP) Symposium*, 28–29 March 2013, Amersfoort, Netherlands.
- Buzzini P., Polston C., Schackmuth M., “On the criteria for the discrimination of inkjet printer inks using micro-Raman spectroscopy”, *Journal of Raman Spectroscopy* 49, 2018, no. 11.
- Cantú A., Prough R.S., “On the relative aging of ink: The solvent extraction technique”, *Journal of Forensic Sciences* 32, 1987.
- Cieśla R., “Questioned document examination with the use of alternative and complementary non-destructive methods”, *Nowa Kodyfikacja Prawa Karnego* 57, 2020.
- Díaz-Santana O.A., Conde-Hardisson F.B., Vega-Moreno D.C., “Comparison of the main methods for six ballpoint pen inks”, *Microchemical Journal* 138, 2018.
- Geiman I., Leona M., Lombardi J.R., “Application of Raman spectroscopy and surface-enhanced Raman scattering to the analysis of synthetic dyes found in ballpoint pen inks”, *Journal of Forensic Science* 54, 2009, no. 4.
- Giorgi, Nevin A., Comelli D., Frizzi T., Alberti R., Zendri E., Piccolo M., Izzo F., “In-situ technical study of modern paintings. Part 2: Imaging and spectroscopic analysis of zinc white in paintings from 1889 to 1940 by Alessandro Milesi (1856–1945)”, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 219, 2019.

- Gorshkova K.O., Rossinskaya E.R., Kirillova N.P., Fogel A.A., Kochemirovskaya S.V., Kochemirovsky V.A., "Investigation of the new possibility of mathematical processing of Raman spectra for dating documents", *Science & Justice* 60, 2020, no. 5.
- Gorshkova K.O., Tumkin I.I., Myund L.A., Tverjanovich A.S., Mereshchenko A.S., Panov M.S., Kochemirovsky V.A., "The investigation of dye aging dynamics in writing inks using Raman spectroscopy", *Dyes and Pigments* 131, 2016.
- Grechukha N., Gorshkova K., Panov M., Tumkin I., Kirillova E., Lukianov V., Kochemirovsky V., "Analysis of the aging processes of writing ink: Raman spectroscopy versus gas chromatography aspects", *Applied Sciences* 7, 2017, no. 10.
- Kunicki M., Fabiańska E., Parczewski A., "Raman spectroscopy supported by optical methods of examination for the purpose of differentiating blue gel pen inks", *Problems of Forensic Sciences* 95, 2013.
- Lee J., Bonaduce I., Modugno F., La Nasa J., Ormsby B., Berg K.J. van den, "Scientific investigation into the water sensitivity of twentieth century oil paints", *Microchemical Journal* 138, 2018.
- Łydzba-Kopczyńska B., Czaja T., Cieśla R., Rusek G., "Application of chemometric methods for the determination of fading and age determination of blue ballpoint inks", *Journal of Raman Spectroscopy* 52, 2021, no. 1.
- Malinowsky E.R., Howery D.G., *Factor analysis in chemistry*, New York 1980.
- Sharma V., Kumar R., "Dating of ballpoint pen writing inks via spectroscopic and multiple linear regression analysis: A novel approach", *Microchemical Journal* 134, 2017.
- Stepanov B.I., *Vvedenie v himiju i tehnologiju organicheskikh krasitelej* [Introduction to chemistry and technology of organic dyes], Moscow 1984
- Tauler R., Juan A. de, "Multivariate curve resolution", [in:] *Practical guide to chemometrics*, ed. P. Gemperline, London 2006.
- Trosman E.A., Bezhanishvili G.S., Arkhangelskaya N.M., Yurova R.A., "Methodology: Age-dating of document attribute entries by the relative content of volatile solvents in strokes", *Теория и практика судебной экспертизы (Theory and Practice of Forensic Science)* 30, 2013, no. 2.
- Weyermann C., Kirsch D., Spengler B., "A GC/MS study of the drying of ballpoint pen ink on paper", *Forensic Science International* 168, 2007.
- Xu Y., Wang J., Yao L., "Dating the writing age of black roller and gel inks by gas chromatography and UV-vis spectrophotometer", *Forensic Science International* 162, 2006.