



Development and evaluation of a simple Raman spectral searching algorithm

Anastasia Rousaki^{1,a}, Emma Paolin^{1,2}, Giorgia Sciutto², Peter Vandenabeele^{1,3} 

¹ Raman Spectroscopy Research Group, Department of Chemistry, Ghent University, S-12, Krijgslaan 281, 9000 Ghent, Belgium

² Department of Chemistry “G. Ciamician”, University of Bologna, Ravenna Campus, Via Guaccimanni 42, 48121 Ravenna, Italy

³ Archaeometry Research Group, Department of Archaeology, Ghent University, Sint-Pietersnieuwstraat 35, 9000 Ghent, Belgium

Received: 10 November 2020 / Accepted: 17 May 2021

© The Author(s), under exclusive licence to Società Italiana di Fisica and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract Raman spectroscopy has grown to a frequently applied technique in the analysis of art and archaeological objects. The growth in applications is further driven by the development of new, small and compact spectrometers that are well suited for in situ analysis. Different spectral databases have been composed, to allow fast identification of the materials at hand. However, the automated identification of the highly similar spectra of contemporary synthetic organic pigments is not straightforward, as spectral libraries that contain several hundreds of spectra are typically recorded on benchtop spectrometers that have different spectral resolutions and sensitivities and often use different excitation laser wavelengths. Therefore, here a simple automated spectral searching algorithm is proposed that is based on the comparison of a series of Raman band positions, to identify the materials. Thus, many of these interferences can be avoided. In this work, the proposed algorithm is tested on spectra that were recorded by using portable Raman instruments on mock-up samples as well as during the in situ analysis of street art.

1 Introduction

During the last decades, Raman spectroscopy has grown to a frequently used technique in archaeometry and art analysis. This evolution was facilitated by instrumental improvements, such as the coupling of the Raman spectrometers with a confocal microscope, the introduction of the fibre optics and the miniaturisation of the systems towards the development of mobile instrumentation for in situ and direct analysis. Along with the broadening of the applications, for comparative reasons, the need of high-quality spectra of known references became clear. Often the laboratories developed their own collection of reference Raman spectra, while some groups published dedicated spectral collections. The first Raman spectral collections dedicated to cultural heritage research focussed mainly on inorganic pigments [1–5] and also on natural binding media and varnishes [6–8]. More extended databases are available for, for example, minerals [9] or biomolecules [10]. Moreover, dedicated papers

^a e-mail: Raman@UGent.be (corresponding author)

focus on discriminating between Raman spectra of a relatively small group of related materials, such as natural silicate glasses [11], Mexican artists' materials, [12, 13], green [14, 15] or black minerals [16]. Concerning the analysis of synthetic organic pigments (SOPs), several groups have worked on this topic and published their results [17–21]. Besides the dedicated research articles on SOPs, a spectral database is openly available [22], thus enhancing the direct comparison and solid identification of such components. This work contributes not only towards the analysis of modern artworks [23], but it has also implications in forensics research [24–26]. In general, the availability of a large collection of reference Raman spectra is a positive aspect for the cultural heritage research community. However, published spectra are not always readily available in an accessible digital format and measurement conditions can influence the appearance of the Raman spectrum (e.g. selected laser wavelength, calibration, spectrometer response, etc.). Moreover, manual comparison of a spectrum against a (large) database of reference spectra is time-consuming and therefore several approaches for automated searching algorithms have been proposed [25, 26]. Often, these spectral searching algorithms have proven to be successful to perform identification in specific groups materials. Different (multivariate) chemometric approaches have been proposed, of which principal components analysis (PCA) seems a frequently used method [27–33]. Other approaches include linear discriminant analysis [34, 35] or more advanced approaches such as fuzzy logic [36, 37].

Unfortunately, multivariate approaches typically have some disadvantages. In general, they use the entire spectra to compare against the spectra in a reference database. As a consequence, variations in the spectral background (e.g. fluorescence or ambient light) may interfere. Moreover, it is of the utmost importance that the reference spectra as well as the spectrum of the unknown are well calibrated [38]. Moreover, intensity fluctuations or differences (between reference spectra and unknown) in spectral resolution may hamper the identification. Especially, if the reference database is recorded with a different excitation laser or more generally speaking, when using different optics or experimental conditions as the spectrum of the unknown, identification is not always straightforward. Chemometric techniques can also be used for data reduction. PCA, for instance, projects all spectra in a multi-dimensional space and recalculates the values to obtain a limited number of orthogonal values. In this case, each time a new reference spectrum is added to the database, all the dataset has to be re-processed for the new data reduction.

To overcome these issues, we propose a simple spectral searching algorithm that is based on the extraction of the Raman band positions and compare the retrieved Raman band positions against those of the reference spectra. The aim of the algorithm is to let the user identify easily spectra from a large database that correspond to the spectrum of the unknown. This approach is evaluated based on the analysis of mock-up samples and is as well applied during the in situ analysis of street art.

2 Experimental

Raman spectra were recorded by using two different portable Raman spectrometers: a Bravo handheld spectrometer (Bruker, Ettlingen, Germany) and a 1064 nm i-Raman EX instrument (BWTek, Newark, USA). The first instrument is a small, battery-operated, fully automated Raman spectrometer that, when started, analyses the sample by irradiating it subsequently with two lasers of different wavelengths (785 and 852 nm), allowing to cover a relatively broad spectral range (between 300 and 3200 cm^{-1}). Moreover, during the automated protocol, the spectrometer records a series of spectra at slightly different laser wavelengths

and performs an automated sequentially shifted wavelength baseline correction, avoiding interferences caused by fluorescence or other features irrelevant to the actual Raman signal. During analysis, the instrument is typically positioned in contact with the paint layer and held by hand during the analysis (typically less than a minute). After analysis, by using the manufacturer's OPUS software, the spectra are converted to Galactic/Thermo.SPC format. The software automatically determines the settings like measurement time and number of accumulations, while the laser power on the sample remains constant (fixed by the manufacturer). Despite the limited possibilities to interfere with the instrument and modify the settings, the spectrometer is easy to operate and in general it yields good quality spectra.

To evaluate the developed simple spectral searching algorithm, spectra of mock-up samples were also recorded with an instrument with a longer laser wavelength: the i-Raman EX spectrometer (BWTek, Newark, USA) is a dispersive spectrometer that uses a Nd:YAG laser (1064 nm) as excitation wavelength. It covers a spectral range of 100–2500 cm^{-1} and

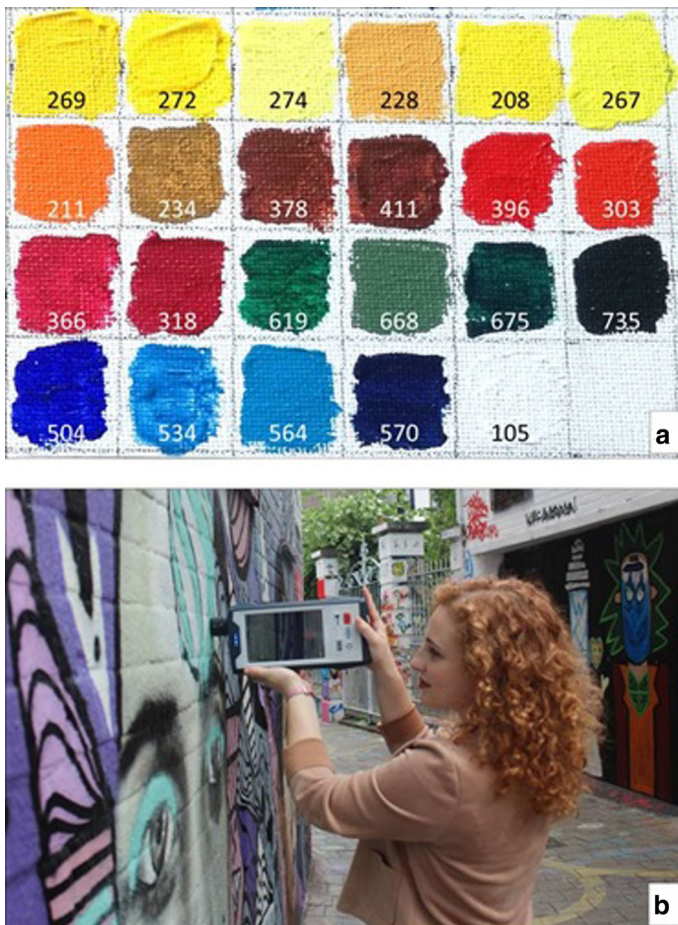


Fig. 1 **a** Mock-up samples of 23 pure paint samples. The overlay numbers refer to the paint numbers as assigned by the manufacturer. **b** Graffiti Street (Werregarenstraat) in Ghent (Belgium), where in situ spectra are recorded from street art using the Bruker Bravo spectrometer

has a spectral resolution of ca. 10 cm^{-1} . This instrument allows us to adjust laser power, measurement time and number of accumulations. Typically, spectra were recorded by using 10 accumulations of 10 s. The fibre optics probe head was positioned in front of the samples by using a manual stage, and the distance to the sample was optimised by performing short measurements.

The in-house written simple spectral searching algorithm was developed using MATLAB (The MathWorks Inc., Natick, MA, USA) version 9.5.0.1033004 (R2018b) and was run under Microsoft Windows 10. Band detection is based on the findpeaks function from the Signal Processing toolbox (Version 8.1). This algorithm is available as electronic supplementary material.

To identify the spectra of the unknown samples, the developed algorithm was used to compare the spectra with the spectra of modern synthetic pigments from the well-known collection of reference Raman spectra from KIK/IRPA (Brussels, Belgium). This collection consists of 325.SPC files that are available in two versions: original spectra and baseline-corrected spectra. For our experiments we worked with the original spectra that were recorded with a Renishaw Raman spectrometer, equipped with a 785 nm excitation laser. The spectra are available under the Creative Commons Attribution-NonCommercial-NoDerivatives licence [17].

To test the developed algorithm on commercial paint samples, simple mock-up samples (Fig. 1a) were made by applying Rembrandt Talens acrylic paint (Apeldoorn, The Netherlands), with the help of clean brushes. The main pigments that are used for the production of this commercial paint are mentioned on the paint tubes and are used for comparison (Table 1, first column).

3 Results and discussion

3.1 Developing a simple spectral searching algorithm

In this research paper a new simple spectral searching algorithm is implemented that is based on the comparison of Raman band positions. Figure 2 provides a schematic overview of the dataflow in the proposed algorithm. The original spectrum of the unknown is loaded into the MATLAB workspace, and an automated polynomial baseline correction is performed [39]. The user can choose to truncate manually the spectrum to a specific spectral range and eventually the sensitivity (related to the relative intensity or signal-to-noise ratio of the spectrum) of the spectral band detection can be modified, and the Raman band positions of the N most intense Raman bands are stored in the so-called peaktable.

A library can be generated from a collection of spectra: all the Raman spectra from a file folder are sequentially loaded into the workspace and arranged into one library. The software is able to read galactic.SPC files or.txt files from the RUFF project [9]. Moreover, a function is implemented that allows the user to add spectral information from literature to the spectral library. This information can for instance be a list of Raman band positions as reported in a research paper. The resulting spectrum consists of a flat line, with sharp, full intensity 'Raman bands' at the specified band positions. After the acquisition of the spectra, the library is transformed to a peaktable library. First, all reference spectra in the library are truncated to the same spectral range as was selected for the peaktable of the unknown (only the Raman bands of references and unknown within a common range are considered), and all reference spectra are automatically baseline corrected before Raman bands are detected. As a consequence, each time an unknown spectrum is investigated, the peaktable library

Table 1 Summary of the results of the simple searching algorithm applied on the Raman spectra as recorded from the mock-up samples

Sample	Number of bands N	Identification (distance d)	Rank (distance d)	Spectrometer
269 (PY74)	11	PY74 (8)	1 (8)	Bravo
269 (PY74)	9	PY74 (11)	1 (11)	i-Raman EX
272 (PY128)	8	PR146 (41)	5 (43)	Bravo
272 (PY128)*	$8 - 2 = 6$	PY128 (1)	1 (1)	Bravo
272 (PY128)	6	PR17 (40)	9 (44)	i-Raman EX
272 (PY128)*	$6 - 2 = 4$	PY128 (6)	1 (6)	i-Raman EX
267 (PY3)	11	PY73 (36)	10 (49)	Bravo
267 (PY3)*	$11 - 2 = 9$	PY73 (15)	3 (29)	Bravo
267 (PY3)	9	PY2 (40)	44 (67)	i-Raman EX
267 (PY3)*	$9 - 2 = 7$	PY73 (14)	8 (30)	i-Raman EX
396 (PR254)	8	PR254 (7)	1 (7)	Bravo
396 (PR254)	9	PR254 (8)	1 (8)	i-Raman EX
366 (PV19)	8	PR206 (23)	2 (53)	Bravo
366 (PV19)*	$8 - 2 = 6$	PV19 (4)	1 (4)	Bravo
366 (PV19)	13	PR54 (43)	143 (83)	i-Raman EX
366 (PV19)*	$13 - 2 = 11$	PO17 (47)	193 (109)	i-Raman EX
675 (PG7)	6	PG7 (11)	1 (11)	Bravo
675 (PG7)	7	PG7 (20)	1 (20)	i-Raman EX
570 (PB15)	6	PB15 (6)	1 (6)	Bravo
570 (PB15)	No bands detected	i-Raman EX		
318 (PR23)	8	PR23 (31)	1 (31)	Bravo
318 (PR23)*	$8 - 1 = 7$	PR23 (22)	1 (22)	Bravo
318 (PR23)	9	PY49 (40)	2 (43)	i-Raman EX
318 (PR23)*	$9 - 2 = 7$	PR23 (23)	1 (23)	i-Raman EX
274 (PY53)	8	PR181 (67)	–	Bravo
274 (PY53)	7	PR285 (68)	–	i-Raman EX
228 (PBr24)	6	PO86 (56)	–	Bravo
228 (PBr24)	4	PB15:4 (91)	–	i-Raman EX
208 (PY35)	4	PR209 (106)	–	Bravo
208 (PY35)	7	PR173 (50)	–	i-Raman EX
211 (PO20)	4	PR209 (66)	–	Bravo
211 (PO20)	4	PR209 (80)	–	i-Raman EX
234 (PY42)	5	PB1:2 (89)	–	Bravo
234 (PY42)	4	PO74 (60)	–	i-Raman EX
378 (PR101)	9	PR190 (61)	–	Bravo
378 (PR101)	8	PR173 (70)	–	i-Raman EX
411 (PR101)	7	PY95 (56)	–	Bravo
411 (PR101)	11	PV53 (37)	–	i-Raman EX

Table 1 continued

Sample	Number of bands N	Identification (distance d)	Rank (distance d)	Spectrometer
303 (PR108)	7	PR209 (89)	–	Bravo
303 (PR108)	10	PY109 (77)	–	i-Raman EX
668 (PG17)	8	PR123 (51)	–	Bravo
668 (PG17)	8	PR173 (54)	–	i-Raman EX
735 (PBk11)	9	PR190 (71)	–	Bravo
735 (PBk11)	9	PY63 (39)	–	i-Raman EX
504 (PB29)	5	PB16 (64)	–	Bravo
504 (PB29)	7	PB1:2 (67)	–	i-Raman EX
534 (PB35)	8	PB16 (76)	–	Bravo
534 (PB35)	9	PV2 (53)	–	i-Raman EX
105 (PW6)	6	PB15:2 (65)	–	Bravo
105 (PW6)	6	PB15:2 (65)	–	i-Raman EX
619 (PG7, PY 128)	8	PG7 (12)	1(12)	Bravo
619 (PG7, PY 128)	9	PB15:3 (43)	7 (56)	i-Raman EX
564 (PB15, PG7, PW6)	10	PB15:3 (34)	1 (34)	Bravo
564 (PB15, PG7, PW6)	No bands detected	i-Raman EX		

For each sample, those pigments are mentioned which are declared by the manufacturer. The number N of bands detected depends on the number of Raman active vibrations, the spectral quality and the sensitivity of the peak-picking function. For two samples, no bands were detected and those spectra could not be used for further analysis. For the other samples, the pigment identified (*i.e.* the first result) and the calculated Euclidean distance d are mentioned, as well as the rank of the best matching spectrum (if present in the database), as well as the corresponding Euclidean distance d . For pigment mixtures, the first mentioned pigment (main component) was used. Some spectra (marked with *) were recalculated after deletion of 1 (1086 cm^{-1}) or 2 (712 and 1086 cm^{-1}) Raman bands, associated with CaCO_3

needs to be generated. The findpeaks function is run with the highest sensitivity (minimum prominence [44] is set to 0) and subsequently the algorithm selects the $N_{\max} = N \times 1.8$ (rounded to the nearest integer) most intense Raman bands. The factor of 1.8 is an arbitrary factor that usually allows to keep computational time and memory use manageable, while still permitting that the algorithm is not limited to the minimum of N most intense bands. Thus, it is possible to cover, to some extent, differences in relative band intensities between the references and the unknowns: the algorithm can select the N band positions (out of N_{\max}) that best correspond to the N bands of the unknown. For each spectrum in the library, the algorithm generates an (N -by- C) array that contains all possible unique combinations of N band positions out of a pool of N_{\max} band positions. As a consequence C equals $C = \binom{N_{\max}}{N} = \frac{N_{\max}!}{N!(N_{\max}-N)!}$. Thus, for each spectrum in the reference library, the algorithm selected a series of N band positions out of the N_{\max} most intense Raman bands. This way, the algorithm is not limited to the N most intense bands for each reference, but has some built-in flexibility to deal with slightly different band intensities that might be caused by different excitation lasers, different detector sensitivities, crystal orientation, bad signal-to-noise ratios or even impurities. It should be remarked that the algorithm requires a minimal separation of

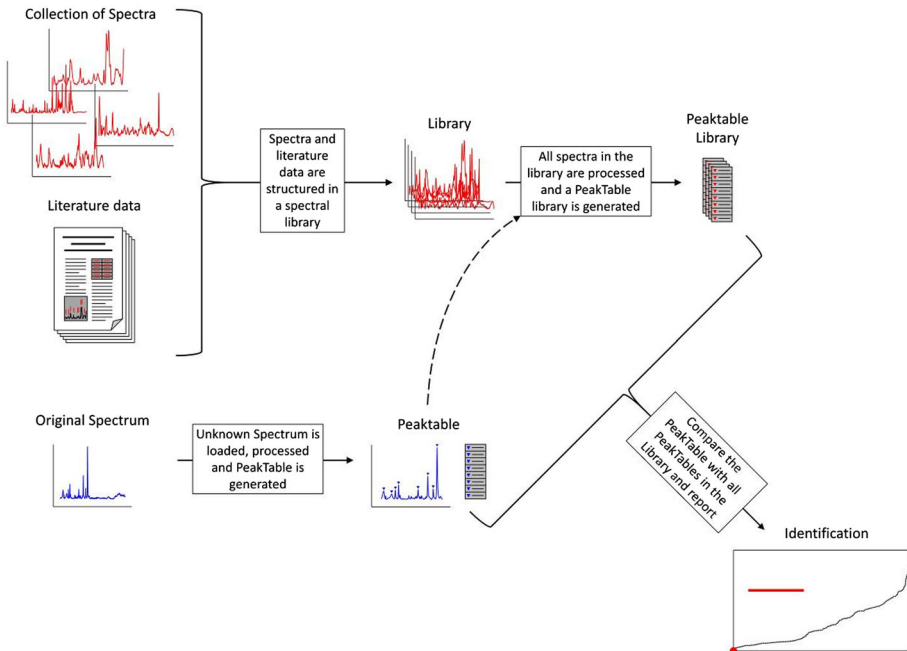


Fig. 2 Schematic overview of the dataflow in the proposed simple spectral searching algorithm

30 cm^{-1} between two detected bands. Thus, a difference in spectral resolution between the reference spectra and the spectrum of the unknown is taken into account: confusion between the presence of two nearby bands and a shoulder is avoided.

In the final stage, the software calculates the Euclidean distance of band positions for each reference spectrum between the peaktable of the unknown and each of the C selected combinations of N Raman bands. For each spectrum the best possible combination is considered and the Euclidean distance d , as calculated in N -dimensional space, is reported. If this should be considered useful, other (dis-)similarity measures can be implemented easily. Next to making all numeric values available in the workspace, the algorithm provides a graphical output as well (Fig. 3). It reports a graph with the distance d plotted for all reference spectra. A further plot is generated with the spectrum of the unknown overlaid on top of the reference. Finally, the selected Raman bands of both unknown and reference are marked in the report. It was not possible to determine an absolute limit value to mark a trustworthy identification as this is highly dependent on both the composition of the reference database and the Raman spectrum of the unknown on hand. In different situations a broad range of distances for correct identifications can be found. Therefore, a tool was provided to facilitate easy evaluation of the results: the user can scroll through the reference spectra, updating the overlay plots accordingly. Apart from the speed of the computer, the speed of the identification algorithm is dependent on the number of bands considered and the size of the reference database. Typically a result was obtained in few minutes time. The MATLAB code for the functions of this algorithm is available as supplementary material (SI).

3.2 Evaluating the simple spectral searching algorithm based on the Raman analysis of mock-up samples

The spectral searching algorithm is evaluated based on a series of Raman spectra that were recorded from 23 mock-up samples of pure paint samples, applied on prepared canvas stretched on cardboard (Fig. 1a). These samples comprise cases where the pigment is present in the reference collection and cases where the pigments are not present, as well as mixtures of pigments. Spectra were recorded by using the Bruker Bravo instrument and the i-Raman EX spectrometer and compared against the 325 spectra in the KIK/IRPA database [17]. As we aim to identify spectra that are recorded under entirely different circumstances (e.g. different excitation lasers, spectral preprocessing, different spectrometer response curves) than the reference spectra, our algorithm does not bring the (relative) Raman intensities into account, only the band positions. In Fig. 3, the spectrum plotted in blue is the spectrum of the mock-up sample and the red spectrum is the corresponding reference spectrum of the yellow pigment PY74. Due to the use of different instruments, calibration and acquisition parameters than the database (spectral resolution and excitation wavelength), the spectra present differences.

In general, upon the analysis of the mock-up samples, the simple spectral searching algorithm yielded interesting results that are summarised in Table 1. We have truncated all recorded spectra between 300 and 1800 cm^{-1} , to avoid interference, caused by TiO_2 , which is commonly used as white pigment in contemporary art and that may dominate the low-wavenumber spectral range. TiO_2 is not included in the spectral database of contemporary organic pigments.

Table 1 gives an overview of the results from the mock-up samples. For each spectrum (S2) the number of Raman bands N that are incorporated in these calculations are mentioned, as well as the identification (the name of the reference with the shortest Euclidean distance d). We also mentioned the rank (in a list of 325 reference spectra ordered from the shortest

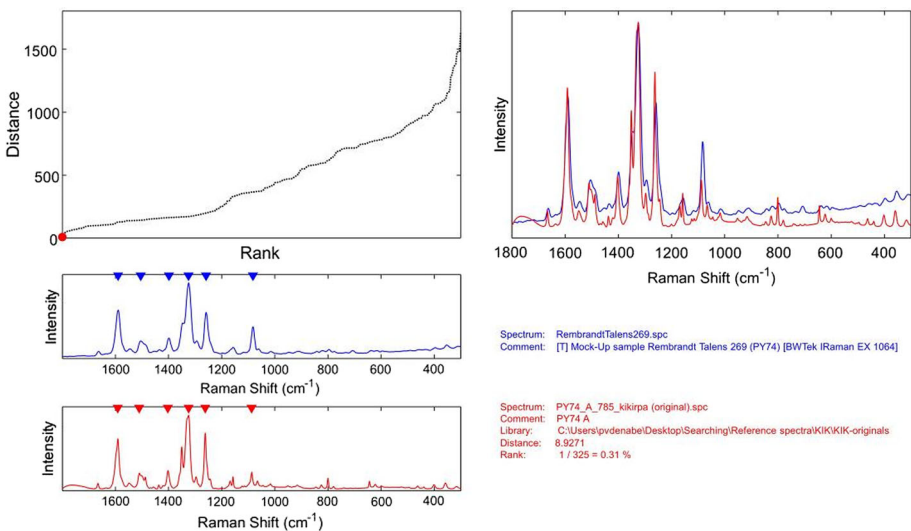


Fig. 3 Example of the output of the simple spectral searching algorithm. The unknown spectrum of a mock-up sample from Rembrandt Talens was recorded with a BWTek i-Raman EX (1064 nm excitation) spectrometer. The correct pigment (PY74) was identified from the KIK/IRPA spectral collection [17]

Euclidean distance to the highest) of the correct identification: the reference spectrum that corresponds to the pigment composition as mentioned by the manufacturer. Three groups of mock-up samples were used for the evaluation of the approach: (1) paint samples where the pigment is present in the reference database, (2) paint samples where the pigment is not present in the reference database, and (3) paint samples that consist of a mixture of pigments. All spectra were recorded with two different spectrometers. Spectral matching was performed by comparing the Euclidean distances d . It should be noted, however, that these are influenced by several parameters, including spectral resolution and calibration of the spectrometers. In general, it seems that the Euclidean distances are slightly higher for the i-Raman EX spectrometer, compared to the Bravo instrument. When the searching algorithm correctly classifies the paint sample, the Euclidean distance is typically less than 12 (for Bravo), except for sample 318, while for the i-Raman EX instrument these values reach a value up to 20. When judging the group of misclassified spectra, the distance values are typically higher than 30 for Bravo and even higher (37) for the i-Raman EX spectrometer. In general, the number of misclassifications is slightly increased with the i-Raman EX instrument, compared to the Bravo spectrometer. From the misclassified spectra no. 272, 267, 366 and 318, it can be seen that these spectra have significant contributions of CaCO_3 , which is either used as a filler in the paint, or as a whitener that was used on the support where the paint samples were applied. This is reflected in intense Raman bands at 1086 cm^{-1} and often also observed at 712 cm^{-1} . As these bands are relatively intense and as they are not present in the reference samples, their presence can hamper the correct assignment. Therefore, we have deleted these bands in the corresponding peaktables, and the results of this subtraction are reported in the table (marked with a *). These omissions significantly improve the results of the simple algorithm, but also illustrate that this approach can be sensitive to the presence of other Raman scatterers. When examining the results of samples that contain several pigments, we remark that sometimes it was possible to identify the main component (e.g. the analysis of samples no. 564 and 619 with the Bravo instrument). However, the algorithm is not designed to identify all components present in a mixture. Manual examination of the overlaying spectra reveals in these cases some Raman bands of the unknown spectrum that are not present in the reference material. It is worth to note that the identification strongly depends on the spectral quality of the recorded Raman spectrum. Indeed, if weak bands and a high background are present, the peak-picking algorithm may erroneously consider the background noise as Raman bands, resulting in wrong assignments.

It should also be noted that the Raman spectra of synthetic organic pigments in general have many, relatively sharp Raman bands. Several classes can be distinguished, but for pigments which present similar molecular structure, the Raman spectra may be highly similar [18, 28]. Thus, to support the further evaluation of the results, the algorithm provides a graphical representation, where the spectra of the unknown component and of the standard are shown as overlay (Fig. 3). However, in the specific case of differentiating between the polymorphs of the same pigment (e.g. copper phthalocyanine pigment (PB15)) a more advanced approach is needed to differentiate between these highly similar spectra [33, 34].

The algorithm automatically selects a number of intense bands in the spectrum of the unknown (based on their intensity and the amount of noise in the spectrum), but the sensitivity can be manually modified. Then, the algorithm finds in the reference library many possible combinations of detected bands, and selects the best fit. The need for this is illustrated in Fig. 4: in the right column the algorithm selects the best possible combination, whereas in the left column only the most intense bands are selected. However, if different relative Raman band intensities are observed between the reference and the spectrum of the unknown pigment, the algorithm may not select the most appropriate bands. These differences in intensities can

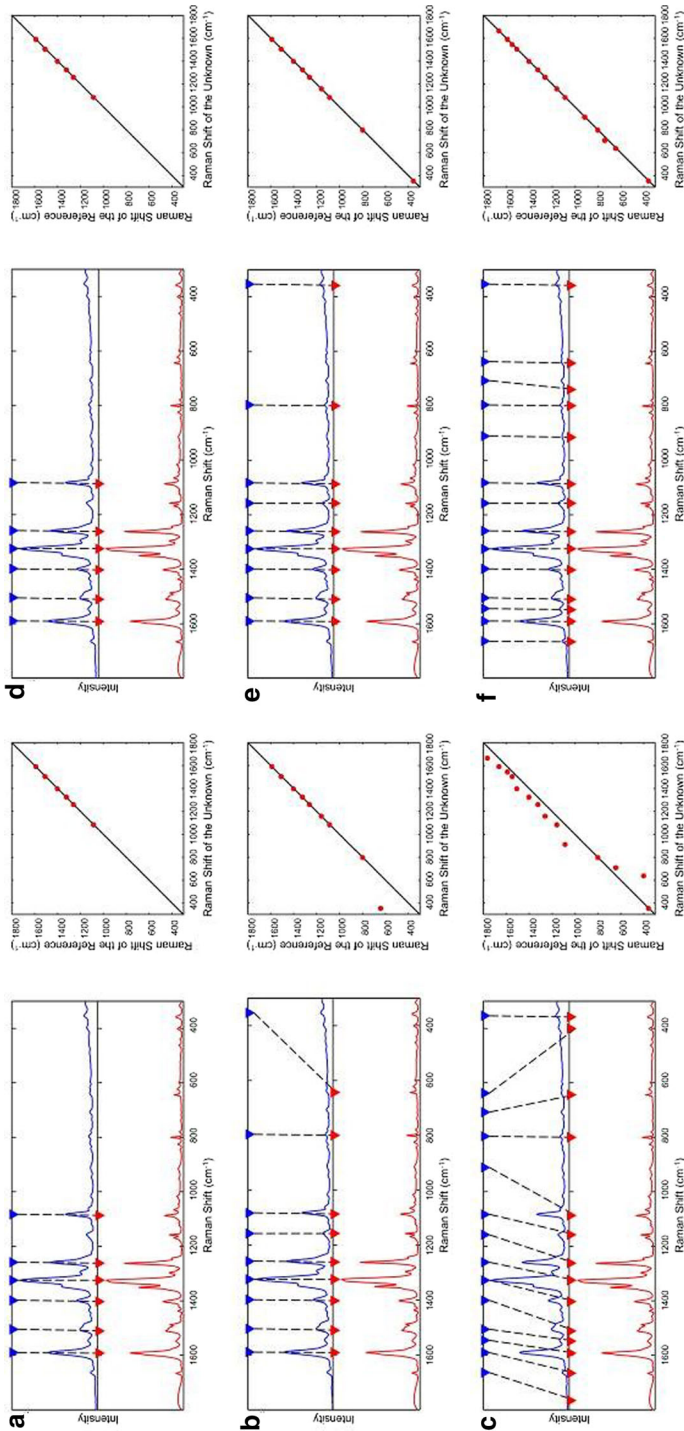


Fig. 4 Results of the analysis of a yellow mock-up sample, with different settings of the peak-picking sensitivity (a,d: $N = 6$ bands; b,e: $N = 9$ bands; c,f: $N = 14$ bands). For the results a-c only the N most intense Raman bands were selected, while for the results d-f all possible combinations of N bands selected out of the N_{\max} most intense bands were selected. Correspondences between the bands are marked with a dashed line. All bands are also plotted in the corresponding scatter plot. The diagonal line is only presented for clarity

be caused by different reasons, such as orientation, different excitation lasers, background correction, detector sensitivity, etc., or even thermal decomposition (that might affect also the Raman band positions). These drawbacks may be overcome by selecting the best fit from a large set of combinations of band positions. On the other hand, the approach may cause a serious increase in calculation time and memory problems when a large number of Raman bands is involved.

3.3 Simple spectral searching algorithm applied on in situ analysis of street art

After having tested the approach on mock-up samples, the approach is applied during the in situ analysis of street art. In the pedestrian zone of downtown Ghent, there is a small alley (Werregarenstraat, Ghent, Belgium), dedicated to graffiti, and commonly called ‘Graffiti Street’. In this alley, street artists are allowed to express themselves on the walls of the adjacent buildings. These walls are covered with graffiti, and are regularly overpainted by other artists. As a consequence, the appearance of the street changes very frequently. We recorded Raman spectra of different scenes in this street (Fig. 1b), to evaluate the simple spectral searching algorithm. In this study, the spray-painted surfaces were analysed by performing in situ Raman spectroscopy—a non-destructive approach that is frequently used in conservation science, next to methods that require a sample, such as benchtop Raman spectroscopy, Fourier-transform infrared spectroscopy (FTIR) or scanning electron microscopy (SEM–EDX). The current study does not aim at fully identifying all components of the spray-paint, such as the binding medium that can be investigated using gas chromatography, but only wants to test usefulness of developed the searching algorithm for the identification of synthetic organic pigments.

When using spectra collected from a real case study, as those from street art, it is of utmost importance to be very critical towards the outcome of the simple searching algorithm. Typically, the spectra as recorded in Graffiti Street are more noisy, as compared to the spectra of the mock-ups. Moreover, sometimes mixtures are present, or spectra may be overwhelmed by contributions of inorganic components, such as CaCO_3 or TiO_2 . Usually, the contribution of the binding media is much weaker than the Raman signal of the pigments. Although serious fluorescence can be present, the Bravo instrument’s shifted wavelength baseline correction algorithm should correct for this [40–42]. To evaluate the correctness of the identification, the user can evaluate the calculated Euclidean distance d . The smaller the Euclidean distance is, the better the band positions of the unknown and the reference spectrum correspond. Moreover, the spectrum of the unknown can be (manually) compared with the reference spectra and be examined for missing Raman bands, or prominent changes in relative intensities, as this also can indicate the presence of pigment mixtures.

Figure 5 represents some spectra in Graffiti Street as recorded with the Bruker Bravo instrument. As can be seen from spectra 5a, c, it is possible to achieve a good identification, which is also supported by the short Euclidean distances d . On the other hand, in the case of the analysis of the red paint from a graffiti painting (Fig. 5d) the identification as PR2 is wrong. The band positions in the region between 1800 and 1000 cm^{-1} at first sight seem to correspond, and the differences in relative intensities could eventually be assigned to a different spectral resolution. However, when looking at the low wavenumber range, it is clear that this spectrum does not correspond. When more bands are included (Fig. 5e), the assignment correctly falls to pigment PR112. It is remarkable that both pigments (PR2 and PR112) belong to the naphthol AS pigments and differ only by an addition of an extra Cl- and CH_3 - group [43]. In the examples in Figs. 5b, f, the pigments seem to be assigned correctly (*i.e.* assigned to the same pigments as would have been done by manual comparison against

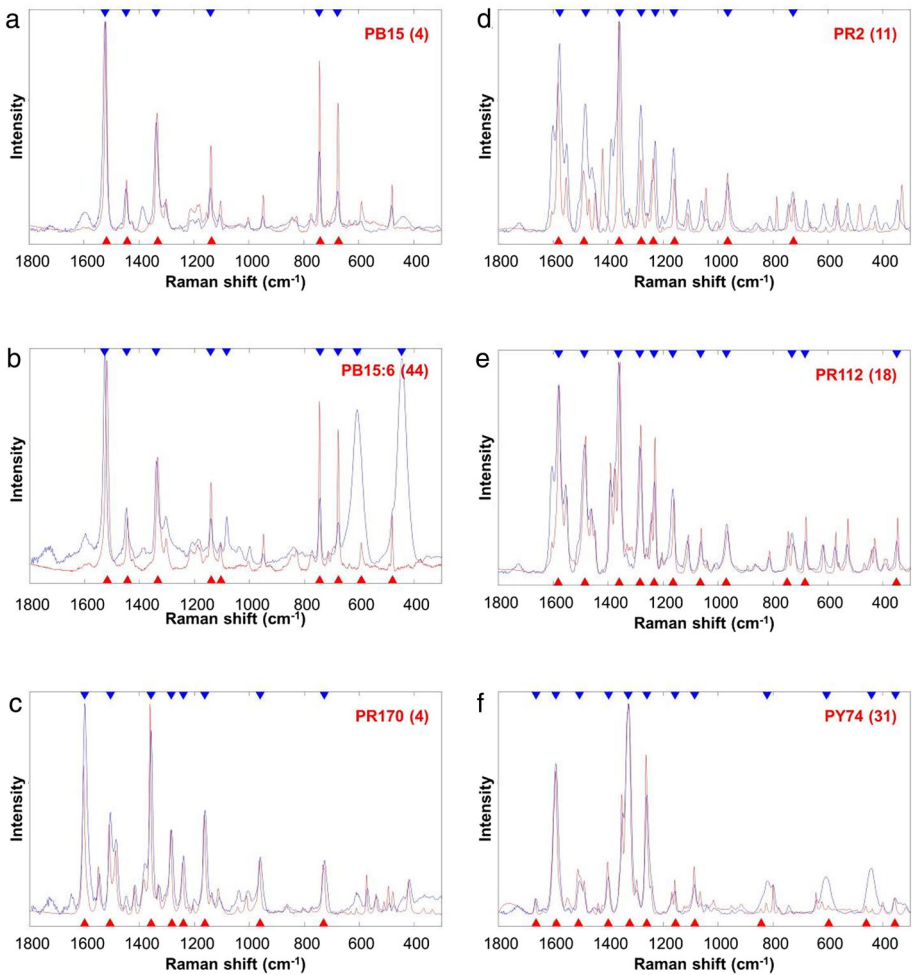


Fig. 5 Examples of spectra as recorded with the Bravo instrument, from Graffiti Street (blue), along with the assigned reference spectrum (red). The triangular markers in red and blue indicate the included band positions of the unknown and reference spectrum, respectively

the library). However, in the low wavenumber range some intense and relatively broad bands are observed that can be assigned to the presence of titanium white (TiO_2), which is the most common white pigment that is now on the market. Despite this interference, it was possible to assign the pigments, although the Euclidean distance is significantly higher.

4 Conclusions

In this paper, we propose a novel, simple spectral searching algorithm that was used successfully for the analysis of pigments in contemporary street art. The algorithm is based on the extraction of the most prominent Raman bands from the spectrum of the unknown, and subsequent comparison against the extracted Raman bands from a spectral library of synthetic

organic pigments. A simple Euclidean distance is used to determine the spectrum with the highest similarity. As it is based on the comparison of the band positions and as it does not take the intensities into account, the main advantage of this algorithm is that it experiences minimal interferences from background radiation, as well as from spectra that are recorded with different wavelengths or at a different spectral resolution, compared to the reference database. The algorithm is extensively tested on mock-up samples and on Raman spectra of street art that were recorded in situ by using portable spectrometers, and successes as well as failures are described.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1140/epjp/s13360-021-01577-8>.

Acknowledgements A.R. thanks the Research Foundation Flanders/FWO-Vlaanderen for her postdoctoral Grant (12X1919N). E.P. is grateful to Fondazione Flaminia and to the University of Bologna for her traineeship grant.

Data availability statement This manuscript has associated data in a data repository. [Authors' comment: The data that support the findings of this study are available from the corresponding author upon reasonable request.]

References

1. I.M. Bell, R.J. Clark, P.J. Gibbs, Raman spectroscopic library of natural and synthetic pigments (pre-~1850 AD). *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **53**(12), 2159–2179 (1997)
2. L. Burgio, R.J. Clark, Library of FT-Raman spectra of pigments, minerals, pigment media and varnishes, and supplement to existing library of Raman spectra of pigments with visible excitation. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **57**(7), 1491–1521 (2001)
3. K. Castro, M. Pérez-Alonso, M.D. Rodríguez-Laso, L.A. Fernández, J.M. Madariaga, On-line FT-Raman and dispersive Raman spectra database of artists' materials (e-VISART database). *Anal. Bioanal. Chem.* **382**(2), 248–258 (2005)
4. G. Burrafato, M. Calabrese, A. Cosentino, A.M. Gueli, S.O. Troja, A. Zuccarello, ColoRaman project: Raman and fluorescence spectroscopy of oil, tempera and fresco paint pigments. *J. Raman Spectrosc.* **35**(10), 879–886 (2004)
5. G. Marucci, A. Beeby, A.W. Parker, C.E. Nicholson, Raman spectroscopic library of medieval pigments collected with five different wavelengths for investigation of illuminated manuscripts. *Anal. Methods* **10**(10), 1219–1236 (2018)
6. P. Vandenabeele, B. Wehling, L. Moens, H. Edwards, M. De Reu, G. Van Hooydonk, Analysis with micro-Raman spectroscopy of natural organic binding media and varnishes used in art. *Anal. Chim. Acta* **407**(1–2), 261–274 (2000)
7. C. Daher, C. Paris, A.S. Le Hô, L. Bellot-Gurlet, J.P. Échard, A joint use of Raman and infrared spectroscopies for the identification of natural organic media used in ancient varnishes. *J. Raman Spectrosc.* **41**(11), 1494–1499 (2010)
8. H.G. Edwards, E.M. Ali, Raman spectroscopy of archaeological and ancient resins: Problems with database construction for applications in conservation and historical provenancing. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **80**(1), 49–54 (2011)
9. B. Lafuente, R.T. Downs, H. Yang, N. Stone, The power of databases: the RRUFF project. In *Highlights in mineralogical crystallography* (pp. 1–29). Walter de Gruyter GmbH (2016)
10. J. De Gelder, K. De Gussem, P. Vandenabeele, L. Moens, Reference database of Raman spectra of biological molecules. *J. Raman Spectrosc.* **38**(9), 1133–1147 (2007)
11. D. Giordano, D. González-García, J.K. Russell, S. Raneri, D. Bersani, L. Fornasini et al., A calibrated database of Raman spectra for natural silicate glasses: implications for modelling melt physical properties. *J. Raman Spectrosc.* **51**, 1822–1838 (2019)
12. P. Vandenabeele, D.M. Grimaldi, H.G. Edwards, L. Moens, Raman spectroscopy of different types of Mexican copal resins. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **59**(10), 2221–2229 (2003)
13. P. Vandenabeele, M. Ortega-Aviles, D.T. Castelleros, L. Moens, Raman spectroscopic analysis of Mexican natural artists' materials. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **68**(4), 1085–1088 (2007)

14. A. Coccato, D. Bersani, A. Coudray, J. Sanyova, L. Moens, P. Vandenabeele, Raman spectroscopy of green minerals and reaction products with an application in Cultural Heritage research. *J. Raman Spectrosc.* **47**(12), 1429–1443 (2016)
15. B. Gilbert, S. Denoël, G. Weber, D. Allart, Analysis of green copper pigments in illuminated manuscripts by micro-Raman spectroscopy. *Analyst* **128**(10), 1213–1217 (2003)
16. A. Coccato, J. Jehlicka, L. Moens, P. Vandenabeele, Raman spectroscopy for the investigation of carbon-based black pigments. *J. Raman Spectrosc.* **46**(10), 1003–1015 (2015)
17. W. Fremout, S. Saverwyns, Identification of synthetic organic pigments: the role of a comprehensive digital Raman spectral library. *J. Raman Spectrosc.* **43**(11), 1536–1544 (2012)
18. P. Vandenabeele, L. Moens, H.G. Edwards, R. Dams, Raman spectroscopic database of azo pigments and application to modern art studies. *J. Raman Spectrosc.* **31**(6), 509–517 (2000)
19. N.C. Scherrer, Z. Stefan, D. Francoise, F. Annette, K. Renate, Synthetic organic pigments of the 20th and 21st century relevant to artist's paints: Raman spectra reference collection. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **73**(3), 505–524 (2009)
20. C. Defeyt, P. Vandenabeele, B. Gilbert, J. Van Pevenage, R. Cloots, D. Strivay, Contribution to the identification of α -, β - and ϵ -copper phthalocyanine blue pigments in modern artists' paints by X-ray powder diffraction, attenuated total reflectance micro-fourier transform infrared spectroscopy and micro-Raman spectroscopy. *J. Raman Spectrosc.* **43**(11), 1772–1780 (2012)
21. Infrared and Raman users group: <http://irug.org>
22. <https://soprano.kikirpa.be>
23. S. Saverwyns, Russian avant-garde... or not? A micro-Raman spectroscopy study of six paintings attributed to Liubov Popova. *J. Raman Spectrosc.* **41**(11), 1525–1532 (2010)
24. J. De Gelder, P. Vandenabeele, F. Govaert, L. Moens, Forensic analysis of automotive paints by Raman spectroscopy. *J. Raman Spectrosc. Int. J. Orig. Work Asp. Raman Spectrosc. Incl. High. Order Processes Brillouin Rayleigh Scatter.* **36**(11), 1059–1067 (2005)
25. C. Muehlethaler, G. Massonnet, P. Esseiva, The application of chemometrics on Infrared and Raman spectra as a tool for the forensic analysis of paints. *Forensic Sci. Int.* **209**, 173–182 (2011)
26. D. Lambert, C. Muehlethaler, P. Esseiva, G. Massonnet, Combining spectroscopic data in the forensic analysis of paint: application of a multiblock technique as chemometric tool. *Forensic Sci. Int.* **263**, 39–47 (2016)
27. A. Nevin, I. Osticioli, D. Anglos, A. Burnstock, S. Cather, E. Castellucci, Raman spectra of proteinaceous materials used in paintings: a multivariate analytical approach for classification and identification. *Anal. Chem.* **79**(16), 6143–6151 (2007)
28. P. Vandenabeele, A. Hardy, H.G. Edwards, L. Moens, Evaluation of a principal components-based searching algorithm for Raman spectroscopic identification of organic pigments in 20th century artwork. *Appl. Spectrosc.* **55**(5), 525–533 (2001)
29. A. Nevin, I. Osticioli, D. Anglos, A. Burnstock, S. Cather, E. Castellucci, The analysis of naturally and artificially aged protein-based paint media using Raman spectroscopy combined with principal component analysis. *J. Raman Spectrosc.* **39**(8), 993–1000 (2008)
30. N. Navas, J. Romero-Pastor, E. Manzano, C. Cardell, Raman spectroscopic discrimination of pigments and tempera paint model samples by principal component analysis on first-derivative spectra. *J. Raman Spectrosc.* **41**(11), 1486–1493 (2010)
31. N.S. Daly, M. Sullivan, L. Lee, K. Trentelman, Multivariate analysis of Raman spectra of carbonaceous black drawing media for the in situ identification of historic artist materials. *J. Raman Spectrosc.* **49**(9), 1497–1506 (2018)
32. C. Daher, L. Bellot-Gurlet, A.S. Le Hô, C. Paris, M. Regert, Advanced discriminating criteria for natural organic substances of Cultural Heritage interest: spectral decomposition and multivariate analyses of FT-Raman and FT-IR signatures. *Talanta* **115**, 540–547 (2013)
33. J. González-Vidal, R. Pérez-Pueyo, M.J. Soneira, Automatic classification system of Raman spectra applied to pigments analysis. *J. Raman Spectrosc.* **47**, 1408–1414 (2016)
34. C. Defeyt, J. Van Pevenage, L. Moens, D. Strivay, P. Vandenabeele, Micro-Raman spectroscopy and chemometrical analysis for the distinction of copper phthalocyanine polymorphs in paint layers. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **115**, 636–640 (2013)
35. P. Vandenabeele, L. Moens, Micro-Raman spectroscopy of natural and synthetic indigo samples. *Analyst* **128**(2), 187–193 (2003)
36. M. Castanys Tutzó, R. Perez-Pueyo, M.J. Soneira, S. Ruiz Moreno, Fuzzy logic: a technique to Raman spectra recognition. *J. Raman Spectrosc.* **37**(10), 1003–1011 (2006)
37. R. Perez-Pueyo, M.J. Soneira, S. Ruiz-Moreno, A fuzzy logic system for band detection in Raman spectroscopy. *J. Raman Spectrosc.* **35**(8–9), 808–812 (2004)

38. D. Hutsebaut, P. Vandenabeele, L. Moens, Evaluation of an accurate calibration and spectral standardization procedure for Raman spectroscopy. *Analyst* **130**(8), 1204–1214 (2005)
39. P. Vandenabeele, *Practical Raman Spectroscopy: An Introduction* (Wiley, 2013), pp. 114–115
40. C. Conti, A. Botteon, M. Bertasa, C. Colombo, M. Realini, D. Sali, Portable sequentially shifted excitation Raman spectroscopy as an innovative tool for in situ chemical interrogation of painted surfaces. *Analyst* **141**(15), 4599–4607 (2016)
41. A. Culka, J. Hyršl, J. Jehlička, Gem and mineral identification using GL gem raman and comparison with other portable instruments. *Appl. Phys. A* **122**(11), 959 (2016)
42. A. Rousaki, M. Costa, D. Saelens, S. Lycke, A. Sánchez, J. Tuñón, B. Ceprián et al., A comparative mobile Raman study for the on field analysis of the Mosaico de los Amores of the Cástulo Archaeological Site (Linares, Spain). *J. Raman Spectros.* **51**, 1913 (2019)
43. W. Herbst, K. Hunger, *Industrial Organic Pigments: Production, Properties, Applications* (Wiley, 2006).
44. <https://www.mathworks.com/help/signal/ug/prominence.html>