How Chemometrics Revives the UV-Vis Spectroscopy Applications as an Analytical Sensor for Spectralprint (Nontargeted) Analysis

Rocío Ríos-Reina and Silvana M. Azcarate

Abstract: In recent years, methodologies based on spectral analysis, using ultraviolet–visible (UV-Vis) radiation, have experienced an amazing development and have been widely applied in various fields such as agricultural, food, pharmaceutical, and environmental sciences. This straightforward technique has re-emerged with novel and challenging proposals to solve, in a direct and fast way, a wide variety of problems. These reaches would not have been possible without the essential support of chemometrics. In this sense, under the general background of the development in data and computer science, and other technologies, the emergence of innovative ideas, approaches, and strategies endows UV-Vis spectroscopy with a new vitality as an analytical sensor with the capability of significantly improving both the robustness and accuracy of results. This review presents modern UV-Vis spectral analysis, which is on the rise, associated with comprehensive chemometric methods that have become known in the last six years, especially from the perspective of practicability, including spectral preprocessing, wavelength (variable) selection, data dimension reduction, quantitative calibration, pattern recognition, and multispectral data fusion. Most importantly, it will foresee future trends of UV-Vis spectroscopy as an analytical sensor for a spectralprint (nontargeted) analysis.

Keywords: UV-Vis; chemometrics; spectralprint; quantification; pattern recognition

1. Introduction

Ultraviolet–visible (UV-Vis) spectroscopy is one of the most straightforward spectroscopic techniques that has undergone an amazing increase over the years from a plethora of methodologies’ developments and applications to an endless number of samples in various fields such as agricultural, food, pharmaceutical, and environmental sciences, among others.

This technique is based on the absorption measurement of the electromagnetic radiation from the ultraviolet and visible regions, which can provide valuable chemical information from the band positions, intensities, and shapes, indicating the presence or absence of specific structural properties or functional groups [1]. Nevertheless, UV-Vis spectra usually contain only a few broad absorbance bands and are often quite broad and difficult to associate with individual chromophores. For this reason, its first approaches have been focused on identification and quantification analysis of monocomponent systems, generally by implementation of colorimetric methodologies that are still currently applied.

However, with the fast technological progress of analytical instrumentation techniques, the advent of sensitive and affordable array detectors in the 1980s and 1990s has enhanced the measurement capacities in the UV-Vis band, making it possible to produce, almost instantly, an entire UV-Vis spectra of a sample [2]. Thus, this technique has re-emerged...
with novel and challenging proposals to solve, in a direct and fast way, a wide variety of problems changing the paradigm to multicomponent systems analysis. Recently, the nontargeted spectroscopic or fingerprinting approach has also been called *spectralprint* analysis [3]. It aims to collect as many compounds or features as technically possible from the whole spectra to obtain an overview of the sample composition for qualitative and quantitative analyses, that is, multivariate techniques are applied in the UV-Vis spectra [3]. Notwithstanding, these reaches would not have been possible without the essential support of chemometrics. In the last 20 years, the advance of chemometrics has also contributed to the use of UV-Vis technology for more complicated chemical matrices than was possible in earlier times, enabling large amounts of spectral data to be analyzed and reducing them to useful information, such as the concentration of one or more chemical species [3].

In light of these considerations, with the essential aid of chemometrics, this spectroscopic technique is no longer a simple data provider, but it has become a provider of chemical information for a complex system and even a direct participant and solver of chemical problems. Moreover, the use of multivariate calibration techniques allows the spectral information of the component of interest to be extracted, enabling the determination of multicomponent content without chemical separation, and thus significantly avoiding interference from coexisting components and complex backgrounds [4]. In this sense, under the general background of the development in technology, computing, data handling, and data analysis, the emergence of innovative ideas, approaches, and strategies endows UV-Vis spectroscopy with a new vitality as an analytical sensor with the capability of significantly improving both the robustness and accuracy of results. Thus, Figure 1 shows the significant progression in the use and application of this technique in combination with chemometric methods in the last twenty years, showing a significant increase mainly from 2010 up to now.

![Figure 1. Publication per year in Scopus Database of the last 20 years searching the keywords “UV-Vis spectroscopy” and “chemometrics”.](image)

This review presents the booming evolution of modern UV-Vis spectral analysis associated with thorough chemometric methods that have been reported since 2017, especially from the perspective of practicability, including spectral preprocessing, wavelength selection, data reduction, quantitative and pattern recognition approaches, and multispectral data fusion. Owing to the extensive number of recently published works on these topics, this review only focuses on those *spectralprint* approaches that can be considered most representative or relevant regarding the results found and the feasibility of their application. Finally, it will provide for future trends of UV-Vis spectroscopy as an analytical sensor for a *spectralprint* (nontargeted) analysis.
2. Advances of UV-Vis Spectrometric Systems and Analysis

UV-Vis spectroscopy is an analytical technique able to monitor and measure the UV and visible light interactions with a plethora of molecules in the specific ranges of 200–350 and 350–700 nm, respectively [5]. This technique exploits some physical phenomena such as absorption, scattering, diffraction, refraction, and reflection occurring between the light and compound/s within the sample (Figure 2). Particularly, UV-Vis light absorption is limited to specific chromophores with defined molecular functional groups, being mainly affected by their composition and concentration. From this phenomenon, the Beer–Lambert law has been described to correlate the quantity of the incident light absorbed by the absorbing compound or molecule present in a matrix, its concentration, and the light path length [6,7]. Thus, an endless number of quantitative analytical methods have been developed to determine and quantify the concentration of the target molecule in a wide variety of matrices [8].

![Figure 2. Energy of incident light and phenomena occurrence when interacting with a sample.](image)

The basic principles of UV-Vis spectroscopy and related instrumentation have been fully described in the literature [5–7]. However, continuous advances in the development and improvement of instrumentation have enhanced its analytical capabilities. For instance, photodiode arrays and charge-coupled devices are the most commonly used detectors in spectrophotometers nowadays since they are capable of producing a complete UV-Vis spectrum of a sample almost instantaneously. Additionally, owing to their compact geometry, these detectors have also offered the possibility to build portable devices [9]. These modern UV-Vis spectrophotometers allow the acquisition of inexpensive and compact equipments with a high scanning speed that can reach outstanding sensitivity and robustness, becoming an analytical tool of choice for a wide range of applications.

The UV-Vis spectrum of a sample is characterized by two major parameters, namely, the position of the maximum of the absorption bands and the intensity of the bands. Thus, the maxima and intensities of absorption bands differ relative to the molecular structure of the compounds, that is, depending on the sample or substance. Moreover, the amount of light absorbed by the interacting molecules also depends on their concentration in the sample. As a result of all this, a unique and specific relationship exists between the substance and its UV-Vis spectrum. Therefore, the full spectrum can then be used for quantitative (i.e., to determine the amounts of certain substances) or qualitative (i.e., to determine the presence of certain substances) analyses or to determine the physical and chemical properties of a sample because this information is contained in the bands’ positions, intensities and shapes [5,10,11]. Thus, the identification and/or quantification of a compound is easy in a pure component; however, in complex matrices, which contain mixtures of many compounds that can absorb in the UV-Vis range, e.g., food, the acquired spectra generally present a few broad absorbance bands that are often hard to associate to single chromophores [8,12]. However, at the same time, it generates a unique UV-Vis
spectrum, the *spectralprint*, that is mainly useful for nontargeting or untargeted analysis, as will be discussed in this review.

Regarding the kind of sample, its nature can present challenging aspects since the sample can be inappropriate for spectroscopic analysis. Most commercially available UV-Vis spectrometers are designed for liquid samples analysis. However, spectrophotometers are presented in different configurations since a wide variety of accessories and sample holders are available for UV-Vis measurements. Thus, their combinations offer different measurement capabilities and sample types and/or different measurement conditions, allowing solids, liquids, and gases to be analyzed by UV-Vis spectroscopy. Various sampling modes and cell holders have been designed to adapt different path-length configurations based on conventional rectangular cells, including fiberoptic-based immersion probes, flow cells, microwell plate configurations, and automated sample changers, among others. Therefore, the sampling device must be optimized for specific applications evaluating sample volume, measurement speed, and reproducibility of sample presentation.

Absorbance is the measurement usually implemented on solutions of the substance in liquid-holding cells. On the one hand, the most common cell path length used for liquid samples is 10 mm. However, other path lengths from 0.01 to 100 mm were employed. For instance, in order to acquire an improved sensitivity for samples with low absorbance, the cell path length was increased, e.g., the theoretical absorbance of a solution in a 100 mm cell is greater by a factor of ten compared to the same solution in a 10 mm cell. Conversely, samples that produce a signal saturation were measured by decreasing the cell path length, avoiding sample dilution.

On the other hand, flow cells are often used in fiberoptic applications. They can be connected directly to a side sampling loop or even into a process of the flow path, usually in the ‘Z’ configuration. In the cases in which it is not possible to connect a cell, an insertion probe is an option of accessing the sample. Moreover, several diffuse reflection cells are used to measure the light reflected from a solid surface or powder. Some probes are also utilized, and they are immune to specular reflections from the sample and detect only diffusely reflected light [2]. These configurations have become increasingly important since they have allowed for the extension of the use of UV-Vis spectroscopy in quality monitoring and process control as a real-time analytical sensor in biological, pharmaceutical, and food applications [1,2,13–15].

Furthermore, to decrease the consumption of valuable and scarce samples, miniaturized UV-Vis spectrometric systems have been developed, enabling the development of a new generation of UV-Vis spectrometric systems for novel technologies applications [13,16].

3. Strengths and Weaknesses of UV-Vis Spectroscopy

UV-Vis spectroscopy is a basic and widely used technique that has many advantages and a few main strengths that makes it popular, but it also has some disadvantages, as do many other analytical techniques.

Regarding its potential strengths, the ones related to the instrumentation are its quick analysis ability and ease of handling and use, allowing easy integration into experimental protocols and requiring little user training. In addition, it is a nondestructive technique, being considered a green technique, which allows the sample to be reused or further processed or analyzed, thus saving samples and protecting the environment, and it is generally an inexpensive or cost-effective instrument to acquire and operate, making it accessible for many laboratories. Moreover, a small amount of sample is needed, it does not contaminate, and there is the availability of portable equipment.

This technique, concerning the data provided, also presents the advantage of requiring an easy data analysis, with minimal processing, compared to other spectroscopic techniques.

Finally, in terms of its usefulness, this technique makes it possible to characterize the absorbance or transmittance through a liquid or the reflectance of a surface over a range of wavelengths, as well as to identify and determine the concentration of a particular molecule
in a solid or liquid sample, to measure the color of a material, and to study chemical reactions or biological processes [1,17].

Despite the several strengths of this technique, there are also certain weaknesses. However, good accuracy and precision in UV-Vis measurements can be achieved if precautions are taken to avoid errors.

Regarding the sample, UV-Vis spectroscopy works well on liquids and solutions, but if the sample has a suspension of solid particles in liquid, there will be light scattering, and therefore data will be skewed. However, this limitation is solved with a previous filtration of the sample. This scattering also occurs with the presence of bubbles in the cuvette or sample, resulting in irreproducible results. Another problem related to the sample is the interference from multiple absorbing species that will have overlapping spectra and make the identification and quantification of specific compounds difficult. The color of the sample will also affect the measurement; thus, very dark samples may lead to saturation of the spectrum. In addition, it should be considered that only the molecules with chromophores are the ones analyzed in the sample, that is, UV-Vis is unable to analyze compounds that do not interact with light in the UV and visible areas of the spectrum; hence, samples without those molecules do not provide any UV-Vis signal. In addition, the results of the absorption can be also affected by pH and temperature.

Concerning the instrument, selecting the most suitable sample holder, solvent, and instrument parameters is critical for the measurement. The appropriate cuvette material needs to be selected to avoid the optical interaction from the light source with it, which alters the absorbance intensity of the sample. Moreover, another problem related to the instrument is stray light, that is, the small amount of light from a wide wavelength range that can be transmitted from the light source to the environment or loosely fitted compartment, possibly causing serious measurement errors.

Finally, the geometrical parameters, that is, the alignment to the same orientation and placement in the same position for every component in the instrument, should be considered for each measurement [1,7,17].

4. UV-Vis Spectral-Chemometric Platforms
4.1. UV-Vis Spectral Data Processing

Regarding the data obtained by UV-Vis, they are a vector of numerical values of absorbance units. Thus, UV-Vis spectroscopy normally provides first-order data [3]. As with many other spectroscopies, there is not a specific rule for preprocessing data of this kind. It will depend on the problems that data present, which will come from the sample, instrument, etc. The same occurs with the selection of the best algorithm for data regression or classification. It will depend on the samples and the aim. In this section, several preprocessing methods as well as quantitative and qualitative approaches most used in the literature of the last 6 years will be discussed. For that, in this study, 58 publications on the subject were chosen, analyzed, and summarized in Table 1.
Table 1. Summary of the areas and aims of application and the analytical and chemometric methods used for UV-Vis fingerprinting in the studies published in the last six years (2017–2022).

<table>
<thead>
<tr>
<th>Year</th>
<th>Ref.</th>
<th>Area</th>
<th>Aim</th>
<th>Sample</th>
<th>Sampling</th>
<th>UV-Vis Analysis (Range/Resolution/Cuvette Path Length)</th>
<th>Preprocessing Method</th>
<th>Multivariate Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>[18]</td>
<td>PR</td>
<td>A, F, and B</td>
<td>Saffron</td>
<td>Extract solutions</td>
<td>200–700 nm/1 nm/10 mm</td>
<td>AS</td>
<td>PCA, LDA</td>
</tr>
<tr>
<td></td>
<td>[19]</td>
<td>PR</td>
<td>PR</td>
<td>Plant food supplements</td>
<td>Dilution 1:10 (v:v)</td>
<td>190–1100 nm/1 nm/1 mm</td>
<td>1D + AS</td>
<td>PCA, SIMCA</td>
</tr>
<tr>
<td></td>
<td>[20]</td>
<td>PR</td>
<td>Coffe</td>
<td>Extract solutions</td>
<td>190–700 nm/1 nm/10 mm</td>
<td>-</td>
<td>PCA, PLS-DA, SIMCA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[21]</td>
<td>PR and Q</td>
<td>coffe</td>
<td>Dilution 1:120 (v:v)</td>
<td>190–700 nm/n.d./10 mm</td>
<td>-</td>
<td>PCA-LDA, PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[22]</td>
<td>Q</td>
<td>Extra virgin olive oil</td>
<td>Direct analysis</td>
<td>190–1100 nm/1 nm/10 mm (transmittance mode)</td>
<td>-</td>
<td>PLS, PLS-JK, SPA-MLR, SW-MLR, GA-MLR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[23]</td>
<td>Q</td>
<td>Palm Civet Coffee</td>
<td>Extraction and dilution 1:20 (v:v)</td>
<td>190–700 nm/1 nm/10 mm</td>
<td>MSC + SNV</td>
<td>PCA, PLS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[25]</td>
<td>Q</td>
<td>5-Hydroxymethylfurfural 8-methoxypsoralen and trypsin</td>
<td>Direct analysis of solutions</td>
<td>200–300 nm/n.d./10 mm</td>
<td>MC</td>
<td>MCR-ALS, PLS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[26]</td>
<td>Q</td>
<td>Sewage</td>
<td>Direct analysis of solutions</td>
<td>230–350 nm/n.m/n.d.</td>
<td>-</td>
<td>MCR-ALS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[27]</td>
<td>Q</td>
<td>Cough syrup</td>
<td>Direct analysis of solutions</td>
<td>230–800 nm/n.d./10 mm</td>
<td>-</td>
<td>MCR-ALS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[28]</td>
<td>A, F, and B</td>
<td>PR</td>
<td>Sparkling Wines</td>
<td>Dilution 1:5 (v:v)</td>
<td>200–600 nm/2 nm/10 mm</td>
<td>-</td>
<td>PCA</td>
</tr>
<tr>
<td></td>
<td>[30]</td>
<td>PR</td>
<td>PR</td>
<td>Olive oil</td>
<td>Direct analysis</td>
<td>200–800 nm/1 nm/1 mm</td>
<td>MCR-ALS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[32]</td>
<td>Q</td>
<td>Wheat straw extracts</td>
<td>Extraction and dilution 6:10 (v:v)</td>
<td>190–450 nm/1 nm/10 mm</td>
<td>1D, 2D</td>
<td>PLS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[33]</td>
<td>Q</td>
<td>Cough syrup</td>
<td>Direct analysis of solutions</td>
<td>220–300 nm/2 nm/10 mm</td>
<td>CWT, DWT</td>
<td>PLS, PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[34]</td>
<td>Q</td>
<td>Rare earth elements</td>
<td>Direct analysis of dilutions</td>
<td>200–800 nm/10 nm/n.d.</td>
<td>-</td>
<td>MCR-ALS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[35]</td>
<td>A, F, and B</td>
<td>PR</td>
<td>Chili Powder</td>
<td>Extract solutions</td>
<td>200–800 nm/0.5 nm/10 mm</td>
<td>SNV</td>
<td>PCA, DA</td>
</tr>
<tr>
<td></td>
<td>[36]</td>
<td>PR</td>
<td>PR</td>
<td>Wine vinegars</td>
<td>Diluted 1:10 (v:v)</td>
<td>180–890 nm/2 nm/10 mm</td>
<td>-</td>
<td>PCA, PLS-DA, SIMCA</td>
</tr>
<tr>
<td></td>
<td>[37]</td>
<td>PR</td>
<td>PR</td>
<td>Red wine</td>
<td>Direct analysis</td>
<td>190–800 nm/1 nm/1 mm</td>
<td>-</td>
<td>PCA, PLS-DA, LDA</td>
</tr>
<tr>
<td></td>
<td>[38]</td>
<td>A, F, and B</td>
<td>PR</td>
<td>Green tea</td>
<td>Dilution 1:25 (v:v)</td>
<td>200-800 nm/1 nm/10 mm</td>
<td>MC + PS</td>
<td>PCA, HCA, PLS-DA, SIMCA</td>
</tr>
<tr>
<td></td>
<td>[39]</td>
<td>PR</td>
<td>PR</td>
<td>Tea</td>
<td>Dilution 1:10 (v:v)</td>
<td>190–800 nm/1 nm/n.d.</td>
<td>-</td>
<td>PCA, PCA-LDA, PCA-MLR</td>
</tr>
<tr>
<td></td>
<td>[40]</td>
<td>PR</td>
<td>PR</td>
<td>Coffea arabica L. leaves</td>
<td>Extraction and dilution 1:20 (v:v)</td>
<td>200–800 nm/1 nm/10 mm</td>
<td>-</td>
<td>PCA, OSC-PLS-DA</td>
</tr>
<tr>
<td>Year</td>
<td>Ref.</td>
<td>Area</td>
<td>Aim</td>
<td>Sample</td>
<td>Sampling</td>
<td>UV-Vis Analysis (Range/Resolution/Cuvette Path Length)</td>
<td>Preprocessing Method</td>
<td>Multivariate Method</td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>---------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[45]</td>
<td>A, F, and B</td>
<td>PR</td>
<td>Herbs</td>
<td>Direct analysis of powders 200–800 nm/1 nm/diffuse reflectance mode</td>
<td>AS, CWT, SG-S</td>
<td>PCA, ELM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[46]</td>
<td>Q</td>
<td>Red wine</td>
<td>200–700 nm/1 mm/10 mm</td>
<td>-</td>
<td>PLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[47]</td>
<td>Q</td>
<td>Vinegars</td>
<td>180–890 nm/2 mm/10 mm</td>
<td>-</td>
<td>PLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[49]</td>
<td>Q</td>
<td>Effluent sewage</td>
<td>Direct analysis 190–1100 nm/1 nm/n.d.</td>
<td>SG-S, MSC + SVN</td>
<td>PLS, SVM, BP-NN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[50]</td>
<td>Q</td>
<td>Excipients</td>
<td>Direct analysis of solutions 190–600 nm/0.5 nm/optical fibre</td>
<td>SNV, D1, D2</td>
<td>PCA, PLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[51]</td>
<td>Q</td>
<td>Interaction of iron(III) and tannic acid</td>
<td>Direct analysis of solutions 350–600 nm/10 mm</td>
<td>-</td>
<td>MCR-ALS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[52]</td>
<td>PR</td>
<td>Beer</td>
<td>Direct analysis 190–1100 nm/1 nm/10 mm</td>
<td>-</td>
<td>PCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[53]</td>
<td>PR</td>
<td>Honey</td>
<td>Direct analysis of solutions 190–400 nm/1 nm/10 mm</td>
<td>OFF, LBC, OFF + LBC, 1D, SG-S</td>
<td>PCA, OC-PLS, DD-SIMCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[54]</td>
<td>PR</td>
<td>Mint species</td>
<td>Extracted solutions 190–400 nm/1 nm/10 mm</td>
<td>1D + SG-S + PQN</td>
<td>SIMCA, PLS-DA, SVM, HCA, SIMCA, PLS-DA, PCA, SO-PLS, SO-COvSel, PLS-DA, DF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[55]</td>
<td>PR</td>
<td>Wine vinegars</td>
<td>Dilution 1:10 (v:v) 180–400 nm/2 mm/10 mm</td>
<td>SMTH + MC + SG-1D</td>
<td>PCA, SIMCA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[56]</td>
<td>PR</td>
<td>Fruit</td>
<td>Direct analysis 190–400 nm/1 nm/10 mm</td>
<td>1D + SG-S + PQN</td>
<td>adulthood, PLS-DA, PCA, SO-PLS, SO-COvSel, PLS-DA, DF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[57]</td>
<td>PR</td>
<td>Fish species</td>
<td>Extraction and dilution 1:80 (v:v)</td>
<td>190–400 nm/1 nm/10 mm</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[58]</td>
<td>PR and Q</td>
<td>PR</td>
<td>Vinegar</td>
<td>Direct analysis 200–700 nm/2 mm/2 mm</td>
<td>SNV, MSC, 1D, 2D, SMTH + SNV +1D</td>
<td>LDA, PLS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[59]</td>
<td>Q</td>
<td>Coffee</td>
<td>Extracted solutions 250–400 nm/1 nm/10 mm</td>
<td>SNV, MSC, 1D, 2D, SMTH + SNV +1D</td>
<td>PCA, PLS, MLR, PCR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Ref.</td>
<td>Area</td>
<td>Aim</td>
<td>Sample</td>
<td>UV-Vis Analysis (Range/Resolution/Cuvette Path Length)</td>
<td>Preprocessing Method</td>
<td>Multivariate Method</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>------</td>
<td>------------</td>
<td>------------------------------</td>
<td>---------------------------------</td>
<td>------------------------------------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[60]</td>
<td>Q</td>
<td>Whole wheat</td>
<td>Extraction and dilution 1:80 (v:v)</td>
<td>240–600 nm/5 nm/microplate</td>
<td>-</td>
<td>PCA, PCR, PLS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[62]</td>
<td>PR</td>
<td>Drainage</td>
<td>Direct analysis</td>
<td>220–680 nm/2.5 nm/5 mm</td>
<td>-</td>
<td>PCA, FNN, MD-CNN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[63]</td>
<td>PR</td>
<td>Plant leaves</td>
<td>Extraction and dilution 2.5:10 (v:v)</td>
<td>200–800 nm/0.5 nm/10 mm</td>
<td>SMTH + SNV</td>
<td>PCA, DA, SIMCA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[64]</td>
<td>PR and Q</td>
<td>Spices</td>
<td>Direct analysis of solutions</td>
<td>200–800 nm/2 nm/10 mm</td>
<td>-</td>
<td>PCR, PLS, sPLS-DA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[65]</td>
<td>Q</td>
<td>Phenolics</td>
<td>Direct analysis of solutions</td>
<td>200–420 nm/0.1 nm/n.d.</td>
<td>-</td>
<td>MCR-ALS, PARAFAC</td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td>[67]</td>
<td>PR</td>
<td>Sappanwood</td>
<td>Extraction and dilution 0.25:5 (v:v)</td>
<td>200–800 nm/n.d./10 mm</td>
<td>SG-S</td>
<td>PCA, DA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[68]</td>
<td>PR</td>
<td>Vinegar</td>
<td>Dilution 5 times</td>
<td>200–550 nm/n.d./96-well plate</td>
<td>1D, 2D, 3D, SNV, MSC, OSC, WCTS, WDTS</td>
<td>PLS-DA, OPLS-DA, ANN</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[70]</td>
<td>PR</td>
<td>Pummelo extracts</td>
<td>Extracted solutions</td>
<td>200–600 nm/n.n./10 mm</td>
<td>-</td>
<td>PCA, PLS-DA, sPLS-DA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[71]</td>
<td>PR</td>
<td>Saffron</td>
<td>Dilution 100-fold</td>
<td>200–700 nm/5 nm/96-well plate</td>
<td>MC + PS</td>
<td>PCA, HCA, OPLS-DA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[72]</td>
<td>Q</td>
<td>Carotenoids from fruit</td>
<td>Extraction and dilution 1:10 (v:v)</td>
<td>250–600 nm/0.5 nm/n.d.</td>
<td>-</td>
<td>MCR-ALS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[73]</td>
<td>C, P, and E</td>
<td>Heterogeneous supernatants</td>
<td>Direct analysis of solutions</td>
<td>240–450 nm/1 nm/n.d.</td>
<td>SG-S</td>
<td>PLS</td>
<td></td>
</tr>
</tbody>
</table>

Moreover, as can be observed in Figure 3, among these papers, in general, a similar proportion of them aim at the application of UV-Vis *spectralprinting* for pattern recognition and quantification, with a slightly higher number of papers of pattern recognition. This could be explained by the fact that there UV-Vis uniparametric analysis is still more common or simpler, that is, the quantification of a compound by analyzing only at a specific wavelength and not using the total spectra or, in other cases, the difficulty of use of algorithms such as multiple curve resolution to solve multicomponent systems.

![Figure 3. Percentage of publications about UV-Vis *spectralprint* approaches comparing pattern recognition and quantification applications in the last 6 years.](image)

4.1.1. Signal Preprocessing, Wavelength (Variable) Selection, and Data Dimension Reduction

The first and crucial chemometric step for a reliable spectroscopic data analysis is data preprocessing. It helps to remove the non-useful information by correcting the deviations caused by several different factors such as instrumental drift, light scattering, interferences, etc., (which impact the spectral data quality), and in turn, highlighting the truly useful information. This step will transform the spectrum to the best-fit conditions, ensuring the optimum performance in the following steps of data analysis.

Although there are many different preprocessing methods, the selection of a specific one depends on the nature and characteristics of the data. In particular, one of the advantages of UV-Vis data is that it is a kind of spectroscopic data that often requires little or no data preprocessing. By contrast, data obtained by other spectroscopic techniques, such as the vibrational one, need to be preprocessed almost always, because they are affected by numerous unwanted sources of variability [3].

In fact, it has been seen that UV-Vis has often been found to be very repetitive over time and shows no spectral variation between analysis, which means that it does not need to be preprocessed to correct this problem [37,75–77]. Thus, as can be observed in Figure 4, 52% of the works summarized in Table 1 did not use any preprocessing of the UV-Vis data, and, when comparing the results obtained from the original spectra and those obtained with the preprocessed data, the use of the raw spectra gave better results in many cases [53,58,64]. Thus, the comparison among different preprocessing methods in the same study in order to select the most appropriate one that minimizes the effect of baseline shifts and noise in the spectra is a common practice in the works considered in Table 1.
The first and second derivatives of the original spectra can be useful in qualitative analysis, either for characterizing the number of bands or for identification purposes. Moreover, such a transform has increased complexity of the derivative spectra, which can be useful in qualitative analysis, either for characterizing materials or for identification purposes [80].

However, the transformation by derivatives not only modifies the visual shape of the signal profiles, but it also modifies, often radically, the significance of variables in the data matrix, which could have an important implication in the final quantification or classification results. For this reason, another preprocessing method commonly applied to UV-Vis spectra, by itself or in combination with other preprocessing methods such as the standard normal variate (SNV), is Savitzky–Golay smoothing (Figure 4) [29,63,64,67]. It overcomes noise enhancement, and it has shown, in many cases, an improvement in the quantification or classification results. The first and second derivatives have also been used in mathematical preprocessing for the UV-Vis spectra in order to enlarge the differences between samples [41,43,50,55,58,59,64,78,79]. Thus, the first derivative corrects for baseline shifts, while the second derivative corrects for both shifts and drifts. In addition, when spectra are very similar and UV-Vis is used to confirm the identity of a substance, derived spectra can be used where spectra are highly similar, because the number of bands can increase with higher orders of derivatives. This increased complexity of the derivative spectra can be useful in qualitative analysis, either for characterizing materials or for identification purposes [80].

Following the use of the raw or original spectra, another preprocessing method commonly applied to UV-Vis spectra, by itself or in combination with other preprocessing methods such as the standard normal variate (SNV), is Savitzky–Golay smoothing (Figure 4) [29,63,64,67]. It overcomes noise enhancement, and it has shown, in many cases, an improvement in the quantification or classification results. The first and second derivatives have also been used in mathematical preprocessing for the UV-Vis spectra in order to enlarge the differences between samples [41,43,50,55,58,59,64,78,79]. Thus, the first derivative corrects for baseline shifts, while the second derivative corrects for both shifts and drifts. In addition, when spectra are very similar and UV-Vis is used to confirm the identity of a substance, derived spectra can be used where spectra are highly similar, because the number of bands can increase with higher orders of derivatives. This increased complexity of the derivative spectra can be useful in qualitative analysis, either for characterizing materials or for identification purposes [80].

However, the transformation by derivatives not only modifies the visual shape of the signal profiles, but it also modifies, often radically, the significance of variables in the data matrix, which could have an important implication in the final quantification or classification results. For this reason, another preprocessing method widely applied in UV-Vis data that modifies the spectra but, in this case, to a lower degree, is SNV (as is shown in Table 1 and Figure 4). The SNV algorithm was proposed by Barnes and coworkers in 1989 to jointly fix the baseline vertical shifts’ correction and global intensity effects [81]. It has been used to preprocess UV-Vis spectra prior to mean centering in matrices such as vinegar [12,36,68], corn [59], pharmacological herbals [63], etc., for both quantitative and pattern recognition purposes. Moreover, such a transform has also been extensively used as a preference choice for other types of spectroscopic data, such as vibrational spectroscopy [3]. However, it should be considered that despite SNV normalization fulfilling the aim of removing unwanted variations from spectra, it has been demonstrated that it also causes a loss of information related to original variables. Thus, Oliveri et al. (2019) demonstrated how SNV-transformed UV-Vis spectra showed misleading results by presenting irrelevant loading values in principal component analysis

Figure 4. Percentage of studies in the last 6 years that applied each preprocessing method for UV-Vis data. Note: SMTH: Savitzky–Golay smoothing; 1D/2D: first and second derivative; SNV: standard normal variate; MSC: multiple scatter correction; CWT: continuous wavelet transforms; OSC: orthogonal signal correction.
for bands that were really related to spectral regions’ characteristics for this problem, and, conversely, bands that were not actually relevant in the samples were the ones that made the difference after the transformation [77]. Moreover, it must be remarked that most normalization strategies are affected by the same problems. Finally, there are other relatively newer preprocessing methods such as multiplicative scatter correction (MSC), probabilistic quotient normalization (PQN), continuous wavelet transforms (CWT), and orthogonal signal correction (OSC) that are still less used (Figure 4), appearing only in a few studies considered in this review (Table 1).

To summarize, the application of preprocessing can increase the model accuracy and repeatability, although there is no guarantee that it will actually work. This means that sometimes it will be necessary and sometimes it will not. In fact, how preprocessing affects the UV-Vis data is highly dependent on the relative intensities and correlations between spectral variables and is not always easy to detect, especially in spectral printing signals. As in other spectroscopies, the correct preprocessing method will be the one that enables the best quantification or classification results according to the specific spectral problems of the dataset.

4.1.2. Exploratory and Pattern Recognition Approaches

The relationship between samples and variables in a UV-Vis dataset is revealed by using suitable tools of multivariate analysis or chemometrics. Because spectral data such as UV-Vis are multivariate and high-dimensional, the following step before preprocessing is usually to reduce the dimensionality of the data and, in turn, extract the important characteristics included in it, and, then, data are fed to classification analyses (i.e., pattern recognition analyses) [82].

One of the most widespread multivariate methods applied in spectral analysis of any kind of spectroscopic data for this exploration and data reduction is principal component analysis (PCA). PCA is commonly used as a first step in spectroscopic data analysis in order to explore the data, finding possible similarities and differences among samples and identifying clusters or patterns. In addition, it is also implemented to reduce the dimensionality of the spectral data to a smaller number of components, determining which variables are important to represent the system, facilitating the subsequent analysis, and reducing the risk of incorrect interferences. It also helps to detect outliers [3]. PCA is a multivariate method applied to convert the spectral measures (i.e., correlated data) into linearly uncorrelated variables that describe or predict meaningful patterns from complex spectral fingerprints. Since the UV-Vis spectrum is not often very informative by itself, i.e., by its direct visualization, PCA is a very useful method due to the fact that it also helps to determine important wavelengths that have contributed to a further discrimination of samples.

In pattern recognition, exploratory data analysis is usually performed first, followed by the classification process. This can also be seen in Table 1 and Figure 5a, showing that PCA was applied in almost all the UV-Vis studies considered in this review, many times as a first step prior to a classification approach and, in other many cases, as the unique chemometric method applied in the study. This is explained by the fact that although most classification algorithms can handle an enormous quantity of data, their efficiency decreases as dimensionality grows, so exploratory analysis such as PCA can sometimes increase it [82]. Moreover, based on the trend over the last 6 years, Figure 5a showed that its use has increased year over year, especially in the last two years.
Table 1

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>5</td>
</tr>
<tr>
<td>2018</td>
<td>2</td>
</tr>
<tr>
<td>2019</td>
<td>7</td>
</tr>
<tr>
<td>2020</td>
<td>5</td>
</tr>
<tr>
<td>2021</td>
<td>4</td>
</tr>
<tr>
<td>2022</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 5.** Number of papers per year from the last 6 years that used each algorithm considered for (a) pattern recognition and (b) quantitative approaches.

Thus, once the spectral data have been cleaned up by preprocessing or the key characteristics have been retrieved, they are ready to be modeled by an appropriate pattern recognition technique. Pattern recognition develops classification models in which a sample with an unspecified class is assigned a predefined class based on its measured characteristics, or samples are grouped into clusters based on their similarities, even without any previously defined class [82]. In the specific case of spectralprint techniques such as UV-Vis, each category or class is assigned using the collective information from a group of samples with common spectral characteristics, which distinguishes them from any other set of samples.

In this context, several papers on pattern recognition in spectral data have been published in the last decade. However, many of them were concentrated on infrared (IR) spectroscopies [83–86], while UV-Vis spectroscopy has received less attention until the last six years, when it has seemed to increase [82]. Thus, Figure 5a shows the trend of use of the different pattern recognition algorithms in the last 6 years in UV-Vis spectralprinting. Several classification algorithms have been applied, with the most commonly used being the partial least squares–discriminant analysis (PLS-DA) and linear discriminant analysis (LDA), followed by soft independent modeling by class analogy (SIMCA) (Table 1 and Figure 5a).

On the one hand, PLS-DA was the one that showed an upward growth up to now, with a slight decline in 2020 that also occurred with the other ones. It was originally proposed for multivariate calibration, and then it was used to solve classification problems by combining the features of PLS regression with the power of a discrimination technique. PLS-DA mathematical bases have been widely reported in the literature [87,88]. To sum up, the data matrix contains the independent variables, while the categories form the dependent variable, with each class of samples coded numerically as integers (usually zeros and ones). PLS-DA provides scores that explain sample location in each known latent variable. The predicted values for the unknown samples are then also numerical values between zero and one, which will be assigned to a specific class according to a defined and optimized threshold. Moreover, in order to improve the model performance, their orthogonal and sparse variants, OPLS-DA and sPLS-DA, respectively, were recently reported [64,68,70].

The advantage of this technique over conventional DA or LDA is that it can be used whether the number of samples is small, or the number of variables is large. In fact,
it should be noticed that all the papers in which LDA is used have combined it with PCA [18,21,39], as it is necessary to reduce the number of variables to be able to apply LDA. Despite this disadvantage, which is easily and successfully corrected by combining PCA-LDA, LDA has the advantage of minimizing factors that have little effect on group differences. Thus, it reduces the dimensions of the dataset from a large number of main variables to a small number of canonical functions while retaining pertinent information from the original dataset [82]. These new variables are created by linear combinations of the original ones, so, considering two categories, the delimiter provided by LDA is a linear function, which creates a straight line or a plane in the case of two or three independent variables, respectively, and so on.

Finally, SIMCA, also used in UV-Vis pattern recognition [19,36,38], is a supervised class modeling method that is based on this case in PCA, which has the advantage of being fast-to-compute. In this algorithm, each class is modeled by a separate PCA and is based on the similarity of the samples within each class. The main disadvantages are that the number of principal components in each model must be perfectly defined and that it is quite common to classify a sample into different classes or none at all, giving poor results when there are little differences across classes.

All these aforementioned methods are also the most commonly used in other spectroscopic techniques in the field of spectral printing [3,82].

On the other hand, as can be seen from most of the papers listed in Table 1, most of them focus more on the applications than on the development or testing of new algorithms for classification. Furthermore, looking at Table 1, it can be seen that a comparison of different pattern recognition algorithms for the same target is not usually made, unlike UV-Vis preprocessing tools that are often compared to select the most suitable one. This does not mean that it is always necessary to explore and perform new or multiple pattern recognition algorithms, due to the fact that if the selected one or the common one work perfectly, it will be enough.

Finally, it is curious that no research in the last 6 years was found that used other commonly applied and extremely simple methods, such as k-nearest neighbors (KNN). This algorithm is simply based on distance between samples. In addition, researchers that use other methods, such as support vector machine (SVM), random forest (RF) or even deep learning methods, are still scarce in the literature from the last 6. The simplest explanation could be that this technique does not need these more sophisticated algorithms to provide satisfactory results.

4.1.3. Quantitative Approaches

The quantification of specific compounds contained in a sample is a routine operation in many laboratories and industries. To fulfill this goal, UV-Vis spectroscopic data have also been used for quantitative proposes by combining spectral measurements with chemometric regression methods. This group is formed by multivariate techniques that seek a relationship between the analytical signal (full/preprocessed spectra or a group of selected wavelengths) and some properties of the sample.

As was discussed in a previous section, preprocessing analysis is a first and fundamental step in chemometric data processing. Then, measured spectral responses can be correlated to the quality traits of interest, such as grape-must caramel in high-quality wine and balsamic vinegars [47], or more common properties such as soluble solids content, dry matter content, or acidity, among others [89]. Thus, a mathematical function between the predictor or independent variables (UV-Vis spectra) and the predicted or dependent variables (traits) is sought to predict the traits of interest from the acquired spectra. For this function, the parameters are estimated from a calibration/training set of samples for which both spectra and the quality traits of interest measured with reference methods were acquired.

This calibration stage presents several challenges, such as overcoming non selectivity, i.e., broad and overlapping absorption or emission peaks. In this context, multivariate data
analysis techniques can solve this problem because they can be used to extract and combine the relevant information contained by multiple variables. In addition, due to the broad and overlapping peaks, the situation can occur where many spectral variables are related to the same absorption peaks and to different absorption peaks for the same component, with the information contained by different variables being very similar. This can complicate the estimation of the multivariate calibration model (known as the collinearity problem) [90].

On the other hand, to obtain a calibration model that will perform well in future samples, careful calibration data selection should be performed. In this stage, calibration samples should encompass the natural range of variability and possible combinations of influencing factors that can be expected in the future. Furthermore, outliers should be treated with care, as on the one hand, samples from outside the target population can have a negative impact on the calibration model, but on the other hand, extreme samples from the target population can be very informative in the model building phase.

Despite most multivariate calibration methods assume a linear relation between the independent and dependent variables, this situation should be corroborated. If the relation is nonlinear, the first recommendation is to linearize the relationship by preprocessing the variables (e.g., logarithmic transform to linearize the exponential relation between absorbance and concentration according to the Beér–Lambert law). Another possibility is to use a nonlinear model instead of a linear model. This makes the calibration with this spectroscopic sensor a challenge, requiring considerable care to avoid overoptimistic results that cannot be reproduced for future samples.

Quantitative approaches using UV-Vis spectroscopy as a spectralprint sensor have been scarcely reported in the last six years, as can be observed in Figure 5b and Table 1. Only a few works applying regression methods have been combined with this spectroscopy technique in order to develop rapid and nondestructive methodologies for the quantification of ingredients in some samples. The reason could also be that many of the works still use one specific wavelength to perform the quantification of a compound.

In this regard, only linear approaches have been considered. Consequently, the ordinary least-squares approach is the most straightforward way to estimate model parameters, and the corresponding method is called multiple linear regression (MLR), although it is not the preferred choice for the spectral sensors’ calibration. Generally, the multicollinearity problem is presented in the spectral matrix since the number of predictor variables may be larger than the number of samples making the covariance matrix singular and noninvertible, which prevents one from finding a unique solution from least-squares estimation [91]. This is similar to the aforementioned problem of directly applying LDA to UV-Vis spectra in pattern recognition, which both can be overcome by increasing the number of samples in the calibration set or by selecting a subset of variables. However, with this solution, the prediction performance may still be very poor, owing to the high correlation among the spectral variables. As can be observed in Figure 5b and Table 1, only one work using several variable selection algorithms has been reported since 2017. Due to these limitations, in the case that standard MLR is not appropriate to build regression models, alternative approaches have been proposed in the literature. One of them has been principal component regression (PCR), which involves two stages to achieve the calibration model. First, PCA is used to compress the data matrix information onto a reduced set of relevant scores. Then, these scores are used as the predictor matrix to perform a multiple linear regression. PCRs have been used very infrequently, probably owing to drawback that this algorithm has to face. The criterion that guides the development of the regression coefficients is not the same than the one to extract the scores from the data matrix. Thus, the directions of maximum explained variance may not be relevant for the prediction of the dependent variables, especially when there are many uninformative sources of variability in the data [92]. To overcome this disadvantage, the partial least squares algorithm (PLS) is presented as an alternative approach to component-based regression. It is probably the most widely used calibration method in chemometrics and has been the most applied method on UV-Vis spectral data (Figure 5b and Table 1). PLS regression was developed as
an alternative method to be used to calculate reliable regression models with ill-conditioned matrices. As occurs with PCR, it is based on the extraction of a scores set by projecting the data onto a subspace of latent variables which are relevant to solve the calibration problem. In this case, for the definition of the scores, it is explicitly considered that the components not only explain a significant part of the variance of the data, but also that they are predictive of the response. In fact, in PLS, the directions onto which the data are projected (i.e., latent variables) are defined in such a way that the covariance between the corresponding scores and the responses are maximized, which allows one to obtain scores that both describe a significant part of the variance of the data and are correlated with the responses [92,93].

At this point, it should be emphasized that to achieve accurate and reliable estimations by a PLS (or PCR) model, the selection of an appropriate number of latent variables to describe the data is a crucial step. In fact, if a low number of components are selected, it can an underfitting can occur, which would not explain all the relevant variance of the data. Otherwise, if too many are captured, overfitting can occur, resulting in a model that is very good at predicting the samples it was calculated on but that performs poorly with the new ones. To reduce this risk, the optimal number of latent variables is selected by an appropriate validation strategy, which leads to the minimum error during the validation steps (usually cross-validation). This proper validation has to also be applied in pattern recognition approaches, such as, for example, in PLS-DA models, as model overfitting can also occur.

Other methods that have been used mainly for a quantitative approach that differ from the aforementioned ones are the curve resolution methods, such as multivariate curve resolution–alternating least square (MCR-ALS) or parallel factor analysis (PARAFAC). They are two-way and three-way data analysis methods, respectively, useful and popular for solving mixed analysis problems, such as peaks overlap, by matrix decomposition [65,66]. Thus, they are subjected to the decomposition of a higher-order dataset into a set of components of analyzed systems for quantitation, classification, or characterization, among other aims. As is well-known, the UV-Vis spectrum provides a vector of values per sample (first-order data), so the critical stage here is the data generation. Consequently, in order to acquire second-order data, it is mandatory to perform some modifications in the measurement conditions. Thus, a vast number of works have carried out pH variation on samples by obtaining pH-dependent spectra, generating one more mode [51,65,66]. Moreover, different temperature conditions, storage times, and oxidative process control of a specific compound have been other strategies reported for such purpose [30,74]. In general terms, the general adopted strategies have consisted in the data collection (e.g., matrices of the pH-ultraviolet absorbance dataset), which have been arranged in a three-way array (cube) and then decomposed by PARAFAC or augmented in a two-way data matrix and then decomposed by MCR-ALS.

MCR-ALS is a method based on a bilinear model that assumes that the observed spectra are a linear combination of spectra of pure components [94]. It aims to achieve a bilinear decomposition of the experimental absorbance matrix D (m,n) into two new matrices, one, C (m,k), with the concentrations of the individual compounds, and the other, ST (k,m), with their normalized spectra. For this decomposition, the number of components k should be determined or estimated by rank analysis methods, such as the singular value decomposition (SVD) [94]. Then, when the initial estimate of C and ST matrices have as many profiles as the number of components determined (k), an iterative resolution process using the ALS algorithm is performed until the error is minimum and the variation of results between consecutive iterations goes below a preset threshold value. Constraints are added to improve the results, with the most common in UV-Vis data being the non-negativity, which can be applied to both the pure concentration profiles (i.e., no negative concentrations) and the pure spectra (no negative signals). Other constraints such as unimodality (useful for letting the reconstituted signals to have only one maximum) and
normalization (useful for normalizing the pure spectra or the concentration profiles of the components to a reference value) can be also applied [94].

This method could be considered to be an alternative of PLS when the information of possible interferences is limited as well as when there is a reduced calibration set [25]. It also has the advantage of providing the pure spectrum of the absorbance response of overlapping analytical components, which is useful in the study of molecule interactions [26,51,66] or in the simultaneous determination and quantification of different compounds in a mixture [27,34,42,72].

On the other hand, PARAFAC is a trilinear model, that is, it is used for the chemometric resolution of three-way data arrays, which is also used in UV-Vis data. However, it is still less applied in UV-Vis data than MCR-ALS, as can be seen in Table 1 [65,66], being more commonly applied in fluorescence excitation-emission data according to the literature since this type of data fulfill the trilinear property, which is often broken in the mode in which the second-order UV-Vis spectroscopic data are generated. It decomposes the complex raw datasets into pure component profiles of samples as MCR-ALS but, in this case, from a three-way data matrix. Thus, it decomposed a data matrix X into three new matrices, Aif, Bjf, and Ckf and Eijk, related to the error data array. In this case, the appropriate selection of the initial parameters and restrictions (e.g., non-negativity) are also needed.

5. Applications for Spectralprint (Nontargeted) Analysis

UV-Vis has demonstrated to be a versatile technique that has been applied to a wide range of fields, from initial uniparametric use to today’s more multiparametric use, i.e., the spectral print. In this section, only the fingerprint application in different matrices and with different aims will be discussed.

5.1. Agriculture, Food, and Beverages

UV-Vis spectralprinting is an easy-to-adopt, cheap, and fast approach, useful for discrimination and classification of characteristics and quality of food products. In fact, the use of a spectrophotometric technique such as UV-Vis for discriminating food products could be of special interest to the industry, as most laboratories have a UV-Vis spectrophotometer for other routine analyses. Moreover, the possibility of using a portable UV-Vis device is also of great benefit for the food industry and control agencies. Moreover, due to its simplicity and reliability, UV-Vis has already been used in several food science and food processing areas [10,11]. Thus, UV-Vis absorption spectroscopy has been extensively implemented to analyze a wide range of food samples, such as beverages, dairy products, processed foods, oils, wines and vinegars, spices, honey, fruits, and vegetables, among others (Table 1). However, despite all this, it is still not well-recognized as a technique to discriminate or authenticate food samples.

For instance, in the case of wines and vinegars, it has been seen that UV-Vis spectralprint has high relevance due to its ability to analyze relevant compounds in wine, such as phenolic compounds. They present distinct UV fingerprints (i.e., they present π bonding and conjugated double bonds on which UV relies), while ignoring the most abundant components of wine, such as water, alcohol, organic acids, and sugars, because they do not have any absorbance in the used UV wavelength range (200–600 nm) [28]. Thus, it has demonstrated to be an effective tool in combination with chemometrics to wine authentication and discrimination as well as for quantification of these wine compounds [28,46,95,96]. Furthermore, its usefulness in wine discrimination according to grape variety has been compared with other spectroscopic techniques, such as FT-IR, demonstrating that UV-Vis spectroscopic techniques worked better, possibly due to the fact that the differences between wine varieties could be attributed to the colored phenolic compounds that absorb in the UV-Vis region [37]. Another study showed a comparison again between UV-Vis and FT-IR spectroscopic data coupled to chemometrics, showing that they both have, in general, similar success rates in determining the adulteration of vinegar with specific adulterants [68]. Moreover, UV-Vis has recently shown satisfactory results in the classifi-
cation of wine vinegars by the use of hierarchical classification models with PLS-DA and SIMCA algorithms by using a UV-Vis portable device [36], with a quality control tool for high-quality wine vinegars even being proposed for development [12].

The comparison of UV-Vis with another spectroscopic technique has been also commonly addressed [46, 54, 68], and many times UV-Vis has given better results, such as is shown for example in the work of Mannu et al. (2022). In this study, although both FT-IR and UV-Vis in reflectance mode analyses were able to provide a good differentiation between waste and edible oils, the UV-Vis results were of particular interest, as it could be conducted in a portable device that could have an important practical application in the field of waste vegetable oils management [69]. There are also works that, in addition to comparing UV-Vis with other techniques, also test the combination of them by data fusion, which results in an improvement in the results obtained in the majority of the cases [29, 41, 55]. Finally, there is also the possibility of obtaining a second-order data matrix with UV-Vis data that, combined with MCR-ALS, can be used for, for example, monitoring the oxidative stability of extra virgin olive oils [30].

5.2. Chemical, Pharmaceutical, and Environmental Sciences

The use of UV-Vis mainly in the control of process, kinetics, and quantification of substances is very common in the pharmaceutical and chemical industry. This can be observed in Table 1, where almost all the studies found in the literature from 2017 regarding this area address the quantification and control of processes. Hence, UV-Vis spectral print has shown to have a very prominent position in the online analysis and monitoring of complex processes.

Thus, regarding kinetics and control processes, UV-Vis coupled to PLS has been applied to speed up process development for high-throughput selective protein crystallization screenings, which may encourage alternative process development to well-established chromatography-based processes [73]. This combination of UV-Vis and PLS has also been used in the in situ monitoring of organosolv pretreatments [32], as well as to evaluate dissolution kinetics via batch tests [44], which is essential for an adaption of process parameters and a constant product quality. This on-line monitoring is incompatible with traditional wet chemistry methods, such as HPLC, since they are expensive and time-consuming. In fact, the long time required for analysis could limit their use for monitoring in-process component concentrations.

Furthermore, using the UV-Vis spectral print coupled to multivariate PLS regression has shown the availability to determine, at the same time and quickly, several products that could not be addressed by using data only from a single wavelength. Thus, regarding the quantification approach, the combination of UV-Vis and PLS has been also used, for example, in the control of water samples [79], as an effective solution for the determination of traces of mixed organic acids in aqueous solutions [31], or for the simultaneous quantifications of propanil and bromoxynil herbicides [24]. This combination has demonstrated to be very useful for situations of simultaneously determining multicomponents, which is difficult to address with the traditional UV-Vis approach.

Moreover, new approaches that applied MCR-ALS or PARAFAC showed successful results in the simultaneous determination and quantification of multiple components in a complex matrix and even the possibility to study the interaction between different molecules or compounds by UV-Vis spectroscopy [27, 34, 42, 65, 66, 72].

Table 1 also showed that there are also a few pharmaceutical, chemical, or environmental studies whose aim was to discriminate or differentiate samples (i.e., pattern recognition) by UV-Vis spectral print. Hence, the utility of this tool, for example, in the detection and identification of narcotic drugs in urine has been seen [48] and for drug differentiation or adulteration [63].
6. Future Perspectives and Final Remarks

In this work, the most recurrent studies using UV-Vis spectroscopy as spectralprint analysis were compiled. This assessment was based on the absorption of compounds in the UV-Vis region and with the improvement of measurements using multivariate techniques, signal-preprocessing algorithms, or both at the same time. The main aim was to clarify, regarding the fundamental uses, the existing bottlenecks for the implementation and acceptance of all possible ways of using UV-Vis spectral data allied to chemometrics. Chemometrics provide a wealth of techniques for exploratory analysis of multivariate data, as well as for creating reliable calibration models capable of predicting quantitative responses, as well as for developing classification strategies for predicting qualitative responses, all based on the experimental profiles collected from the samples.

Evidently, spectralprint represents an indispensable and highly versatile sensor for analysis at all fields, whether qualitatively or quantitatively, and is an approach of great potential, as explained in the works described here. In combination with the ease of handling, the instrument is relatively easy to access in research laboratories, and, with the development of more robust analytical methods, the results obtained are increasingly satisfactory. Concomitant to these applications, the diffusion of multivariate statistics, as well as the continued implementation of increasingly sophisticated computational tools, has brought numerous gains to the use of UV-Vis spectroscopy as a sensor to predict chemical compounds. For this reason, the application of multivariate calibration models to predict compounds has numerous applications, and the determination of quality parameters using this method should become a recurrent practice. The gains related to the use of this methodology for a diverse area are reflected both in terms of measuring the variables of interest and in the potential decrease in time, cost, and waste generated, which is often toxic.

The development of portable spectroscopic devices can meet the demand of future industrial trends, becoming an important tool for control analysis due to their several advantages. Thus, they are devices with a considerably low price, which allows their acquisition by untrained analysts; they have a small size for easy movement and handling; they are robust, thus being useful at all stages of the supply chain; they are easy to apply with minimum or no sample preparation, preventing sample damage and allowing more samples to be scanned; and they provide high precision analysis, being sensitive and selective. In addition to the characteristics mentioned above, portable UV-Vis devices are quite manageable (i.e., the size of a cell phone) and user-friendly, with interfaces that allow quick understanding and results in a short time, reducing computational processing time. Moreover, they can be used in a variety of samples.

The continuous improvements in hardware and software designs suggest that, in the near future, UV-Vis spectralprint techniques can respond effectively to any demand, directly in situ and in real-time, being performed as routine methods for process control and monitoring. Therefore, the UV-Vis spectralprint application could continue growing toward new horizons, inquiring in novel proposals regarding data acquisition, with the possibility to acquire higher-order data and data analysis with the opportunity to be applied to more complex algorithms, such as those related to deep learning.

Author Contributions: Conceptualization, investigation, writing—original draft preparation, and funding acquisition, R.R.-R. Conceptualization, investigation, writing—review and editing, supervision, and funding acquisition, S.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Scientific and Technological Research Projects (Project-PICT 2018-04496), and “Fondo Europeo de Desarrollo Regional (FEDER)” and “Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía, Programa Operativo FEDER 2014-2020” under the project “US-1380830 US/JUNTA/FEDER, UE”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.
Data Availability Statement: Not applicable.

Acknowledgments: The authors express their gratitude to CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica) from Argentina, and “Fondo Europeo de Desarrollo Regional (FEDER)” and “Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía, Programa Operativo FEDER 2014-2020”.

Conflicts of Interest: The authors declare no conflict of interest.

References


49. Li, P.; Qu, J.; He, Y.; Bo, Z.; Pei, M. Global Calibration Model of UV-Vis Spectroscopy for COD Estimation in the Effluent of Rural Sewage Treatment Facilities. RSC Adv. 2020, 10, 20691–20700. [CrossRef] [PubMed]


51. Berto, S.; Alladio, E. Application of Chemometrics Tools to the Study of the Fe(III)–Tannic Acid Interaction. Front. Chem. 2020, 8, 614171. [CrossRef]


57. Chai, Z.; Wang, C.; Bi, H. Rapid Identification between Two Fish Species Using UV-Vis Spectroscopy for Substitution Detection. Molecules 2021, 26, 6529. [CrossRef] [PubMed]


59. Yulia, M.; Suhandy, D. Quantification of Corn Adulteration in Wet and Dry-Processed Peaberry Ground Roasted Coffees by UV–Vis Spectroscopy and Chemometrics. Molecules 2021, 26, 6091. [CrossRef] [PubMed]

60. Tian, W.; Chen, G.; Gui, Y.; Zhang, G.; Li, Y. Rapid Quantification of Total Phenolics and Ferulic Acid in Whole Wheat Using UV-Vis Spectrophotometry. Food Control 2021, 123, 107691. [CrossRef]


70. Petretto, G.L.; Di Pietro, M.E.; Piromdi, M.; Pintore, G.; Mannu, A. Classification of Pummelo (Citrus grandis) Extracts through UV-VIS-Based Chemical Fingerprint. Beverages 2022, 8, 34. [CrossRef]
