

Analysis of UV spectral bands using multidimensional scaling

J. A. Tenreiro Machado · Erdal Dinç ·
Dumitru Baleanu

Abstract This study describes the change of the ultraviolet spectral bands starting from 0.1 to 5.0 nm slit width in the spectral range of 200–400 nm. The analysis of the spectral bands is carried out by using the multidimensional scaling (MDS) approach to reach the latent spectral background. This approach indicates that 0.1 nm slit width gives higher-order noise together with better spectral details. Thus, 5.0 nm slit width possesses the higher peak amplitude and lower-order noise together with poor spectral details. In the above-mentioned conditions, the main problem is to find the relationship between the spectral band properties and the slit width. For this aim, the MDS tool is used to recognize the hidden information of the ultraviolet spectra of sildenafil citrate by using a Shimadzu UV–VIS 2550, which is in the world the best double monochromator instrument. In this study, the proposed mathematical approach gives the rich findings for the efficient use of the spectrophotometer in the qualitative and quantitative studies.

Keywords Spectral bands · Multidimensional scaling ·
Visualization · Quantitative analysis

the slit width on the spectral bands are very important.

1 Introduction

Ultraviolet–visible spectroscopy (UV–VIS) region for the electromagnetic radiation includes the range of 100–800 nm corresponding to the electronic transition after the radiation of a molecule.

Ultraviolet–visible spectroscopy (UV–VIS) spectrometer is a very important tool for the analysis of the chemical compounds having chromophore groups providing the electronic transition from background energy state to an excited energy state. In addition, UV–VIS spectrophotometry has been widely used in qualitative and quantitative evaluation of the content of the samples due to the providing rapid and repeatable spectral registration. In this domain, various spectrophotometers have been manufactured at the different marks, e.g., Shimadzu UV–VIS2550.

A specified spectrometer possesses a spectral bandwidth that characterizes how monochromatic the light is. As a result, if this bandwidth is comparable to the width of the absorption features, then the measured extinction coefficient will be altered. In practical measurements, the instrument bandwidth is kept below the width of the spectral lines. When a new material is about to be measured, we have to test and verify whether the bandwidth is sufficiently narrow. On the other hand, it is known that the effects of the changes of

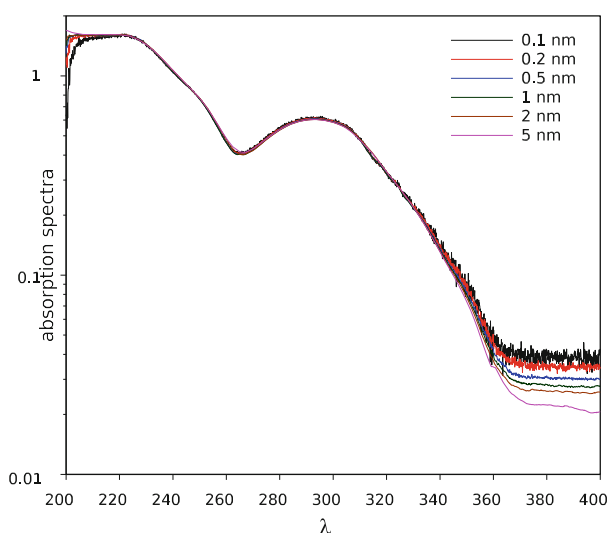


Fig. 1 The absorption spectra of sildenafil citrate (25 µg/ml) at slit widths {0.1, 0.2, 0.5, 1.0, 2.0, 5.0} nm

In our case, the UV absorption spectra of the sildenafil citrate (SC) at the constant concentration level (25 µg/ml) were recorded between 200 and 400 nm at the different slit width set {0.1, 0.2, 0.5, 1.0, 2.0, 5.0} nm.

SC represents a drug for the treatment of the erectile dysfunction and pulmonary arterial hypertension [5,8,11]. SC is a white to off-white crystalline powder possessing a solubility of 3.5 mg/ml in water as well as a molecular weight of 666.7.

Multidimensional scaling (MDS) is a technique for visualization information in the perspective of exploring similarities in data [1,3,7,9,10,13,17,18]. MDS arranges points in a space with a given number of dimensions, so as to reproduce the similarities observed in the measurements. Often, instead of similarities are considered dissimilarities, or distances, between the objects. For two or three dimensions, the resulting locations may be displayed in a “map”. If we rotate or translate the chart, the similarities between items remain the same. Therefore, the center of the portrait and the final orientation of axes in space are mostly the result of a subjective decision by the researcher, and the analysis of the “map” must be in the perspective of comparing which points are close and which are distant.

For the evaluation of the spectrophotometer performance, we obtained six different UV for each slit width. In this application, the main aim is to find the higher spectral signal/noise ratio for the registrations of the spectral bands. In other words, slit width gives us better spectral resolution.

The manuscript is organized as follows. Section 2 describes the experimental setup. Section 3 deals with the theoretical aspects of MDS. Section 4 is devoted to the

definition of the comparison measures. Section 5 presents the results of several visualization tools, namely dendograms and MDS maps. Finally, Sect. 6 outlines the main conclusions.

provides 2,000 points with a step size of $\Delta\lambda = 0.1$ nm.

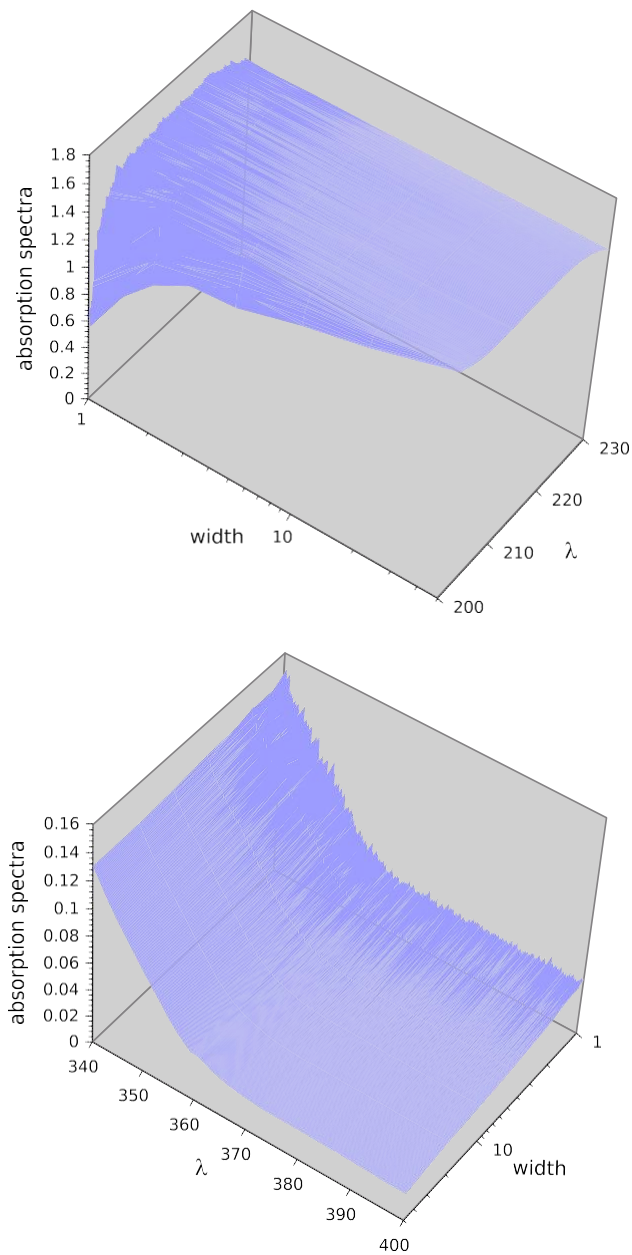


Fig. 2 Details of the absorption spectra of sildenafil citrate (25 μg/ml) versus λ and slit width at the lower and upper limits of the domain

2 Experimental setup

A Shimadzu UV-2550 UV-VIS spectrophotometer connected to computer having Shimadzu UVProbe 2.32 software and a HP Laser Jet P1102 printer were used for the registration of the absorption spectra.

In our experiment, we keep the concentration of SC constant at 25 μg/ml, we consider six slit widths {0.1, 0.2, 0.5, 1.0, 2.0, and 5.0} nm, and, for each slit width, we repeat the experiment five times. In each case, the Shimadzu UV-2550 UV-VIS spectrophotometer

Fig. 3 Dendrogram of the 30 measurements based on the angular distance between spectra (2)–(3)

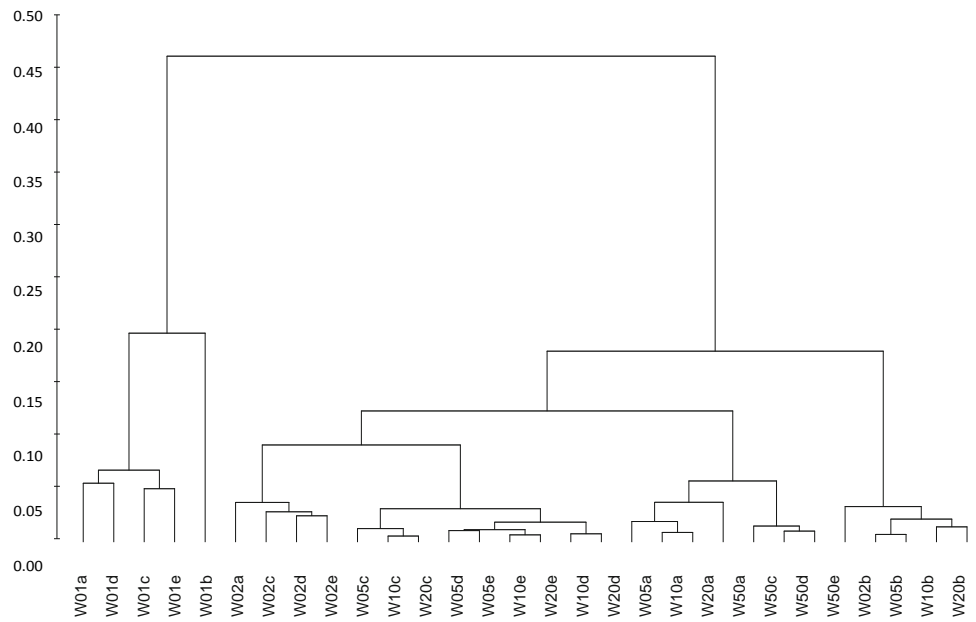
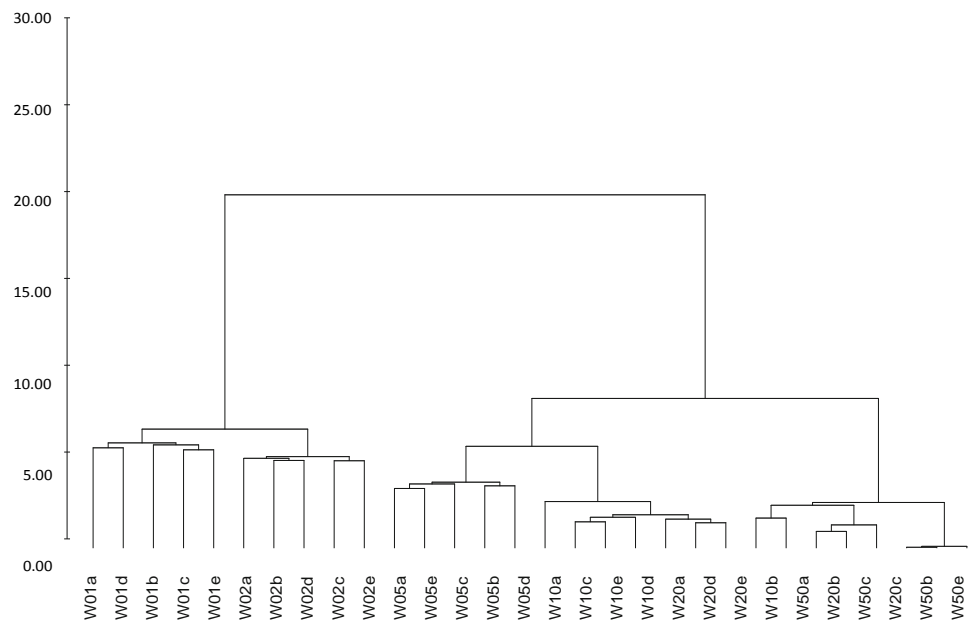


Fig. 4 Dendrogram of the 30 measurements based on the Lorentzian distance and the Fourier transform (5)–(6)



For the purpose of the analysis with the MDS method, we organize the set of measurable data in a 6×5 matrix \mathbf{A} :

$$\mathbf{A} = \begin{bmatrix} W_{01a} & W_{01b} & W_{01c} & W_{01d} & W_{01e} \\ W_{02a} & W_{02b} & W_{02c} & W_{02d} & W_{02e} \\ W_{05a} & W_{05b} & W_{05c} & W_{05d} & W_{05e} \\ W_{10a} & W_{10b} & W_{10c} & W_{10d} & W_{10e} \\ W_{20a} & W_{20b} & W_{20c} & W_{20d} & W_{20e} \\ W_{50a} & W_{50b} & W_{50c} & W_{50d} & W_{50e} \end{bmatrix} \quad (1)$$

where the subscripts {01,02,05,10,20,50} and {a, b, c, d, e} denote the six widths and the five measurements, respectively. In other words, \mathbf{A} denotes the absorbance measurements at different slit widths in the spectral

the middle range of λ revealing only differences at the low and high values, where some noise occurs. Figure 2 shows details of the absorption spectra versus λ and slit width for those two cases. Due to the presence of noise, it is considered the median of the absorption spectra of the five distinct measurements.

region.

The graph of the collected spectra is depicted in Fig. 1. We observe that the spectra superimposed considerably over

3 Multidimensional scaling

An MDS algorithm starts by defining a measure of similarity (or, alternatively, of distance), for constructing a square matrix of item-to-item similarities. In classical MDS, the matrix is symmetric and its main diagonal is composed of "1" for similarities (or "0" for distances). MDS tries to rearrange

Fig. 5 Three-dimensional MDS map of the 30 measurements based on the angular distance between spectra (2)–(3) at slit widths {0.1, 0.2, 0.5, 1.0, 2.0, 5.0}nm

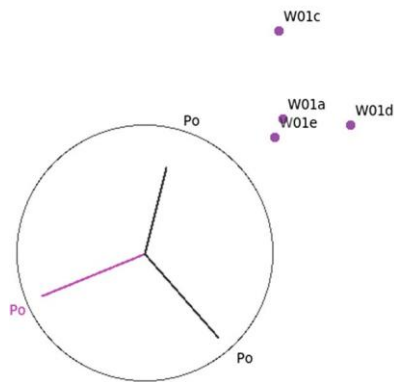
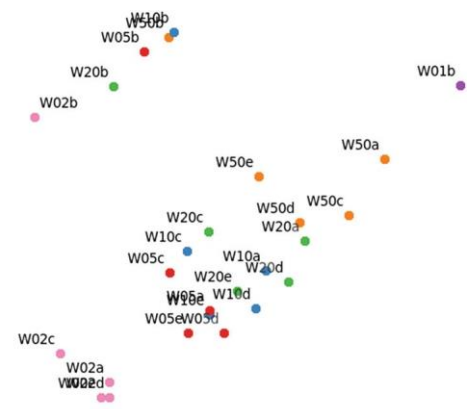
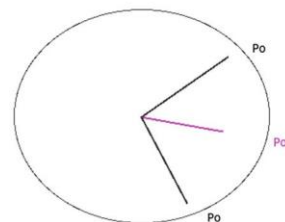
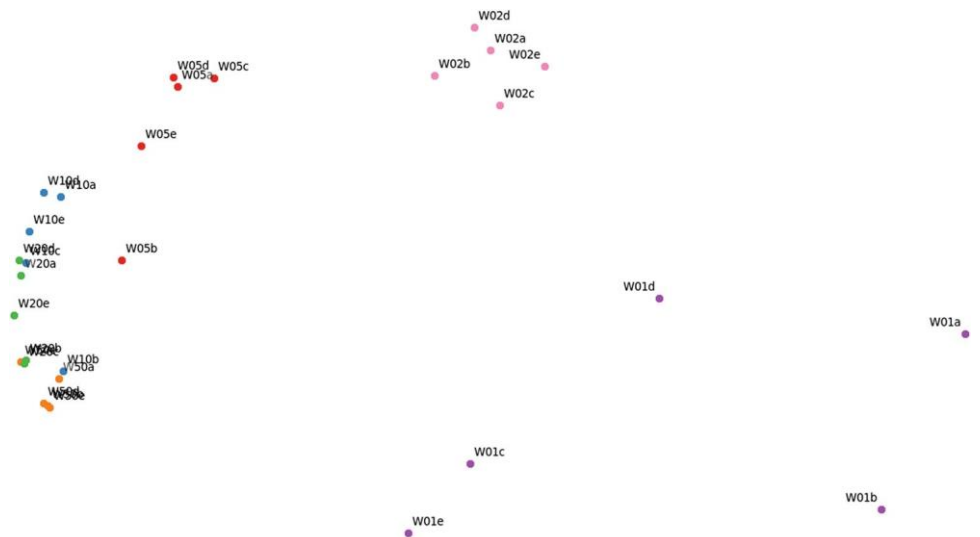


Fig. 6 Three-dimensional MDS map of the 30 measurements based on the Lorentzian distance with the Fourier transform (5)–(6) and slit widths {0.1, 0.2, 0.5, 1.0, 2.0, 5.0}nm



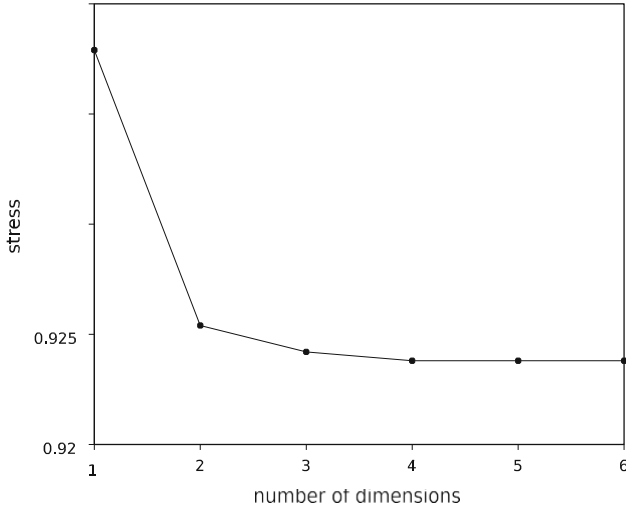


Fig. 7 Stress versus number of dimensions of the MDS map for the angular distance between spectra (2)–(3)

the points in the map so as to arrive at a configuration that best approximates the observed similarities (or observed distances). For this purpose, MDS uses a function minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit. The most common measure that is used to evaluate how well a particular configuration reproduces the observed distance matrix is

raw stress measure defined by $S = [d_{jk} - f(\delta_{jk})]^2$, where d_{jk} and δ_{jk} represent the reproduced distances (for a given number of map dimension) and the input data, respectively. The expression $f(\delta_{jk})$ corresponds to a monotone transformation of the input data. One measure commonly used is the sum of squared deviations of observed from the reproduced distances. Consequently, the smaller the stress value S , the better the fit.

Plotting S versus the number of map dimensions usually leads to a monotonic decreasing curve. The “best dimension” is a compromise between stress reduction and number of required dimension for the representation. In practical terms, we chose a low dimension at the point where there is no significant further reduction of S . Alternatively, plotting the reproduced distances (for a given number of dimensions)

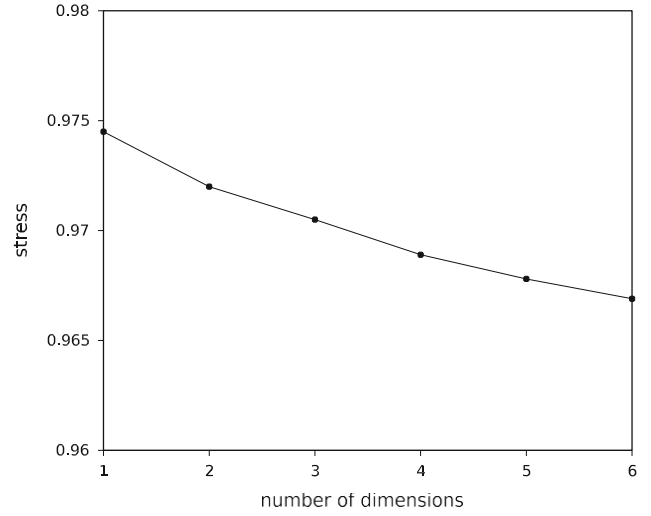


Fig. 8 Stress versus number of dimensions of the MDS map for the Lorentzian distance and the Fourier transform (5)–(6)

normalized wavelength correlation and one based on the normalized Fourier transform of the wavelength measurements [12,14,15].

The cosine correlation α_{jk} is defined as [2]:

$$\alpha_{jk} = \frac{|\sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} x_j(\lambda)x_k(\lambda)|}{\sqrt{\sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} x_j^2(\lambda) \sum_{\lambda=\lambda_{\min}}^{\lambda_{\max}} x_k^2(\lambda)}}, \quad (2)$$

where λ denotes wavelength, $x_j(\lambda)$ represents the j th signal, and $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ is the interval of variation of λ under analysis. Therefore, the angular distance between spectra c_{jk} is defined as:

$$c_{jk} = \arctan \sqrt{\frac{1}{\alpha_{jk}^2} - 1} \quad (3)$$

The distances between spectra inspires a second measure, namely the normalized Fourier transform index c_{jk} defined as:

$$X_j(t\omega) = \int_{\lambda_{\min}}^{\lambda_{\max}} x_j(\lambda)e^{-i\omega\lambda} d\lambda \quad (4)$$

$$\alpha_{jk} = \frac{\sum_{\omega=\omega_{\min}}^{\omega_{\max}} \frac{[\operatorname{Re}\{X_j(t\omega)\} - \operatorname{Re}\{X_k(t\omega)\}]^2}{[\operatorname{Re}\{X_j(t\omega)\}]^2 + [\operatorname{Re}\{X_k(t\omega)\}]^2} + \frac{[\operatorname{Im}\{X_j(t\omega)\} - \operatorname{Im}\{X_k(t\omega)\}]^2}{[\operatorname{Im}\{X_j(t\omega)\}]^2 + [\operatorname{Im}\{X_k(t\omega)\}]^2}}{\sum_{\omega=\omega_{\min}}^{\omega_{\max}} \frac{[\operatorname{Re}\{X_j(t\omega)\} - \operatorname{Re}\{X_k(t\omega)\}]^2}{[\operatorname{Re}\{X_j(t\omega)\}]^2 + [\operatorname{Re}\{X_k(t\omega)\}]^2} + \frac{[\operatorname{Im}\{X_j(t\omega)\} - \operatorname{Im}\{X_k(t\omega)\}]^2}{[\operatorname{Im}\{X_j(t\omega)\}]^2 + [\operatorname{Im}\{X_k(t\omega)\}]^2}} \quad (5)$$

versus the input data leads to another type of plot denoted as ^{0.94} Shepard diagram. Therefore, a narrow scatter of the points around a 45° line indicates a good fit of the distances to the dissimilarities.

4 Data analysis ^{0.93}

In this section, we define two indices for comparing the experimental data. For that purpose, we start by establishing two measures of comparison, namely one based on the

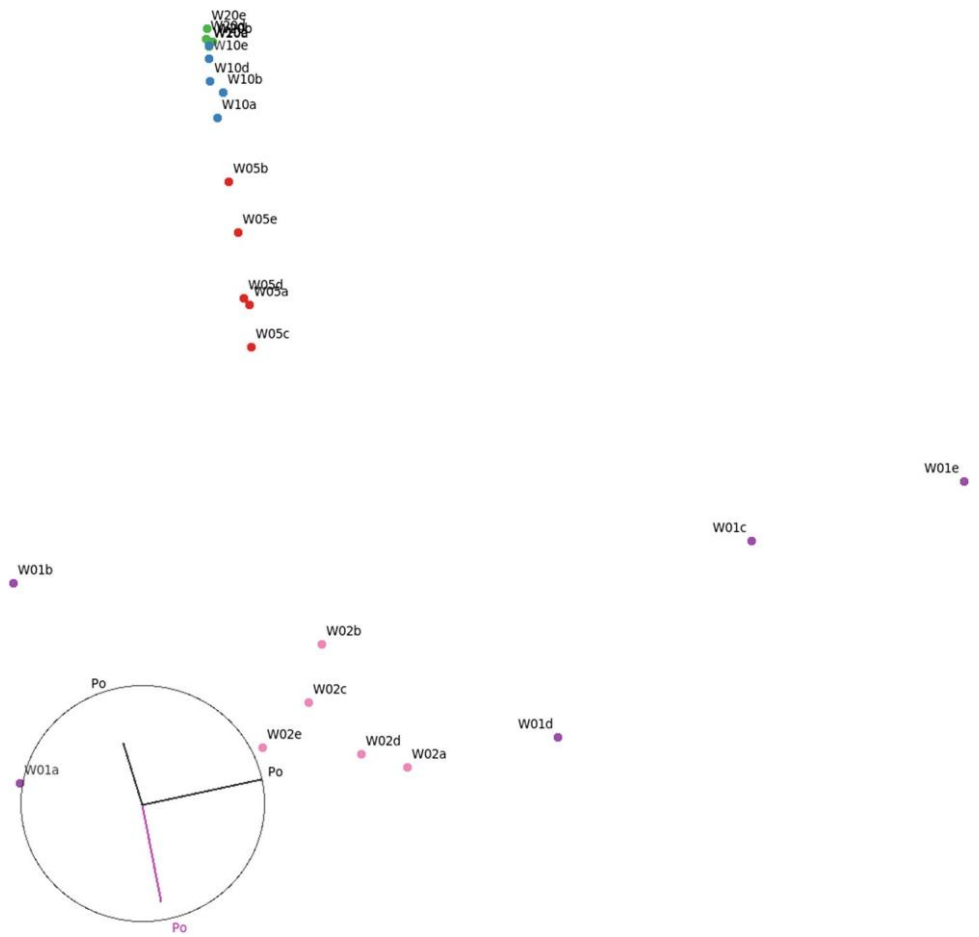
ω_{\max}

This expression can feed the Lorentzian distance [4]:

$$c_{jk} = \ln(1 + \alpha_{jk}), \quad (6)$$

where $\alpha = \frac{\nu}{\omega}$, $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ denote the real and imaginary components, and ω can be loosely defined as the “frequency”, but having units inverse of the wavelength. The

Fig. 9 Three-dimensional MDS map of the 25 measurements based on the Lorentzian distance with the Fourier transform (5)–(6) and slit widths {0.1, 0.2, 0.5, 1.0, 2.0, 5.0}nm



objective is to distinguish the different slit widths using the information at the spectra, particularly at the low and high values of λ , while overcoming the problems associated with the presence of noise.

With these measures, we can now implement two alternative matrices $C = [c_{jk}]$, of dimension 30×30 , that feed the MDS algorithm for constructing the “maps”.

measures handled by indices (2)–(3) and (5)–(6). The resulting matrix $C = [c_{jk}]$ is treated by the package MultiDendogramshierar-

5 Data visualization

In this section, we apply several visualization methods for constructing data “maps”. Section 5.1 analyses the performance of dendograms, and Sect. 5.2 addresses the MDS method.

5.1 Dendograms

In this section, we construct dendogram “maps” [6,16]. Our starting point will be the set of experimental SC

chical clustering package [16], and the results are visualized in Figs. 3 and 4, respectively, for the two alternative indices.

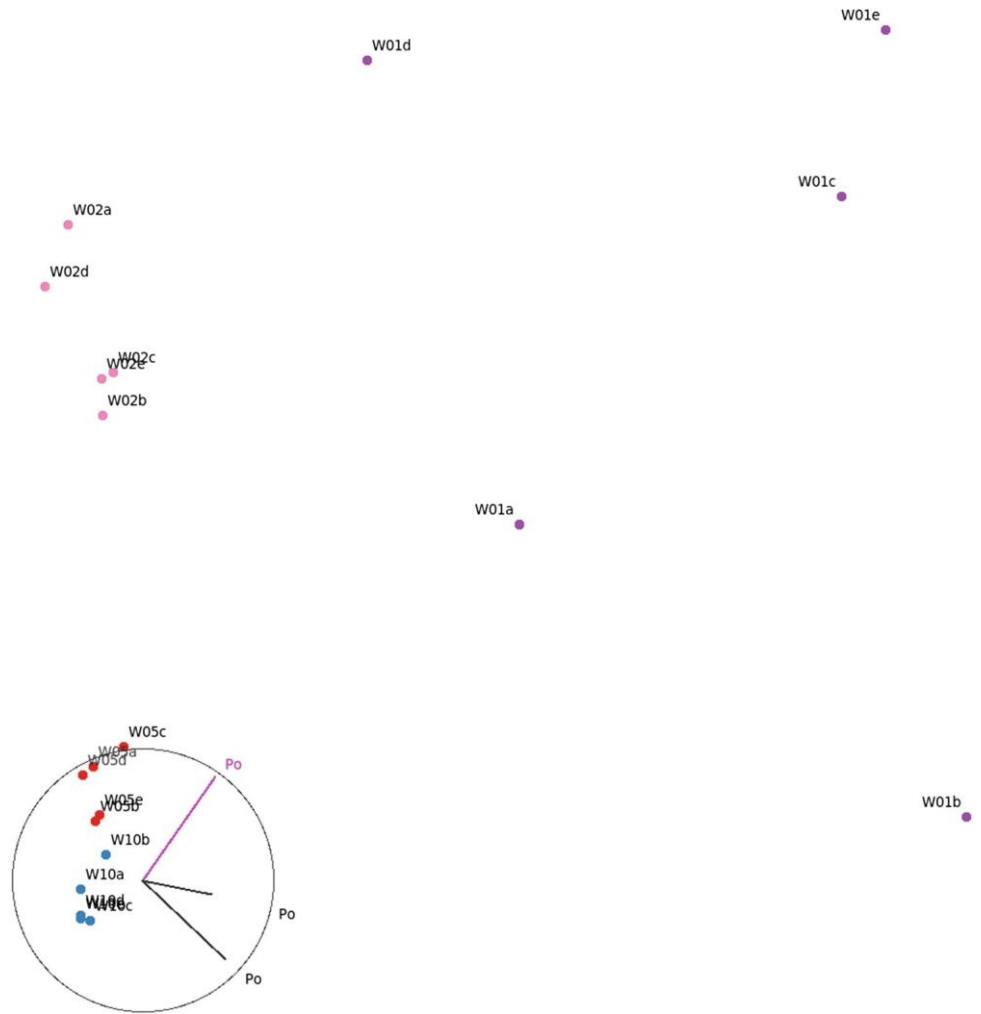
We observe a clear separation of W_{01} but some overlap particularly for the cases with larger widths.

5.2 Multidimensional scaling

In this subsection, we construct MDS “maps” using the MDS package GGobi-Interactive and dynamic graphics [7]. As described previously, the starting point will be the matrix $C = [c_{jk}]$ based on indices (2)–(3) and (5)–(6). Figures 5 and 6 depict the MDS maps for two alternative indices. Figures 7 and 8 represent the corresponding plots of stress versus number of dimensions of the MDS visualization map.

For both indices, the three-dimensional representation establishes a good compromise between feasibility and accuracy. The stress plots confirm this observation. Another observation is that the MDS maps (Figs. 5, 6) are more intuitive than the dendograms (Figs. 3, 4) for visualizing the information since they use more efficiently the graphical portrait.

Fig. 10 Three-dimensional MDS map of the 20 measurements based on the Lorentzian distance with the Fourier transform (5)–(6) and slit widths {0.1, 0.2, 0.5, 1.0, 2.0, 5.0}nm



It should be noted that for the two indices, we obtain distinct maps, since measures (2)–(3) and (5)–(6) capture distinct dynamical characteristics. However, in both cases, we get a representation that makes sense when we compare the position of the points representing the distinct systems.

In what concerns the data clustering the second index is superior to the first one. For both cases, we verify the formation of groups in the direction $W_{01} \rightarrow W_{02} \rightarrow W_{05} \rightarrow W_{10} \rightarrow W_{20} \rightarrow W_{50}$, but some difficulties in separating the five experiments in cases W_{20} and W_{50} . This effect is due to the relatively smaller difference between W_{20} and W_{50} as can be seen in the spectra represented in Figs. 1 and 2. Another aspect is the scattering and the larger distance toward the five points of W_{01} . The spectra of Figs. 1 and 2 reveal W_{01} to have the higher level of noise and to be the most apart for small values of λ , where most of the energy of the signal is located. This reasoning can be

tested in Figs. 9 and 10 that depict the MDS plots for the slit widths {0.1, 0.2, 0.5, 1.0, 2.0}nm and {0.1, 0.2, 0.5, 1.0}nm making 25 and 20 measurements. These cases do not correspond to a simple zoom of the initial

plot, since the MDS algorithm calculates independently the points. In these visualization maps, we observe a much clear separation of the clusters.

6 Conclusion

As it can be seen from Fig. 1, to understand visually the differentiation of the UV spectra of SC in accordance with the change of slit width from 0.1 to 2.0 nm is a difficult problem with respect to the 5.0 nm case. For these reasons, we apply the MDS method to the absorbance data matrix having different slit widths with experimental repetition to uncover the spectral changes between 0.1 and 2.0 nm slit widths. Also, we verify that MDS constitutes a mathematical tool capable of representing and discriminating the spectral band analysis of SC. Moreover, further measuring indices can be explored and different signals can be investigated with this technique.

Acknowledgments This work was done within the Chemometric Laboratory of Faculty of Pharmacy, and it was supported by the sci-

entific research Project No. 10A3336001 of Ankara University. The authors thank the anonymous reviewers for their valuable comments that helped improving the paper.

References

1. Borg, I., Groenen, P.J.: *Modern Multidimensional Scaling-Theory and Applications*. Springer, New York (2005)
2. Cha, S.: Taxonomy of nominal type histogram distance measures. In: *Proceedings of the American conference on applied mathematics*, pp. 325–330. Harvard, Massachusetts, USA (2008)
3. Cox, T.F., Cox, M.A.A.: *Multidimensional Scaling*. Chapman & Hall/CRC, Boca Raton (2001)
4. Deza, M.M., Deza, E.: *Encyclopedia of Distances*. Springer, Berlin (2009)
5. Elshafeey, A.H., Bendas, E.R., Mohamed, O.H.: Intranasal microemulsion of sildenafil citrate: in vitro evaluation and in vivo pharmacokinetic study in rabbits. *AAPS PharmSciTech* **10**(2), 361–367 (2009)
6. Fernández, A., Gómez, S.: Solving non-uniqueness in agglomerative hierarchical clustering using multidendrograms. *J. Classif.* **25**(1), 43–65 (2008)
7. GGobi.<http://www.ggobi.org/>
8. Harriett, J., Broderick, W.C.: Utilization patterns of sildenafil citrate in a senior managed care population. *Value in Health* **8**(3), 302–302 (2005)
9. Kruskal, J.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**(1), 1–27 (1964)
10. Kruskal, J.B., Wish, M.: *Multidimensional Scaling*. Sage, Newbury Park (1978)
11. Lee, H.G., Kim, W.M., Choi, J.I., Yoon, M.H.: Roles of adenosine receptor subtypes on the antinociceptive effect of sildenafil in rat spinal cord. *Neurosci. Lett.* **480**(3), 182–185 (2010)
12. Machado, J.T.: Multidimensional scaling analysis of fractional systems. *Comput. Math. Appl.* **64**(10), 2966–2972 (2012)
13. Martinez, W.L., Martinez, A.R.: *Exploratory Data Analysis with MATLAB*. Chapman & Hall/CRC, Boca Raton (2005)
14. Machado, J.T., Duarte, F.B., Duarte, G.M.: Analysis of stock market indices through multidimensional scaling. *Commun. Nonlinear Sci. Numer. Simul.* **16**(12), 4610–4618 (2011)
15. Machado, J.T., Costa, A.C., Quelhas, M.D.: Analysis and visualization of chromosome information. *Gene* **491**(1), 81–87 (2012)
16. Multidendrograms. <http://deim.urv.cat/sgomez/multidendrograms.php>
17. Shepard, R.N.: The analysis of proximities: multidimensional scaling with an unknown distance function. *Psychometrika* **27**(I and II), 219–246 and 219–246 (1962)
18. Torgerson, W.: *Theory and Methods of Scaling*. Wiley, New York (1958)