



INVESTIGAÇÃO DE ABORDAGEM ESTATÍSTICA DE ANÁLISE MULTIVARIADA NO DESEMPENHO DE MÉTODOS TEÓRICOS PARA CALCULAR AS PROPRIEDADES DE ESPALHAMENTO RAMAN

Tatiane de Freitas Gomes^{1*}, Lucas Gabriel Póvoas Silva¹; Raquel Maria Trindade Fernandes¹, Pedro Antônio Muniz Vazquez², Alamgir Khan^{1**}

¹*Universidade Estadual do Maranhão – ftatiane593@gmail.com**
*alamgir@cecen.uema.br***

²*Universidade Estadual de Campinas*

1. INTRODUÇÃO

Propriedades ópticas Raman não lineares dinâmicas e estáticas de moléculas têm sido um assunto de contínuo interesse para os pesquisadores experimentais e teóricos (CASTRO et al, 2017). A importância do cálculo teórico das propriedades Raman resultou do fato de que as propriedades elétricas moleculares são difíceis de medir (ANASTASOPOULOU). Atualmente, as implementações computacionais para o cálculo de intensidades Raman estáticas são bastante frequentes em pacotes de química quântica, onde esta propriedade é obtida através de métodos de *ab initio* ou Teoria do Funcional da Densidade (PAPAYANNIS, 2017). A correlação eletrônica é considerada muito importante para o cálculo da propriedade Raman, sendo necessários métodos mais sofisticados para se obter uma concordância quantitativa com os dados experimentais (WILKINS et al, 2019). Com isso, há aumento no interesse no desenvolvimento de novos métodos de química quântica computacional, que sejam capazes de cálculos de baixo custo, em termos de tempo e recursos computacionais, obtendo propriedades ópticas Raman acurada e precisas.

Desta forma, o presente trabalho teve como objetivo aplicar a metodologia do ECP (pSBKJC e pStuttgart) desenvolvido por Vazquez e Vidal (2009), para avaliar seu desempenho na computação das propriedades moleculares, ou seja, atividades e intensidades de espalhamento Raman, para um conjunto de moléculas poliatômicas (H_2O , NH_3 , CH_4 , CH_2O e C_2H_2) usando os método *ab initio* (LR-CCSD) e DF, além da utilização do método estatístico multivariado de análise de componentes principais (PCA) para diagnosticar e investigar a eficiência da metodologia proposta, como também explorar as relações entre o conjunto de dados Raman computados.

2. METODOLOGIA

A geometria de equilíbrio em fase gasosa, a matriz hessiana e modo normal da vibração das moléculas foram calculadas com o conjunto de bases cc-pVDZ e de cálculos híbridos da Teoria Funcional da Densidade (DFT) pelo método B3LYP, utilizando o programa eletrônico de estrutura Gamess. A polarizabilidade molecular estática, assim como a polarizabilidade dinâmica em comprimentos de onda de 1064 a 337 nm de excitação foram calculados utilizando o módulo de Resposta Linear (LR) do código DALTON-2.0, usando funções de onda DFT (PBE0, LB94 e CAMB3LYP) com o conjunto de base pSBKJC e pStuttgart ECP.

Após a computação das propriedades moleculares de interesse, utilizou-se o método estatístico multivariado, Análise de Componentes Principais (PCA) no programa Statistic versão 8.0 (VIDAL; VAZQUEZ, 2012), para diagnosticar e investigar a eficiência da metodologia proposta. Com isso, foi estabelecido uma comparação entre os resultados esperados obtidos com a proposta do trabalho de ECP (pSBKJC e pStuttgart) e o conjunto de base Sadlej-pVTZ como referência.

A análise multivariada de componentes principais foi utilizada para investigar a avaliar os desempenhos dos métodos teóricos aplicados para computar as propriedades espectroscópicas de Raman. Foi utilizado um banco de dados das intensidades Raman computadas para as moléculas poliatômicas nas frequências de Laser de 337 nm a 1064 nm.

A informação referente a análise de dados, foi obtida interpretando os parâmetros PCA como loadings, scores, autovetores, e autovalores das componentes principais (PCs) uma vez que, seus autovalores identificam as mudanças e inclinações mais importantes nos conjuntos de dados (StatSoft, 2004).

3. RESULTADOS E DISCUSSÃO

Para a análise explanatória dos dados utilizou-se o método multivariado PCA para todo o conjunto de dados, incluindo as bandas de intensidade Raman forte, média e fraca, considerando os protocolos teóricos proposto de pseudo-potencial (pSBKJC e pStuttgart) e as bases de Sadlej estabelecidas nos níveis das teorias CCSD e TDDFT. Os autovalores e as variabilidade nas componentes principais coletaram para o conjunto de dados 140 casos (intensidades Raman) e o conjunto de dados individuais 35 casos para cada intensidade Raman forte, média e fraca usando 12 métodos computacionais.

A análise multivariada de componentes principais (PCA) foi realizada para reduzir o problema de dimensionalidade e possível análise exploratória dos métodos teóricos propostos utilizados no cálculo das propriedades Raman. As informações foram obtidas pela interpretação dos parâmetros multivariados loadings, scores, autovetores e autovalores. Onde, as componentes principais (PCs) e seus autovalores identificaram e exploraram as mudanças e inclinações baseados nos valores de computação utilizando os conjuntos de dados (EUERBY; PETERSSON, 2003).

Na Figura 1, observou-se que as três primeiras componentes principais (PCs) foram as componentes mais significativas, representando 99,938% da variância nas propriedades Raman calculadas (ou seja, PC1: 99,417%, PC2: 0,339%, PC3: 0,182%) para o conjunto de dados Raman completo.

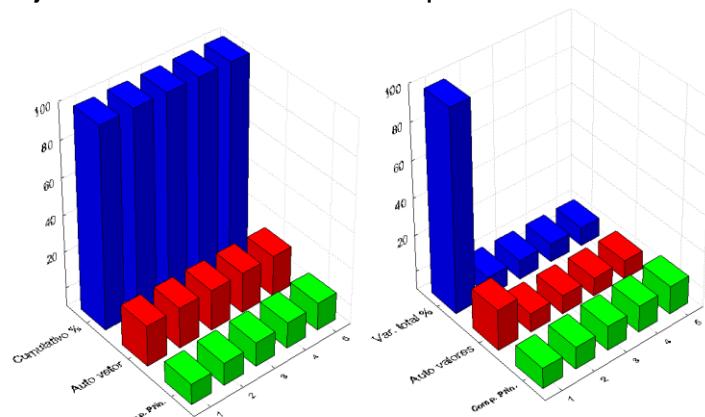


Figura 1 - Autovalor, variância total (%) e Cumulativo dos primeiros 5 componentes principais (Fonte: A autora, 2021.)

Na PCA, as componentes principais (PC) são ajustadas a um conjunto de dados, de forma que a primeira PC possa descrever o máximo possível da variação original entre os casos. Enquanto, a segunda PC é ajustada de forma ortogonal a primeira PC, e é direcionada para caracterizar o máximo das variações remanescentes e assim por diante (NASR; ZAHRAN, 2016).

Na presente análise, a primeira e a segunda componentes principais mostram uma variância geral de 99.756% para o conjunto de dados Raman completo, conforme mencionado.

Na figura-1 tem-se os scores, os quais revelam as projeções das intensidades Raman para cada método computacional ao longo da linha da PC destacando as semelhanças e diferenças entre eles, resultando na formação de grupos. Com isso, o score da PC1xPC2 do conjunto de dados expressou sucesso na distribuição e classificação das intensidades de espalhamento Raman estático e dinâmico. Destacou-se também a formação de outliers, os quais são considerados os cálculos superestimados dos funcionais DFT para função de onda LB94 nas frequências excitadas do laser entre 632 a 337 nm. A análise de loadings das componentes principais é sugerida como ferramenta muito útil para explorar a importância das variáveis e investigar suas eficiências, com base nas correlações entre o espalhamento Raman, o qual pode ser visto na figura 1.

Na análise de PCA, as variáveis correlacionadas têm pesos semelhantes nos vetores de loadings, e eles aparecem próximos nos gráficos de todas as dimensões. Foram analisados também os dados de intensidades fortes, médios e fracos conforme a figura 1. Observamos, que os protocolos de pseudo-potencial pSBKJC e pStuttgart demonstraram a semelhanças nos resultados comparados ao referência de Sadlej-pVTZ. Exceto no caso de LB94/Sadlej-pVTZ observamos uma peculiaridade na computação da mesma. Com isso, revelou-se que a função de base LB94, um método não deu resultados satisfatório.

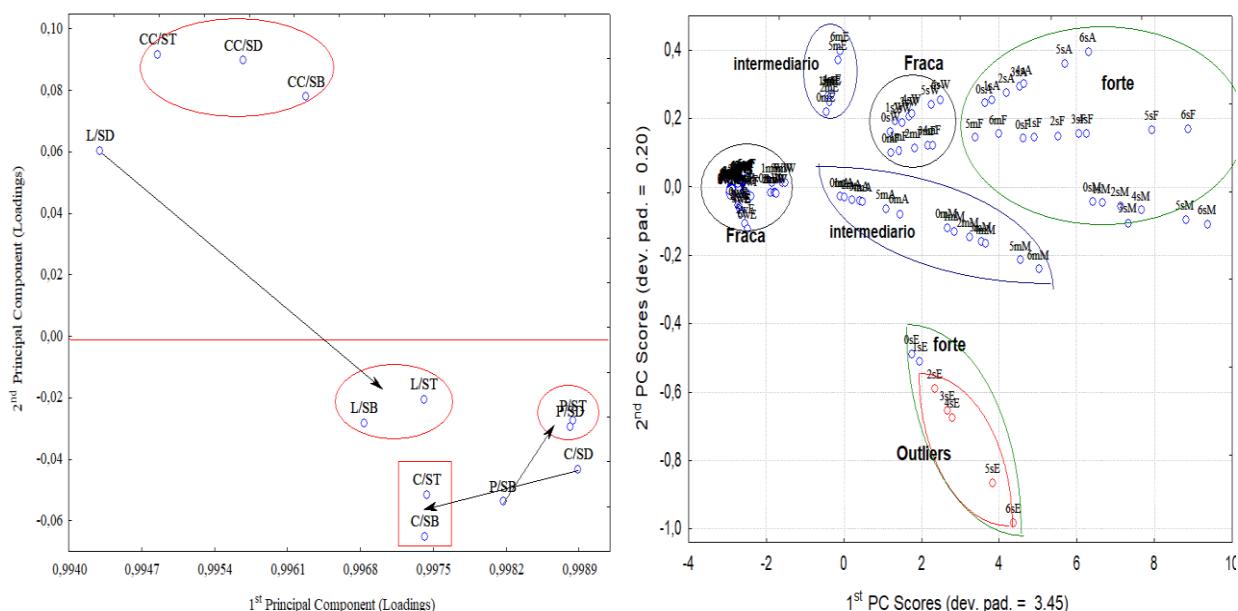


Figura 2 - a) scores plot entre os PC1 e PC2 para as intensidades Raman; b) Loadings-coefficient entre componentes principais 1 e 2. (Fonte: A autora, 2021.)



4. CONCLUSÕES

Com a análise de PCA proposta, foi possível identificar e agrupar com sucesso os métodos computacionais aplicados baseados na computação de propriedade Raman vibracional.

Além disso, os gráficos de scores diagnosticaram e exploraram o efeito da excitação do laser de energias nas intensidades Raman, onde um aumento na intensidade Raman dos valores foi observado com o aumento da potência de excitação do laser. A PCA também conseguiu expressar a eficiência na exploração dos dados teóricos obtidos pelos métodos de interesse utilizando os bi-plots entre os parâmetros de loadings e scores das componentes principais.

Sugere-se então que a PCA é um método estatístico robusto que pode ser aplicado para avaliar a eficiência na área de química computacional, uma vez que, investigou e avaliou com sucesso os desempenhos dos protocolos de ECP (pSBKJC e pStuttgart) para o cálculo das propriedades Raman.

5. REFERÊNCIAS BIBLIOGRÁFICAS

A.N. Castro, F.A.P. Osório, R.R. Ternavisk, H.B. Napolitano, C. Valverde, B. Baseia Theoretical investigations of nonlinear optical properties of two crystalline acetamides structures including polarization effects of their environment **Chem. Phys. Lett.**, 681 (2017), pp. 110-123, 10.1016/j.cplett.2017.05.066.

D.M. Wilkins, A. Grisafi, Y. Yang, K.U. Lao, R.A. Distasio, M. Ceriotti Accurate molecular polarizabilities with coupled cluster theory and machine learning **Proc. Natl. Acad. Sci.**, 11 (6) (2019), p. 201816132, 10.1073/pnas.1816132116

L.N. Vidal, P.A.M. Vazquez A new ECP basis set for accurate calculations of dynamic Raman intensities **Revista Brasileira de Aplicações de Vácuo**, 2009, 28 (2009), 10.17563/rbav.v28i1-2.476 21.

L.N. Vidal, P.A.M. Vazquez Electrically polarized valence basis sets for the SBKJC effective core potential developed for calculations of dynamic polarizabilities and Raman intensities **Theor. Chem. Accounts**, 131 (2012), 10.1007/s00214-012-1111-4 1111.

M. Anastasopoulou, K.C. Vasilopoulos, D. Anagnostopoulos, I. Koutselas, D.K.

M. Nasr, H.F. Zahran. Performance evaluation of agricultural drainage water using modeling and statistical approaches. **The Egyptian Journal of Aquatic Research**, 42 (2) (2016), pp. 141-148.

M.R. Euerby, P. Petersson Chromatographic classification and comparison of commercially available reversedphase liquid chromatographic columns using principal component analysis **J. Chromatogr. A**, 994 (1–2) (2003), pp. 13-36, 10.1016/s0021-9673(03)00393-5.

Papayannis, M.A. Karakassides Structural and theoretical study of strontium borophosphate glasses using raman spectroscopy and ab initio molecular orbital method **J. Phys. Chem. B**, 121 (17) (2017), pp. 4610-4619, 10.1021/acs.jpcb.7b01563.

StatSoft Inc., (2004). “Statistica,” Data Analysis Software System, version 7, 2004.