BJPS

Brazilian Journal of
Pharmaceutical Sciences

http://dx.doi.org/10.1590/s2175-97902024e23487

Article

# Development of spectroscopic methods for assessing polymorphic content of efavirenz

**Talita Atanazio Rosa[1]; Marcos Victor Gregório de Oliveira[1]; Leandro de Moura França[1]; Maria Joanellys dos Santos Lima[1*]; Pollyne Amorim Silva[1]; Rosali Maria Ferreira da Silva[1]; Larissa Araújo Rolim[2]; Maria Fernanda Pimentel[3]; Pedro José Rolim Neto[1]**

*[1]Department of Pharmacy, Federal University of Pernambuco, Recife, Pernambuco, Brazil, [2]Collegiate of Pharmaceutical Sciences, Federal University of Vale do São Francisco, Petrolina, Pernambuco, Brazil, [3]Department of Chemical Engineering, Federal University of Pernambuco, Recife, Pernambuco, Brazil.*

The present work aimed to develop a new method for assessing the content of mixtures of polymorphic forms I and II of the drug Efavirenz (EFV) by vibrational spectroscopic techniques such as Middle (MIR) and Near (NIR) infrared and Raman, using multivariate calibration models. Benchtop and handheld instruments were used for NIR and Raman and a benchtop instrument for MIR. In addition, VIP scores and iPLS, variable selection methods were employed. The infrared techniques showed the best models, with Root Mean Squares Error (RMSE) around 5% (w/w). When MIR and portable NIR instruments were used, this value was lowered to 4% (w/w) with selection of variables by iPLS. Raman spectroscopy showed higher error, even with selection of variables, possibly due to the spot laser size used by instruments and the lack of uniformity of the particle size in the samples. The infrared methods developed were shown to be effective in quantifying polymorphic mixtures of EFV. Given the ease of use of handheld instruments, they may be applied as tools of process analytical technology for monitoring quality control during industrial processing.

**Keywords:** Efavirenz. Infrared spectroscopy. Near-infrared spectroscopy. Partial least squares. Polymorph(s). Raman spectroscopy.

## INTRODUCTION

Acquired immunodeficiency syndrome (AIDS) has as its main symptom the appearance of infections due to opportunistic diseases, a consequence of the immunodepression caused by HIV (Sankoh *et al.*, 2014). The use of non-nucleoside reverse transcriptase inhibitors has been of great importance in controlling the viral load in patients infected with HIV-1, with one of these inhibitors being the drug Efavirenz (Ravikumar, Sridhar, 2009; Bastos *et al.*, 2016).

Efavirenz (EFV), (4~{S})-6-chloro-4-(2-cyclopropylethynyl)-4-(trifluoromethyl)-1~{H}-3,1-benzoxazin-2-one (Figure 1), is an active pharmaceutical ingredient from the class of non-nucleoside reverse transcriptase inhibitors, patented as Sustiva by Bristol-Myers Squibb in 1998. It is still used in the fight against HIV-1 and has been studied as treatment for other diseases (Mbuagbaw *et al.*, 2016; Radesca *et al.*, 2004; Vrouenraets *et al.*, 2007; Costa, Vale, 2023). It is described as a slightly pinkish white powder with a melting point between 139-141 °C and pKa of 10.2. EFV belongs to class II in the biopharmaceutical classification system, due to its low solubility in water and high permeability (Marques *et al.*, 2017).

***Correspondence:** M. Joanellys. Departamento de Farmácia. Universidade Federal de Pernambuco. Av. Professor Artur de Sá, s/n, Cidade Universitária. CEP 50740-525, Recife, PE, Brasil. E-mail: joanellys.lima@ufpe.br. Phone: +55 81 996575023. ORCID: https://orcid.org/0000-0002-1880-5267

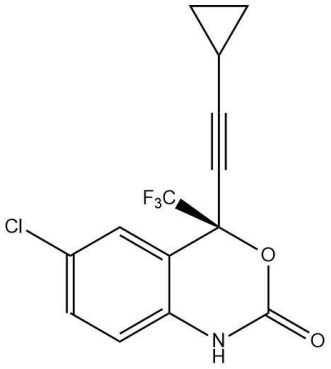


**FIGURE 1 -** Molecular structure of EFV.

In common with EFV, several active pharmaceutical ingredients (API) are administered as solids mainly due to ease of storage. These solid inputs, however, may result from various arrangements in space, such as polymorphs, solvates, salts, co-crystals and amorphous solids, with each arrangement exhibiting distinct physicochemical properties (Brittain, 2007). These different forms affect many aspects of the finished pharmaceutical product, modifying stability, bioavailability and industrial processing methods (Qiu *et al.*, 2015). For these reasons, regulatory agencies have set acceptance criteria for the polymorphic forms in order to establish whether the different forms can affect the performance of the drug product (International Council for Harmonization, 1999; Food and Drug Administration, 2007).

EFV has five different polymorphic forms, as previously described in its patent. The patent also provides other information about the drug, such as methods of production of its polymorphic forms and their characterization by differential scanning calorimetry (DSC) and X-ray diffraction (XRD) techniques (Radesca *et al.*, 2004; Reddy *et al.*, 2013).

The most relevant polymorphs are forms I and II (Marques *et al.*, 2017). While polymorph I is the API used in the pharmaceutical industry, polymorph II has been demonstrated as a candidate for pharmaceutical formulations. It has been reported that form II is ten times more soluble than form I. In addition, thermodynamic studies have shown form II to be more stable at room temperature (Fandaruff *et al.*, 2014).

Besides these two forms, there are several others that have been studied such as synthesized EFV conformers, solvates and co-solvates (Marques *et al.*, 2017; Ravikumar, Sridhar, 2009). These, however, except for polymorph II, have not demonstrated the stability necessary or any improvement in physicochemical characteristics compared to the form already in use as the active pharmaceutical ingredient (API).

Since it is important to guarantee the characteristics of an API and the final products derived from it, development of analytical methods to assess polymorphic content is necessary. The performance of the method on any raw material depends on the correct choice of technique and analytical parameters, according to chemical and physical properties inherent in the API. Special attention must be paid to vibrational spectroscopic techniques, given their speed and precision as well as their being non-destructive, and needing neither sample preparation nor reagents, thus, attending to the requirements of green chemistry (Esclusa-Diaz *et al.*, 1996).

Kang *et al.* (2018) validated a method using Raman spectroscopy as being as precise and accurate as another method using XRD to quantify two forms of entecavir. Silva *et al.* (2015) showed the feasibility of Near (NIR) and Middle (MIR) infrared spectroscopy for quantification of three polymorphs of mebendazole in pharmaceutical raw materials. Farias and Carneiro (2014) developed a method to quantify mixtures of three polymorphs of carbamazepine using Raman Spectroscopy. Guo *et al.* (2017) found that Fourier Transform Infrared (FTIR) spectroscopy was the best technique for assaying binary polymorphic mixtures of fusidic acid compared with FT-NIR and portable Raman spectroscopy.

Portable and miniaturized instruments make it possible to use different methods of spectral acquisition for process monitoring, such as in-, on-, and at-line methods. These differ in some features and make the production process more efficient and assuring better quality. Some differences are the following: at-line methods need sampling and the analysis is done close to, but separated from the process stream; on-line methods have automatic sampling without stopping the process; in-line methods realize measurements directly in the process stream and there is no need for sampling, making this a real time monitoring method (Kim *et al.*, 2021).

Although, portable instruments, in general, have a lower spectral resolution and signal-to-noise ratio than benchtop instruments, they can achieve a similar performance. Silva *et al.* (2017) evaluated the performance of a NIR benchtop and three different portable NIR instruments to assess the content of ternary mixtures of polymorphs of mebendazole. The results demonstrated the feasibility of using portable devices. In this application, the performance of one of the tested portable NIR instrument was comparable to the benchtop instrument.

For EFV, Fandaruff *et al*. (2014) characterized forms I and II using benchtop instruments, including Raman and MIR spectroscopy. Differences between the spectra of the two polymorphs were shown and discussed. Discussions about NIR spectra of EFV forms I or II, on the other hand, have not been found in the literature up to the date of our study. Although the advantages of using the polymorph II have been noted, reports about methods to quantify the content of each polymorphic form in mixtures have also not been found up to now.

The present work aimed to develop analytical methods using vibrational spectroscopy (NIR, MIR and Raman) associated with multivariate calibration techniques to quantify binary mixtures containing the forms I and II of EFV. The results obtained using benchtop and portable instruments are discussed and compared.

## MATERIAL AND METHODS

### Chemicals

The EFV polymorph I was donated by Cristália® (lots 1289/07 and 86/16). The EFV polymorph II was obtained by recrystallization of polymorph I in methanol HPLC grade (Merck®) at around 4 °C.

### Samples Preparation

EFV polymorphs I and II were ground using mortar and pestle, then mixed for 5 minutes using a vortex mixer model MA-162 (Marconi) in the following concentrations of polymorph II: 0, 5, 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100% (w/w). Each sample weighed 220 mg.

### X-Ray Powder Diffraction

X-ray powder diffraction (XRPD) patterns were recorded on a Shimadzu® diffractometer (XRD-7000), equipped with copper anode, and on a Bruker® diffractometer (D8 Advance), equipped with Cu K(alfa) tube and nickel filter, operating with 40 kV and 40 mA. The Soller divergent and antiscattering slits used were 0.04 rad and 0.25°, respectively. XRPD measurements were carried out using 2θ range from 5° to 45°, scan step size of 0.008°, and scan step time of 20 s.

### Mid-Infrared Spectroscopy

The spectra were obtained using a Spectrum 400 (PerkinElmer®) with an Attenuated Total Reflectance (ATR) device with zinc selenide crystal. Before analyzing each sample, the clean crystal was scanned to obtain the background spectrum in order to eliminate environmental interference such as humidity and presence of $CO_2$. Samples were transferred directly to the ATR device compartment, where 16 scans were carried out between 4000 and 650 $cm^{-1}$ with a resolution of 4 $cm^{-1}$. For each concentration, five spectra were obtained from which two spectral means were calculated. For all spectroscopic techniques evaluated, the spectra replicates were acquired by resampling the mixtures.

### Near- Infrared Spectroscopy

The NIR spectra were obtained in two spectrometers: a Spectrum 400 (PerkinElmer®) with a Near Infrared Reflectance Accessory (NIRA), spectral region of 750-2500 nm (10000-4000 $cm^{-1}$) and spectral resolution of 4 $cm^{-1}$; and a MicroNIR 1700 (Viavi®), equipped with a Linear Variable Filter (LVF) with an InGaAs detector array of 128-pixel. The MicroNIR operational range is from 11012.0 to 5965.8 $cm^{-1}$ (950-1650 nm) and has a nominal spectral resolution of 6.25 nm.

The samples were placed in petri dishes for analysis. For the NIR benchtop, spectra were obtained from an average of 32 scans. For the portable NIR, the results were obtained from an average of 50 scans. Polytetrafluoroethylene (PTFE) was employed as

background. For each concentration analyzed, eight spectra were obtained from which two spectral means were calculated.

### Raman Spectroscopy

The Raman spectra were obtained using two spectrometers: a confocal Raman microscope (SENTERRA, Bruker, Germany) with a laser beam at 785 nm, 100 mW, 20x objective, spot laser of 2 μm, spectral resolution of 3-5 $cm^{-1}$, spectral range of 60-2622 $cm^{-1}$ and time exposure of 30 s with 2 coadditions; and a portable Mira DS model (Metrohm), with a laser at 785 nm, spectral range of 400-2300 $cm^{-1}$, spectral resolution of 8-10 $cm^{-1}$ and spot laser of 2-2.5 mm. The spectra were acquired in the smart mode, which adjusts laser power and time exposition automatically. Anterior to analysis, the equipment was subjected to an automatic background correction. For each concentration, five spectra were obtained from which two spectral means were calculated.

### Data Treatment

As mentioned, for each concentration, 5 to 8 measurements were performed (resampling the mixtures), and two averages were calculated, obtaining 44 samples in total. These samples were separated into calibration (~55%) and validation (~45%) data sets, by means of the KSXY algorithm (KSXY) (Galvão *et al*., 2005).

Different preprocessing techniques were evaluated to minimize undesired effects in the spectra that could be generated by instrumental variation, scattering, spurious radiation and fluorescence (Luo *et al*., 2005; Fearn *et al*., 2009; Rinnan, Van Den Berg, Engelsen, 2009). For infrared data sets, the following were tested: Savitzky-Golay (SG) smoothing, Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), baseline offset, first and second derivative (SG, second polynomial order and different window sizes). In the case of the Raman data sets, Automatic Weighted Least Squares (AWLS) and Asymmetric Least Squares (ALS), besides normalization and the other mentioned preprocessing methods were tested (Eigenvector, 2019; Eilers, Boelens, 2005).

Tools for selection of variables such as forward interval Partial Least Squares (iPLS) and Variable Importance in Projection (VIP) scores were evaluated (Chong, Jun, 2005; Nørgaard, 2000). The selection using VIP was made choosing variables whose VIP scores were above 1. For forward iPLS, an automatic number of intervals was used, five was taken as the maximum number of latent variables and the interval size varied accordingly to the size of the regular band in the spectra.

The models for quantification of the polymorphs were built using PLS regression. To select the number of latent variables, venetian blind cross-validation was used. Plots of scores, residuals and leverage of the models were utilized to detect and eliminate outliers.

The performance of the achieved models was evaluated by the following parameters: Root Mean Square Error of Cross Validation (RMSECV) and Prediction (RMSEP), bias and $R^2$. Bias values distant from zero indicate the presence of systematic errors, while the RMSEP values summarize the accuracy of the predictions. The F test at 95% confidence was used to evaluate significant differences between the RMSEP values obtained. To test the presence of significant bias, a t-test, as recommended by ASTM E-1655-05 was carried out.

The software tools used to perform the data treatment were Matlab® R2015a 8.5 (Mathworks) and PLS toolbox 8.5.2. (Eigenvector®).

## RESULTS

### Obtention of the Polymorph II

Polymorph II acquisition by recrystallization was confirmed by XRD (Figure 2), Infrared and Raman Spectroscopy (Figure 3).

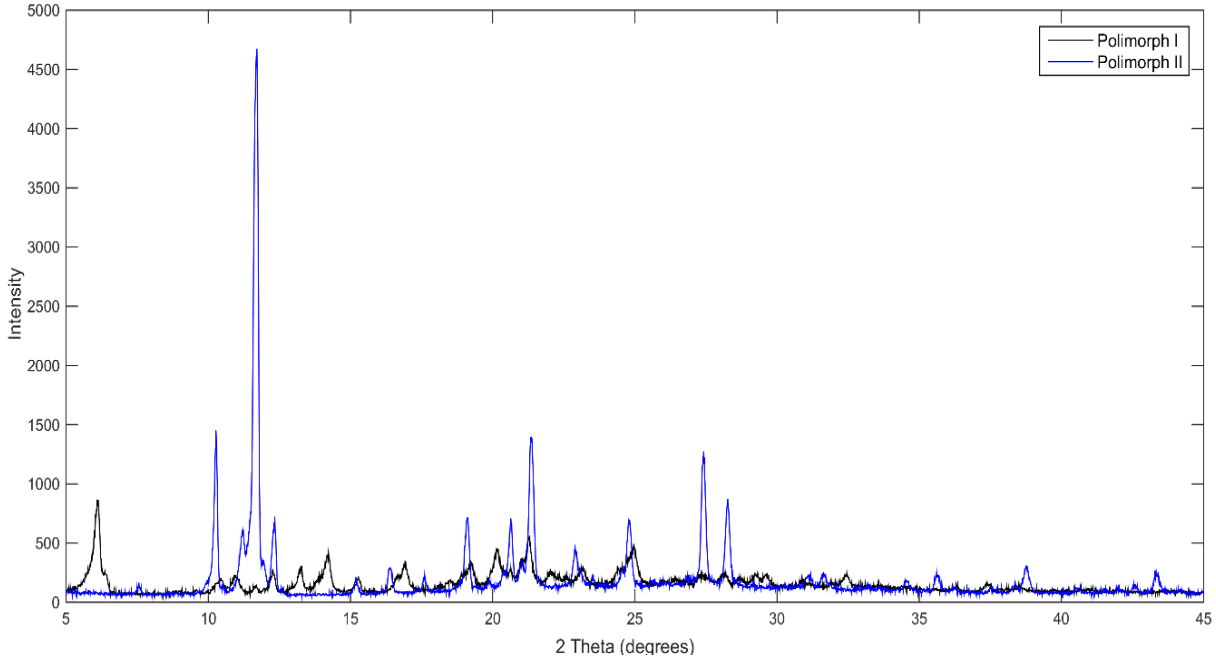


**FIGURE 2 –** XRPD patterns of the polymorph I and the recrystallized polymorph II of EFV.

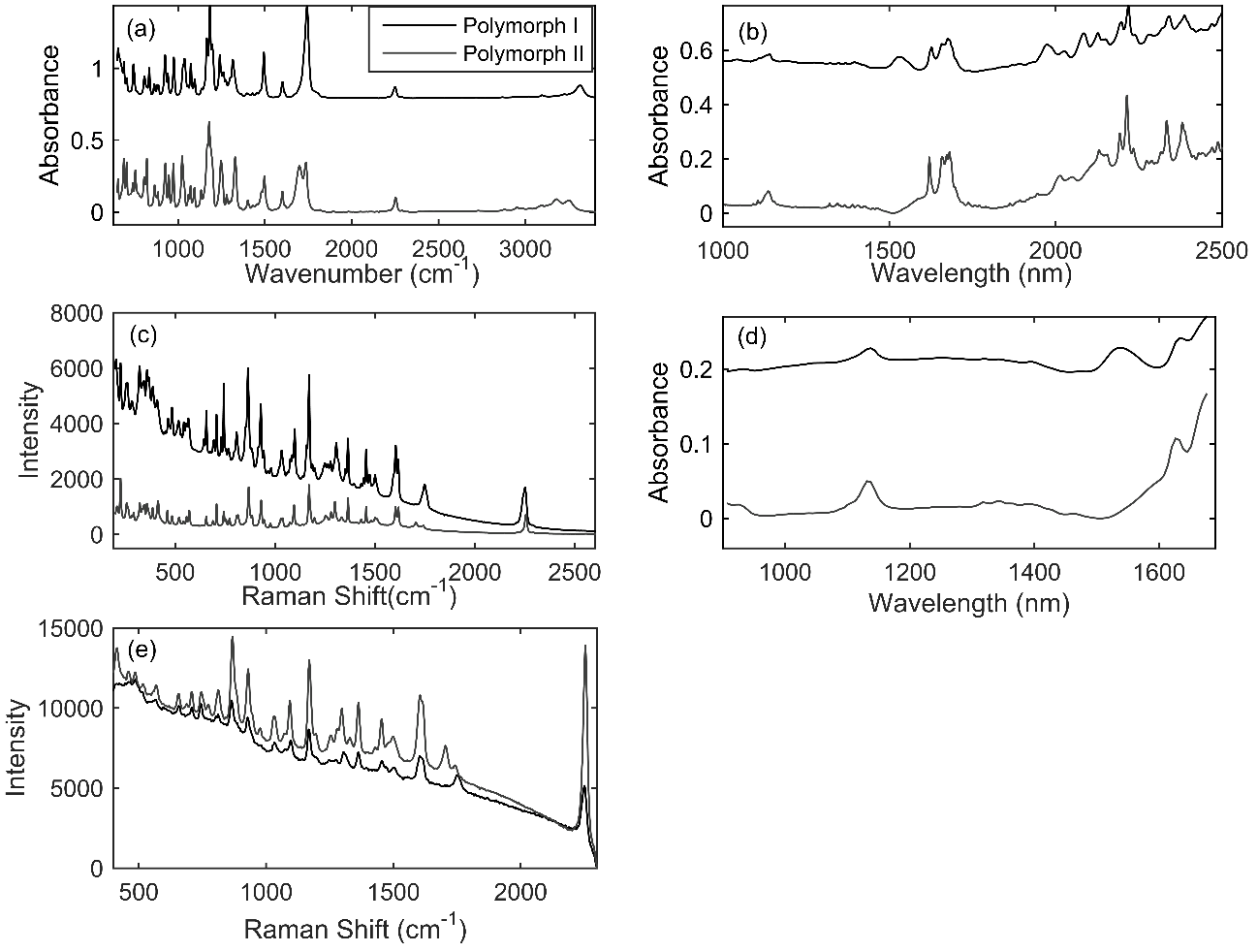


**FIGURE 3 –** Infrared and Raman spectra of polymorphs I and II of EFV. (a) MIR; (b) benchtop NIR; (c) benchtop Raman; (d) portable NIR; (e) portable Raman.

## Spectral data

The spectra of the polymorphic mixtures are presented in Figure 4.

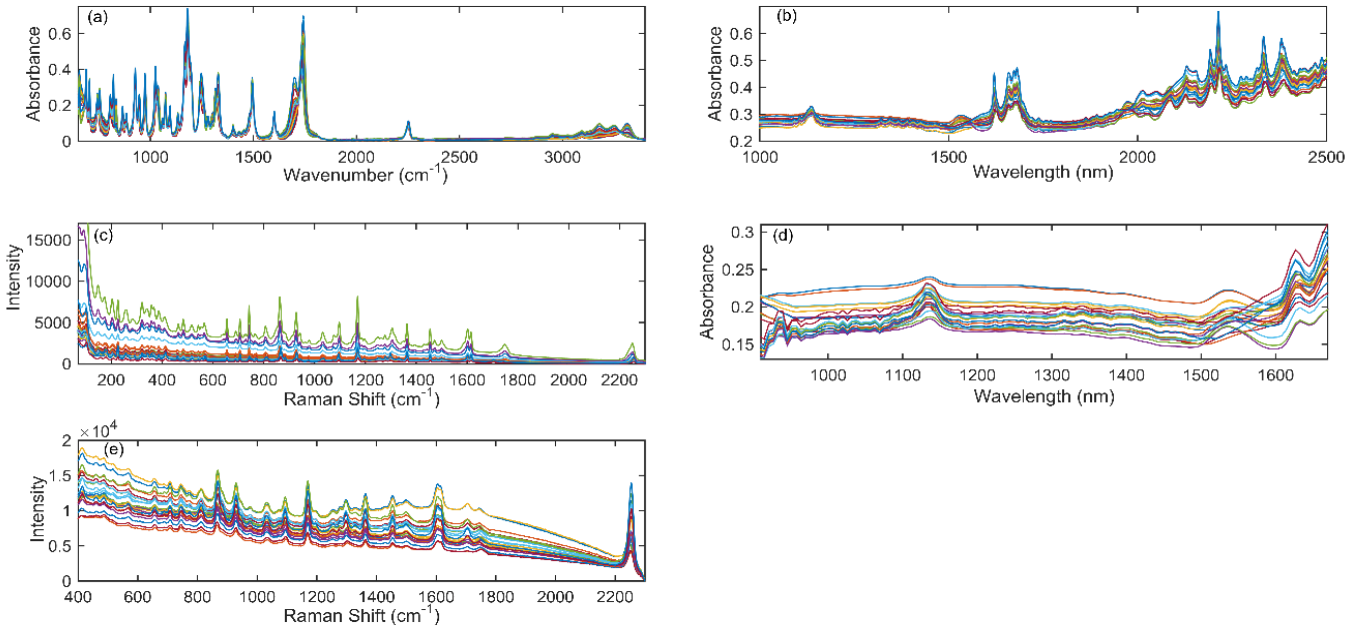


**FIGURE 4 -** Infrared and Raman spectra of the polymorphic mixtures of EFV. (a) MIR; (b) benchtop NIR; (c) benchtop Raman; (d) portable NIR; (e) portable Raman.

## Data Treatment

The best preprocessing techniques and the parameters of the PLS models are shown in Table I. The RMSEP values for the infrared data sets were around 5% (w/w); for Raman data sets, the RMSEP values were higher than 5% (w/w).

**TABLE I -** The best PLS models obtained for EFV polymorphs. The numbers in parentheses are the latent variables

| Instruments | Models | $R^2$ | RMSE | Bias | Pre-proc. |
|---|---|---|---|---|---|
| MIR | CV | 0.97 | 5.3 (3) | 0.4 | Smooth. + MSC |
| | Pred | 0.98 | 4.3 | -0.7 | |
| NIR benchtop | CV | 0.97 | 5.0 (5) | -0.2 | 1st derivative |
| | Pred | 0.97 | 5.5 | 0.1 | |
| Portable NIR | CV | 0.97 | 5.4 (4) | -0.1 | 1st derivative |
| | Pred | 0.98 | 5.0 | 1.6 | |
| Raman benchtop | CV | 0.80 | 14.1 (7) | 0.8 | Smooth. + MSC |
| | Pred | 0.73 | 15.3 | 0.8 | |
| Portable Raman | CV | 0.95 | 7.0 (5) | 0.6 | Smooth. + SNV |
| | Pred | 0.92 | 8.6 | 0.8 | |

Considering η = 19 for MIR and NIR benchtop spectra and η = 17 for portable NIR spectra, the performed F test (at 95% confidence level) revealed that there was no statistically significant difference between the RMSEP values of models built using data sets obtained from these infrared instruments. In relation to $R^2$ values, values higher than 0.95 were obtained, indicating a good fit of the models (Antonio, Maggio, 2018).

Figure 5 shows the reference versus predicted values from the PLS models. Regarding the bias values, according to the t-test (as recommended by ASTM E-1655-17), there was no significant bias (at 95% confidence) (ASTM Standard, 2017; Lindon *et al.*, 2017).

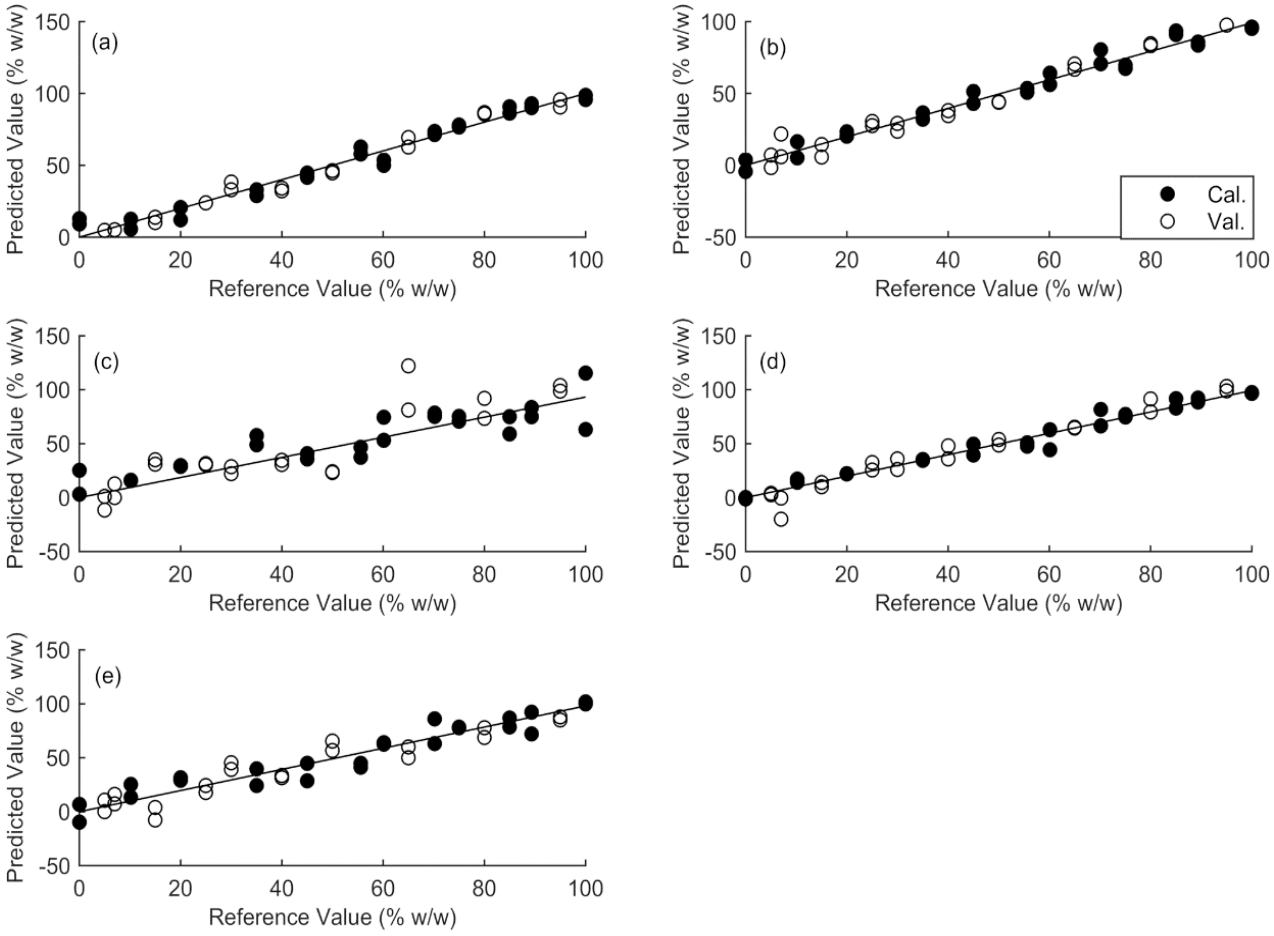


**FIGURE 5 –** Graph with the values predicted from the PLS models. (a) MIR; (b) NIR benchtop; (c) Raman benchtop; (d) Portable NIR; (e) Portable Raman.

After developing PLS models, tools for selection of variables were used in attempting to improve these models. Figure 6 shows a graph with the variables selected by VIP score higher than one.

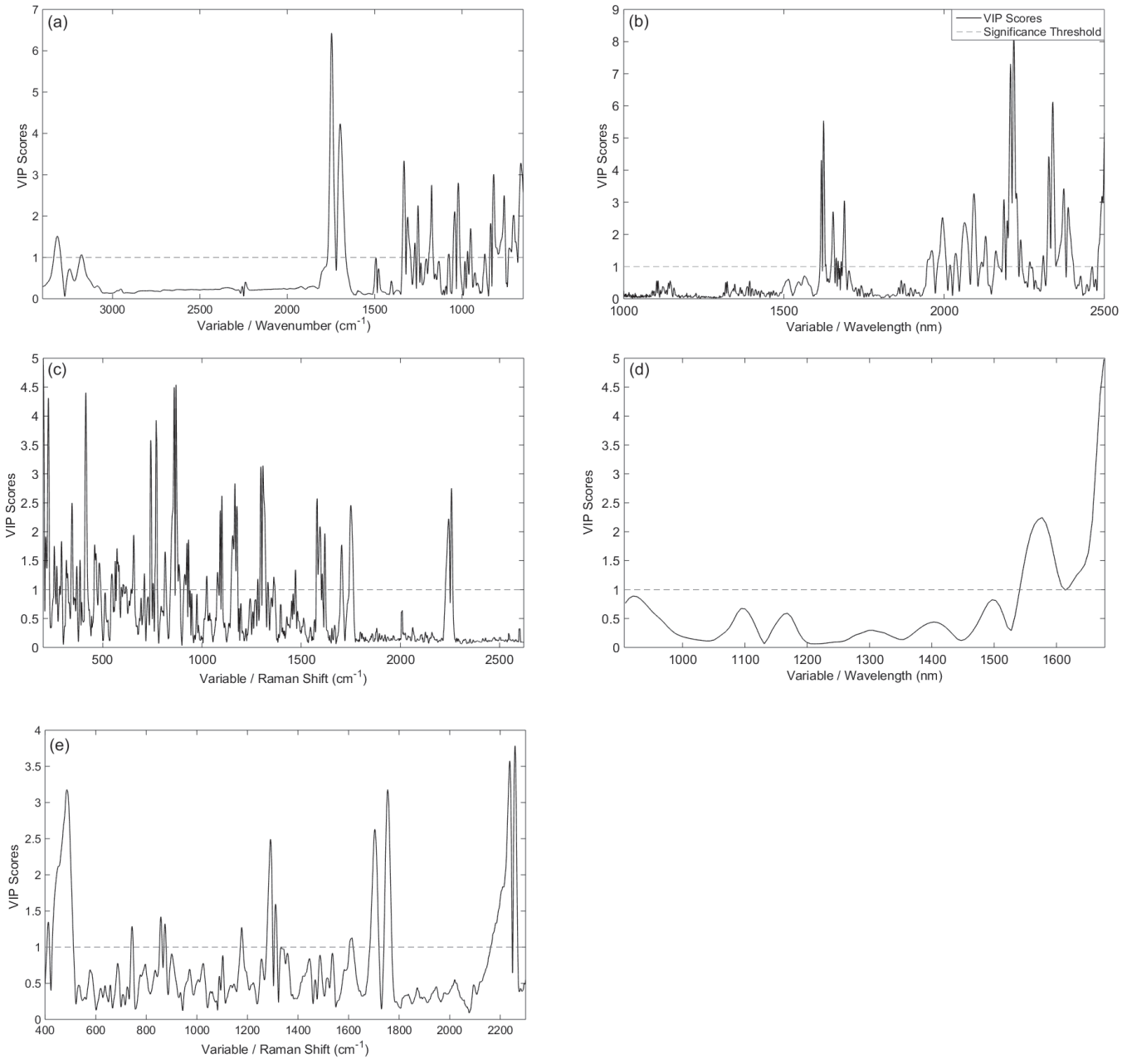


**FIGURE 6 –** VIP scores graphs. (a) MIR; (b) NIR benchtop; (c) Raman benchtop; (d) Portable NIR; (e) Portable Raman.

Figure 7 shows the spectral data selected by iPLS. The PLS models with the variables selected by iPLS resulted in best parameters for portable NIR equipment: predicted error lowered to around 4% w/w and the number of latent variables was reduced to 2. The F tests (at 95% confidence level) indicated that RMSEP from iPLS models was not considered statistically different from the RMSEP with no variable selection, but the decrease of the RMSECV value was considered statistically different (Table II).

**TABLE II -** The iPLS models obtained for EFV polymorphs. The numbers in parentheses are the latent variables

| Instruments | Models | $R^2$ | RMSE | Bias | Pre-proc. |
|---|---|---|---|---|---|
| MIR | CV | 0.98 | 3.9 (2) | 0.1 | Smooth. + MSC |
| | Pred | 0.98 | 4.1 | -0.6 | |
| NIR benchtop | CV | 0.99 | 2.9 (4) | 0.1 | 1st derivative |
| | Pred | 0.96 | 6.2 | -0.3 | |
| Portable NIR | CV | 0.99 | 3.3 (5) | -0.01 | 1st derivative |
| | Pred | 0.98 | 4.2 | 1.4 | |
| Raman benchtop | CV | 0.71 | 19.0 (7) | -3.7 | Smooth. + MSC |
| | Pred | 0.73 | 19.6 | 0.6 | |
| Portable Raman | CV | 0.97 | 5.1 (5) | 0.1 | Smooth. + SNV |
| | Pred | 0.92 | 9.0 | -1.0 | |

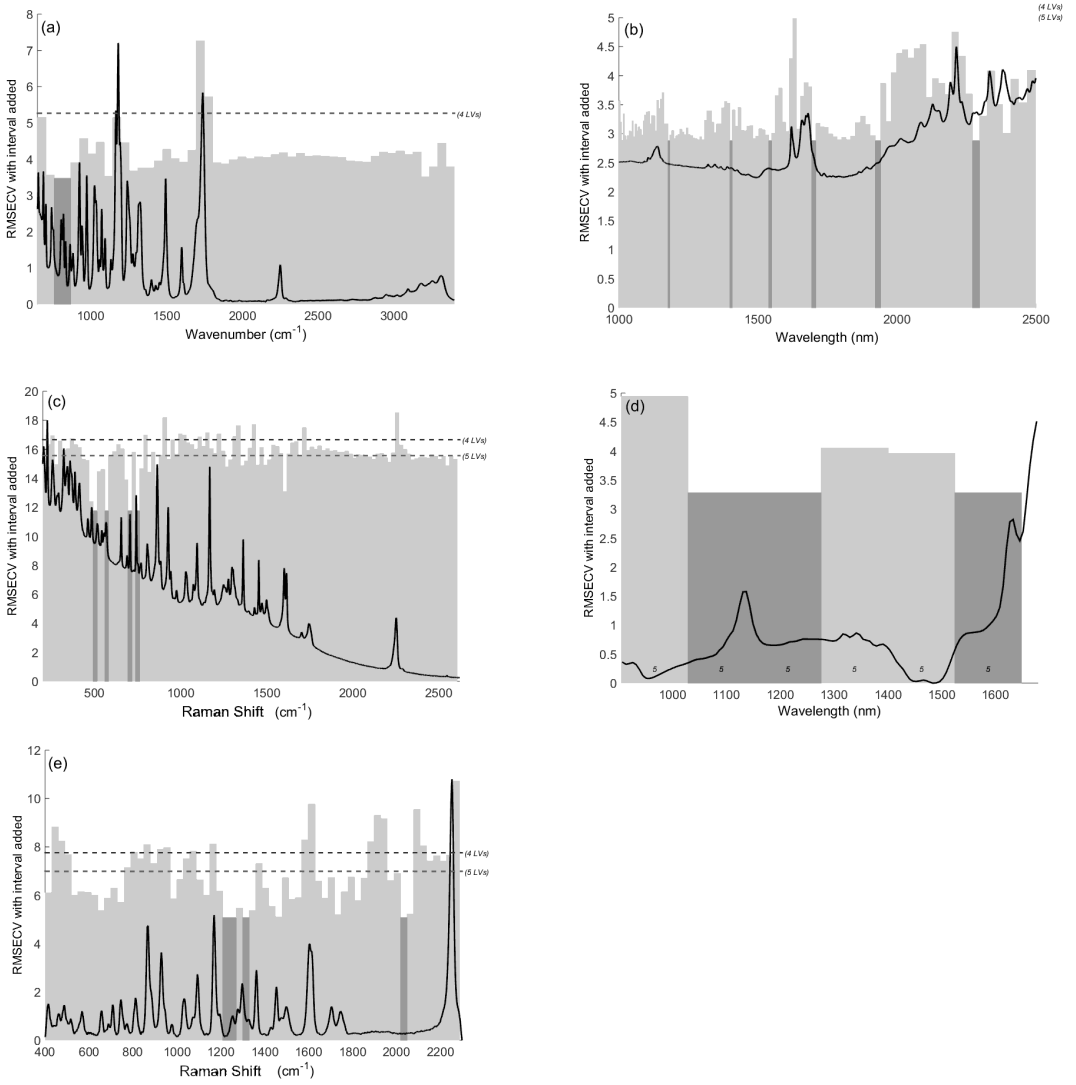


**FIGURE 7 –** Forward iPLS results. (a) MIR; (b) NIR benchtop; (c) Raman benchtop; (d) Portable NIR; (e) Portable Raman.

## DISCUSSION

The literature reports that form II of EFV can be obtained in seven days by recrystallization using methanol (2 mg.ml$^{-1}$ of EFV form I) at freezing temperature (Fandaruff *et al.*, 2014). As the freezing temperature of methanol is -97.6 °C, recrystallization in our study was initially conducted in an ultra-freezer in the range of -91 to -82 °C. Under these conditions the time to obtain the form II took an average of four weeks. We also verified that recrystallization occurred at a concentration of 20 mg.ml$^{-1}$ of EFV, which would decrease the volume of methanol needed.

Other conditions were tried using higher temperatures and different superficial areas of evaporation. The results indicated that some of the conditions that greatly accelerated evaporation, such as the use of temperatures above 4 °C, were unsuitable. However, it was possible to obtain three grams of the polymorph II in only six days at a temperature ranging between -4.5 and 4 °C.

MIR spectra (Figures 3a and 4a) showed the characteristic bands reported in the literature. There is a single band (1743 cm$^{-1}$) attributed to carbonyl group in polymorph I spectra and a coupling of modes (1735 and 1698 cm$^{-1}$) in the spectra of polymorph II. Other couplings of modes can be seen at 3249 cm$^{1}$ and at 3178 cm$^{1}$ (polymorph II), corresponding to a symmetric and asymmetric stretching of the nitrogen of the amide group, while polymorph I shows just one mode at 3314 cm$^{-1}$. According to previous works, this difference may have been due to the occurrence of synthon A for polymorph I and synthon C for polymorph II, both formed by different molecular arrangements caused by hydrogen bonds (-NH--O). Thus, synthon A would form a two-fold axis and synthon C would form a screw axis in the unit cell (Fandaruff *et al.*, 2014; Marques *et al.*, 2017; Mahapatra *et al.*, 2010; Mishra *et al.*, 2012).

NIR spectra of the EFV polymorph II (Figures 3b and 4b) have not been reported yet in literature. Organic compounds, such as EFV, that contain carboxyl and amide groups, can have the position and shape of their respective bands influenced by hydrogen bonds. Thus, the bands around 1531 nm, 1972 nm, 2084 nm, corresponding respectively to regions of overtone of amide, carbonyl and combination of amide stretching, can distinguish between forms I and II of EFV (Stuart, 2004). The spectrum from the portable NIR was also able to identify the two polymorphic forms, although the spectral range was narrower (Figure 3d).

The Raman spectra (Figures 3c and 4c) presented a baseline variation that was corrected posteriorly using ALS correction. The shape is similar to that described by Fandaruff and colleagues (2014); differences in intensity and in the form of bands of the two polymorphs can be observed (Fandaruff *et al.*, 2014). In particular, two modes that appear around 1740 cm$^{-1}$ in the polymorph II spectrum, while polymorph I shows only one mode nearly (~1752 cm$^{-1}$), which would correspond to a deformation of the amide group and stretching of the aromatic ring (Marques *et al.*, 2017). The portable Raman (Figure 3e) covered a shorter spectral range, but also included important vibrational modes attributed to all functional groups of the EFV (Mishra *et al.*, 2012).

Although the spectra of the two polymorphs have differences, none of the evaluated techniques demonstrated a unique band, without overlapping, that could be used to build univariate calibration models to quantify the content of either form. As stated by Siddiqui *et al.* (2013), the data obtained from spectroscopic instruments are multivariate in nature, so multivariate methods such as PLS provide information easier to extract.

PLS models were developed (Table I) and the performance of the portable NIR was comparable to the benchtop equipment. This result is relevant, considering that the main features of the handheld equipment are low cost and lightweight, providing a mobility that permits quality monitoring *in loco* sampling in places such as in the manufacturing production area.

In general, the Raman spectroscopy methods performed worse than the infrared spectroscopy methods. The portable Raman had a better performance (Figure 5e), presenting a lower RMSEP and a higher $R^2$ than with the benchtop device. These better values may be due to the size of the laser spot. The small size used (2 μm) by the benchtop instrument may not have been representative of the bulk proportions of the polymorphic mixtures, even after taking the means from five spectra. The portable

instrument used a much larger spot size (2-2.5 mm). In the literature, Paiva and collaborators (2018) have already described a higher error of micro-Raman in quantitative analysis of polymorphs of mebendazole compared with macro-Raman, indicating that the spot size of the laser is an important factor to be considered.

Another influence that must be considered is the particle size of the samples. Studies in the literature have reported that particle size is related to changes in spectral slope, baseline offset and signal intensity (Gómez, Coello, Maspoch, 2019). This could introduce variables into the model and, consequently, errors. The two polymorphs evaluated had different particle sizes, due to a tendency of aggregation of form I, observed at the stage of the particle calibration of the polymorphs.

Tools for variable selection were also evaluated aiming to improve the models as well as to help interpretation of the importance of each wavelength for building the models (Andersen, Bro, 2010).

The VIP score is a method that analyzes the importance of each variable by weight and the sum of squares of the explained variance. The model using MIR spectra was analyzed and the bands that were considered relevant by the VIP scores were vibrations of the nitrogen and the carbonyl of the amide group and the carbons from the rings (Figure 6a). Vibrations of fluorine and chlorine atoms were not considered relevant for the construction of the model. With the use of wavelengths with VIP scores higher than one, the number of variables was reduced from 2751 to 167, after two applications of VIP scores.

For the model built with NIR spectra, the number of variables was reduced from 3001 to 32, after selecting variables by the VIP scores. The important regions were the first overtone of CH, the aryl-CH group and combination regions of NH, CH and $CH_2$ (Figure 6b). The portable NIR had the lowest number of variables for all the equipment tested, with only 125 variables; selection by VIP scores reduced this number to 21. The selected area comprised the first overtone of CH and the aryl-CH group (Figure 6d).

For Raman spectra, the selection of variables by VIP scores reduced 4845 variables to 1038. The range of wavelength that was selected still contained information about all the functional groups. After two repeated selections by VIP scores, some bands between 200 and 1310 $cm^{-1}$ were still considered important (Figure 6c). These bands represented vibrations of the carbons at the three rings and bonds of carbon and atoms of nitrogen, fluorine and chlorine. For the portable instrument, the number of variables was lower (1901); with one application of the VIP scores selection, the number was reduced to 355. Different from the benchtop equipment, the VIP scores of the portable equipment considered the vibration of carbonyl important but did not consider as important the vibrations of groups containing atoms of nitrogen and chlorine (Figure 6e).

The other method, forward iPLS, is based on the evaluation of the intervals of fixed width in the spectral data that shows the best values for RMSECV, $R^2$ and slope (Nørgaard, 2000). For the MIR spectra, 55 $cm^{-1}$ was chosen as the interval size. The iPLS selected a fingerprint region between 870 and 761 $cm^{-1}$, which characteristically represents an area with out-of-plane CH bending vibrations from alkenes, alkynes, or aromatic groups (Figure 7a). Spectral analysis of polymorph I (Figure 3a) found three typical bands of the cyclopropane ring: 864 $cm^{-1}$ (vibrations of carbon and hydrogen bonds), 835 $cm^{-1}$ and 807 $cm^{-1}$ (rocking vibration from $CH_2$) (Silverstein, Webster, Silverstein, 2005). Polymorph II differed from polymorph I only with respect to the occurrence of a band at 821 $cm^{-1}$ instead of at 835 $cm^{-1}$, which could have been caused by the different spatial arrangements of the polymorphs. When a PLS model was built to this restricted wavelength, there was a slight improvement in the model performance parameters.

For the NIR benchtop, an interval size of 25 nm was set. The iPLS resulted in five intervals, one of them at 1183-1176 nm, without any spectral information (Figure 7b). The other intervals provided information about combination of vibrations of C-H (1398-1408 nm), a band of methine (1538-1550 nm), first overtone CH stretching (1695-1709 $cm^{-1}$), first overtone of CH (1695-1709), combination bands of NH (1923-1940 nm) and combination bands of CH and $CH_2$ in the aromatic ring (2274-2298 nm) (Stuart, 2004; Weyer, Workman, 2007). The model obtained using these wavelength intervals (Table II) did not show improvements. The RMSECV and RMSEP values were significantly different (2.9% and 6.2%) indicating a possible overfitting.

For the portable device, the iPLS considered as important two wavelength intervals, using 20 nm for interval size: 1029-1277 nm (second overtone region of CH) and 1524-1648 nm (band of methine) (Figure 7d). The iPLS resulted in an improved model, with lower RMSE and bias values and a higher $R^2$ value (Table II).

For the benchtop Raman, an interval size of 45 $cm^{-1}$ was used. Four intervals were selected as important (Figure 7c): 492-515 $cm^{-1}$ (deformations of $CF_3$ and OCO), 560-582 $cm^{-1}$ (bonds of carbon at cyclopropane ring, CNH and C≡CC), 695-717 $cm^{-1}$ and 741-762 $cm^{-1}$ (symmetric deformation of $CF_3$ and torsions of CCH) (Marques *et al*., 2017). The intervals selected did not improve the model (Table II). On the other hand, for the portable device, using an interval size of 30 $cm^{-1}$, iPLS considered three intervals as important (Figure 7e): 1210-1270 $cm^{1}$ (deformation CCH of the aromatic ring, deformation CN and stretching of $CC_{CF3}$), 1300-1330 $cm^{1}$ (stretching of the aryl ring, NC, CO and $CC_{CF3}$) and 2020-2050 $cm^{1}$ (no spectral information) (Marques *et al*., 2017). In this case, the models were not improved (Table II).

In general, there were no improvements in the models when VIP scores were used, but there were slight improvements with forward iPLS.

## CONCLUSION

In the present work, models were developed using PLS regression for quantifying polymorphic mixtures of forms I and II of EFV, using MIR and NIR spectroscopic techniques with error around 5% (w/w). Raman spectroscopy presented the worst performance, with error higher than 5% (w/w) and inadequate $R^2$ value, even with variable selection methods, possibly due to the lack of uniformity in the particle size of the samples. Besides this, the size of the laser spot was noted as making a difference, since the results were even worse for the benchtop instrument with a laser spot around 1000 times smaller than the portable instrument.

For infrared, the best results were obtained after refinement with application of variable selection methods. The forward iPLS demonstrated better performance than VIP scores in reducing the number of variables used to construct the models with lower error, which was around 4% for both the benchtop MIR and the portable NIR instrument. The adequate performance of the NIR portable instrument, beside costing less than the benchtop, facilitates the use for quality control of forms I and II of EFV in raw material and its process monitoring.

## ACKNOWLEDGEMENTS

This work was supported by NUQAAPE/FACEPE [Process No.: APQ-0346-1.06/14]; INCTAA [CNPQ grants 573894/2008-6 and 465768/2014-8, and FAPESP grants 2008/57808-1 and 2014/50951-4]; CNPq (Process No.: 428891/2018-7).

The English text of this paper has been revised by Sidney Pratt, Canadian, MAT (The Johns Hopkins University), RSAdip - TESL (Cambridge University).

## REFERENCES

Andersen CM, Bro R. Variable selection in regression-a tutorial. J Chemom, 2010;24:728–737. https://doi.org/10.1002/cem.1360.

Antonio M, Maggio RM. Assessment of mefenamic acid polymorphs in commercial tablets using chemometric coupled to MIR and NIR spectroscopies. Prediction of dissolution performance. J Pharm Biomed Anal, 2018;149:603–611. https://doi.org/10.1016/j.jpba.2017.11.053.

ASTM Standard E1655. Standard practices for infrared multivariate quantitative analysis. ASTM Internacional, West Conshohocken, PA, 2017.

Bastos MM, Costa CCP, Bezerra TC, Silva FC, Boechat N. Efavirenz a nonnucleoside reverse transcriptase inhibitor of first-generation: Approaches based on its medicinal chemistry. Eur J Med Chem, 2016; 108:455–465.

Brittain H. On the physical characterization of pharmaceutical solids. United States: Informa Healthcare USA, Inc. 2007

Chong I, Jun C. Performance of some variable selection methods when multicollinearity is present. Chemom Intell Lab Syst, 2005; 78:103–112. https://doi.org/10.1016/j.chemolab.2004.12.011.

Costa B, Vale N. Efavirenz: History, Development and Future. Biomolecules. 2023;13(1):88. https://doi.org/10.3390/biom13010088.

Eilers PHC, Boelens HFM. Baseline Correction with Asymmetric Least Squares Smoothing. Life Sci, 2005:1–24. https://doi.org/10.1021/ac034173t.

Esclusa-Diaz MT, Guimaraens-Méndez M, Pérez-Marcos MB, Vila-Jato JL, Torres-Labandeira JJ. Characterization and in vitro dissolution behaviour of ketoconazole/β- and 2-hydroxypropyl-β-cyclodextrin inclusion compounds. Int J Pharm, 1996; 143:203–210. https://doi.org/10.1016/S0378-5173(96)04704-7.

Fandaruff C, Rauber GS, Araya-Sibaja AM, Pereira RN, Campos CEM, Rocha HVA, et al. Polymorphism of anti-HIV drug efavirenz: Investigations on thermodynamic and dissolution properties. Cryst Growth Des, 2014; 14:4968–4975. https://doi.org/10.1021/cg500509c.

Farias MAS, Carneiro RL. Simultaneous Quantification of Three Polymorphic Forms of Carbamazepine using Raman Spectroscopy and Multivariate Calibration. Anal Lett, 2014; 47:6. https://doi.org/10.1080/00032719.2013.860537.

FDA - Food and Drug Administration. 2007. Draft Guidance for Industry. ANDAs: Pharmaceutical Solid Polymorphism. Chemistry, Manufacturing and Controls Information.

Fearn T, Riccioli C, Garrido-Varo C, Guerrero-Ginel JE. On the geometry of SNV and MSC. Chemom. Intell Lab Syst, 2009; 96:22–26. https://doi.org/10.1016/j.chemolab.2008.11.006.

Galvão RKH, Araujo MCU, José GE, Pontes MJC, Silva EC, Saldanha TCB. A method for calibration and validation subset partitioning. Talanta. 2005; 67:736–740. https://doi.org/10.1016/j.talanta.2005.03.025.

Gómez DA, Coello J, Maspoch S. The influence of particle size on the intensity and reproducibility of Raman spectra of compacted samples. Vib Spectrosc. 2019;100:48–56. https://doi.org/10.1016/j.vibspec.2018.10.011

Guo C, Luo X, Zhou X, Shi B, Wang J, Zhao J, Zhang X, et al. Quantitative analysis of binary polymorphs mixtures of fusidic acid by diffuse reflectance FTIR spectroscopy, diffuse reflectance FT-NIR spectroscopy, Raman spectroscopy and multivariate calibration. J Pharm Biomed Anal. 2017;140:130–136. https://doi.org/10.1016/j.jpba.2017.02.053.

ICH - International Council for Harmonization, Specifications: Test procedures and aceptance Criteria for New Drug Substances and New Drug Products: Chemical Substances Q6A, in: Intergovernmental Panel on Climate Change (Ed.), ICH Harmon Tripart Guidel, Cambridge: Cambridge University Press;1999:1–30.

Kang Y, Shao Z, Wang Q, Hu X, Yu D. Quantitation of polymorphic impurity in entecavir polymorphic mixtures using powder X-ray diffractometry and Raman spectroscopy. J Pharm Biomed Anal. 2018;158:28–37. https://doi.org/10.1016/j.jpba.2018.05.026.

Kim EJ, Kim JH, Kim M-S, Jeong SH, Choi DH. Process Analytical Technology Tools for Monitoring Pharmaceutical Unit Operations: A Control Strategy for Continuous Process Verification. Pharmaceutics. 2021;13(6):919. https://doi.org/10.3390/pharmaceutics13060919

Lindon JC, Tranter GE, Koppenaal D. 2017. Encyclopedia of Spectroscopy and Spectrometry. Elsevier ltd.

Luo J, Ying K, Bai J. Savitzky-Golay smoothing and differentiation filter for even number data. Signal Processing. 2005;85:1429–1434. https://doi.org/10.1016/j.sigpro.2005.02.002.

Mahapatra S, Thakur TS, Joseph S, Varughese S, Desiraju GR. New solid state forms of the anti-HIV drug efavirenz. Conformational flexibility and high Z′ issues. Cryst Growth Des, 2010;10:3191–3202. https://doi.org/10.1021/cg100342k.

Marques MM, Rezende CA, Lima GC, Marques ACS, Prado LD, Leal KZ, et al. New solid forms of efavirenz: Synthesis, vibrational spectroscopy and quantum chemical calculations. J Mol Struct, 2017;1137:476–484. https://doi.org/10.1016/j.molstruc.2017.02.061.

Mbuagbaw L, Mursleen S, Irlam JH, Spaulding AB, Rutherford GW, Siegfried N. Efavirenz or nevirapine in three-drug combination therapy with two nucleoside or nucleotide-reverse transcriptase inhibitors for initial treatment of HIV infection in antiretroviral-naïve individuals. Cochrane Database Syst Rev. 2016;12(12):CD004246. doi:10.1002/14651858.CD004246.pub4.

Mishra S, Tandon P, Ayala AP 2012. Study on the structure and vibrational spectra of efavirenz conformers using DFT: Comparison to experimental data. Spectrochim Acta - Part A Mol Biomol Spectrosc, 2012; 88:116–123. https://doi.org/10.1016/j.saa.2011.12.010.

Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. Appl Spectrosc, 2000; 54:413–419. https://doi.org/10.1366/0003702001949500.

Paiva EM, Silva VH, Poppi RJ, Pereira CF, Rohwedder JJR. Comparison of macro and micro Raman measurement for reliable quantitative analysis of pharmaceutical polymorphs. J Pharm Biomed Anal, 2018; 157:107–115. https://doi.org/10.1016/j.jpba.2018.05.010.

Qiu J, Li G, Sheng Y, Zhu M. Quantification of febuxostat polymorphs using powder X-ray diffraction technique. J Pharm Biomed Anal, 2015;107:298–303. https://doi.org/10.1016/j.jpba.2015.01.005.

Radesca LA, Maurin MB, Rabel SR, Moore JR, inventors. Bristol Myers Squibb Pharma Co, assignee. *Crystalline Efavirenz*. US Patent 6673372B1. 2004-01-06. [ONLINE] Avaiable at: https://patents.google.com/patent/US6673372B1/

en?q=Crystalline&q=Efavirenz&oq=Crystalline+Efavirenz [Accessed on March 25, 2024].

Ravikumar K, Sridhar B. Molecular and crystal structure of efavirenz, a potent and specific inhibitor of HIV-1 reverse transcriptase, and its monohydrate. Mol Cryst Liq Cryst, 2009;515:190–198.

Reddy BP, Reddy KR, Reddy RR, Reddy DM, Reddy KSC, inventors. Hetero Drugs Ltd, assignee. *Polymorphs of efavirenz*. US patent 8466279B2. 2013-06-18. [ONLINE] Avaiable at: https://patents.google.com/patent/US8466279 [Accessed on March 25, 2024].

Rinnan Å, Van Den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC - Trends Anal Chem, 2009; 28:1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.

Eigenvector. Advanced Preprocessing: Noise, Offset, and Baseline Filtering. 2019. Avaiable at: http://wiki.eigenvector.com/index.php?title=Advanced_Preprocessing:_Noise,_Offset,_and_Baseline_Filtering [Accessed on March 25, 2024].

Sankoh O, Arthur S, Nyide B, et al. Prevention, treatment and future challenges of HIV/AIDS: A decade of INDEPTH research. HIV AIDS Rev, 2014;14:1–8.

Siddiqui A, Rahman Z, Sayeed VA, et al. Chemometric evaluation of near infrared, fourier transform infrared, and raman spectroscopic models for the prediction of nimodipine polymorphs. J Pharm Sci, 2013; 102:4024–4035. https://doi.org/10.1002/jps.23712

Silva VH, Gonçalves JL, Vasconcelos FVC, et al. Quantitative analysis of mebendazole polymorphs in pharmaceutical raw materials using near-infrared spectroscopy. J Pharm Biomed Anal, 2015;115:587–593. https://doi.org/10.1016/j.jpba.2015.08.018.

Silva VH, Silva JJ, Pereira CF. Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer. J Pharm Biomed Anal, 2017;134:287–294. https://doi.org/10.1016/j.jpba.2016.11.036.

Silverstein KD, Webster RM, Silverstein FX. Spectrometric Identification of Organic Compounds. New York: John Wiley & Sons, Inc. 2005.

Stuart BH. Infrared Spectroscopy: Fundamentals and Applications. John Wiley & Sons. 2004.

Vrouenraets SM, Wit FW, Van Tongeren J, Lange JMA. Efavirenz: a review. Expert Opin Pharmacother, 2007; 8:851–871. https://doi.org/10.1517/14656566.8.6.851.

Weyer L, Workman Jr J. Practical Guide to Interpretive Near-Infrared Spectroscopy. CRC Press. 2007. https://doi.org/10.1201/9781420018318.

Received for publication on 27th June 2023
Accepted for publication on 21st March 2024