Comparison of different spectroscopic data and chemometric tools for determining citalopram content in the medicinal product

J.Cruz [1], G.Gergov [2], E.Kirilova[2], S.Kasabov [3], O.Kostadinova [4] , F.Svetoslavov [5]

[1] Escola Universitària Salesiana de Sarrià Passeig Sant Joan Bosco 74 08017 Barcelona, Catalonia.
[2] Bulgarian Academy of Sciences, Institute of Chemical Engineering, Acad. Georgi Bontchev Str., Bl.103, Sofia 1113, Bulgaria.
[3] Medical University of Sofia, Faculty of Pharmacy, Dunav 2, 1000 Sofia, Bulgaria.
[4] Institute of Electrochemistry and Energy Systems "Acad. E. Budevski", "Acad. Georgi Bonchev" Str. Bl.10, 1113 Sofia, Bulgaria.
[5] F&L Pharmascience, Tintyava Str. 126,Sofia 1113,Bulgaria.
E-mail (jcruz@euss.cat)

Background

Citalopram hydrobromide salt is an antidepressant belonging to a group of drugs called selective serotonin reuptake inhibitors (SSRIs), and it is used to treat major depressive disorders. In the investigation of Durby et al. [1], Near-Infrared (NIR) transmittance and Raman spectroscopy chemometric calibrations of the active substance content of a pharmaceutical tablet were developed using partial least squares regression (PLS) combined with suitable preprocessing and variable selection. In the present work, various dosage values for active substance were measured (120 tablets in total) and manufactured by pilot production scale.

Aims

Use individually and combine the Raman and NIR spectra to:

1) Perform single block NIR after suitable preprocessing using SPORT
2) Perform single block Raman after finding suitable preprocessing
3) Explore a low-level SO-PLS data fusion (Raman and NIR spectra)
4) Explore SO-Cov Sel and find suitable wavelengths for analysis
5) Compare SO-PLS and SO-Cov Sel

Materials and methods

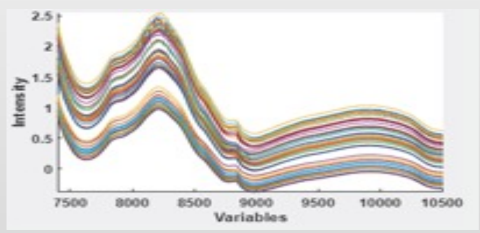
Medicinal product Citalopram IR film coated tablets with four strengths are used for this scientific work. 10, 15, and 20 mg are dose proportional (8% w/w) API per tablet and 5 mg (5,6% w/w) API per tablet. 12 pilot batches are used for analysis, shown in the Table below.

Nominal content of active substance per tablet (mg)	Nominal tablet weight (mg)	Nominal API content per tablet (%)	Number of batches
5	90	5.6	3 pilot scale
10	125	8.0	3 pilot scale
15	188	8.0	3 pilot scale
20	250	8.0	3 pilot scale

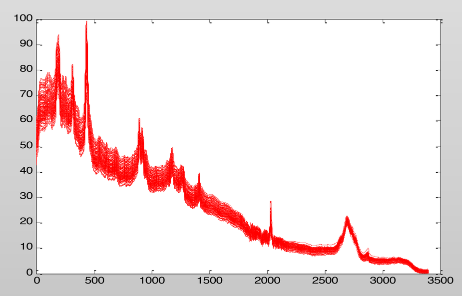
The product composition included common excipients with various properties like Microcrystalline Cellulose, Mannitol, Colloidal Silicone Dioxide, Magnesium stearate, and others. HPMC or PVA based materials are used For film coating.

Spectra collection:

- NIR transmittance
ABB Bomem FT-NIR model MB-160 : InGaAs detector



- Raman reflectance Perkin Elmer system 2000 NIR
FT Raman: YAG laser 1064.4 nm 200-3600 cm^{-1}(interval $1cm^{-1}$), Resolution: 8 cm^{-1}



- Reference method: HPLC

Software used:

1. Pretreatment selection: **SPORT**(sequential preprocessing through orthogonalization)
2. Calibration and validation data set selection: **Duplex** algorithm: 60 samples training, 60 samples testing set
3. Spectra multiplicative effects, and baseline correction: **VSN** (Variable sorting for normalization), **PQN** (probabilistic quotient normalization) preprocessing), **airPLS** (asymmetrically reweighted penalized least squares) baseline removal)
4. Spectra fusion (NIR+Raman): **SO-PLS** (Sequential and Orthogonalized PLS).
5. Variable selection: **SO-CovSel** (Sequential and Orthogonalized Covariance Selection)

Note: All these algorithms written in Matlab.

Total spectra used:
NIR: 120 (samples) at 404 wavelengths
Raman: 120 (samples) at 3401 wavelengths

Total spectra after fusion:
Raman and NIR - 3805 wavelengths

Prediction accuracy:

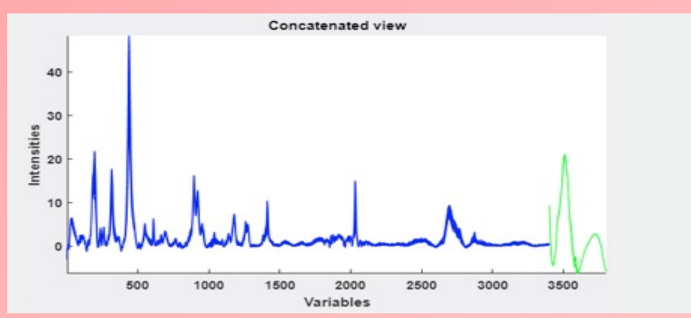
$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

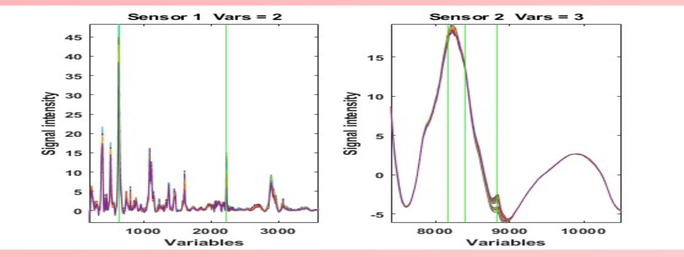
$$\mathrm{RSEP}\,(\%) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i)^2}} \times 100$$

Results and discussion

Fusion data strategies

Concatenation Raman NIR:





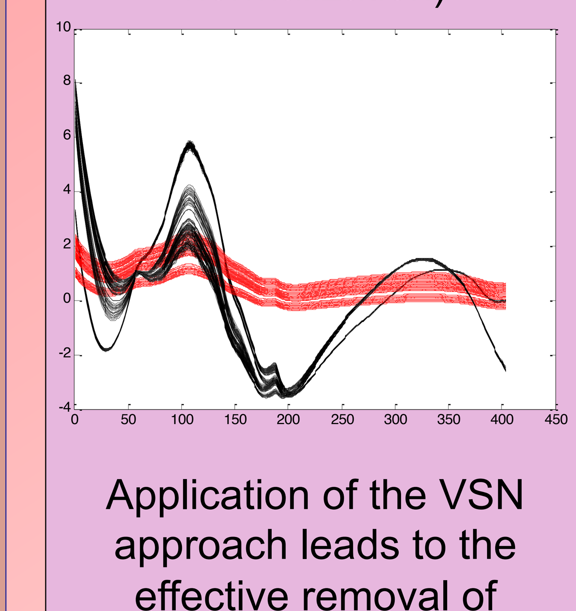
Raman vibrational band, cm−1	Raman assignment	IR vibrational band, cm^{-1}	IR assignment	NIR vibrational band, cm−1	NIR assignment
518	B_{1g} in TiO_2				
763	C-F stretching	1390	C-F stretching		
1056	C-O vibration in furan				
2230	-C≡N symmetric stretching	2238	-C≡N sym. stretching		
2838	CH_3 symmetric stretching	2853	C–H str. (aliphatic)		
3063	C-H symmetric stretching	3062	C–H str. (aromatic)		
3074	C-H symmetric stretching of benzene upon fluorine substitution	3090	C–H str. (aromatic)		
				8830	Second overtone C–H str. (aromatic)

Model	LVs	R^2c	RMSEC	R^2p	RMSEP	RSEP%
Raman PLS	6	0.92	0.298	0.53	0.827	11.21
NIR PLS VSN	3	0.95	0.232	0.93	0.329	4.46
Multiblock C PLS (NIR+RAMAN)	6	0.92	0.292	0.55	0.807	10.94
SO PLS (Raman PQN/airPLS) (NIR VSN)	2-4	0.96	0.199	0.91	0.363	4.92
SO PLS (Raman PQN/airPLS) (NIR VSN) COV SEL	5	0.94	0.263	0.92	0.344	4.66

Raman NIR SO-PLS gives comparable results to a single NIR PLS analysis.

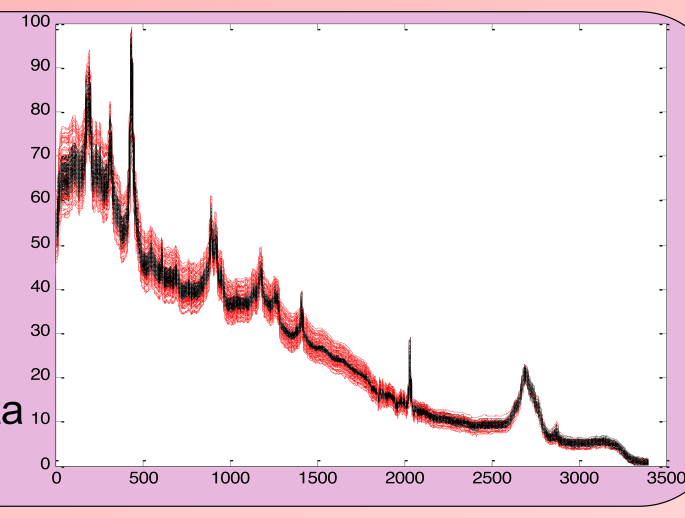
Preprocessing strategy for NIR and Raman

VSN
(Variable sorting for normalization)

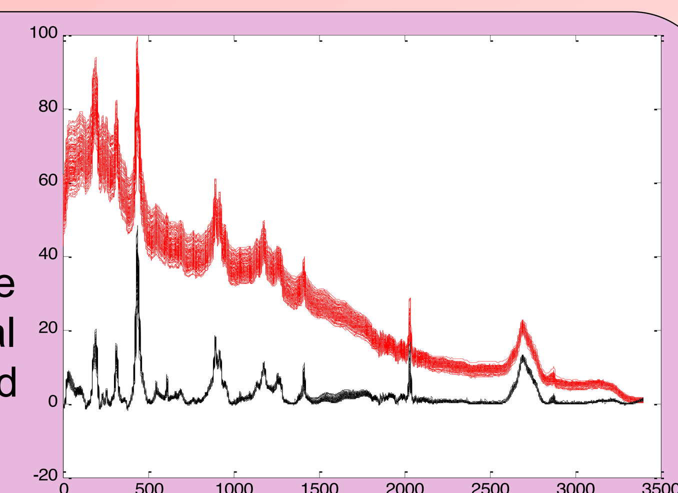


Application of the VSN approach leads to the effective removal of multiplicative effects [3].

PQN
(probabilistic quotient normalization) preprocessing – original and preprocessed data [3].



airPLS
(asymmetrically reweighted penalized least squares) baseline removal – original and preprocessed data [2].



Conclusions

- Duplex algorithm was applied to the reference values "y" instead of spectral data X to select the calibration and test dataset. The performance of the Duplex algorithm outperformed classical Kennard Stone.
- Appropriate preprocessing had to be found before the fusion of Raman and NIR data.
- VSN outperformed SNV (Standard Normal Variate) in removing multiplicative effect in NIR spectra tablet [3].
- Combining PQN and airPLS is the best preprocessing to remove baseline shift in Raman spectra in tablets [2,3].
- Multiblock C-PLS (simple concatenation) does not improve models compared to a single Raman and NIR PLS models [4].
- Multiblock data fusion method SO-PLS improve models compared to single Raman PLS model and performs similarly to a single NIR PLS model [5].
- SO-CovSel performs variable selection so that extraction of the variables from the consecutive block improves the model [6].

Future work

- Remove the range in which excipient TiO2 absorbs in order to improve single Raman calibration
- Development of film coating system with replacement of TiO2 with another opacifier to improve single Raman calibration and to fulfil the future requirements of TiO2 free products.
- Perform middle level SO PLS Raman NIR data fusion

References

1.Dyrby, M., et al. "Chemometric quantitation of the active substance (containing C≡ N) in a pharmaceutical tablet using near-infrared (NIR) transmittance and NIR FT-Raman spectra." Applied spectroscopy 56, no. 5 (2002): 579-585.
2.Zhang, Z. M., et al. (2010). Baseline correction using adaptive iteratively reweighted penalized least squares. Analyst, 135(5), 1138-1146.
3. J.-M. Roger, J.-C. Boulet, M. Zeaiter, D.N. Rutledge, in: second ed., in: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, vol. 3, Elsevier, Amsterdam, The Netherlands, 2020, pp. 1–75.
4.Azcarate, S. M., et al. (2021). Data handling in data fusion: Methodologies and applications. TrAC Trends in Analytical Chemistry, 143, 1163.
5. A. Biancolillo, T. Næs, The sequential and orthogonalized PLS regression for multiblock regression: theory, examples, and extensions, in: M. Cocchi (Ed.), Data Fusion Methodologies and Applications, Data Handling in Science and Technology, vol. 31, Elsevier, Oxford, UK, 2019, pp. 157–177
6. A. Biancolillo, F. Marini, J.-M. Roger, SO-CovSel, A novel method for variable selection in a multi-block framework, J. Chemometr. 34 (2020), e3120.

Acknowledgments

The work was supported in part by collaboration with REDTPI4.0-320RT0006 CYTED program.