HUE UNIVERSITY
UNIVERSITY OF SCIENCES

Nguyen Duy Luu

SIMULTANEOUS DETERMINATION OF SEVERAL PHARMACEUTICAL COMPOUNDS WITH OVERLAPPING UV-VIS ABSORPTION SPECTRA USING THE PARTIAL LEAST SQUARES METHOD

Major: Analytical Chemistry
Code: 9440118

PhD DISSERTATION ABSTRACT

Supervisors:
1. Assoc. Prof Tran Thuc Binh
2. Assoc. Prof Nguyen Dinh Luyen

HUE, 2025

The dissertation has been complete at: .....

.....

.....

Supervisors: .....

.....

.....

Reviewer 1:.....

.....

.....

Reviewer 2:.....

.....

.....

Reviewer 3: .....

.....

.....

The dissertation was defended in front of the dissertation committee of: .....

.....

at :…….hours on the ..... day of ……., ……. year

The dissertation will be found at the libraries:……

……

INTRODUCTION

According to a report by the World Health Organization (WHO) in 2017, it was estimated that 1 in 10 medical products circulating in low- and middle-income countries is substandard or falsified. This means that people are using medicines that are unable to treat or prevent diseases. WHO identified that counterfeit drugs may account for about 50% of the global pharmaceutical market, with many of them originating from developing countries. According to reports submitted to WHO from 20 different countries, most counterfeit drugs fall into one of the following three categories: (1) products containing no active ingredients — about 30%; (2) products containing incorrect amounts of active ingredients — about 20%; and (3) products containing the wrong ingredients — about 20%. Therefore, Quality Control is an essential and indispensable factor in the production, circulation, and use of pharmaceuticals. With the rapid development of the pharmaceutical industry, the demand for drug quality has become increasingly stringent. Regulatory agencies worldwide — including the U.S. Food and Drug Administration, the European Medicines Agency, and others — emphasize the necessity of strict quality control throughout the manufacturing and distribution processes. This is to ensure that pharmaceutical products not only meet therapeutic efficacy standards but also guarantee absolute safety for users.

With the increasing diversity and complexity of pharmaceutical products today, the development of rapid, accurate, and reliable analytical methods has become an essential requirement. Currently, in the monographs for the analysis of active pharmaceutical ingredients specified in various pharmacopoeias, most are individual analytical procedures. This means that the analytical process must be accompanied by an extraction or separation step when dealing with multicomponent drugs, making the overall process time-consuming, complex, and costly in terms of reagent consumption. Meanwhile, simultaneous analysis not only saves time and cost but also minimizes errors arising from the extraction and separation processes. This situation highlights the necessity for regulatory authorities and researchers to study, improve, and develop more simultaneous

analytical methods to enhance efficiency in pharmaceutical quality control.

It can be observed that the group of physicochemical analytical methods has been increasingly developed and widely applied in pharmaceutical analysis, accounting for the majority of analytical procedures in pharmacopoeial monographs. The clear advantages of these methods have been well demonstrated — including high accuracy, precision, and reliability. However, techniques such as high-performance liquid chromatography, gas chromatography, and mass spectrometryinvolve complex, expensive instruments, strict analytical procedures, and require operation by highly skilled professionals. In this context, the UV-Vis spectrophotometric method can overcome several disadvantages of those instrumental techniques. Nevertheless, one major challenge of the UV-Vis method arises when analyzing multicomponent pharmaceutical samples whose spectra overlap or when interferences from impurities and excipients are present, making it difficult to accurately determine the concentrations of individual components in the sample.

In 1972, Swede Svante Wold and Bruce R. Kowalski introduced the term chemometrics, which was defined as a branch of chemistry that employs mathematical and statistical methods to design or select measurement procedures, optimize experiments, and maximize the extraction of chemical information through the analysis of chemical data. Since the 1980s, with the advent of personal computers, a new era began for the collection, processing, and interpretation of chemical data. Up to now, the rapid and powerful development of computer science, along with the emergence of artificial intelligence (AI) models, has made data processing easier and more accurate. This has provided significant advantages in the creation, improvement, and advancement of analytical methods. The limitations of UV-Vis spectrophotometric analysis can be thoroughly and accurately addressed by combining spectral data with mathematical modeling. A variety of mathematical models have been applied to spectral data processing, ranging from Classical Least Squares (CLS), Inverse Least Squares (ILS), Principal Component Regression (PCR), and Partial Least Squares (PLS) to more advanced machine learning approaches

such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Genetic Algorithms (GA)...

When measuring the UV absorbance of a solution, it is possible to accurately and rapidly obtain its entire spectrum with low noise, compared to measuring absorbance at a single wavelength. The absorption spectrum obtained from the sample measurement provides information about all the components present in the sample as well as the interactions between them. This represents a significant advantage over univariate analysis, in which only information about a single component is obtained while the interactions among coexisting substances in a mixture are ignored. This is one of the key advantages of multivariate analysis.

The Partial Least Squares (PLS) model is often described as a principal regression technique used when dealing with multivariate data. Its successful applications in fields such as Quantitative Structure–Activity Relationship (QSAR) studies, metabolomics, and psychometrics have demonstrated its effectiveness. Subsequently, PLS has been widely applied in various spectroscopic techniques such as near-infrared (NIR), Raman, UV-Vis spectroscopy, as well as in chromatographic and electrochemical methods, highlighting the importance of this approach in data processing and result computation. The advantages of the PLS model include: data dimensionality reduction, handling multicollinearity, eliminating noise effects, and the ability to analyze individual components within a mixture without requiring complete prior knowledge of the analytical sample.

In particular, in the analysis of pharmaceutical mixtures with overlapping UV-Vis spectra, the PLS method allows the simultaneous determination of multiple active ingredients without the need for separation or extraction, by constructing a regression model based on spectral data and reference concentrations. As a result, the combination of PLS with UV-Vis spectroscopy not only enables spectral overlap resolution but also enhances accuracy, reduces cost, and shortens analysis time. These outstanding advantages make this approach a highly practical and promising research direction in the quality control of multicomponent pharmaceutical formulations.

For the above reasons, we have chosen the dissertation topic: "**Simultaneous determination of several pharmaceutical compounds with overlapping uv-vis absorption spectra using the partial least squares method**"

Chapter 1. OVERVIEW

1.1. UV-Vis spectrophotometric method
1.2. Chemometric methods
1.3. Some multivariate analysis methods
1.4. Overview of the application of the PLS method combined with spectroscopic techniques in compound analysis
1.5. Limit of detection in multivariate analysis
1.6. Wavelength selection in spectroscopic analysis
1.7. Introduction to Amlodipine besylate, Indapamide, Atorvastatin calcium, Spironolactone, Furosemide, Rifampicin, Isoniazid, Pyrazinamide, Metronidazole, and Spiramycin.

Chapter 2. RESEARCH CONTENTS AND METHODS

2.1. Research content

- Development of multicomponent quantitative models using PLS–UV for two- and three-component systems with overlapping absorption spectra; calibration sets are designed based on the Central Composite Design (CCD), latent variables are selected using the PRESS criterion, and model validation is performed through statistical parameters (RMSEC, RMSECV, RMSEP, R^2, residual analysis, recovery, and repeatability).
- Applying the PLS–UV models to commercial drugs (AgimDogyl, Zoamco-A, Spirosemid-40, Natrixam, Turbezid) for simultaneous quantification of active ingredients, with results compared to HPLC using t-tests.
- Comparing PLS performance with PCR and CLS to assess advantages and noise-reduction in multicomponent UV-Vis analysis.
- Optimizing wavelength selection for PLS using SR, BVE, and GA algorithms to reduce input variables without compromising prediction accuracy.
- Proposing and standardized a practical procedure for applying PLS–UV in multicomponent pharmaceutical analysis, supporting implementation in QC/QA laboratories.

2.2. Research methods

2.2.1. Experimental Design

2.2.2. Cross-Validation Method

2.2.4. Limit of Detection in the PLS Model

2.2.5. Wavelength Selection by the Selectivity Ratio Method in PLS

2.2.6. Wavelength Selection by the Backward Variable Elimination Method in PLS

2.2.7. Wavelength Selection by the Genetic Algorithm Method in PLS

2.3. Experimental Work

Chapter 3. RESULTSS AND DISCUSSION

3.1 Experimental Design and Spectral Characteristics of the Studied Active Ingredients

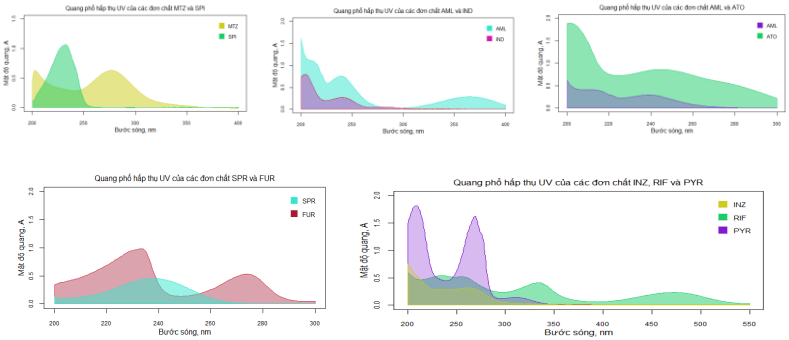


Figure 3.1. Absorption spectra of the individual components in the pharmaceutical samples

The concentrations of the substances in the calibration set used for model construction were designed according to the Central Composite Design (CCD), with the central concentrations corresponding to the ratio of active ingredients present in the pharmaceutical samples. The absorption spectra of the individual components in the analyzed samples exhibit a high degree of overlap, as shown in Figure 3.1.

3.2 PLS Combined with UV-Vis Absorption Spectroscopy

3.2.1 Results of the PLS Algorithm

3.2.1.1 Score Plot in the PLS Model

The PLS method is an approach in which the original data are represented in a latent variable space. The latent variables are expressed as a score matrix, which provides a summarized representation of the original data. In this case, the dimensionality of the data has been reduced from 401 wavelengths to 25 principal components, with each principal component corresponding to a new dimension in the latent space.

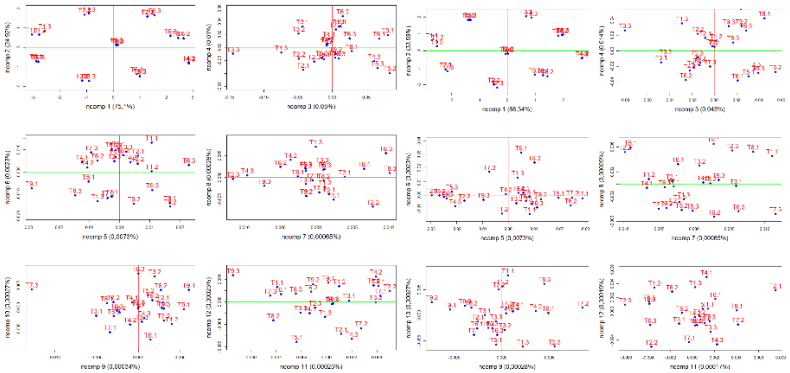


Figure 3.2. Score plot of the first 12 latent variables for MTZ and SPI

The score plot in Figure 3.2 shows that the PLS model can transform the original data into the latent variable space, revealing the relationship between the concentration information of the analytes and the latent variables, while also explaining the noise factors present in the data.

3.2.1.2. Loading (P), weight loading (W), projection (P_{proj}) and regression coefficients of the PLS model

The spectra of the projection vectors of the spectral data onto the latent space for MTZ and SPI show a strong similarity to the spectra of the pure components (first latent variable), indicating that the characteristic properties of each compound in the original mixture can be well represented. When the number of latent variables increases, the spectra exhibit stronger fluctuations, suggesting uncertainty that may arise from noise in the original data. This demonstrates the critical importance of selecting an appropriate number of latent variables to retain when constructing the regression

model in PLS. This observation is also consistent with the regression coefficient plot, which illustrates the relationship between the independent variables and the concentrations: when too many latent variables are included, the regression coefficients fluctuate significantly, indicating that even small variations in latent variable data can lead to large changes in the predicted concentrations.

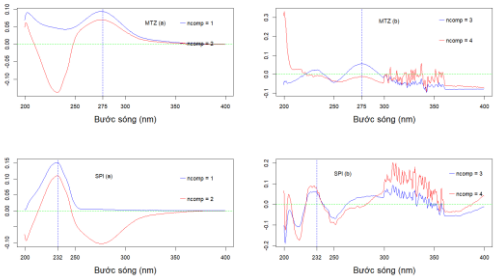


Figure *3.1.* *Spectra of the projection vectors of the original spectral data onto the PLS latent space for MTZ and SPI using four latent variables.*

3.2.2 Results of determining the number of LVs in the PLS model

3.2.2.1 Selection of the number of latent variables based on the minimum PRESS value

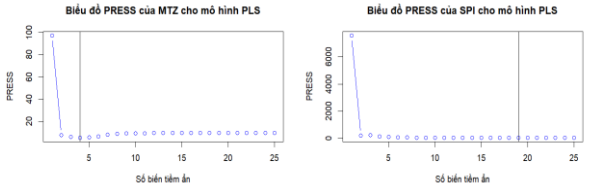


Figure **3.2.** PRESS value plot in the PLS model for MTZ and SPI

The results of the PLS model show that the original data were reduced to 25 latent variables, with each latent variable corresponding to one PRESS value. In this model, the minimum PRESS values correspond to the 4th and 19th latent variables for MTZ and SPI, respectively.

3.2.2.2 *Selection of the Number of Latent Variables Based on Statistical Testing for the Calibration Set*

According to Haaland and Thomas, the latent variable with the smallest F-ratio having a probability less than 0.75 is selected as the optimal latent variable for building the PLS model. The p-values in the Fisher distribution were calculated with degrees of freedom equal

to the number of calibration samples (27) for both the numerator and denominator. As a result, the number of latent variables selected for constructing the PLS model was 3 for MTZ and 8 for SPI.

3.2.2.3 Selection of the Number of Latent Variables Based on Statistical Testing for the Validation Set

Similar to the selection based on the PRESS value for the calibration set, the F-ratio was also calculated to determine the optimal latent variables for the validation set. The results showed that the number of latent variables selected for MTZ and SPI were 2 and 7, respectively.

3.2.3 Evaluation results of the PLS model for the simultaneous determination of MTZ and SPI

Table 3.1. *Statistical parameters of the PLS model for MTZ and SPI at the selected latent variables*

LVs		n = 2	n = 3	n = 4	n = 7	n = 8	n = 19
RMSECV (µg/mL)		0,540	0,474	0,446	1,60	1,46	1,37
RMSEC (µg/mL)		0,483	0,407	0,347	0,294	0,294	0
Shapiro-Wilk	W	0,928	0,921	0,896	0,977	0,960	-
	p	0,061	0,043	0,011	0,789	0,373	-
R^2		0,9894	0,9925	0,9945	0,9995	0,9999	1
LOD_{min} (µg/mL)		0,926	0,867	0,816	2,63	2,73	1,18
LOD_{max} (µg/mL)		1,02	1,14	1,07	3,19	3,29	1,50

The PLS model was evaluated using parameters such as RMSECV, RMSEC, residual analysis, determination coefficient (R^2), and limit of detection (LOD). The results showed that MTZ required two latent variables, while SPI required seven, to achieve statistically optimal values. The RMSE values were less than 5% compared to the mean concentration, the residuals followed a normal distribution, and the determination coefficients (R^2) were greater than 0.99. The low LOD values indicated that the developed PLS model was capable of detecting very small amounts of the analytes.

The established PLS model was then applied to predict the concentrations of samples in the validation set, which were prepared in the laboratory with varying concentration ratios. The average recoveries of MTZ and SPI were 100.9% and 99.3%, respectively, with corresponding RSD% values of 2.06% and 1.658%, meeting the

AOAC acceptance criteria. The RMSEP values were below 2%, demonstrating the excellent predictive performance of the model.

3.2.4 Validation of the PLS method combined with UV-Vis spectroscopy for the simultaneous quantification of MTZ and SPI.

3.2.4.1 Accuracy

The accuracy of the developed PLS method for the simultaneous quantification of MTZ and SPI was assessed using two approaches:

Method 1: Evaluation based on recovery by spiking known amounts of MTZ and SPI at three concentration levels — low, medium, and high. The results showed average recoveries for MTZ of 93.0% (RSD = 3.08), 98.6% (RSD = 1.50), and 96.3% (RSD = 1.01), and for SPI of 100.2% (RSD = 1.05), 101.4% (RSD = 2.08), and 104.9% (RSD = 0.71). These results comply with AOAC requirements for concentrations below 100 µg/mL, which specify acceptable recovery values between 90% and 107%.

Method 2: Comparison with the HPLC method for MTZ quantification and the microbiological method for SPI quantification, as specified in the Vietnamese Pharmacopoeia V. The results indicated no statistically significant difference between the mean values obtained by the two methods ($p > 0.05$).

3.2.4.2 Precision

Precision was assessed through repeatability studies conducted over two different days. On each day, six replicate analyses were performed, starting from sample weighing. Day 1: $\bar{m}_{MTZ} = 14.96$ µg/mL (RSD = 1.303%), $\bar{m}_{MTZ} =$ 88.86 IU/mL (RSD = 1.511%). Day 2: $\bar{m}_{MTZ} = 14.98$ µg/mL (RSD = 1.275%), $\bar{m}_{MTZ} = 88.06$ IU/mL (RSD = 0.907%). All RSD values were below the AOAC acceptance limit of 5.3%, indicating good repeatability of the analytical results.

3.2.5 Quantitative Results of MTZ and SPI in AgimDogyl Tablets

The quantitative determination of MTZ and SPI in AgimDogyl tablets (labeled content: 125 mg / 750,000 IU): m_{MTZ} =124,38 ± 1,26 mg and m_{SPI} = 735971 ± 3370 IU. These results indicate that the contents of both MTZ and SPI are within the acceptable limits specified in the Vietnamese Pharmacopoeia V.

3.2.6. Application of the PLS Method for Simultaneous Quantification of Active Ingredients in Multicomponent Drugs

3.2.6.1 For two-component Drugs

a. Development of PLS Models for Two-Component Formulations

Bảng 3.2. *Statistical parameters used to evaluate the calibration sets of Spirosemid, Zoamco-A, and Natrixam using the PLS method*

	Spirosemid		Zoamco-A		Natrixam	
	SPR	FUR	AML	ATO	IND	AML
Wavelength range (nm)	200 - 400		230-260		200 - 400	
Scan (Δλ)	0,5		0,2		0,2	
n	2	2	3	2	4	2
RMSECV (µg/mL)	0,190	0,212	0,129	0,489	0,128	0,385
RMSEC (µg/mL)	0,260	0,280	0,057	0,314	0,109	0,354
Shapiro- W	0,8842	0,9742	0,9062	0,8676	0,9093	0,8871
Wilk test p	0,1747	0,9278	0,2903	0,116	0,2395	0,128
R^2	0,9980	0,9977	0,9991	0,9929	0,9902	0,9992
LOD_{min} (µg/mL)	0,375	0,272	1,00	4,69	0,269	1,263
LOD_{max} (µg/mL)	0,420	0,320	1,06	4,83	0,338	1,708

The PLS models for the calibration sets of Spiromid-40, Zoamco-A, and Natrixam were constructed, and the statistical parameters were calculated and presented in Table 3.12. The results show that all statistical parameters used to assess model accuracy met the required criteria and were applied to simultaneously determine the components present in the two-component pharmaceutical samples.

The developed PLS models were applied to quantify the components in the test samples. The recovery ranges were as follows: SPR and FUR: 92.8–106.3% and 97.0–108.0%; AML and ATO: 97.3–104.5% and 98.0–105.2%; IND and AML: 96.9–103.9% and 97.3–106.3%. The RMSE values were less than 2% or within the range of 2–5%.

b. *Evaluation results of the PLS model applied to the simultaneous quantification of active ingredients in two-component drugs*

Accuracy: For Spiromid, the recovery rates of SPR and FUR were in the ranges of 98.4 – 107.8% and 101.4 – 105.3%, respectively; For Zoamco-A, the recovery rates of AML and ATO were 101.6 –

108.1% and 99.1 – 107.2%, respectively, For Natrixam, the recovery rates of IND and AML were 96.7 – 104.7% and 101.1 – 104.6%, respectively.

Repeatability: : For Spiromid: SPR: $\overline{m}$ = 9,76 µg/mL (RSD = 2,54), FUR: $\overline{m}$ = 7,76 µg/mL (RSD = 2,18); For Zoamco-A: AML $\overline{m}$ = 10,44 µg/mL (RSD = 1,48), ATO $\overline{m}$ = 20,81 µg/mL (RSD = 2,49); For Natrixam: IND: $\overline{m}$ = 4,35 µg/mL (RSD = 1,32), AML: $\overline{m}$ = 14,45 µg/mL (RSD = 0,65).

The results show that the PLS models for Spiromid-40, Zoamco-A, and Natrixam met the required criteria for both accuracy and precision.

c. Quantitative results

For Spiromid tablets:: m_{SPR} = 49.65 ± 3.21 mg/ tablet; m_{FUR} = 39.39 ± 1.93 mg/ tablet; For Zoamco-A tablets: m_{AML} = 5.23 ± 0.24 mg/ tablet; m_{ATO} = 10.39 ± 0.77 mg/ tablet; For Natrixam 1.5/5 tablets: 1,5/5: m_{IND} = 1.47 ± 0.05 mg/ tablet; m_{AML} = 4.83 ± 0.10 mg/ tablet (confidence interval 95%). The percentage content relative to the labeled amount of each drug was also calculated. The obtained values for all active ingredients were within the 90–110% limits specified by the Vietnamese Pharmacopoeia V.

A comparison between the mean results obtained by the PLS method and those obtained by the HPLC method showed no statistically significant difference ($t_{cal} < t_{table} = 2.775$).

3.2.6.2 For three-component drugs

a. Construction of PLS models for three-component drugs

Statistical parameters were calculated to evaluate the performance of the PLS models, and the results are presented in Table 3.18.

The established PLS model was applied to determine the concentrations of RIF, INZ, and PYR in laboratory-prepared samples. The average recoveries were 100.0%, 101.5%, and 100.2% for RIF, INZ, and PYR, respectively. The low RMSEP values indicate the strong predictive capability of the PLS method combined with UV–Vis spectroscopy, demonstrating its applicability for the simultaneous quantification of RIF, INZ, and PYR in multicomponent pharmaceutical formulations.

Table 3.3. Statistical parameters for evaluating calibration sets using the PLS method.

		RIF	INZ	PYR
Wavelength range (nm)		220-440	220 - 425	220 - 300
Scan (Δλ)			0,5	
LVs		3	4	3
RMSECV (µg/mL)		0,136	0,227	0,367
RMSEC (µg/mL)		0,107	0,184	0,317
Shapiro-Wilk test	W	0,9825	0,9816	0,9452
	p	0,888	0,8661	0,1254
R^2		0,9944	0,9836	0,9929
LOD_{min} (µg/mL)		0,43	0,87	1,30
LOD_{max}(µg/mL)		0,45	0,97	1,35

b. Evaluation of the PLS model applied for simultaneous quantification of active pharmaceutical ingredients in multicomponent formulations

The PLS method combined with UV–Vis spectroscopy successfully determined the concentrations of RIF, INZ, and PYR in the analyzed samples, with recoveries ranging from 97.6–102.4%, 95.4–108.2%, and 95.9–102.1%, respectively. These results comply with AOAC performance criteria.

The mean concentrations of RIF, INZ, and PYR in the repeatability assessment on Day 1 were 9.13, 4.57, and 24.53 µg/mL, respectively; and on Day 2 were 9.10, 4.60, and 24.41 µg/mL, respectively. The repeatability results were satisfactory, with RSD (%) values lower than those calculated using the Horwitz equation. Comparison of the mean values between the two trials revealed no statistically significant differences (t_calculated < t_table), indicating that the analytical method meets the precision requirements.

c. Quantitative results

m_{RIF} = 154,11 ± 6.36 mg/table; m_{INZ} = 77.39 ± 2.74 mg/table; m_{PYR} = 407.83 ± 10.40 mg/tabke. The percentage of the measured content relative to the labeled claim complies with the requirements of the Vietnamese Pharmacopoeia V.

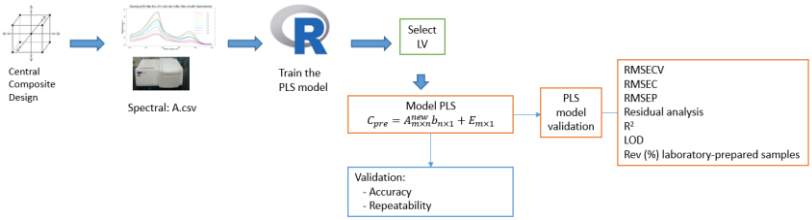
Comparison of the mean values obtained by the two methods shows that the average concentrations of RIF, INZ, and PYR

determined using the PLS–UV–Vis spectrophotometric method are not significantly different from those obtained by the HPLC reference method.

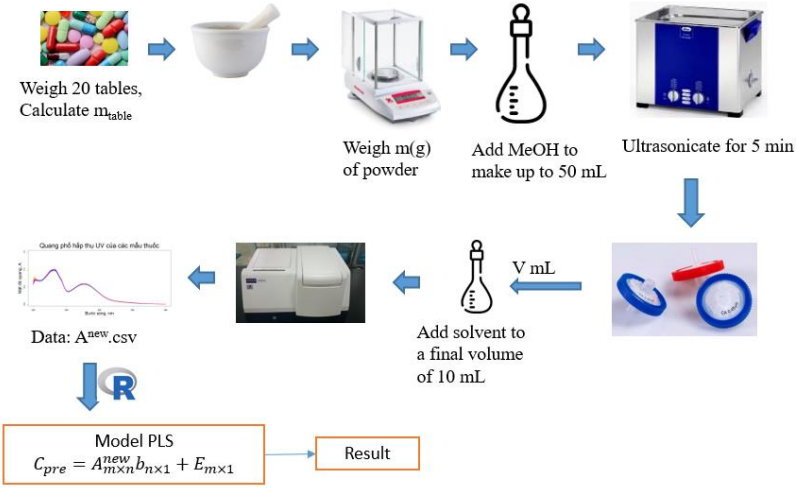
3.2.7 Conclusion of the PLS Method

We propose a general procedure for using the PLS method combined with UV–Vis absorption spectroscopy for the simultaneous quantification of compounds with overlapping spectra. The procedure consists of two stages:

*** Stage 1:** PLS model train:



*** Stage 2:** Determination of analyte concentration in real sample



3.3 Wavelength selection in PLS

3.3.1 Wavelength selection using the selectivity ratio combined with Partial Least Squares method (SR-PLS)

3.3.1.1 *SR calculation results for each wavelength*

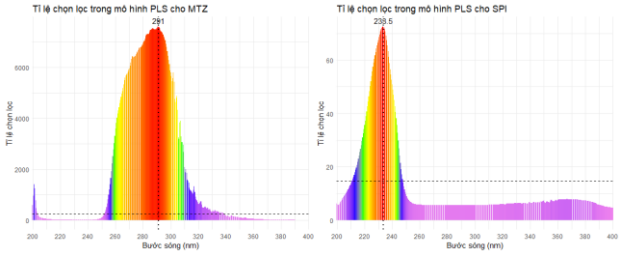


Figure 3.5. *Results of SR value determination for MTZ and SPI in the established PLS model***.**

For MTZ, the SR values are mainly concentrated in the wavelength range of 250–350 nm. For SPI, the SR values in the wavelength range of 210–247 nm are much higher than those at other wavelengths (Figure 3.5).

A similar evaluation was performed for IND and AML in Natrixam 1.5/5 tablets. The SR values at different wavelengths for each compound, IND and AML, are presented in Figure 3.6. The results indicate that for wavelengths above 320 nm, the SR values for IND are relatively low, which aligns with the fact that pure IND does not absorb in this wavelength range. In contrast, for AML, the SR values in the 320–400 nm range are very high, while they remain very low at other wavelengths.

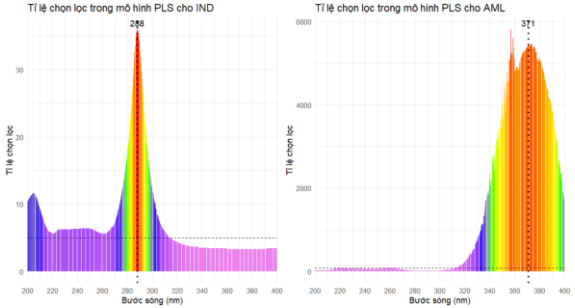


Figure 3.6. *SR values for IND and AML*

3.3.1.2 *Selection of Important Wavelengths Using the SR Method*

Table 3.4. *Wavelengths selected by the SR method for MTZ and SPI in the PLS model*

Drug	Analyte	number of selcted wavelengths	initial number of wavelengths	ratio (%)
AgimDogyl	MTZ	181 (252 – 338,5 and 200 – 203 nm)	401	45,1%
	SPI	74 (211 – 247,5 nm)	401	18,4%
Natrixam 1,5/5	IND	569 (200 – 313,6 nm)	1001	56,8%
	AML	514 (219 – 222 and 300,6 – 400 nm)	1001	51,3%

3.3.2 Wavelength Selection Using the Backward Variable Elimination Combined with Partial Least Squares (BVE-PLS) Method

3.3.2.1 *VIP Value Calculation Results*

In the backward variable elimination method, we applied the procedure to three drug samples: AgimDogyl, Natrixam 1.5/5, and Turbezid. The calculated VIP values are presented in Figures 3.7, 3.7, and 3.9.

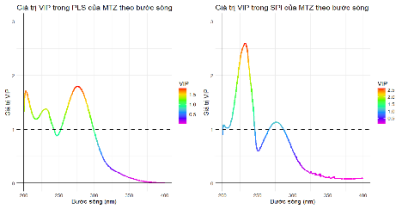


Figure 3.7. *VIP values calculated for MTZ and SPI in the PLS model, corresponding to 2 and 7 latent variables, respectively.*

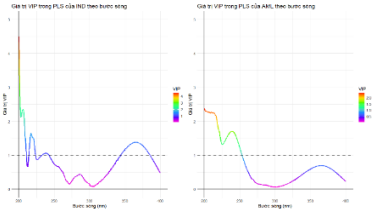


Figure 3.8. *VIP values calculated for INZ and AML in the PLS model, corresponding to 4 and 2 latent variables, respectively.*

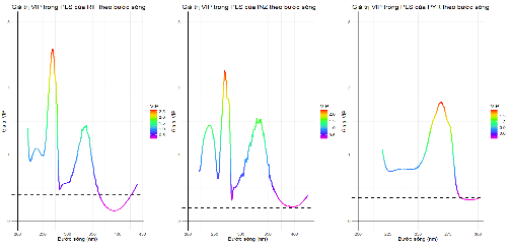


Figure 3.9. *VIP values calculated for RIF, INZ, and PYR in the PLS model, corresponding to 3, 4, and 3 latent variables, respectively.*

The results indicate that, compared with the absorption spectra of the individual components, the high VIP values are all located in the strong absorption regions corresponding to those pure components.

3.3.2.2 Wavelength Selection Results Using the BVE-PLS Method

Table 3.5. *Wavelengths selected by the BVE-PLS method for the drug samples AgimDogyl, Natrixam 1.5/5, and Turbezid*

	Analyte	number of selcted wavelengths	initial number of wavelengths	ratio (%)
AgimDogyl	MTZ	180 (*200 – 241,5 nm*)	401	44,9%
	SPI	90 (*200,5 – 245,5; 265 – 288 nm*)	401	22,4%
Natrixam 1,5/5	IND	370 (*200-210,4; 214,2 – 223,8; 231- 243,8; 344,2 – 384,6 nm*)	1001	37,0%
	AML	268 (*200-253,4 nm*)	1001	26,7%
Turbezid	RIF	300 (*220 – 362,5; 428,5 – 440*)	441	68,0%
	INZ	401 (*220-395,5; 401 – 425 nm*)	411	97,6%
	PYR	151 (*220 – 285,5; 291 – 300 nm*)	161	93,8%

3.3.3 Wavelength Selection Using the Genetic Algorithm Combined with Partial Least Squares (GA-PLS)

3.3.3.1 Genetic Algorithm Results

Figures 3.10, 3.11, and 3.12 show the optimization process for wavelength selection across different generations. The results indicate that initially, the RMSE values are high; as the genetic algorithm

progresses through successive generations, the RMSE values gradually decrease and eventually stabilize.

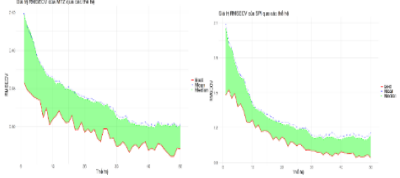
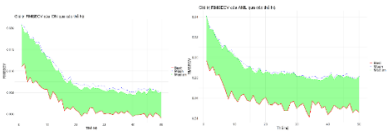
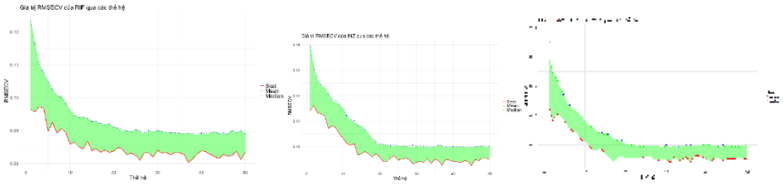


Figure 3.3. *RMSECV values over 50 genetic algorithm generations for MTZ and SPI*



Hình 3.4. *RMSECV values over 50 genetic algorithm generations for IND and AML*



Hình 3.5. *RMSECV values over 50 genetic algorithm generations for RIF, INZ and PYR*

3.3.3.2Wavelength Selection Results Using the Genetic Algorithm

The results show that the proportion of wavelengths selected by the GA-PLS method relative to the total initial number of wavelengths is relatively small. The retained wavelength ratios for MTZ and SPI in the AgimDogyl dataset are 12.2% and 8.5%, respectively; for IND and AML in the Natrixam dataset, both are 10.5%; and for RIF, INZ, and PYR in the Turbezid dataset, the ratios are 9.6%, 9.5%, and 8.7%, respectively.

3.3.4 Results of Applying Wavelength Selection Methods for Determining the Content of Compounds in Multicomponent Drug Samples

3.3.4.1 Application of Wavelength Selection Methods for Simultaneous Determination of MTZ and SPI in AgimDogyl Tablets

a. *Construction of PLS Models Based on Wavelength Selection Algorithms*

The PLS method was applied to build calibration models for the AgimDogyl calibration sets using wavelengths selected by the SR-PLS, BVE-PLS, and GA-PLS methods. The results show that all

models met the statistical evaluation criteria, with the SR-PLS method demonstrating superior accuracy and better data fit.

The performance of the SR-PLS, BVE-PLS, and GA-PLS methods was further evaluated based on predictions for the test set. The recovery percentages of the methods showed no significant differences. For MTZ, the recovery ranges were: SR-PLS: 100.4–104.8%, BVE-PLS: 98.9–104%, GA-PLS: 98.7–104.1%. For SPI, the recovery ranges at different concentrations were: SR-PLS: 97.4–102.1%, BVE-PLS: 97.1–101.8%, GA-PLS: 97.6–103.0%. The RMSEP values for MTZ using SR, BVE, and GA methods were 0.5313 (3.5%), 0.330 (2.2%), and 0.336 (2.2%), respectively. These values are small relative to the mean concentration of the test set (<5%), indicating good predictive performance. For SPI, the RMSEP percentages relative to the mean test set concentration were 1.3%, 1.6%, and 2.4% for SR, BVE, and GA methods, respectively, showing very good prediction for SR and BVE and good prediction for GA.

b. *Method Validation*

* The recovery of MTZ and SPI, evaluated using the standard addition method, met the AOAC requirements.

- Repeatability was assessed over two different days on AgimDogyl tablet samples.

+ Day 1: MTZ: $m_{TB}^{SR} = 14{,}58$ µg/mL (RSD = 1,85), $m_{TB}^{BVE} = 15{,}04$ µg/mL (RSD = 0,93), $m_{TB}^{GA} = 14{,}87$ µg/mL (RSD = 0,95); SPI: $m_{TB}^{SR} = 94{,}28$ IU/mL (RSD = 1,17), $m_{TB}^{BVE} = 93{,}25$ IU/mL (RSD = 1,13), $m_{TB}^{GA} = 93{,}24$ IU/mL (RSD = 1,67).

+ Day 2: MTZ: $m_{TB}^{SR} = 14{,}35$µg/mL (RSD = 0,70), $m_{TB}^{BVE} = 15{,}08$ µg/mL (RSD = 0,86), $m_{TB}^{GA} = 14{,}87$µg/mL (RSD = 0,84); SPI: $m_{TB}^{SR} = 94{,}76$ IU/mL (RSD = 0,96), $m_{TB}^{BVE} = 93{,}74$ IU/mL (RSD = 0,69), $m_{TB}^{GA} = 91{,}73$IU/mL (RSD = 0,26).

Comparison of Mean Results Between Data Sets: The mean values from the repeatability tests over two days were compared using the t-test. The calculated t-values were all less than the critical t-value, indicating that the methods provided highly accurate results across the different data sets.

* *Quantitative Results*

The quantified MTZ content with a confidence interval (α = 0.05) was: 122.06 ± 3.62 mg, 124.21 ± 2.67 mg, and 123.54 ± 2.20 mg for the SR-PLS, BVE-PLS, and GA-PLS methods, respectively. The quantified SPI content with a confidence interval (α = 0.05) was: 779,805 ± 11,446 IU, 774,601 ± 15,712 IU, and 776,592 ± 14,223 IU for the SR-PLS, BVE-PLS, and GA-PLS methods, respectively. The percentage of content relative to the label claim, as regulated by the Vietnamese Pharmacopoeia V (acceptable range: 90–110%), shows that the results in Tables 3.32 and 3.33 are reliable.

The ANOVA comparison of results among the SR-PLS, BVE-PLS, GA-PLS methods and the reference method shows: For MTZ, calculated F = 2.20 < critical F = 4.07, p = 0.166 > 0.05, indicating no significant difference in mean MTZ values across the groups. For SPI, calculated F = 1.98 < critical F = 4.07, p = 0.196 > 0.05, indicating no significant difference in mean SPI values across the groups. These results confirm that the developed methods possess good accuracy.

3.3.4.2 Application of Wavelength Selection Methods for Simultaneous Determination of IND and AML in Natrixam Tablets

a. *Construction of PLS Models Based on Wavelength Selection Algorithms*

For the calibration set, statistical parameters assessing accuracy and reliability of SR-PLS, BVE-PLS, and GA-PLS all met the required standards.

When applied to predict IND and AML contents in laboratory-prepared samples, all methods provided satisfactory results, with recoveries close to 100%.

b. *Method Evaluation*

* Accuracy:

-The recovery for IND and AML using the BVE-PLS method did not meet the requirements, indicating that the wavelengths selected by BVE-PLS were not optimal.

- The recoveries for IND and AML using SR-PLS and GA-PLS met the AOAC criteria.

- Comparison with the HPLC method:

+SR-PLS: The calculated t-values were 3.88 and 5.17 for IND and AML, respectively, both greater than t(0.05;4) = 2.78, indicating statistically significant differences in means.

+ BVE-PLS: The mean value for AML showed significant difference compared to HPLC.

+ GA-PLS: The quantified values for IND and AML showed no significant difference compared to HPLC.

* Precision

- Repeatability was evaluated by analyzing Natrixam samples multiple times. All RSD values were less than half of the Horwitz-calculated RSD, indicating good precision.

Thus, for Natrixam tablets, only GA-PLS met the requirements for both accuracy and precision for the simultaneous determination of IND and AML.

c. *Quantitation Results*

Quantification using the GA-PLS method yielded: $m_{IND} = 1.47 \pm 0.038$ mg/tablet; $m_{AML} = 4.83 \pm 0.080$ mg/tablet

3.3.4.3 Application of Wavelength Selection Methods for Simultaneous Determination of RIF, INZ, and PYR in Turbezid Tablets

a. *Construction of PLS Models Based on Wavelength Selection Algorithms*

In this study, two wavelength selection methods — BVE-PLS and GA-PLS — were applied to construct calibration models for the simultaneous determination of rifampicin (RIF), isoniazid (INZ), and pyrazinamide (PYR) in Turbezid tablets. The models were evaluated using statistical parameters on the calibration set, including RMSEC, RMSECV, R^2, and residual distribution assessed by the Shapiro-Wilk test. Results showed that both methods produced accurate and reliable models with $R^2 > 0.98$ and normally distributed residuals.

On the test set, both models achieved average recoveries close to 100% and low RMSEP, indicating good predictive ability on real samples. However, the BVE-PLS model demonstrated higher repeatability and accuracy, particularly for INZ determination, while GA-PLS used fewer selected wavelengths, resulting in reduced sensitivity and larger errors for some analytes.

b. *Analytical Method Evaluation*

Regarding accuracy, the recovery values for all analytes were within the specified range (80–110%) with low RSD, meeting pharmacopoeial requirements. At low concentration levels, INZ quantified by GA-PLS showed repeatability below the Horwitz standard.

Comparison with HPLC results from the testing center indicated that BVE-PLS provided results comparable to HPLC, whereas GA-PLS showed significant errors for RIF and INZ. Furthermore, evaluation against labeled content revealed that only BVE-PLS met the 90–110% requirement according to the Vietnamese Pharmacopoeia V, while GA-PLS underestimated INZ.

c. *Quantitative results*

sing the BVE-PLS method, m_{RIF} = 158,11 ± 7,42 mg/tablet; m_{INZ} = 75,55 ± 2,36 mg/table; m_{PYR} = 408,45 ± 10,15 mg/table.

3.4 Evaluation and Comparison of the Performance of Multivariate Analytical Methods

3.4.1 PLS Compared to Other Multivariate Analytical Methods

The comparison of CLS, PCR, and PLS shows that PLS is generally superior, with lower RMSEP, higher R^2, and better recoveries, especially for complex datasets. CLS is the least effective due to its reliance on pure spectra, which are difficult to obtain, resulting in large errors. PCR performs comparably to PLS in many cases; however, PLS can still provide more accurate predictions in certain situations. Overall, PLS is the most reliable choice, whereas PCR is suitable only in specific cases.

3.4.2 PLS and Wavelength Selection Methods

Comparing PLS with wavelength selection methods reveals differing efficiencies. SR-PLS often improves RMSEP in some cases but shows instability across different drugs. BVE-PLS can sometimes increase RMSEP by removing important variables, although it may improve RSD. GA-PLS has the potential to reduce RMSECV and RMSEC through genetic optimization, but its predictive performance (RMSEP, RSD) is inconsistent due to parameter dependency. In general, standard PLS remains the most reliable method with high R^2

and good recoveries, while wavelength selection methods are useful only in specific applications.

CONCLUSION

Based on the experiments conducted, this dissertation has demonstrated that the Partial Least Squares (PLS) regression method combined with UV-Vis absorption spectroscopy can be developed and applied for the simultaneous quantification of active pharmaceutical ingredients (APIs) with overlapping spectra in drug formulations.

(1) The dissertation successfully established PLS models for two- and three-component systems (MTZ–SPI, AML–ATO, SPR–FUR, IND–AML, RIF–INZ–PYR) using calibration data designed by CCD. The selection of the number of latent variables based on the PRESS criterion allowed optimization of the models, with the number of latent variables ranging from 2 to 7.

(2) Validation parameters (RMSEC, RMSECV, RMSEP, R^2, residual analysis, recovery, repeatability) all met the required standards: RMSE < 5% of the average concentration; residuals followed a normal distribution (Shapiro-Wilk $p > 0.05$); R^2 approximately 1; recoveries within 90–110% (AOAC). The t-test indicated no statistically significant difference between the results obtained by PLS–UV and the standard HPLC method ($p > 0.05$).

(3) The method was applied to quantify APIs in commercial tablets including AgimDogyl, Zoamco-A, Spirosemid-40, Natrixam, and Turbezid. The quantified contents were all within the acceptable limits according to the Vietnam Pharmacopoeia V.

(4) Comparison with other regression algorithms showed that PLS exhibited superior accuracy, owing to its better ability to eliminate noise compared with PCR and CLS.

(5) The dissertation proposed and standardized a practical PLS–UV workflow for systems with overlapping spectra and successfully investigated variable reduction mechanisms using SR, BVE, and GA algorithms for optimal wavelength selection. This approach reduces the number of input variables without compromising model quality.

In conclusion, PLS combined with UV-Vis spectral data is a feasible, low-cost method that can serve as an alternative to HPLC in suitable cases for multicomponent quantification in pharmaceuticals.

RECOMMENDATION

1. Continue developing and validating the PLS method combined with UV-Vis absorption spectra across various dosage forms and active pharmaceutical ingredients, particularly those with overlapping absorption spectra. This aims to evaluate the method's broad applicability, accuracy, and reliability in drug testing, moving toward practical implementation in pharmaceutical production and quality control.

2. Expand and test multiple wavelength selection strategies to improve the efficiency of PLS models. Diversifying variable selection methods will enhance accuracy, reduce complexity, and shorten computational time in spectral analysis.

3. Investigate the integration of machine learning and deep learning techniques for processing and predicting absorption spectral data. This approach has strong potential to maximize information extraction from spectral data while improving modeling capability, accuracy, and generalizability of non-separative quantitative analysis methods.

4. Develop software tools: create a laboratory analysis application (e.g., Shiny/R web app) to input UV-Vis spectra, automatically build PLS models, perform evaluation and validation, and generate reports – thereby increasing practical applicability and facilitating technology transfer to QC/QA laboratories.

LIST OF PUBLICATIONS

1. **Nguyen Duy Luu**, Le Van Thuan, Nguyen The Khang, Nguyen Hung Nhat Duy, Pham Phu Quoc, Vo Thi Thanh Truc, Nguyen Dinh Luyen, Tran Thuc Binh (2025). Backward variable elimination algorithm, Principal Component Regression, Partial Least Squares: simultaneous determination of Rifampicin, Isoniazid, and Pyrazinamide in tablets, *Journal of Science and Technology, University of Science, Hue University*, Volume 28, Issue 2, pp. 1–15.

2. **Nguyen Duy Luu**, Nguyen Thi Thanh Tram, Pham Phu Quoc, Nguyen Hung Nhat Duy, Nguyen Dinh Luyen, Tran Thuc Binh (2024), Simultaneous determination of Amlodipine and Atorvastatin in multi-component pharmaceuticals using UV-Vis spectroscopy combined with multivariate analysis, *Hue University Journal of Science: Natural Sciences*, Volume 133, Issue 1C, pp. 39–52.

3. Nguyen Dinh Luyen, Nguyen Quynh Giao, Pham Phu Quoc, Nguyen Hung Nhat Duy, Nguyen Duy Luu, Tran Thuc Binh (2024), Wavelength selection method for UV-Vis spectra combined with partial least squares for simultaneous determination of Amlodipine and Indapamide in tablet form, *Journal of Chemical, Physical and Biological Analysis*, Volume 30, Issue 2A, pp. 1–12.

4. **Nguyen Duy Luu**, Tran Thi My Duyen, Nguyen Dinh Luyen, Tran Thuc Binh (2024), A novel method of the simultaneous determination of Spironolactone and Furoseimide in pharmaceutical tablets, *Journal of Analytical Chemistry*, Vol. 79, No. 7, pp. 923-933. (SCIE, Q3, IF = 1,0). https://doi.org/10.1134/S1061934824700291

5. **Nguyen Duy Luu**, Nguyen Hung Nhat Duy, Pham Phu Quoc, Nguyen The Khang, Le Thanh Huyen, Nguyen Dinh Luyen, Tran Thuc Binh (2025), Simultaneous quantification of Metronidazole and Spiramycin by Ultraviolet-Visible method: Comparison of wavelength selection techniques, *Journal of Analytical Chemistry*, Vol. 80, No. 6, pp. 1005 – 1016. (SCIE, Q3, IF = 1,0). https://doi.org/10.1134/S1061934824605486

ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

Nguyễn Duy Lưu

NGHIÊN CỨU XÁC ĐỊNH ĐỒNG THỜI MỘT SỐ DƯỢC CHẤT CÓ PHỔ HẤP THỤ UV-VIS XEN PHỦ NHAU BẰNG PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU RIÊNG PHẦN

Ngành: Hóa phân tích
Mã số: 9440118

TÓM TẮT LUẬN ÁN TIẾN SĨ HÓA PHÂN TÍCH

Người hướng dẫn khoa học:
1. PGS.TS Trần Thúc Bình
2. PSG.TS Nguyễn Đình Luyện

HUẾ, NĂM 2025

Công trình được hoàn thành tại: .....

.....

.....

Người hướng dẫn khoa học: .....

.....

.....

Phản biện 1: .....

.....

.....

Phản biện 2: .....

.....

.....

Phản biện 3: .....

.....

.....

Luận án sẽ được bảo vệ tại Hội đồng chấm luận án cấp Đại học Huế họp tại: .....

.....

Vào lúc :…….giờ.....ngày……. tháng …….năm…….

Có thể tìm hiểu luận án tại thư viện:…… ……

……

MỞ ĐẦU

Theo một báo cáo của Tổ chức Y tế Thế giới (WHO) vào năm 2017, ước tính 1 trong 10 sản phẩm y tế lưu hành tại các quốc gia có thu nhập thấp và trung bình là sản phẩm kém chất lượng hoặc giả mạo. Điều này có nghĩa là mọi người đang sử dụng các loại thuốc không thể điều trị hoặc phòng ngừa bệnh. WHO đã xác định rằng các sản phẩm thuốc giả có thể chiếm khoảng 50% thị trường thuốc trên toàn cầu; nhiều trong số này xuất phát từ các quốc gia đang phát triển. Theo các báo cáo gửi tới WHO từ 20 quốc gia khác nhau, phần lớn các loại thuốc giả rơi vào một trong ba nhóm sau: (1) các sản phẩm không chứa thành phần hoạt chất: khoảng 30%; (2) các sản phẩm chứa hàm lượng hoạt chất không chính xác: khoảng 20%; và (3) các sản phẩm chứa sai thành phần: khoảng 20%. Chính vì vậy, Kiểm soát chất lượng (Quality Control - QC) là một yếu tố thiết yếu và không thể thiếu trong lĩnh vực sản xuất, lưu thông, và sử dụng thuốc. Với sự phát triển mạnh mẽ của ngành công nghiệp dược phẩm, yêu cầu về chất lượng thuốc ngày càng trở nên khắt khe hơn. Các cơ quan quản lý dược phẩm trên toàn thế giới, bao gồm Cục Quản lý Thực phẩm và Dược phẩm Hoa Kỳ (FDA), Cơ quan Dược phẩm châu Âu (EMA), và các cơ quan quản lý khác, đều nhấn mạnh sự cần thiết của việc kiểm soát chất lượng nghiêm ngặt trong suốt quá trình sản xuất và phân phối thuốc. Điều này nhằm đảm bảo rằng các sản phẩm thuốc không chỉ đạt được các tiêu chuẩn về hiệu quả điều trị mà còn đảm bảo an toàn tuyệt đối cho người sử dụng.

Với sự đa dạng và phức tạp của các sản phẩm dược phẩm hiện nay, nhu cầu phát triển các phương pháp phân tích nhanh chóng, chính xác và tin cậy là một đòi hỏi tất yếu. Hiện nay, trong các chuyên luận về phân tích các hoạt chất trong dược phẩm được quy định trong Dược điển, hầu hết là các quy trình phân tích riêng lẽ, điều này đòi hỏi quá trình phân tích phải kèm theo quá trình tách chiết các thành phần nếu là thuốc đa thành phần; làm cho quá trình phân tích tốn nhiều thời gian, quy trình phức tạp, tốn kém về hóa chất. Trong khi đó, phân tích đồng thời không chỉ giúp tiết kiệm thời gian và chi phí mà còn giảm thiểu sai số do quá trình tách chiết các thành phần. Vấn đề này đặt ra cho các nhà quản lý sự cần thiết phải nghiên cứu, cải tiến và phát triển nhiều phương pháp phân tích đồng thời nhằm tăng hiệu quả trong công tác quản lý chất lượng.

Có thể thấy rằng, nhóm các phương pháp phân tích hóa lý ngày càng được phát triển và ứng dụng mạnh mẻ trong phân tích thuốc, trọng số của các phương pháp này chiếm hầu hết trong các chuyên luận phân tích; Các ưu điểm rõ ràng của các phương pháp này đã được chỉ ra: có độ đúng, độ chính xác, độ tin cậy cao. Tuy nhiên, các phương pháp như sắc ký lỏng hiệu năng cao, sắc ký khí, phương pháp khối phổ… đều là các thiết bị phức tạp, đắt tiền, quy trình phân tích khắt khe và yêu cầu sự vận hành bởi các chuyên gia có kỹ năng cao. Trong bối cảnh đó, phương pháp UV-Vis có thể khắc phục được các nhược điểm của các phương pháp phân tích đó. Một vấn đề khó khăn trong phương pháp UV-Vis là đối với mẫu thuốc đa thành phần và có phổ xen phủ nhau hoặc ảnh hưởng của tạp chất, tá dược có trong mẫu thì phương pháp UV-Vis thường gặp khó khăn trong việc xác định chính xác hàm lượng của các chất có trong mẫu.

Năm 1972, Swede Svante Word và Bruce R. Kowalski đã giới thiệu thuật ngữ *chemometric,* được định nghĩa là ngành hóa học sử dụng các phương pháp toán học và thống kê để thiết kế hoặc lựa chọn các quy trình đo lường, tối ưu hóa thí nghiệm và để cung cấp tối đa hóa thông tin hóa học bằng cách phân tích dữ liệu hóa học. Từ những năm 1980 với sự xuất hiện máy tính cá nhân bắt đầu cho một kỷ nguyên mới cho việc thu thập, xử lý và giải thích dữ liệu hóa học. Và cho đến nay, tốc độ phát triển vô cùng mạnh mẽ của khoa học máy tính cùng với các mô hình AI cho phép việc xử lý các dữ liệu trở nên dễ dàng hơn, chính xác hơn . Chính điều này đem lại những lợi thế không nhỏ cho việc hình thành, cải tiến và phát triển các phương pháp phân tích. Các bất lợi của phương pháp phân tích quang phổ UV-Vis có thể được giải quyết một cách triệt để, chính xác khi kết hợp giữa các dữ liệu quang phổ với các mô hình toán học [15]. Một loạt các mô hình toán học được áp dụng vào trong xử lý dữ liệu quang phổ bao gồm từ các phương pháp bình phương tối thiểu cổ điển (CLS), bình phương tối thiểu nghịch đảo (ILS), hồi quy cấu tử chính (PCR), bình phương tối thiểu riêng phần (PLS) cho đến những phương pháp áp dụng trong mô hình học máy (machine learning) như mạng notron nhân tạo (ANN), máy vecto hỗ trợ (Support vector machine - SVM), thuật toán di truyền (GA) ....

Khi chúng ta đo độ hấp thụ UV của một dung dịch, có thể dễ dàng đo được toàn bộ phổ của nó một cách chính xác và nhanh chóng với độ nhiễu thấp so với đo phổ của nó ở một bước sóng riêng lẽ. Phổ hấp thụ thu được khi đo mẫu chúng ta có được thông tin của toàn bộ các chất có trong mẫu và sự tương tác giữa các thành phần trong mẫu phân tích, điều này là một lợi thế hơn hẳn trong phân tích đơn biến – chúng ta chỉ có được thông tin của 1 thành phần mà bỏ qua sự tương tác giữa các chất với nhau khi chúng cùng có mặt trong một hỗn hợp. Đây là một trong những lợi thế của phân tích đa biến.

Mô hình PLS thường được trình bày là kỹ thuật hồi quy chính được sử dụng khi có dữ liệu đa biến. Sự thành công trong các lĩnh vực QSAR, khoa học y sinh (metabolomic), khoa học đo lường đặc điểm tâm lý con người (Psychometrics) khi áp dụng mô hình PLS ; sau đó là áp dụng trong các lĩnh vực quang phổ như phổ hồng ngoại gần (NIR), phổ Raman, phổ UV-Vis hoặc các phương pháp sắc ký, điện hóa để nói lên được tầm quan trong của phương pháp này trong việc xử lý dữ liệu và tính toán kết quả. Các lợi thế của mô hình PLS bao gồm: giảm chiều dữ liệu, xử lý vấn đề đa cộng tuyến, loại bỏ ảnh hưởng của nhiễu, có thể phân tích một chất riêng lẽ trong hỗn hợp mà không cần phải biết tất cả các thông tin của mẫu phân tích.

Đặc biệt, trong phân tích dược phẩm hỗn hợp có phổ UV-Vis xen phủ, PLS cho phép xác định đồng thời hàm lượng của nhiều hoạt chất mà không cần tách chiết, bằng cách xây dựng mô hình hồi quy từ dữ liệu phổ và nồng độ chuẩn. Nhờ đó, phương pháp PLS kết hợp phổ UV-Vis không chỉ giúp phân giải chồng phổ, mà còn tăng độ chính xác, giảm chi phí và rút ngắn thời gian phân tích. Các ưu điểm vượt trội này khiến phương pháp trở thành hướng nghiên cứu có ý nghĩa thực tiễn cao trong kiểm soát chất lượng thuốc đa thành phần.

Vì những lí do trên, chúng tôi chọn đề tài luận án là "**Nghiên cứu xác định đồng thời một số dược chất có phổ hấp thụ UV-VIS xen phủ nhau bằng phương pháp bình phương tối thiểu riêng phần".**

Chương 1. TỔNG QUAN TÀI LIỆU

1.1. Phương pháp quang phổ uv-vis

1.2. Phương pháp chemometrics

1.3 Một số phương pháp phân tích đa biến

1.4 Tổng quan về áp dụng phương pháp pls kết hợp các phương pháp phổ trong phân tích các chất

1.5 Giới hạn phát hiện trong phân tích đa biến

1.6 Lựa chọn bước sóng trong phân tích quang phổ

1.7 Giới thiệu Amlodipin besylat, Indapamid, Atorvastatin calcium, Spironolacton, Furosemid, Rifampicin, Isoniazid, Pyrazynamid, metronidazol, spiramycin.

Chương 2. NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Nội dung nghiên cứu

Để đạt được mục tiêu đề ra, luận án triển khai các nội dung nghiên cứu chính sau đây:

- Xây dựng mô hình định lượng đa cấu tử bằng PLS–UV cho hệ hai và ba thành phần có phổ hấp thụ xen phủ; thiết kế tập hiệu chuẩn theo phương pháp CCD, lựa chọn biến tiềm ẩn theo tiêu chuẩn PRESS và thẩm định mô hình qua các chỉ tiêu thống kê (RMSEC, RMSECV, RMSEP, R^2, phân tích phần dư, độ thu hồi, độ lặp lại).
- Ứng dụng mô hình PLS–UV vào mẫu thực: định lượng đồng thời các hoạt chất trong thuốc thương mại (AgimDogyl, Zoamco-A, Spirosemid-40, Natrixam, Turbezid) và so sánh kết quả với phương pháp HPLC chuẩn bằng kiểm định t.
- So sánh hiệu năng của PLS với các thuật toán hồi quy khác (PCR, CLS) nhằm đánh giá ưu thế và khả năng khử nhiễu của PLS trong phân tích đa cấu tử từ dữ liệu phổ UV-Vis.
- Tối ưu hóa lựa chọn biến đầu vào cho PLS bằng các thuật toán hiện đại (SR, BVE, GA) để rút gọn phổ hấp thụ, giảm số biến nhưng vẫn đảm bảo hoặc nâng cao độ chính xác dự báo.
- Đề xuất và chuẩn hóa quy trình ứng dụng PLS kết hợp phổ UV-Vis cho phân tích đa cấu tử trong dược phẩm, làm cơ sở triển khai trong công tác kiểm nghiệm QC/QA.

2.2. Phương pháp nghiên cứu

2.2.1. Thiết kế thí nghiệm

2.2.2. Phương pháp thẩm định chéo

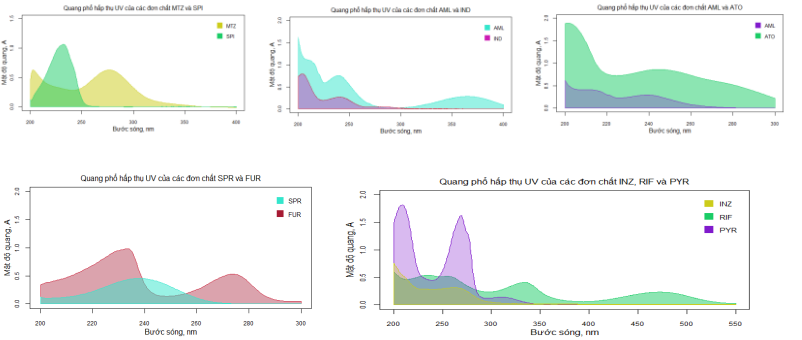
2.2.4. Giới hạn phát hiện trong mô hình PLS
2.2.5. Lựa chọn bước sóng bằng phương pháp tỉ lệ chọn lọc trong PLS
2.2.6. Lựa chọn bước sóng bằng phương pháp loại bỏ biến ngược trong PLS
2.2.7. Lựa chọn bước sóng bằng phương pháp thuật toán di truyền trong PLS

2.3 Thực nghiệm

Chương 3. KẾT QUẢ VÀ THẢO LUẬN

3.1 Thiết kế thí nghiệm và tính chất quang phổ của các hoạt chất cần nghiên cứu

Nồng độ các chất trong tập hiệu chuẩn để xây dựng các mô hình được thiết kế theo CCD, với nồng độ tâm tương ứng với tỉ lệ nồng độ của các hoạt chất có trong các mẫu thuốc. Phổ hấp thụ của các đơn chất trong các mẫu phân tích có sự xen phủ nhau lớn, thể hiện trong hình 3.1



Hình 3.1. *Phổ hấp thụ của các đơn chất trong các mẫu thuốc*

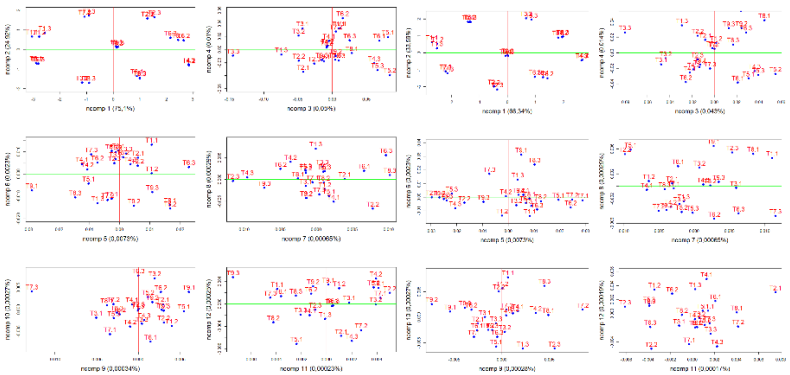
3.2 Phương pháp PLS kết hợp phổ hấp thụ UV-Vis

3.2.1 Kết quả thuật toán PLS

3.2.1.1 Biểu đồ điểm trong mô hình PLS

Phương pháp PLS là phương pháp mà dữ liệu thực được thu gọn trong một không gian biến tiềm ẩn. Các biến tiềm ẩn được đại diện là một ma trận điểm (scores), đây là dữ liệu khái quát cho dữ liệu gốc ban đầu với số chiều đã giảm từ 401 (bước sóng) còn 25 (thành

phần chính), mỗi thành phần chính tương ứng với 1 chiều không gian mới.

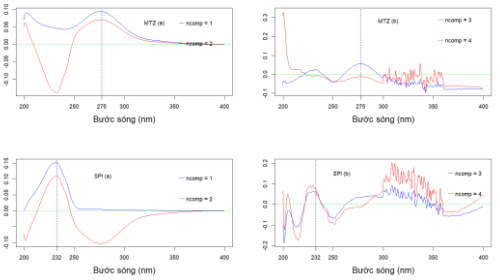


Hình 3.2. *Biểu đồ điểm của 12 biến tiềm ẩn đầu tiên của MTZ vàSPI*

Biểu đồ điểm trong hình 3.2 cho thấy mô hình PLS có thể tách dữ liệu gốc ban đầu thành các dữ liệu trong không gian biến tiềm ẩn PLS và giải thích được mối liên quan giữa thông tin về nồng độ của chất phân tích với tiến tiềm ẩn; đồng thời giải thích được các yếu tố nhiễu trong dữ liệu.

3.2.1.2. Hệ số tải (P), hệ số trọng số tải (W), hệ số chiếu (P_{proj}) của phổ và hệ số hồi qui của mô hình PLS

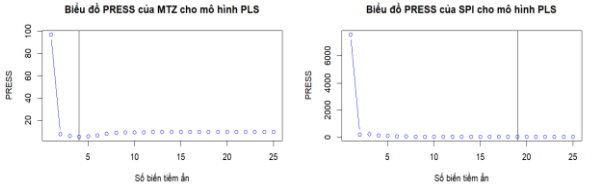
Phổ của vector chiếu dữ liệu quang phổ lên không gian biến tiềm ẩn của MTZ và SPI thể hiện được sự đồng dạng với phổ của đơn chất (biến tiền ẩn thứ 1) cho thấy có thể giải thích được tính chất đặc trưng của mỗi chất trong hỗn hợp ban đầu. Khi số biến tiềm ẩn lớn, được phổ dao động mạnh cho thấy sự không chắc chắn, có thể do nhiễu trong dữ liệu gốc. Điều này cho thấy rằng tầm quan trọng của số lượng biến tiềm ẩn được giữ lại để xây dựng mô hình hồi qui trong mô hình PLS là rất quan trọng. Điều này cũng phù hợp với đường cong của hệ số hồi qui, thể hiện mối quan hệ giữa biến độc lập với nồng độ, khi số biến tiềm ẩn lớn thì mức độ dao động của hệ số hồi qui cũng rất manh, chứng tỏ khi sự thay đổi nhỏ của dữ liệu trong các biến tiềm ẩn có thể dẫn đến sự thay đổi rất lớn về nồng độ.



Hình 3.3. *Phổ của các vector chiếu dữ liệu quang phổ gốc lên không gian tiềm ẩn PLS của MTZ và SPI với 4 biến tiềm ẩn đầu tiên.*

3.2.2 Kết quả xác định số biến tiềm ẩn trong mô hình PLS

3.2.2.1 Chọn số biến tiềm ẩn dựa vào giá trị PRESS nhỏ nhất



Hình 3.4. Biểu đồ giá trị PRESS trong mô hình PLS của MTZ và SPI

Kết quả của mô hình PLS đã thu gọn dữ liệu gốc ban đầu thành 25 biến tiềm ẩn, tương ứng với một biến ẩn có 1 giá trị PRESS. Trong mô hình này, giá trị PRESS nhỏ nhất tương ứng với MTZ và SPI lần lượt là 4 và 19.

3.2.2.2 Chọn số biến tiềm ẩn dựa vào kiểm định thống kê cho tập hiệu chuẩn

Theo Haaland và Thomas, giá trị tỉ lệ F ở biến tiềm ẩn nhỏ nhất có xác suất nhỏ hơn 0,75 được chọn làm biến tiềm ẩn để xây dựng mô hình PLS. Các giá trị phần trăm xác suất p trong phân phối Fisher được tính với bậc tự do bằng số mẫu hiệu chuẩn là 27 cho cả tử và mẫu. Kết quả số biến tiềm ẩn được chọn để xây dựng mô hình PLS đối với MTZ và SPI lần lượt là 3 và 8

3.2.2.3 Lựa chọn số biến tiềm ẩn dựa vào kiểm định thống kê cho tập kiểm tra

Tương tự cách chọn biến từ giá trị PRESS cho tập hiệu chuẩn, giá trị F cũng được tính để so sánh lựa chọn biến trên tập kiểm tra. Kết quả cho thấy số biến tiềm ẩn được chọn cho MTZ và SPI lần lượt là 2 và 7.

3.2.3 Kết quả đánh giá mô hình PLS để xác định đồng thời MTZ và SPI

Bảng 3.1. *Kết quả các thông số thống kê của mô hình PLS cho MTZ và SPI ở các biến tiềm ẩn được chọn*

Số biến tiềm ẩn		n = 2	n = 3	n = 4	n = 7	n = 8	n = 19
RMSECV (µg/mL)		0,540	0,474	0,446	1,60	1,46	1,37
RMSEC (µg/mL)		0,483	0,407	0,347	0,294	0,294	0
Shapiro-Wilk	W	0,928	0,921	0,896	0,977	0,960	-
	p	0,061	0,043	0,011	0,789	0,373	-
R^2		0,9894	0,9925	0,9945	0,9995	0,9999	1
LOD_{min} (µg/mL)		0,926	0,867	0,816	2,63	2,73	1,18
LOD_{max} (µg/mL)		1,02	1,14	1,07	3,19	3,29	1,50

Mô hình PLS được đánh giá qua các tham số RMSECV, RMSEC, phân tích phần dư, hệ số xác định, giới hạn phát hiện. Kết quả cho thấy, MTZ có số biến tiềm ẩn bằng 2 cho các giá trị thông kê phù hợp trong khi đó SPI là 7. Giá trị RMSE < 5% so với nồng độ trung bình, phần dư có phân phối chuẩn, hệ số xác định $R^2 > 0,99$. Mô hình PLS đã xây dựng có LOD thấp cho thấy mô hình có thể xác định được một lượng nhỏ hàm lượng các chất cần phân tích.

Từ mô hình PLS đã xây dựng được sử dụng để dự đoán nồng độ của các mẫu trong tập kiểm tra có các tỉ lệ nồng độ khác nhau được chuẩn bị trong phòng thí nghiệm, kết quả có độ thu hồi trung bình của MTZ và SPI lần lượt là 100,9% và 99,3% với giá trị RSD% tương ứng là 2,06% và 1,658%, phù hợp với quy định của AOAC; giá trị RMSEP nhỏ hơn 2% cho thấy mô hình có hiệu quả dự đoán rất tốt.

3.2.4 Kết quả thẩm định phương pháp PLS kết hợp phổ hấp thụ uv-vis để định lượng đồng thời MTZ và SPI.

3.2.4.1 Đánh giá độ đúng

Độ đúng của phương pháp PLS đã xây dựng để định lượng đồng thời MTZ và SPI được đánh giá qua hai cách:

Cách 1: đánh giá qua độ thu hồi bằng cách thêm chuẩn MTZ và SPI lần lượt ở 3 mức thấp – trung bình – cao. Kết quả cho thấy độ thu hồi trung bình tương ứng với 3 mức cho MTZ lần lượt là 93,0% (RSD = 3,08); 98,6% (RSD = 1,50) và 96,3% (RSD = 1,01), với SPI là 100,2% (RSD = 1,05); 101,4% (RSD = 2,08) và 104,9% (RSD =

0,71), kết quả này phù hợp với quy định của AOAC cho mức nồng độ dưới 100 µg/mL là từ 90-107%.

Cách 2: so sánh với phương pháp HPLC đối với định lượng MTZ và phương pháp vi sinh đối với định lượng SPI theo qui định của Dược điển Việt Nam V. Kết quả cho thấy giá trị trung bình giữa hai phương pháp không có sự khác biệt có ý nghĩa thống kê ($p > 0,05$)

3.2.4.2 Đánh giá độ chính xác

Độ chính xác được đánh giá thông qua độ lặp lại được thực hiện trong hai ngày khác nhau, mỗi ngày được làm lặp lại 6 lần giống nhau bắt đầu từ cân mẫu. Ngày 1: $\overline{m}_{MTZ}$ = 14,96 µg/mL (RSD = 1,303%), $\overline{m}_{SPI}$ =88,86 IU/mL (RSD = 1,511%); ngày thứ 2 $\overline{m}_{MTZ}$ = 14,98 µg/mL (RSD = 1,275%), $\overline{m}_{SPI}$ = 88,06 IU/mL (RSD = 0,907%). Kết quả các giá trị RSD đều nhỏ 5,3 (AOAC), điều này cho thấy kết quả phân tích cho độ lặp lại tốt.

3.2.5 Kết quả định lượng MTZ và SPI trong viên nén Agimdogyl

Kết quả định lượng MTZ và SPI trong viên nén AgimDogyl với hàm lượng 125mg/750000IU m_{MTZ} =124,38 ± 1,26 mg và m_{SPI} = 735971 ± 3370 IU. Kết quả cho thấy, hàm lượng của MTZ và SPI đều nằm trong giới hạn theo quy định của Dược điển Việt Nam V.

3.2.6. Ứng dụng phương pháp PLS định lượng đồng thời các hoạt chất trong thuốc đa thành phần

3.2.6.1 Đối với thuốc 2 thành phần

a. Xây dựng mô hình PLS cho các thuốc 2 thành phần

Bảng 3.2. *Các thông số thống kê đánh giá các tập hiệu chuẩn Spirosemid, Zoamco-A và Natrixam bằng phương pháp PLS*

	Spirosemid		Zoamco-A		Natrixam	
	SPR	FUR	AML	ATO	IND	AML
Khoảng bước sóng (nm)	200 - 400		230-260		200 - 400	
Khoảng quét (Δλ)	0,5		0,2		0,2	
n	2	2	3	2	4	2
RMSECV (µg/mL)	0,190	0,212	0,129	0,489	0,128	0,385
RMSEC (µg/mL)	0,260	0,280	0,057	0,314	0,109	0,354
Shapiro- W	0,8842	0,9742	0,9062	0,8676	0,9093	0,8871
Wilk test p	0,1747	0,9278	0,2903	0,116	0,2395	0,128
R^2	0,9980	0,9977	0,9991	0,9929	0,9902	0,9992

LOD_{min} (µg/mL)	0,375	0,272	1,00	4,69	0,269	1,263
LOD_{max} (µg/mL)	0,420	0,320	1,06	4,83	0,338	1,708

Các mô hình PLS cho các tập hiệu chuẩn Spiromid-40, Zoamco-A, Natrixam đã được xây dựng và các giá trị thống kê đã được tính và thể hiện trong bảng 3.12. Các kết quả cho thấy các tham số thống kê để đánh độ chính xác của mô hình đều đạt yêu cầu và được sử dụng để xác định đồng thời các chất có mặt trong các mẫu thuốc hai thành phần.

Áp dụng mô hình PLS để định lượng các thành phần trong các mẫu kiểm tra. Độ thu hồi của SPR và FUR: 92,8-106,3%, 97,0-108,0%; của AML và ATO: 97,3-104,5%, 98,0-105,2%; của IND và AML: 96,9-103,9%, 97,3-106,3%. Các giá trị RMSEP nhỏ hơn 2% hoặc trong khoảng 2-5%.

b. *Kết quả đánh giá mô hình PLS áp dụng định lượng đồng thời các hoạt chất trong thuốc hai thành phần*

Đánh giá độ đúng: Đối với thuốc Spirosemid độ thu hồi của SPR và FUR nằm trong khoảng 98,4 – 107,8% và 101,4 – 105,3%; đối với thuốc Zoamco-A độ thu hồi của AML và ATO lần lượt là 101,6 – 108,1% và 99,1 – 107,2%, đối với thuốc Natrixam độ thu hồi của IND là 96,7 – 104,7% và của AML là 101,1 – 104,6%.

Đánh giá độ lặp lại: Đối với thuốc Spirosemid: SPR: m_{TB} = 9,76 µg/mL (RSD = 2,54), FUR: m_{TB} = 7,76 µg/mL (RSD = 2,18); Đối với thuốc Zoamco-A: AML có m_{TB} = 10,44 µg/mL (RSD = 1,48), ATO có m_{TB} = 20,81 µg/mL (RSD = 2,49); Đối với thuốc Natrixam: IND có m_{TB} = 4,35 µg/mL (RSD = 1,32), AML có m_{TB} = 14,45 µg/mL (RSD = 0,65).

Kết quả đánh giá mô hình PLS để xác định đồng thời các hoạt chất trong thuốc Spirosemid-40, Zoamco-A, Natrixam đều đạt yêu cầu về độ đúng và độ chính xác

c. Kết quả định lượng

Đối với viên nén Spirosemid: m_{SPR} = 49.65 ± 3.21 mg/viên; m_{FUR} = 39.39 ± 1.93 mg/viên; Đối với viên nén Zoamco-A: m_{AML} = 5.23 ± 0.24 mg/viên; m_{ATO} = 10.39 ± 0.77 mg/viên; Đối với viên nén Natrixam 1,5/5: m_{IND} = 1.47 ± 0.05 mg/viên; m_{AML} = 4.83 ± 0.10 mg/viên với độ tin cậy 95%. Hàm lượng phần trăm so với hàm lượng ghi trên nhãn của các thuốc cũng đã được tính. Hàm lượng % ghi trên

nhãn của các hoạt chất trong các thuốc đều nằm trong khoảng quy định từ 90 – 110% của Dược điển Việt Nam.

So sánh giá trị trung bình kết quả phân tích bằng phương pháp PLS và phương pháp HPLC cho thấy các kết quả không khác nhau có ý nghĩa thống kê ($t_{tính} < t_{bảng} = 2,775$).

3.2.6.2 Đối với thuốc có 3 thành phần

a. Xây dựng mô hình PLS cho các thuốc 3 thành phần

Các tham số thống kê đã được tính để đánh giá khả năng hoạt động của mô hình PLS, kết quả được trình bày trong bảng 3.18.

Áp dụng mô hình PLS đã xây dựng để xác định nồng độ RIF, INZ và PYR trong các mẫu được chuẩn bị trong phòng thí nghiệm. Độ thu hồi trung bình lần lượt là 100,0%, 101,5% và 100,2% tương ứng với RIF, INZ và PYR, các giá trị RMSEP nhỏ chứng tỏ khả năng dự đoán của phương pháp PLS kết hợp phổ UV-Vis tốt, có thể áp dụng để định lượng đồng thời RIF, INZ và PYR trong các mẫu thuốc đa thành phần.

Bảng 3.3. *Các thông số thống kê đánh giá các tập hiệu chuẩn bằng phương pháp PLS*

		RIF	INZ	PYR
Khoảng bước sóng (nm)		220-440	220 - 425	220 - 300
Khoảng quét (Δλ)			0,5	
Số biến tiềm ẩn		3	4	3
RMSECV (µg/mL)		0,136	0,227	0,367
RMSEC (µg/mL)		0,107	0,184	0,317
Shapiro-Wilk test	W	0,9825	0,9816	0,9452
	p	0,888	0,8661	0,1254
R^2		0,9944	0,9836	0,9929
LOD_{min} (µg/mL)		0,43	0,87	1,30
LOD_{max}(µg/mL)		0,45	0,97	1,35

b. Kết quả đánh giá mô hình PLS áp dụng định lượng đồng thời các hoạt chất trong thuốc hai thành phần

Áp dụng phương pháp PLS kết hợp phổ UV đã xác định được nồng độ của RIF, INZ và PYR trong các mẫu với độ thu hồi lần lượt là 97,6%-102,4%, 95,4 – 108,2%, 95,9 – 102,1%. Kết quả này phù hợp với quy định của AOAC.

Giá trị trung bình của RIF, INZ và PYR trong đánh giá độ lặp lại ngày thứ 1 lần lượt là 9,13, 4,57 và 24,53 µg/mL; trong đánh giá độ lặp lại ngày thứ 2 lần lượt là 9,10, 4,60 và 24,41 µg/mL. Kết quả đánh giá độ lặp tốt, giá trị RSD (%) đều nhỏ hơn giá trị RSD tính theo hàm Horwirtz. So sánh kết quả trung bình giữa 2 lần thực hiện cũng cho thấy giá trị trung bình không có sự khác biệt có ý nghĩa thông kê với $t_{tính} < t_{bảng}$. Như vậy phương pháp phân tính đáp ứng yêu cầu về độ chính xác.

c. Kết quả định lượng

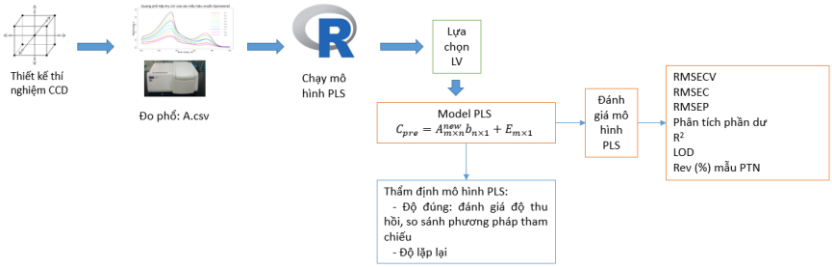
m_{RIF} = 154,11 ± 6.36 mg/viên; m_{INZ} = 77.39 ± 2.74 mg/viên; m_{PYR} = 407.83 ± 10.40 mg/viên. Phần trăm hàm lượng so với hàm lượng ghi trên nhãn lần lượt là 102,7%, 103,2% và 102,0%, kết quả này phù hợp với quy định của Dược điển Việt Nam V.

Kết quả so sánh giá trị trung bình giữa hai phương pháp cho thấy kết quả trung bình xác định RIF, INZ và PYR bằng phương pháp PLS kết hợp phổ hấp thụ UV-Vis không có sự khác biệt đáng kể so với phương pháp HPLC.

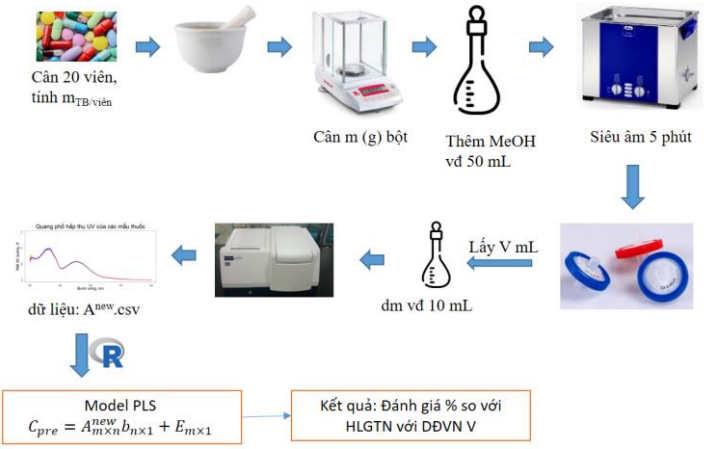
3.2.7 Kết luận phương pháp PLS

Chúng tôi đề xuất một qui trình chung để sử dụng phương pháp PLS kết hợp phổ hấp thụ UV-Vis để định lượng đồng thời các chất cho phổ xen phủ nhau, qui trình bao gồm 2 giai đoạn:

*** Giai đoạn 1:** huấn luyện mô hình PLS:



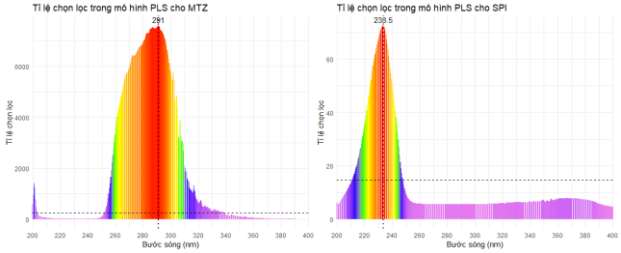
*** Giai đoạn 2:** Xác định nồng độ chất phân tích trong mẫu thực



3.3 Lựa chọn bước sóng trong PLS

3.3.1 Lựa chọn bước sóng bằng phương pháp tỉ lệ chọn lọc kết hợp bình phương tối thiểu riêng phần (SR-PLS)

3.3.1.1 *Kết quả tính SR cho mỗi bước sóng*

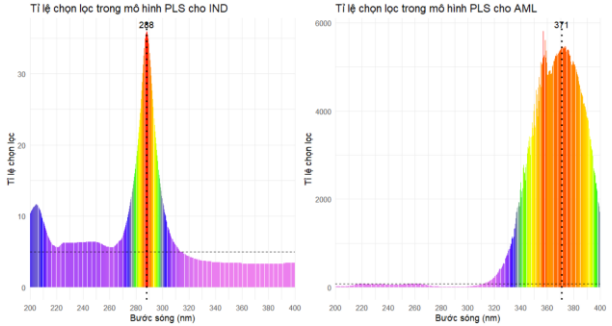


Hình 3.5 *Kết quả xác định các giá trị SR cho MTZ và SPI trong mô hình PLS đã xây dựng.*

Đối với MTZ, giá trị SR lớp tập trung ở khoảng bước sóng từ 250 đến 350 nm. Đối với SPI, giá trị SR trong khoảng bước sóng 210 đến 247 nm lớn hơn rất nhiều đối với các bước sóng còn lại (hình 3.9).

Đánh giá tương tự với IND và AML trong viên nén Natrixam 1,5/5. Kết quả đánh giá giá trị SR tại các bước sóng khác nhau của mỗi chất IND và AML được thể hiện trong hình 3.10. Kết quả cho thấy rằng, ở khoảng bước sóng trên khoảng 320 nm giá trị SR của IND khá

nhỏ phù hợp với thực tế IND đơn chất không hấp thụ trong khoảng bước sóng này, trong khi đó, đối với AML thì giá trị SR ở trong khoảng 320 – 400 nm có giá trị rất lớn trong khi khoảng bước sóng còn lại giá trị này rất nhỏ.



Hình 3.6. *Kết quả giá trị SR cho IND và AML*

3.3.1.2 *Lựa chọn các bước sóng quan trọng trong phương pháp SR*

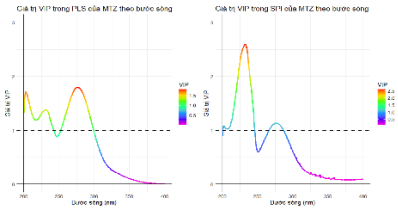
Bảng 3.4. *Kết quả lựa chọn bước sóng theo phương pháp SR cho MTZ và SPI trong mô hình PLS*

Thuốc	Thành phần	Số bước sóng được chọn	số bước sóng ban đầu	tỉ lệ (%)
AgimDogyl	MTZ	181 (252 – 338,5 và 200 – 203 nm)	401	45,1%
	SPI	74 (211 – 247,5 nm)	401	18,4%
Natrixam 1,5/5	IND	569 (200 – 313,6 nm)	1001	56,8%
	AML	514 (219 – 222 và 300,6 – 400 nm)	1001	51,3%

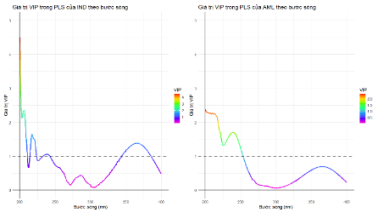
3.3.2 Lựa chọn bước sóng bằng phương pháp loại bỏ biến ngược kết hợp bình phương tối thiểu riêng phần (BVE-PLS)

3.3.2.1 *Kết quả tính giá trị VIP*

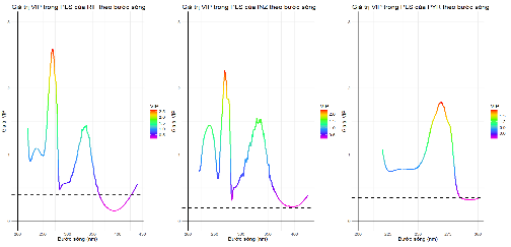
Trong phương pháp loại bỏ biến ngược, chúng tôi áp dụng để tính cho ba mẫu thuốc AgimDogyl, Natrixam 1,5/5 và Turbezid. Các giá trị VIP được tính và thể hiện trong hình 3.11; hình 3.12; hình 3.13.



Hình 3.7. *Các giá trị VIP được tính cho MTZ và SPI trong mô hình PLS tương ứng với số biến tiềm ẩn là 2 và 7.*



Hình 3.8. *Các giá trị VIP được tính cho INZ và AML trong mô hình PLS tương ứng với số biến tiềm ẩn là 4 và 2.*



Hình 3.9. *Các giá trị VIP được tính cho RIF, INZ và PYR trong mô hình PLS tương ứng với số biến tiềm ẩn là 3, 4 và 3.*

Kết quả cho thấy rằng, khi so sánh với độ hấp thụ của các thành phần đơn chất, các giá trị VIP lớn đều nằm trong các vùng hấp thụ mạnh tương ứng với các đơn chất đó.

3.3.2.2 Kết quả lựa chọn bước sóng bằng phương pháp BVE-PLS

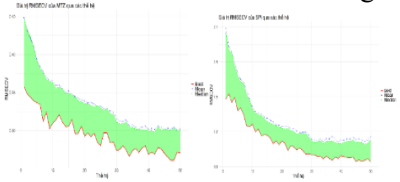
Bảng 3.5. *Kết quả lựa chọn bước sóng bằng phương pháp BVE-PLS cho các mẫu thuốc AgimDogyl, Natrixam 1,5/5 và Turbezid*

	Thành phần	Số bước sóng được chọn	số bước sóng ban đầu	tỉ lệ (%)
Thuốc AgimDogyl	MTZ	180 (*200 – 241,5 nm*)	401	44,9%
	SPI	90 (*200,5 – 245,5; 265 – 288 nm*)	401	22,4%
Thuốc Natrixam 1,5/5	IND	370 (*200-210,4; 214,2 – 223,8; 231- 243,8; 344,2 – 384,6 nm*)	1001	37,0%
	AML	268 (*200-253,4 nm*)	1001	26,7%
Turbezid	RIF	300 (*220 – 362,5; 428,5 – 440*)	441	68,0%
	INZ	401 (*220-395,5; 401 – 425 nm*)	411	97,6%
	PYR	151 (*220 – 285,5; 291 – 300 nm*)	161	93,8%

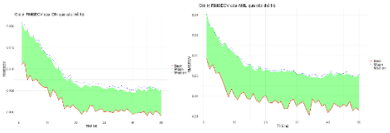
3.3.3 Lựa chọn bước sóng bằng thuật toán di truyền kết hợp bình phương tối thiểu riêng phần

3.3.3.1 Kết quả thuật toán di truyền

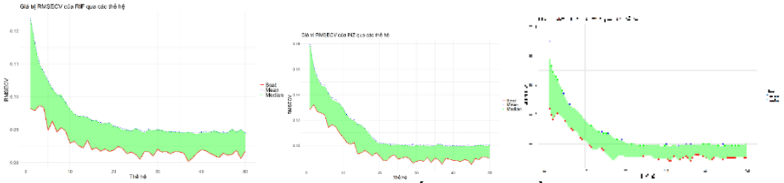
Các hình 3.14; hình 3.15; hình 3.16 thể hiện quá trình tối ưu cho việc lựa chọn bước sóng qua các thế hệ khác nhau. Kết quả cho thấy, ban đầu các giá trị RMSE lớn, qua các thế hệ di truyền giá trị RMSE nhỏ dần và có xu hướng đi ngang.



Hình 3.10. *Giá trị RMSECV qua 50 thế hệ di truyền của MTZ và SPI*



Hình 3.11. *Giá trị RMSECV qua 50 thế hệ di truyền của IND và MTZ*



Hình 3.12. *Giá trị RMSECV qua 50 thế hệ di truyền của RIF, INZ và PYR*

3.3.3.2. Kết quả lựa chọn bước sóng bằng thuật toán di truyền

Kết quả cho thấy, tỉ lệ các bước sóng được chọn bằng phương pháp GA-PLS so với tổng số bước sóng ban đầu tương đối nhỏ. Tỉ lệ bước sóng được giữ lại của MTZ và SPI trong tập dữ liệu AgimDogyl lần lượt là 12,2% và 8,5%; đối với IND và AML trong tập dữ liệu Natrixam đều là 10,5%; đối với RIF, INZ và PYR trong tập dữ liệu Turbezid là 9,6%; 9,5% và 8,7%.

3.3.4 Kết quả áp dụng phương pháp lựa chọn bước sóng để xác định hàm lượng các chất trong các mẫu thuốc đa thành phần

3.3.4.1 Ứng dụng phương pháp lựa chọn bước sóng để xác định đồng thời MTZ và SPI trong viên nén AgimDogyl

a. *Xây dựng mô hình PLS dựa trên các thuật toán lựa chọn bước sóng*

Đã áp dụng phương pháp PLS để xây dựng mô hình hiệu chuẩn cho các tập hiệu chuẩn AgimDogyl với dữ liệu ở các bước sóng được lựa chọn theo các phương pháp SR-PLS, BVE-PLS và GA-PLS. Kết quả cho thấy các mô hình đều đạt yêu cầu về đánh giá các chỉ số

thông kê. Trong đó phương pháp SR cho thấy vượt trội hơn về độ chính xác và khả năng phù hợp với dữ liệu.

Đánh giá hiệu quả các phương pháp SR-PLS, BVE-PLS, GA-PLS dựa trên kết quả dự đoán trên tập kiểm tra. Độ thu hồi của các phương pháp có sự khác biệt không đang kể, đối với MTZ khoảng % độ thu hồi đối với phương pháp SR là 100,4% - 104,8%, phương pháp BVE là 98,9% - 104% và phương pháp GA là 98,7% - 104,1%; tương tự, 97,4% - 102,1%, 97,1% - 101,8% và 97,6% - 103,0% là các khoảng phần trăm độ thu hồi của SPI ở các nồng độ khác nhau của các phương pháp tương ứng. Giá trị RMSEP cho MTZ của các phương pháp SR, BVE, GA lần lượt là 0,5313 (3,5%), 0,330 (2,2%); 0,336 (2,2%), giá trị này nhỏ so với phần trăm nồng độ trung bình của tập kiểm tra đều nhỏ hơn 5% cho thấy phương pháp cho kết quả dự đoán tốt. Trong khi đó, phần trăm RMSEP của SPI so với nồng độ trung bình của SPI trong tập kiểm tra trong các phương pháp SR, BVE và GA lần lượt là 1,3%; 1,6% và 2,4% cho kết quả dự đoán rất tốt (đối với phương pháp SR và BVE) và tốt đối với phương pháp GA.

b. *Đánh giá phương pháp phân tích*

* Đánh giá độ thu hồi MTZ và SPI bằng phương pháp thêm chuẩn đều cho kết quả thỏa mãn theo qui định của AOAC.

- Độ lặp lại được thực hiện với hai ngày khác nhau trên mẫu thuốc AgimDogyl.

+ Ngày 1: với MTZ: $m_{TB}^{SR} = 14{,}58$ µg/mL (RSD = 1,85), $m_{TB}^{BVE} = 15{,}04$ µg/mL (RSD = 0,93), $m_{TB}^{GA} = 14{,}87$ µg/mL (RSD = 0,95); với SPI: $m_{TB}^{SR} = 94{,}28$ IU/mL (RSD = 1,17), $m_{TB}^{BVE} = 93{,}25$ IU/mL (RSD = 1,13), $m_{TB}^{GA} = 93{,}24$ IU/mL (RSD = 1,67).

+ Ngày 2: với MTZ: $m_{TB}^{SR} = 14{,}35$µg/mL (RSD = 0,70), $m_{TB}^{BVE} = 15{,}08$ µg/mL (RSD = 0,86), $m_{TB}^{GA} = 14{,}87$µg/mL (RSD = 0,84); với SPI: $m_{TB}^{SR} = 94{,}76$ IU/mL (RSD = 0,96), $m_{TB}^{BVE} = 93{,}74$ IU/mL (RSD = 0,69), $m_{TB}^{GA} = 91{,}73$IU/mL (RSD = 0,26).

So sánh kết quả trung bình giữa các dãy số liệu giữa kết quả độ lặp lại giữa 2 ngày bằng chuẩn t, kết quả cho thấy các giá trị $t_{tính}$ đều nhỏ hơn $t_{bảng}$, chứng tỏ các phương pháp khi xử lý kết quả trên các tập dữ liệu đều cho kết quả có độ chính xác cao.

* *Kết quả định lượng*

Kết quả định lượng MTZ với khoảng tin cậy (α =0,05): 122,06 ± 3,62 mg; 124,21 ± 2,67 mg; 123,54 ± 2,20 mg tương ứng với các phương pháp SR-PLS; BVE-PLS; GA-PLS. Kết quả định lượng SPI với khoảng tin cậy (α = 0,05): 779805 ± 11446 IU; 774601 ± 15712 IU; 776592 ± 14223 IU tương ứng với các phương pháp SR-PLS; BVE-PLS; GA-PLS. Hàm lượng phần trăm so với hàm lượng ghi trên nhãn do nhà sản xuất công bố được qui định trong Dược điển Việt nam V, trong đó giới hạn phần trăm cho phép từ 90-110%, kết quả từ bảng 3.32 và 3.33 cho thấy các kết quả xác định bởi các phương pháp là đáng tin cậy.

Kết quả phân tích ANOVA kết quả định lượng giữa các phương pháp SR-PLS, BVE-PLS, GA-PLS và phương pháp tham chiếu. Đối với MTZ, kết quả $F_{tính} = 2,20 < F_{bảng} = 4,07$, $p = 0,166 > 0,05$ cho thấy giá trị trung bình kết quả phân tích MTZ của các nhóm không khác nhau đáng kể. Tương tự, đối với SPI có $F_{tính} = 1,98 < F_{bảng} = 4,07$; $p = 0,196 > 0,05$ cho thấy giá trị trung bình kết quả phân tích SPI của các nhóm không khác nhau đán kể. Kết quả này khẳng định các phương pháp đã xây dựng có độ đúng tốt.

3.3.4.2 Ứng dụng phương pháp lựa chọn bước sóng để xác định đồng thời IND và AML trong viên nén Natrixam 1,5/5

a. *Xây dựng mô hình PLS dựa trên các thuật toán lựa chọn bước sóng*

Đối với tập hiệu chuẩn: các tham số thống kê đánh giá độ chính xác, tin cậy các phương pháp SR-PLS, BVE-PLS, GA-PLS đều đạt theo yêu cầu.

Áp dụng để dự đoán hàm lượng IND và AML ở các mẫu được chuẩn bị trong phòng thí nghiệm đều cho kết quả tốt: độ thu hồi đều gần 100%.

b. *Đánh giá phương pháp phân tích*

* Đánh giá độ đúng:

- Độ thu hồi khi định lượng IND và AML đối với phương pháp BVE-PLS chưa đạt yêu cầu, chứng tỏ các bước sóng được lựa chọn theo phương pháp BVE-PLS chưa tối ưu.

- Độ thu hồi khi định lượng IND và AML đối với phương pháp SR-PLS và GA-PLS thỏa mãn theo yêu cầu của AOAC.

- So sánh kết quả định lượng với phương pháp HPLC:

+ Phương pháp SR: các giá trị $t_{tính}$ có được lần lượt là 3,88 và 5,17 cho IND và AML đều lớn hơn giá trị $t_{(0,05;4)}$ = 2,78 cho thấy có giá trị trung bình khác nhau có ý nghĩa thống kê.

+ Phương pháp BVE-PLS: giá trị trung bình của AML có sự khác biệt đáng kể khi so sánh với phương pháp HPLC.

+ PHương pháp GA-PLS: kết quả định lượng IND và AML không có sự khác biệt với với phương pháp HPLC.

* Đánh giá độ chính xác

- Đánh giá độ lặp lại bằng cách phân tích lặp lại các mẫu thuốc Natrixam với kết quả đều cho giá trị RSD nhỏ hơn ½ giá trị RSD tính theo hàm Horwitz

Như vậy, đối với thuốc Natrixam chỉ có phương pháp GA-PLS là đạt yêu cầu về độ đúng và độ chính xác khi xác định đồng thời IND và AML trong viên nén.

c. *Kết quả định lượng*

Kết quả định lượng dựa theo phương pháp GA-PLS: m_{IND} = 1,47 ± 0,038 mg/viên; m_{AML} = 4,83 ± 0,080 mg/viên.

3.3.4.3 Ứng dụng phương pháp lựa chọn bước sóng để xác định đồng thời RIF, INZ và PYR trong viên nén Turbezid

a. *Xây dựng mô hình PLS dựa trên các thuật toán lựa chọn bước sóng*

Trong nghiên cứu này, hai phương pháp lựa chọn bước sóng — BVE-PLS và GA-PLS — đã được áp dụng để xây dựng mô hình hiệu chuẩn xác định đồng thời rifampicin (RIF), isoniazid (INZ) và pyrazinamid (PYR) trong viên nén Turbezid. Các mô hình được đánh giá thông qua các chỉ số thống kê trên tập hiệu chuẩn như RMSEC, RMSECV, R^2 và kiểm định phân phối phần dư bằng Shapiro-Wilk. Kết quả cho thấy cả hai phương pháp đều cho mô hình chính xác, đáng tin cậy với $R^2 > 0,98$ và phần dư phân phối chuẩn.

Trên tập kiểm tra, cả hai mô hình đều cho độ thu hồi trung bình xấp xỉ 100% và RMSEP thấp, cho thấy khả năng dự đoán tốt trên mẫu thực. Tuy nhiên, mô hình BVE-PLS có độ lặp lại và độ đúng cao hơn, đặc biệt trong việc xác định INZ, trong khi GA-PLS có số bước sóng chọn ít hơn, dẫn đến giảm độ nhạy và sai số lớn hơn ở một số chỉ tiêu.

b. *Đánh giá phương pháp phân tích*

Về đánh giá độ đúng, các giá trị độ thu hồi đều nằm trong khoảng quy định (80–110%) với RSD nhỏ, đáp ứng yêu cầu theo Dược

điển. Riêng ở mức nồng độ thấp, INZ xác định bằng GA-PLS có độ lặp lại chưa đạt theo tiêu chuẩn Horwitz.

Khi so sánh với kết quả định lượng bằng HPLC tại trung tâm kiểm nghiệm, phương pháp BVE-PLS cho kết quả tương đương HPLC, trong khi GA-PLS cho sai số đáng kể đối với RIF và INZ. Ngoài ra, đánh giá hàm lượng ghi trên nhãn thuốc cho thấy chỉ BVE-PLS đáp ứng tiêu chuẩn 90–110% theo Dược điển Việt Nam V, trong khi GA-PLS xác định INZ dưới ngưỡng yêu cầu.

c. *Kết quả định lượng*

Theo phương pháp BVE-PLS, m_{RIF} = 158,11 ± 7,42 mg/viên; m_{INZ} = 75,55 ± 2,36 mg/viên; m_{PYR} = 408,45 ± 10,15 mg/viên.

3.4 Đánh giá và so sánh hiệu quả các phương pháp phân tích đa biến

3.4.1 Phương pháp PLS so với một số phương pháp phân tích đa biến khác

So sánh CLS, PCR và PLS cho thấy PLS nhìn chung vượt trội với RMSEP thấp, R^2 cao và độ thu hồi tốt hơn, đặc biệt trong dữ liệu phức tạp. CLS kém hiệu quả nhất do yêu cầu phổ thuần khiết khó đáp ứng, dẫn đến sai số lớn. PCR có hiệu suất tương đương với phương pháp PLS, tuy nhiên trong một số trường hợp, PLS vẫn có khả năng dự đoán chính xác hơn. Nhìn chung, PLS là lựa chọn đáng tin cậy nhất, còn PCR chỉ phù hợp trong một số trường hợp cụ thể.

3.4.2 Phương pháp PLS và phương pháp lựa chọn bước sóng

So sánh PLS với các phương pháp lựa chọn bước sóng cho thấy hiệu quả khác nhau. SR-PLS thường cải thiện RMSEP trong một số trường hợp nhưng không ổn định giữa các thuốc. BVE-PLS đôi khi làm tăng RMSEP do loại bỏ sai biến quan trọng, song có thể cải thiện RSD. GA-PLS có tiềm năng giảm RMSECV và RMSEC nhờ tối ưu di truyền, nhưng dự đoán (RMSEP, RSD) chưa nhất quán do phụ thuộc tham số. Nhìn chung, PLS vẫn đáng tin cậy nhất với R^2 cao và độ thu hồi tốt, trong khi các phương pháp lựa chọn bước sóng chỉ hữu ích trong trường hợp cụ thể.

KẾT LUẬN

Trên cơ sở các thực nghiệm đã tiến hành, luận án đã chứng minh được rằng phương pháp hồi quy bình phương tối thiểu riêng phần (PLS) kết hợp phổ hấp thụ UV-Vis có thể được xây dựng và ứng dụng để định lượng đồng thời các hoạt chất có phổ xen phủ trong chế phẩm dược.

(1) Luận án đã xây dựng thành công các mô hình PLS cho các hệ hai và ba thành phần (MTZ–SPI, AML–ATO, SPR–FUR, IND–AML, RIF–INZ–PYR) với dữ liệu hiệu chuẩn được thiết kế bằng CCD. Việc lựa chọn số biến tiềm ẩn dựa trên tiêu chuẩn PRESS cho phép tối ưu hoá mô hình với số latent variables từ 2 – 7.

(2) Các chỉ tiêu thẩm định (RMSEC, RMSECV, RMSEP, R^2, phân tích phần dư, độ thu hồi, độ lặp lại) đều thoả mãn yêu cầu: RMSE < 5% so với nồng độ trung bình; phần dư tuân theo phân phối chuẩn (Shapiro-Wilk p > 0,05); R^2 xấp xỉ 1; độ thu hồi nằm trong 90 – 110% (AOAC). Kiểm định t cho thấy không có khác biệt có ý nghĩa thống kê giữa kết quả PLS–UV và phương pháp chuẩn HPLC (p > 0,05).

(3) Phương pháp đã được ứng dụng để định lượng các hoạt chất trong thuốc AgimDogyl, Zoamco-A, Spirosemid-40, Natrixam và Turbezid trên thị trường. Kết quả hàm lượng thu được nằm trong giới hạn cho phép theo DĐVN V.

(4) So sánh với các thuật toán hồi quy khác cho thấy PLS có độ chính xác vượt trội do khả năng khử nhiễu tốt hơn so với PCR và CLS.

(5) Luận án đã đề xuất và chuẩn hoá quy trình thực hành PLS–UV cho các hệ phổ xen phủ, đồng thời nghiên cứu thành công cơ chế rút gọn biến bằng các thuật toán SR, BVE và GA để lựa chọn bước sóng tối ưu, giúp giảm số biến đầu vào mà không làm giảm chất lượng mô hình.

Như vậy, phương pháp PLS kết hợp với dữ liệu phổ UV-Vis là một giải pháp khả thi, chi phí thấp, có thể thay thế cho HPLC trong các trường hợp phù hợp để định lượng đa cấu tử trong dược phẩm.

KIẾN NGHỊ

1. Tiếp tục phát triển và thẩm định phương pháp PLS kết hợp phổ hấp thụ UV-Vis trên nhiều dạng thuốc và hoạt chất khác nhau, đặc biệt là các thuốc có phổ hấp thụ xen phủ. Việc này nhằm đánh giá tính ứng dụng rộng rãi, độ chính xác và độ tin cậy của phương pháp trong kiểm nghiệm thuốc, hướng đến việc áp dụng trong thực tế sản xuất và kiểm tra chất lượng dược phẩm.

2. Mở rộng và thử nghiệm nhiều chiến lược lựa chọn bước sóng khác nhau nhằm cải thiện hiệu quả của mô hình PLS. Việc đa dạng hóa phương pháp chọn biến sẽ giúp tăng độ chính xác, giảm độ phức tạp và rút ngắn thời gian tính toán trong phân tích phổ,

3. Nghiên cứu tích hợp các phương pháp học máy (Machine Learning) và học sâu (Deep Learning) vào xử lý và dự đoán dữ liệu phổ hấp thụ. Đây là hướng đi tiềm năng giúp khai thác tối đa thông tin từ dữ liệu phổ, đồng thời nâng cao khả năng mô hình hóa, độ chính xác và khả năng khái quát hóa của các phương pháp phân tích định lượng không phân tách các chất trong mẫu.

4. Xây dựng công cụ phần mềm: phát triển một ứng dụng phân tích dành cho phòng lab (ứng dụng web Shiny/R) để nhập phổ UV-Vis, tự động

xây dựng mô hình PLS, thực hiện đánh giá, kiểm định, báo cáo kết quả – giúp phương pháp có tính ứng dụng thực tế cao hơn, dễ chuyển giao hơn cho phòng QC/QA.

DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ KẾT QUẢ NGHIÊN CỨU CỦA LUẬN ÁN

I. Tạp chí trong nước

1. **Nguyễn Duy Lưu**, Lê Văn Thuận, Nguyễn Thế Khang, Nguyễn Hùng Nhật Duy, Phạm Phú Quốc, Võ Thị Thanh Trúc, Nguyễn Đình Luyện, Trần Thúc Bình (2025), Thuật toán lựa chọn biến ngược, Hồi quy thành phần chính, bình phương tối thiểu riêng phần: định lượng đồng thời Rifampicin, Isoniazid và Pyrazinamid trong thuốc viên nén, *Tạp chí Khoa học và Công nghệ, Trường Đại học Khoa học, Đại học Huế,* Tập 28, Số 2, trang 1-15.

2. **Nguyễn Duy Lưu**, Nguyễn Thị Thanh Trầm, Phạm Phú Quốc, Nguyễn Hùng Nhật Duy, Nguyễn Đình Luyện, Trần Thúc Bình (2024), Định lượng đồng thời Amlodipin và Atorvastatin trong thuốc đa thành phần bằng phương pháp quang phổ UV-Vis kết hợp phân tích đa biến, *Tạp chí Khoa học Đại học Huế: Khoa học Tự nhiên,* Tập 133, Số 1C, trang 39 – 52.

3. Nguyễn Đình Luyện, Nguyễn Quỳnh Giao, Phạm Phú Quốc, Nguyễn Hùng Nhật Duy, **Nguyễn Duy Lưu**, Trần Thúc Bình (2024), Phương pháp lựa chọn bước sóng cho phổ UV-Vis kết hợp bình phương tối thiểu riêng phần để định lượng đồng thời Amlodipin và Indapamid dạng viên nén, *Tạp chí phân tích Hóa, Lý và Sinh học,* Tập 30, số 2A, trang 1 – 12.

II. Tạp chí quốc tế (WoS/Scopus)

1. **Nguyen Duy Luu**, Tran Thi My Duyen, Nguyen Dinh Luyen, Tran Thuc Binh (2024), A novel method of the simultaneous determination of Spironolactone and Furoseimide in pharmaceutical tablets, *Journal of Analytical Chemistry*, Vol. 79, No. 7, pp. 923-933. (SCIE, Q3, IF = 1,0). https://doi.org/10.1134/S1061934824700291

2. **Nguyen Duy Luu**, Nguyen Hung Nhat Duy, Pham Phu Quoc, Nguyen The Khang, Le Thanh Huyen, Nguyen Dinh Luyen, Tran Thuc Binh (2025), Simultaneous quantification of Metronidazole and Spiramycin by Ultraviolet-Visible method: Comparison of wavelength selection techniques, *Journal of Analytical Chemistry*, Vol. 80, No. 6, pp. 1005 – 1016. (SCIE, Q3, IF = 1,0). https://doi.org/10.1134/S1061934824605486