

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 PHỬ NHAU BẰNG PHƯƠNG PHÁP BÌNH
PHƯƠNG TỐI THIỂU RIÊNG PHẦN**

LUẬN ÁN TIẾN SĨ HÓA HỌC

HUẾ, 2025

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 PHỬ NHAU BẰNG PHƯƠNG PHÁP BÌNH
PHƯƠNG TỐI THIỂU RIÊNG PHẦN**

Ngành: Hóa phân tích

Mã số: 944.01.18

LUẬN ÁN TIẾN SĨ HÓA HỌC

Người hướng dẫn khoa học:

- 1. PGS.TS Trần Thúc Bình**
- 2. PSG.TS Nguyễn Đình Luyện**

HUẾ, 2025

LỜI CAM ĐOAN

Tôi xin cam đoan rằng toàn bộ luận án này là kết quả nghiên cứu của bản thân tôi thực hiện dưới sự hướng dẫn của PGS. TS Trần Thúc Bình và PGS. TS Nguyễn Đình Luyện. Tôi khẳng định luận án chưa từng được sử dụng để đăng ký xét cấp bằng tại bất kỳ cơ sở đào tạo nào khác. Các số liệu và kết quả nghiên cứu trong luận án là trung thực, chính xác, được các đồng tác giả cho phép sử dụng và chưa từng được công bố trong bất cứ một công trình nào khác.

Tác giả luận án

Nguyễn Duy Lưu

LỜI CẢM ƠN

Lời đầu tiên, tôi xin gửi lời cảm ơn chân thành và sâu sắc nhất đến PGS. TS. Trần Thúc Bình và PGS. TS. Nguyễn Đình Luyện – những người thầy, người hướng dẫn tận tâm đã dành thời gian, công sức và tâm huyết chỉ dẫn, giúp đỡ tôi trong suốt quá trình thực hiện luận án này.

Tôi cũng vô cùng biết ơn GS. TS. Đinh Quang Khiếu, Trưởng khoa Hóa học cùng toàn thể quý thầy cô trong bộ môn Hóa phân tích, Trường Đại học Khoa học, Đại học Huế đã hỗ trợ nhiệt tình, tạo điều kiện thuận lợi và giúp tôi hoàn thành nghiên cứu một cách tốt đẹp.

Xin gửi lời cảm ơn sâu sắc đến Ban Giám hiệu trường Đại học Kỹ thuật Y - Dược Đà Nẵng, Ban Chủ nhiệm Khoa Dược, các thầy cô trong bộ môn Kiểm nghiệm thuốc cùng các em sinh viên khoa Dược, những người đã tạo điều kiện học tập, nghiên cứu tốt nhất cho tôi trong suốt thời gian qua.

Tôi không thể không nhắc đến sự động viên, hỗ trợ về tinh thần to lớn từ ba mẹ, anh chị trong gia đình và bạn bè thân thiết, những người luôn đồng hành, chia sẻ cùng tôi trong suốt quãng đường học tập và nghiên cứu.

Cuối cùng, tôi xin dành trọn vẹn tình cảm và lời cảm ơn sâu sắc nhất đến vợ Võ Thị Lan Anh – người bạn đời tuyệt vời đã luôn sát cánh, chia sẻ và là nguồn động lực to lớn cho tôi; cùng với hai con yêu ngoan ngoãn, vâng lời, luôn mang đến niềm vui và sự ấm áp trong cuộc sống.

Không có sự hỗ trợ quý báu ấy, tôi sẽ không thể hoàn thành được luận án này.

Thành phố Huế, ngày tháng năm 2025

DANH MỤC CÁC CHỮ CÁI VIẾT TẮT, KÝ HIỆU

Viết tắt	Tiếng anh	Tiếng việt
AI	Artificial Intelligence	Trí tuệ nhân tạo
AML	Amlodipin Besylat	Amlodipin
ANN	Artificial Neural Network	Mạng Nơ-ron nhân tạo
API	Active pharmaceutical ingredient	thành phần hoạt chất trong thuốc
ATO	Atorvastatin calcium	Atorvastatin
BVE	Backward Variable Elimination	Lựa chọn biến ngược
CCD	Central composite design	Thiết kế hỗn hợp trung tâm
CLS	Classical Least Squares	Bình phương tối thiểu cổ điển
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
CV	Cross validation	Thẩm định chéo
CV LOO	Leave-one-out cross-validation	Kiểm tra chéo loại bỏ từng mẫu một
EMA	European Medicines Agency	Cơ quan Dược phẩm châu Âu
FDA	Food and Drug Administration	Cục Quản lý Thực phẩm và Dược phẩm Hoa Kỳ
FUR	Furosemid	Furosemid
GA	Genetic Algorithm	Thuật toán di truyền
HPLC	High Performance Liquid Chromatography	Sắc ký lỏng hiệu năng cao
HPTLC	High-Performance Thin-Layer Chromatography	Sắc ký lớp mỏng hiệu năng cao
ILS	Inverse Least Squares	bình phương tối thiểu nghịch đảo
IND	Indapamid	Indapamid
INZ	Isoniazid	Isoniazid
iPLS	interval Partial Least Squares	PLS theo khoảng
LOD	Limit of Detection	Giới hạn phát hiện
LV	Latent Variable	Biến tiềm ẩn
MLR	Multiple Linear Regression	Hồi quy tuyến tính đa biến
MTZ	Metronidazol	Metronidazol
NIPLS	Non-linear Iterative Partial Least-Squares	Bình phương nhỏ nhất từng phần lặp phi tuyến
NIR	Near-infrared	Phổ hồng ngoại gần
NMR	Nuclear Magnetic Resonance	Phổ cộng hưởng từ hạt nhân
PC	Principal Component	Thành phần chính

Viết tắt	Tiếng anh	Tiếng việt
PCA	Principal Component Analysis	phân tích thành phần chính
PCR	Principal Component Regression	Hồi quy thành phần chính
PLS	Partial Least Squares	Bình phương tối thiểu riêng phần
PRESS	Predicted Residual Error Sum of Squares	Tổng bình phương sai số dư dự đoán
PYR	Pyrazynamid	Pyrazynamid
QC	Quality Control	Kiểm soát chất lượng
QSAR	Quantitative Structure Activity relationship	Mối quan hệ cấu trúc-hoạt tính định lượng
RIF	Rifampicin	Rifampicin
RMSEC	Root Mean Square Error of Calibration	Căn bậc hai sai số trung bình của tập hiệu chuẩn
RMSECV	Root Mean Square Error of Cross-Validation	Căn bậc hai sai số trung bình của kiểm định chéo
RMSEP	Root Mean Square Error of Prediction	Căn bậc hai sai số trung bình của tập kiểm tra
RSD	Relative Standard Deviation	Độ lệch chuẩn tương đối
SD	Standard Deviation	Độ lệch chuẩn
siPLC	Synergy Partial Least Squares	PLS cộng hưởng
SPI	Spiramycin	Spiramycin
SPR	Spironolacton	Spironolacton
SR	Selectivity Ratio	Tỉ lệ chọn lọc
SVD	singula value decomposition	Phân rã giá trị suy biến
SVM	Support Vector Machine	Máy vector hỗ trợ
UIPAC	International Union of Pure and Applied Chemistry	Liên đoàn Quốc tế về Hóa học Cơ bản và Ứng dụng
UV-Vis	Ultraviolet–Visible	bức xạ vùng tử ngoại - khả kiến
VIP	Variable Importance Projection	Phép chiếu tầm quan trọng của biến
WHO	World Health Organization	Tổ chức Y tế thế giới

MỤC LỤC

	Trang
LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC CÁC CHỮ CÁI VIẾT TẮT, KÝ HIỆU	iii
MỤC LỤC	v
DANH MỤC CÁC BẢNG.....	viii
DANH MỤC CÁC HÌNH	xi
MỞ ĐẦU.....	1
CHƯƠNG 1. TỔNG QUAN TÀI LIỆU	5
1.1. PHƯƠNG PHÁP QUANG PHỔ UV-VIS	5
1.2. PHƯƠNG PHÁP CHEMOMETRICS	7
1.3 MỘT SỐ PHƯƠNG PHÁP PHÂN TÍCH ĐA BIẾN	10
1.3.1 Hiệu chuẩn đơn biến và hiệu chuẩn đa biến	10
1.3.2 Phương pháp bình phương tối thiểu cổ điển.....	11
1.3.2 Phương pháp bình phương tối thiểu nghịch đảo.....	12
1.3.4 Phương pháp hồi quy thành phần chính	13
1.3.5 Phương pháp bình phương tối thiểu riêng phần	16
1.4 TỔNG QUAN VỀ ÁP DỤNG PHƯƠNG PHÁP PLS KẾT HỢP CÁC PHƯƠNG PHÁP PHỔ TRONG PHÂN TÍCH CÁC CHẤT	21
1.5 GIỚI HẠN PHÁT HIỆN TRONG PHÂN TÍCH ĐA BIẾN.....	29
1.6 LỰA CHỌN BƯỚC SÓNG TRONG PHÂN TÍCH QUANG PHỔ	30
1.6.1 Sự cần thiết phải lựa chọn bước sóng.....	30
1.6.2 Tổng quan về phương pháp lựa chọn bước sóng trong PLS	32
1.7 GIỚI THIỆU AMLODIPIN BESYLAT, INDAPAMID, ATORVASTATIN CALCIUM, SPIRONOLACTON, FUROSEMID, RIFAMPICIN, ISONIAZID, PYRAZYNAMID, METRONIDAZOL, SPIRAMYCIN.....	38
1.7.1 Amlodipin Besylat.....	38
1.7.2 Indapamid	38
1.7.3 Atorvastatin calcium.....	40
1.7.4 Spironolacton.....	41
1.7.6 Rifampicin	43
1.7.7 Isoniazid	44
1.7.8 Pyrazynamid	44
1.7.9 Metronidazol.....	45
1.7.10 Spiramycin.....	46
CHƯƠNG 2. NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU	48

2.1. NỘI DUNG NGHIÊN CỨU.....	48
2.2. PHƯƠNG PHÁP NGHIÊN CỨU	48
2.2.1. Thiết kế thí nghiệm.....	48
2.2.2. Phương pháp thẩm định chéo	50
2.2.3. Lựa chọn số biến tiềm ẩn tối ưu cho mô hình PLS	51
2.2.4. Giới hạn phát hiện trong mô hình PLS	52
2.2.5. Lựa chọn bước sóng bằng phương pháp tỉ lệ chọn lọc trong PLS	55
2.2.6. Lựa chọn bước sóng bằng phương pháp loại bỏ biến ngược trong PLS	58
2.2.7. Lựa chọn bước sóng bằng phương pháp thuật toán di truyền trong PLS	60
2.3 THỰC NGHIỆM	62
2.3.1. Thiết bị, hóa chất	62
2.3.2. Chuẩn bị mẫu.....	63
2.3.3 Đánh giá độ đúng.....	64
2.3.4. Đánh giá độ lặp lại.....	65
2.4. PHƯƠNG PHÁP ĐÁNH GIÁ THỐNG KÊ	65
2.4.1. Khoảng tin cậy.....	65
2.4.2. So sánh độ lặp lại giữa hai dãy thí nghiệm khác nhau(chuẩn F – Fisher)	65
2.4.3. So sánh giá trị trung bình của 2 dãy thí nghiệm	66
2.4.4. Đánh giá phân phối chuẩn của tập dữ liệu.....	66
Chương 3. KẾT QUẢ VÀ THẢO LUẬN	68
3.1 THIẾT KẾ THÍ NGHIỆM VÀ TÍNH CHẤT QUANG PHỔ CỦA CÁC HOẠT CHẤT CẦN NGHIÊN CỨU.....	68
3.2 KẾT QUẢ PHƯƠNG PHÁP PLS KẾT HỢP PHỔ HẤP THỤ UV-VIS	69
3.2.1 Kết quả thuật toán PLS	69
3.2.2 Kết quả xác định số biến tiềm ẩn trong mô hình PLS	76
3.2.3 Kết quả đánh giá mô hình PLS để xác định đồng thời MTZ và SPI.....	80
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.....	83
3.2.5 Kết quả định lượng MTZ và SPI trong viên nén Agimdogyl.....	85
3.2.6. Ứng dụng phương pháp PLS định lượng đồng thời các hoạt chất trong thuốc đa thành phần	86
3.2.7 Kết luận phương pháp PLS.....	95
3.3 LỰA CHỌN BƯỚC SÓNG TRONG PLS	97
3.3.1 Lựa chọn bước sóng bằng phương pháp tỉ lệ chọn lọc kết hợp bình phương tối thiểu riêng phần (SR-PLS)	97
3.3.2 Lựa chọn bước sóng bằng phương pháp loại bỏ biến ngược kết hợp bình phương tối thiểu riêng phần (BVE-PLS)	100

3.3.3 Lựa chọn bước sóng bằng thuật toán di truyền kết hợp bình phương tối thiểu riêng phần	102
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	106
3.3.5 Kết luận phương pháp lựa chọn bước sóng trong PLS.....	121
3.4 ĐÁNH GIÁ VÀ SO SÁNH HIỆU QUẢ CÁC PHƯƠNG PHÁP PHÂN TÍCH ĐA BIẾN	122
3.4.1 Phương pháp PLS so với một số phương pháp phân tích đa biến khác	122
3.4.2 Phương pháp PLS và phương pháp lựa chọn bước sóng.....	128
KẾT LUẬN	133
KIẾN NGHỊ	134
DANH MỤC CÁC CÔNG TRÌNH CÔNG BỐ KẾT QUẢ NGHIÊN CỨU CỦA LUẬN ÁN	135
TÀI LIỆU THAM KHẢO.....	136

DANH MỤC CÁC BẢNG

Bảng 1.1. Kết quả một số nghiên cứu phân tích định lượng bằng phương pháp quang phổ kết hợp PLS trong Dược phẩm	25
Bảng 1.2 Số lượng các nghiên cứu về lựa chọn bước sóng từ 2000 - 2024.	32
Bảng 1.3. Một số công trình nghiên cứu về định lượng đồng thời Amlodipin và Indapamid trên thế giới.....	39
Bảng 1.4. Một số công trình nghiên cứu xác định đồng thời AML và ATO trên thế giới	41
Bảng 1.5. Một số công trình nghiên cứu về định lượng đồng thời Spironolacton và Furosemid trên thế giới.....	43
Bảng 1.6. Một số công trình nghiên cứu về định lượng đồng thời Spironolacton và Furosemid trên thế giới.....	46
Bảng 2.1. Mã hóa các điểm thiết kế thí nghiệm theo CCD.....	49
Bảng 2.2. Các thông số chuẩn bị mẫu thuốc	64
Bảng 3.1. Nồng độ của các chất trong tập hiệu chuẩn của các mẫu thuốc.....	68
Bảng 3.2. Các tham số được trích xuất trong mô hình PLS	70
Bảng 3.3. Các tham số thống kê của mô hình PLS cho MTZ và SPI ở các biến tiềm ẩn khác nhau	77
Bảng 3.4. Kết quả phân tích thống kê kiểm tra chéo loại bỏ một phần theo Haaland và Thomas cho tập hiệu chuẩn AgimDogyl	78
Bảng 3.5. Kết quả phân tích thống kê giá trị PRESS trên tập hiệu chuẩn AgimDogyl theo Haalaand và Thomas	79
Bảng 3.6. Kết quả các thông số thống kê của mô hình PLS cho MTZ và SPI ở các biến tiềm ẩn được chọn	80
Bảng 3.7. Kết quả định lượng MTZ và SPI đối với các mẫu trong tập kiểm tra.....	82
Bảng 3.8. Kết quả đánh giá độ thu hồi của MTZ và SPI	83
Bảng 3.9. Kết quả đánh giá độ đúng thông qua so sánh với phương pháp tham chiếu	84
Bảng 3.10. Kết quả đánh giá độ lặp lại định lượng đồng thời MTZ và SPI trong viên nén AgimDogyl	85
Bảng 3.11. Kết quả định lượng MTZ và SPI trong viên nén AgimDogyl bằng phương pháp PLS.....	86
Bảng 3.12. Các thông số thống kê đánh giá các tập hiệu chuẩn Spirosemid, Zoamco-A và Natrixam bằng phương pháp PLS	87
Bảng 3.13. Kết quả đánh giá độ thu hồi và RMSEP trên tập kiểm tra của Spirosemid, Zoamco-A và Natrxam	88
Bảng 3.14. Kết quả đánh giá độ thu hồi các thuốc Spirosemid, Zoamco-A và Natrixam	89
Bảng 3.15. Kết quả đánh giá độ lặp lại các thuốc Spirosemid, Zoamco-A và Natrixam	89
Bảng 3.16. Kết quả định lượng (mg/viên) các thành phần trong viên nén Spirosemid; Zoamco-A và Natrixam 1,5/5	90

Bảng 3.17. Kết quả hàm lượng phần trăm so với hàm lượng ghi trên nhãn của các thuốc Spirosemid, Zoamco-A và Natrixam.....	90
Bảng 3.18. Các thông số thống kê đánh giá trên tập hiệu chuẩn Turbezid bằng phương pháp PLS91	
Bảng 3.19. Kết quả định lượng RIF, INZ và PYR trong tập kiểm tra bằng phương pháp PLS	92
Bảng 3.20. Kết quả đánh giá độ thu hồi RIF, INZ và PYR bằng phương pháp PLS ...	93
Bảng 3.21. Kết quả đánh giá độ lặp lại xác định RIF, INZ và PYR trong viên nén Turbezid bằng phương pháp PLS	94
Bảng 3.22. Kết quả định lượng các thành phần trong viên turbezid (mg/viên) bằng phương pháp PLS	95
Bảng 3.23. Kết quả lựa chọn bước sóng theo phương pháp SR cho MTZ và SPI trong mô hình PLS	99
Bảng 3.24. Kết quả lựa chọn bước sóng bằng phương pháp BVE-PLS cho các mẫu thuốc AgimDogyl, Natrixam 1,5/5 và Turbezid.....	102
Bảng 3.25. Các thông số cho GA-PLS để lựa chọn bước sóng	102
Bảng 3.26. Giá trị RMSE ($\mu\text{g/mL}$) tốt nhất nhất qua 50 thế hệ của thuật toán di truyền	103
Bảng 3.27. Số lượng và tỉ lệ các bước sóng được chọn của các tập dữ liệu AgimDogyl, Natrixam và Turbezid bằng phương pháp GA-PLS.....	106
Bảng 3.28. Kết quả đánh giá các tham số thống kê đánh giá trên tập hiệu chuẩn bởi các mô hình lựa chọn bước sóng cho tập hiệu chuẩn AgimDogyl.....	107
Bảng 3.29. Kết quả đánh giá hiệu của hoạt động của các mô hình lựa chọn bước sóng trên tập nồng độ kiểm tra của thuốc AgimDogyl	108
Bảng 3.30. Kết quả đánh giá độ thu hồi của MTZ và SPI bằng các phương pháp lựa chọn bước sóng SR-PLS, BVE-PLS, GA-PLS	109
Bảng 3.31. Kết quả đánh giá độ lặp lại bằng các phương pháp SR-PLS, BVE-PLS và GA-PLS	110
Bảng 3.32. Kết quả xác định hàm lượng MTZ (mg/viên) và phần trăm so với hàm lượng ghi trên nhãn bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC	111
Bảng 3.33. Kết quả xác định hàm lượng SPI (IU/viên) và phần trăm so với hàm lượng ghi trên nhãn bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và vi sinh.....	111
Bảng 3.34. Kết quả phân tích ANOVA kết quả định lượng MTZ và SPI giữa các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC.....	112
Bảng 3.35. Kết quả đánh giá các tham số thống kê đánh giá trên tập hiệu chuẩn bởi các mô hình lựa chọn bước sóng cho tập hiệu chuẩn Natrixam 1,5/5.....	113
Bảng 3.36. Kết quả đánh giá hiệu của hoạt động của các mô hình lựa chọn bước sóng trên tập nồng độ kiểm tra của thuốc Natrixam 1,5/5	114
Bảng 3.37. Kết quả đánh giá độ thu hồi xác định nồng độ IND và AML bằng các phương pháp SR-PLS, BVE-PLS và GA-PLS	115
Bảng 3.38. Kết quả đánh giá độ lặp lại mẫu Natrixam bằng phương pháp SR-PLS, BVE-PLS, GA-PLS	115

Bảng 3.39. Kết quả xác định hàm lượng IND và giá trị phần trăm so với hàm lượng ghi trên nhãn trong viên nén Natrixam 1,5/5 bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC	116
Bảng 3.40. Kết quả xác định hàm lượng AML và giá trị phần trăm so với hàm lượng ghi trên nhãn trong viên nén Natrixam 1,5/5 bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC	116
Bảng 3.41 Kết quả đánh giá các tham số thông kê đánh giá trên tập hiệu chuẩn bởi các mô hình BVE-PLS, GA-PLS cho tập hiệu chuẩn Turbezid	117
Bảng 3.42. Kết quả xác định trên tập kiểm tra bằng phương pháp BVE-PLS và GA-PLS đối với hỗn hợp RIF, INZ và PYR	118
Bảng 3.43. Kết quả đánh giá độ thu hồi cho mẫu hỗn hợp Turbezid	119
Bảng 3.44. Kết quả đánh giá độ lặp lại và độ chính xác trung gian của thuốc Turbezid bằng các phương pháp BVE-PLS, GA-PLS	120
Bảng 3.45. Kết quả xác định hàm lượng RIF, INZ, PYR (mg/viên) trong viên nén Turbezid bằng các phương pháp BVE-PLS, GA-PLS và HPLC	120
Bảng 3.46. Kết quả so sánh với hàm lượng ghi trên nhãn của thuốc Turbezid bằng phương pháp BVE-PLS và GA-PLS	121
Bảng 3.47. Các chỉ số đánh giá mô hình và phân phối dữ liệu phân tích phần dư của các phương pháp PLS, PCR và CLS	123
Bảng 3.48. Các chỉ số đánh giá mô hình và phân phối dữ liệu phân tích phần dư của các phương pháp lựa chọn bước sóng và phương pháp PLS	128

DANH MỤC CÁC HÌNH

Hình 1.1. Sơ đồ hoạt động máy quang phổ UV-Vis hai chùm tia. Đèn vonfram phát ra ánh sáng khả kiến, trong khi đèn D2 tạo ra tia UV .	5
Hình 1.2. Mối quan hệ giữa các công nghệ phân tích hiện đại, phương pháp phân tích đa biến và ứng dụng trong các ngành khoa học đời sống, công nghiệp và y tế	7
Hình 1.3. Sự biến đổi giữa biến thực (real variables) và biến tiềm ẩn (latent variable) bằng cách nén (compression) hoặc giải nén (decompression) thông qua các vector tải (loading)	14
Hình 1.4. Minh họa thuật toán của PLS1	17
Hình 1.5. Minh họa thuật toán của PLS2	18
Hình 1.6. Phân tích hồi quy bình phương tối thiểu riêng phần	19
Hình 1.7. Biểu đồ mô tả số lượng các công trình nghiên cứu liên quan đến phương pháp PLS từ năm 2000 đến 2024	24
Hình 1.8. Các mô hình lựa chọn bước sóng	31
Hình 1.9. Biểu đồ số lượng các công trình nghiên cứu về lựa chọn bước sóng từ năm 2000 – 2024	34
Hình 1.10. Công thức cấu tạo của Amlodipin	38
Hình 1.11. Công thức cấu tạo của Indapamid	39
Hình 1.12. Công thức cấu tạo của Atorvastatin calcium	40
Hình 1.13. Công thức cấu tạo của Spironolacton]	41
Hình 1.14. Công thức cấu tạo của Furosemid	42
Hình 1.15. Công thức cấu tạo của Rifampicin	43
Hình 1.16. Công thức cấu tạo của Isoniazid	44
Hình 1.17. Công thức cấu tạo của Pyrazinamid	45
Hình 1.18. Công thức cấu tạo của Metronidazol	45
Hình 1.19. Công thức cấu tạo của Spiramycin	46
Hình 2.1. Thiết kế thí nghiệm cấu trúc phức hợp với 2, 3 yếu tố	48
Hình 2.2. Biểu đồ minh họa giá trị LOD tối thiểu và tối đa, các quyết định liên quan đến sự có mặt và vắng mặt của chất phân tích trong các phạm vi nồng độ khác nhau.	53
Hình 2.3 Sơ đồ thuật toán SR-PLS	58
Hình 2.4 Sơ đồ thuật toán BVE-PLS	59
Hình 2.5. Sơ đồ thuật toán GA-PLS	61
Hình 3.1. Phổ hấp thụ của các đơn chất trong các mẫu thuốc	69
Hình 3.2. Phổ hấp thụ UV của hỗn hợp MTZ và SPI trong các mẫu hiệu chuẩn	71
Hình 3.3a. Biểu đồ điểm của 12 biến tiềm ẩn đầu tiên của MTZ	72
Hình 3.4b. Biểu đồ điểm của 12 biến tiềm ẩn đầu tiên của SPI	73
Hình 3.5. Phổ của các vector chiếu dữ liệu quang phổ góc lên không gian tiềm ẩn PLS của MTZ và SPI với 4 biến tiềm ẩn đầu tiên	74
Hình 3.6. Phổ của các hệ số tải (P), hệ số trọng số tải (W), hệ số chiếu (P_{proj}) và hệ số hồi qui của MTZ và SPI	75

Hình 3.7. Biểu đồ giá trị PRESS trong mô hình PLS của MTZ và SPI	78
Hình 3.8. Biểu đồ thể hiện sự phân tán của phần dư quanh điểm 0 của MTZ và SPI..	81
Hình 3.9. Đồ thị biểu diễn mức độ chính xác giữa nồng độ dự đoán bởi mô hình PLS và nồng độ thực tế của MTZ và SPI.....	82
Hình 3.10 Kết quả xác định các giá trị SR cho MTZ và SPI trong mô hình PLS đã xây dựng.	97
Hình 3.11. Kết quả giá trị SR cho IND và AML trong mô hình PLS đã xây dựng	98
Hình 3.12. 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.....	100
Hình 3.13. Các giá trị VIP được tính cho IND và AML trong mô hình PLS tương ứng với số biến tiềm ẩn là 4 và 2.....	100
Hình 3.14. 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.....	101
Hình 3.15. Giá trị RMSECV qua 50 thế hệ di truyền của MTZ và SPI	105
Hình 3.16. Giá trị RMSECV qua 50 thế hệ di truyền của IND và MTZ.....	105
Hình 3.17. Giá trị RMSECV qua 50 thế hệ di truyền của RIF, INZ và PYR.....	105
Hình 3.18. So sánh các giá trị RMSE trong các phương pháp PLS, CLS và PLS cho các tập thuốc Zoamco_A ((a) AML; (b) ATO) và Spirosemid-40 ((c) FUR; (d) SPR)	124
Hình 3.19. So sánh các giá trị RMSE trong các phương pháp PLS, CLS và PLS cho các tập thuốc Natrixam 1,5/5 ((e) AML; (f) IND) và AgimDogyl ((g) MTZ; (h) SPI)	125
Hình 3.20. So sánh các giá trị RMSE trong các phương pháp PLS, CLS và PLS cho các tập thuốc Turbezid ((i) RIF, (j) INZ và (k) PYR)	125
Hình 3.21. Giá trị p trong kiểm tra phân phối chuẩn của phần dư trong tập hiệu chuẩn của các phương pháp CLS, PCR và PLS bằng kiểm định Shapiro-Wilk.....	126
Hình 3.22. So sánh các giá trị RMSE trong các phương pháp SR-PLS, BVE-PLS, GA-PLS và PLS cho các tập thuốc Natrixam (a, b); AgimDogyl (c,d) và Turbezid (e, f, g)	129
Hình 3.23. So sánh giá trị p giữa các phương pháp lựa chọn bước sóng với phương pháp PLS.....	131

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 [1]. 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% [2]. 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 [3].

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 lẻ, 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 mẽ 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 [4-6]. Trong bối cảnh đó, phương pháp UV-Vis có thể khắc phục được các nhược điểm của các phương pháp phân tích đó [7].

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 [6, 8-10].

Năm 1972, Swede Svante Wold 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 [11-13]. 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 [14]. 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 neuron nhân tạo (ANN), máy vectơ hỗ trợ (Support vector machine - SVM), thuật toán di truyền (GA) ... [12, 16].

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 lẻ. 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 [17].

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[18], khoa học y sinh (metabolomic) [19, 20], khoa học đo lường đặc điểm tâm lý con người (Psychometrics) [21, 22] 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 [23-26] để nói lên được tầm quan trọng của phương pháp này trong việc xử lý dữ liệu và tính toán kết quả [27]. 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 lẻ trong hỗn hợp mà không cần phải biết tất cả các thông tin của mẫu phân tích [26, 28-31].

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”**.

Mục tiêu luận án:

1. Xây dựng và tối ưu phương pháp bình phương tối thiểu riêng phần nhằm xác định đồng thời các hoạt chất có phổ hấp thụ xen phủ trong hỗn hợp dược phẩm.
2. Lựa chọn bước sóng tối ưu dựa trên thuật toán PLS nhằm rút gọn phổ, tăng độ tin cậy, chính xác cho mô hình PLS.

Những đóng góp mới của luận án:

Luận án đã xây dựng và tối ưu hóa mô hình bình phương tối thiểu riêng phần (PLS) để xác định đồng thời các dược chất có phổ hấp thụ UV-Vis xen phủ nhau, góp phần mở rộng ứng dụng của phương pháp PLS trong phân tích định lượng không phá hủy các chế phẩm dược.

Đề xuất quy trình thực nghiệm hoàn chỉnh bao gồm thiết kế tập hiệu chuẩn theo phương pháp CCD, ghi phổ UV-Vis, lựa chọn số biến tiềm ẩn bằng phân tích PRESS và đánh giá mô hình bằng các chỉ số thống kê (RMSECV, RMSEP, R^2 , kiểm định Shapiro–Wilk), bảo đảm độ chính xác và tính tin cậy của phương pháp.

Lần đầu tiên tại Việt Nam, mô hình PLS đã được ứng dụng thành công để định lượng đồng thời hai và ba hoạt chất trong các thuốc thương mại (AgimDogyl, Zoamco-A, Spirosemid-40, Natrixam, Turbezid), chứng minh tính khả thi và hiệu quả của phương pháp so với HPLC ($p > 0,05$).

Luận án đã đề xuất và áp dụng ba thuật toán lựa chọn bước sóng tối ưu gồm tỉ lệ chọn lọc (SR), lựa chọn biến ngược (BVE) và giải thuật di truyền (GA) kết hợp với PLS, giúp rút gọn phổ, loại nhiễu và nâng cao hiệu suất dự đoán. Các phương pháp lựa chọn bước sóng tối ưu góp phần đa dạng các phương pháp xử lý dữ liệu trước khi thực hiện mô hình PLS.

Các kết quả nghiên cứu đã được công bố trên các tạp chí khoa học quốc tế thuộc danh mục SCIE và các tạp chí chuyên ngành uy tín trong nước.

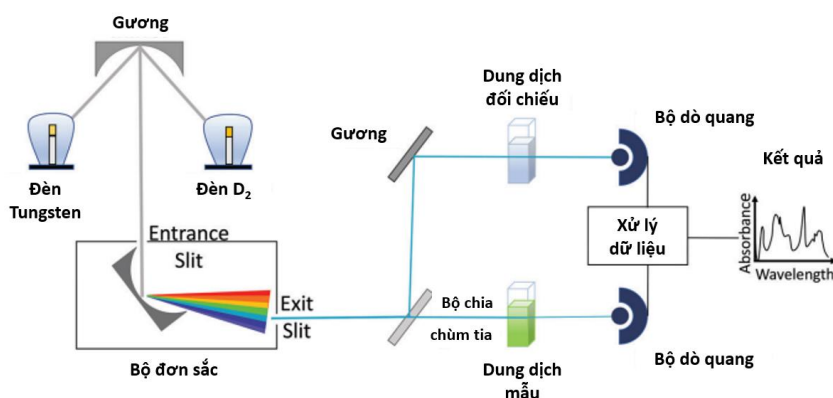
Cấu trúc của Luận án

- Mở đầu;
- Chương 1: Tổng quan tài liệu;
- Chương 2: Nội dung và phương pháp nghiên cứu;
- Chương 3: Kết quả và thảo luận;
- Kết luận;
- Danh mục các công trình công bố kết quả nghiên cứu của luận án;
- Tài liệu tham khảo;
- Phụ lục.

CHƯƠNG 1. TỔNG QUAN TÀI LIỆU

1.1. PHƯƠNG PHÁP QUANG PHỔ UV-VIS

Phương pháp quang phổ hấp thụ phân tử UV-Vis là một phần của nhóm phương pháp phân tích quang học, là một trong những phương pháp phân tích được sử dụng rộng rãi nhất. Nó được sử dụng phổ biến trong phân tích môi trường, phân tích nguồn nước, sử dụng trong nông nghiệp, phân tích các hoạt chất trong dược phẩm... [32-35] Thiết bị cũng dễ dàng tiếp cận, các máy quang phổ UV-Vis kích thước nhỏ tương đối rẻ và dễ vận hành (hình 1.1).



Hình 1.1. Sơ đồ hoạt động máy quang phổ UV-Vis hai chùm tia. Đèn vonfram phát ra ánh sáng khả kiến, trong khi đèn D2 tạo ra tia UV [36].

Nguyên tắc của phương pháp quang phổ UV-Vis là dựa vào sự hấp thụ một lượng bức xạ đơn sắc trong vùng tử ngoại khả kiến của các chất có trong mẫu phân tích được mô tả bởi định luật Lambert-Beer .

$$A = \varepsilon \cdot l \cdot C \quad (1)$$

Trong đó, A là độ hấp thụ; ε là hệ số hấp thụ mol ($M^{-1} \cdot cm^{-1}$), chỉ phụ thuộc vào bản chất của chất phân tích và nhiệt độ; l là độ dày của lớp dung dịch (cm); C là nồng độ mol/L (M). Biểu thức (1) là cốt lõi của phép đo quang phổ khi áp dụng vào hóa học phân tích. Đặt $K = \varepsilon \cdot l$, công thức (1) được viết lại :

$$A = KC \quad (2)$$

Trong phân tích quang phổ, trường hợp đơn giản nhất là mẫu chỉ có một thành phần và dung môi được đo phổ hấp thụ phân tử UV-Vis trong một khoảng bước sóng và chọn bước sóng có độ hấp thụ cực đại để xây dựng mô hình hiệu chuẩn, trong trường hợp này độ hấp thụ tuyến tính với nồng độ và gần như không chịu ảnh hưởng bởi các tác nhân xung quanh như các cấu tử khác, phản ứng hóa học, hiệu ứng cấu trúc...[37]. Đây là trường hợp của phân tích đơn biến.

Trong thực tế thì mẫu phân tích luôn có nhiều chất cùng tồn tại với nhau, việc tạo ra một mẫu phân tích chỉ có 1 thành phần thường rất phức tạp và khó khăn, đòi hỏi quá trình xử lý mẫu bằng các phương pháp tách, chiết, phân lập phức tạp. Một cách tiếp cận khác có thể giải quyết vấn đề, đó là lựa chọn phương pháp phân tích đa biến. Lợi thế của hiệu chuẩn đa biến là có xét đến sự tương tác lẫn nhau giữa các thành phần trong mẫu phân tích ở các nồng độ khác nhau trong quá trình hiệu chuẩn, đồng thời cũng đánh giá được sự ảnh hưởng của thiết bị, môi trường, thí nghiệm cho quá trình xây dựng mô hình hiệu chuẩn. Do đó, khi xây dựng các mô hình phân tích đa biến thì đã có sự góp phần của các yếu tố này vào trong đó, làm cho quá trình hiệu chuẩn chính xác hơn và mô hình thiết lập được gần với thực tế tương tác của các thành phần trong mẫu hơn [38].

Giả sử có p mẫu phân tích, mỗi mẫu có m thành phần khác nhau được đo phổ hấp thụ ở n bước sóng. Định luật Lambert-Beer biểu diễn mối quan hệ giữa độ hấp thụ ở n bước sóng của mẫu thứ 1 như sau:

$$\text{Tại } \lambda_1: \quad A_{11} = K_{11}C_{11} + K_{12}C_{21} + \dots + K_{1m}C_{m1} + E_{11} \quad (3)$$

$$\text{Tại } \lambda_2: \quad A_{21} = K_{21}C_{11} + K_{22}C_{21} + \dots + K_{2m}C_{m1} + E_{21} \quad (4)$$

.....

$$\text{Tại } \lambda_n: \quad A_{n1} = K_{n1}C_{11} + K_{n2}C_{21} + \dots + K_{nm}C_{m1} + E_{n1} \quad (5)$$

Trong đó, $A_{11}, A_{21} \dots A_{n1}$ là độ hấp thụ của mẫu số 1 ở n bước sóng; $K_{11}, \dots K_{nm}$ ($K = \varepsilon \cdot l$) là hệ số của thành phần thứ m ở bước sóng n ; $C_{11} \dots C_{m1}$ là nồng độ của thành phần thứ m trong mẫu số 1; $E_{11} \dots E_{n1}$ là độ hấp thụ của các thành phần khác có trong mẫu hoặc do nhiễu của thiết bị hoặc do ảnh hưởng của các hiện tượng quang học khác ... Biểu diễn dưới dạng ma trận đại số như sau:

$$\begin{bmatrix} A_{11} \\ A_{21} \\ \dots \\ A_{n1} \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1m} \\ K_{21} & K_{22} & \dots & K_{2m} \\ \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & \dots & K_{nm} \end{bmatrix} \times \begin{bmatrix} C_{11} \\ C_{21} \\ \dots \\ C_{m1} \end{bmatrix} + \begin{bmatrix} E_{11} \\ E_{21} \\ \dots \\ E_{n1} \end{bmatrix} \quad (6)$$

Tổng quát cho việc đo p mẫu phân tích, lưu ý rằng K là không đổi khi ta thay đổi nồng độ của chất phân tích khác nhau.

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{np} \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1m} \\ K_{21} & K_{22} & \dots & K_{2m} \\ \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & \dots & K_{nm} \end{bmatrix} \times \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1p} \\ C_{21} & C_{22} & \dots & C_{2p} \\ \dots & \dots & \dots & \dots \\ C_{m1} & C_{m2} & \dots & C_{mp} \end{bmatrix} + \begin{bmatrix} E_{11} & E_{12} & \dots & E_{1p} \\ E_{21} & E_{22} & \dots & E_{2p} \\ \dots & \dots & \dots & \dots \\ E_{n1} & E_{n2} & \dots & E_{np} \end{bmatrix} \quad (7)$$

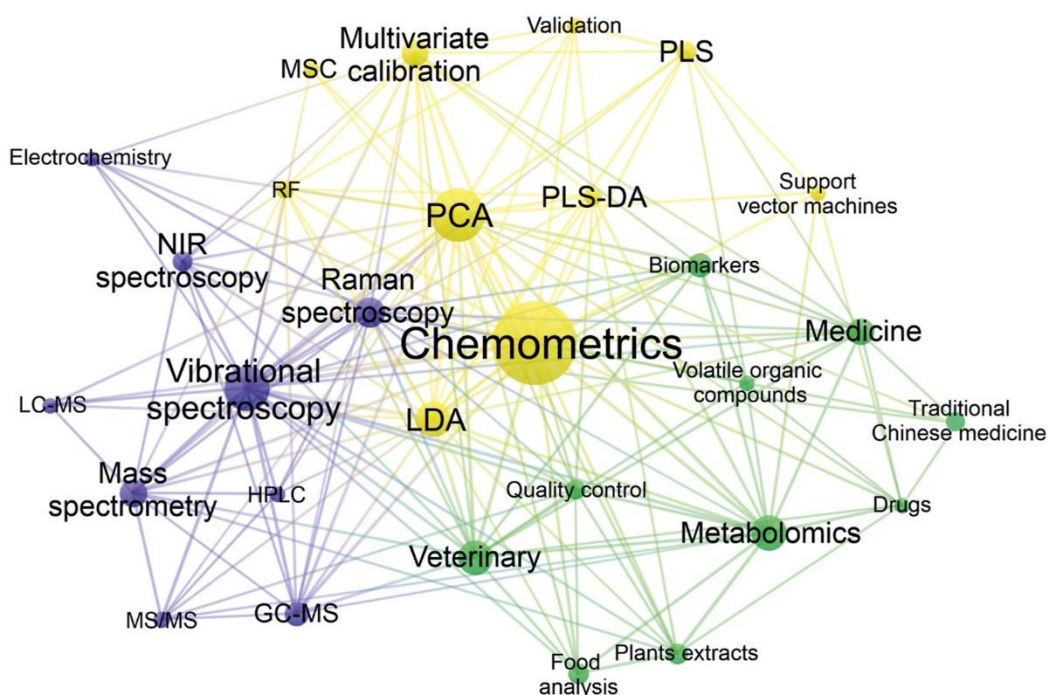
Để thống nhất dữ liệu cho các phương pháp tiếp theo, chúng tôi ký hiệu $\mathbf{A}$ là ma trận độ hấp thụ (n bước sóng, p mẫu); $\mathbf{C}$ là ma trận độ hấp thụ (m thành phần, p mẫu); $\mathbf{K}$ là ma trận hệ số; $\mathbf{E}$: là ma trận n hàng p cột. Phương trình (7) được viết lại như sau:

$$A = K.C + E \quad (8)$$

1.2. PHƯƠNG PHÁP CHEMOMETRICS

Sự phát triển của ngành chemometrics có liên quan chặt chẽ đến việc sử dụng máy tính trong hóa học. Một số nhóm phân tích trong những năm 1970 đã làm việc với các phương pháp thống kê và toán học mà ngày nay được coi là phương pháp chemometrics. Ký hiệu chemometrics được giới thiệu vào năm 1972 bởi Swede Svante Wold và Bruce R. Kowalski. Sự thành lập của Hiệp hội Chemometrics Quốc tế vào năm 1974 đã đưa đến mô tả đầu tiên về ngành học này. Như vậy, có thể định nghĩa về Chemometrics là: “ngành hóa học sử dụng các phương pháp toán học và thống kê, (a) để thiết kế hoặc lựa chọn các quy trình và thí nghiệm đo lường tối ưu, và (b) để cung cấp thông tin hóa học tối đa bằng cách phân tích dữ liệu hóa học” [39].

Chemometrics bắt nguồn từ hóa học, một trong những ứng dụng đầu tiên tập trung vào việc cải thiện hiệu suất định lượng của các thiết bị phân tích, như hiệu chuẩn NIR (hồng ngoại gần), cải thiện độ phân giải HPLC (sắc ký lỏng hiệu năng cao), và giải phân tích UV-VIS, phát triển các mối quan hệ hoạt động cấu trúc định lượng và đánh giá dữ liệu phân tích - hóa học [40]. Dữ liệu lớn được tạo ra bởi các thiết bị phân tích hiện đại là một lý do mà các nhà hóa học phân tích phát triển các ứng dụng của phương pháp chemometrics. Phương pháp chemometrics trong phân tích là ngành học sử dụng các phương pháp toán học và thống kê để thu được thông tin liên quan về hệ thống vật chất [39, 41-44].



Hình 1.2. Mối quan hệ giữa các công nghệ phân tích hiện đại, phương pháp phân tích đa biến và ứng dụng trong các ngành khoa học đời sống, công nghiệp và y tế [45]

Trong hóa học phân tích ngày nay, các phương pháp phân tích hiện đại việc có thể đo cùng một lúc nhiều mẫu và nhiều thông số đo khác nhau là một công việc dễ dàng. Việc này tạo ra một lượng lớn dữ liệu phân tích, vấn đề đặt ra là chúng ta có sử dụng hết nguồn thông tin tạo ra đó hay không, các thông tin đó liên quan đến các thành phần có mặt ở trong mẫu như thế nào? làm thế nào để sử dụng các thông tin đó và tạo ra kết quả phân tích một cách chính xác, tin cậy?. Chính vì vậy, các công cụ toán học và thống kê được sử dụng; các mô hình toán học như đơn biến, đa biến, các phép biến đổi về ma trận, về đại số tuyến tính được sử dụng một cách triệt để; bên cạnh đó, các mô hình thống kê cũng được sử dụng để đánh giá độ tin cậy của mô hình phân tích. Có thể nói rằng, chemometrics cho đến hiện nay có một vai trò rất quan trọng trong hóa học phân tích, và đang được các nhà hóa học phân tích tiếp tục khai thác và mở rộng để đưa ra các mô hình phân tích mới với độ đúng, độ tin cậy ngày càng cao.

Trong nghiên cứu của Peris-Diaz và cộng sự đánh giá tổng quan mối quan hệ giữa các công nghệ phân tích hiện đại, phương pháp phân tích đa biến và ứng dụng trong các ngành khoa học đời sống, công nghiệp và y tế (hình 1.2). Nghiên cứu đã cho thấy xu hướng tích hợp chemometrics để tối ưu hóa quy trình phân tích, nâng cao độ chính xác và hiệu quả trong nghiên cứu và sản xuất [45].

Trong một nghiên cứu của nhóm tác giả M.C Santos và cộng sự đã tổng quan về phương pháp chemometrics trong hóa học phân tích từ năm 2014 đến 2018 cho thấy có 832 nghiên cứu liên quan đến các vấn đề phân loại đa biến, hiệu chuẩn đa biến, hóa học phân tích, mạng neuron nhân tạo (ANN), thiết kế thí nghiệm và thiết kế nhân tố áp dụng trong nhiều lĩnh vực khác nhau: y học và dược phẩm, phân tích thực phẩm, nhiên liệu, sinh học và pháp y [46].

Áp dụng chemometrics trong hóa học xanh được sử dụng như là một công cụ nhằm thiết kế các thí nghiệm, tối ưu hóa quy trình đa mục tiêu, phân loại các dung môi hữu cơ khác nhau và dự đoán tính chất của chúng như độ nhớt, độ hòa tan của cacbon dioxit, độc tính, hệ số tích tụ sinh học... Các phương pháp được sử dụng như phân tích cụm, phân tích thành phần chính để xử lý kết quả nhằm xử lý dữ liệu thiếu, giúp dự đoán các thông số hóa học mà không cần thực hiện các phép đo phức tạp, tốn thời gian và đòi hỏi nhiều nguyên liệu [40, 47].

Năm 2022, Hai-Peng Wang và cộng sự đã đánh giá những tiến bộ bao gồm thuật toán, quy trình và những vấn đề liên quan của các phương pháp hiệu chuẩn chemometrics trong quang phổ hiện đại. Các kỹ thuật phân tích quang phổ hiện đại, như quang phổ tử ngoại-khả kiến (UV-vis), quang phổ hồng ngoại trung (Mid-infrared MIR), quang phổ hồng ngoại gần (Near-infrared NIR), quang phổ Raman, quang phổ terahertz (THz), quang phổ cộng hưởng từ hạt nhân (Nuclear magnetic resonance - NMR), quang phổ phân hủy cảm ứng laser (laser-induced breakdown spectroscopy - LIBS), v.v., đã phát triển nhanh chóng và được áp dụng rộng rãi trong nhiều lĩnh vực

như nông nghiệp, thực phẩm, dược phẩm, dầu khí, công nghiệp hóa chất, thuốc lá, bảo vệ môi trường và y học. Một đặc điểm nổi bật của tất cả các kỹ thuật này là khả năng trích xuất thông tin hóa học hữu ích từ dữ liệu quang phổ một cách chi tiết nhất có thể với sự hỗ trợ của các phương pháp chemometrics, nhằm cải thiện đáng kể cả tính bền vững và độ chính xác của kết quả phân tích. Trên nền tảng chung của sự phát triển trong trí tuệ nhân tạo, dữ liệu lớn, điện toán đám mây và các công nghệ khác, sự xuất hiện của các ý tưởng, phương pháp và chiến lược mới đã mang lại sức sống mới cho chemometrics. Các phương pháp chemometrics khác nhau được áp dụng trong phân tích quang phổ hiện đại, bao gồm tiền xử lý phổ hấp thụ, lựa chọn bước sóng, giảm chiều dữ liệu, hiệu chuẩn định lượng, nhận dạng mẫu, chuyển đổi hiệu chuẩn (Calibration transfer), duy trì hiệu chuẩn (calibration maintenance) và hợp nhất dữ liệu đa quang phổ [24].

Chemometrics đã đạt được những bước tiến đáng kể, đặc biệt trong lĩnh vực phân tích quang phổ hiện đại. Các kỹ thuật như phân tích đa biến và học sâu (ví dụ: mạng nơ-ron tích chập - CNN) đã mở rộng ứng dụng trong công nghiệp, từ phân tích nhanh tại chỗ đến giám sát thời gian thực. Sự tích hợp phương pháp không cần hiệu chuẩn, dựa trên so sánh trực tiếp dữ liệu phổ với thư viện chuẩn, cũng mang lại hiệu quả cho ngành dược và dầu khí. Bên cạnh đó, việc đưa hóa học tính toán vào giảng dạy đại học đang dần phổ cập hóa công nghệ này, tạo nền tảng cho đội ngũ chuyên gia tương lai. Triển vọng của lĩnh vực tập trung vào việc phát triển các thuật toán học sâu tối ưu, giải quyết vấn đề quá khớp (overfitting) và nâng cao khả năng giải thích mô hình. Hướng nghiên cứu mới bao gồm khai thác dữ liệu không nhãn (unlabeled samples) để xây dựng mô hình bán giám sát (semi-supervised), thậm chí không giám sát (unsupervised), kết hợp dữ liệu nhiều loại quang phổ (NIR, Raman, khối phổ) nhằm tăng độ tin cậy, và ứng dụng điện toán đám mây để quản lý dữ liệu quy mô lớn. Các nền tảng này không chỉ hỗ trợ phân tích sâu mà còn cung cấp thông tin thời gian thực cho sản xuất và kiểm soát chất lượng. Phát triển phần mềm thương mại thân thiện cũng là chìa khóa để triển khai rộng rãi các kỹ thuật phức tạp. Tuy nhiên, thách thức lớn vẫn tồn tại. Việc thiết lập và duy trì mô hình hiệu chuẩn đòi hỏi chuyên môn cao, gây hạn chế trong ứng dụng thực tế. Sự chậm trễ trong cập nhật phần mềm hỗ trợ thiết bị cũng cần được giải quyết thông qua hợp tác giữa nhà nghiên cứu và nhà sản xuất. Ngoài ra, cân bằng giữa độ phức tạp của thuật toán và tính khả thi trong công nghiệp đòi hỏi sự sáng tạo liên ngành. Dù vậy, với sự phát triển của AI, dữ liệu lớn và điện toán đám mây, chemometrics đang trở thành trụ cột của phân tích hiện đại. Nó không chỉ nâng cao độ chính xác mà còn thúc đẩy tự động hóa, mở ra kỷ nguyên mới cho nghiên cứu khoa học và tối ưu hóa sản xuất công nghiệp. Để tận dụng tiềm năng này, sự kết hợp giữa cải tiến công nghệ và đào tạo nguồn nhân lực chuyên sâu là yếu tố then chốt [24].

Như vậy, Chemometrics là một lĩnh vực liên ngành đầy tiềm năng, với ứng dụng rộng rãi không chỉ trong hóa học, môi trường, y sinh, kinh tế, mà còn trong thực phẩm,

nông nghiệp và công nghiệp dược phẩm. Sự kết hợp giữa khoa học dữ liệu và phân tích hóa học đã tạo nên những bước tiến vượt bậc, mang lại giá trị thực tiễn cao trong việc giải quyết các bài toán phức tạp.

Về mặt toán học, chemometrics không ngừng phát triển các mô hình tiên tiến nhằm tối ưu hóa khả năng dự đoán và phân tích dữ liệu. Từ phương pháp hồi quy truyền thống đến các thuật toán máy học phức tạp như SVM, Random Forest, hay mạng neural, chemometrics không ngừng cải tiến để nâng cao độ chính xác, giảm thiểu nhiễu, và xử lý bài toán đa biến trong điều kiện dữ liệu phức tạp. Sự phát triển không ngừng của các thuật toán này là động lực thúc đẩy sự đổi mới trong lĩnh vực.

Về phương tiện phân tích, chemometrics tận dụng các công nghệ tiên tiến như quang phổ UV-Vis, phổ huỳnh quang, hồng ngoại, phổ khối, điện hóa, sắc ký... Sự cải tiến của các thiết bị này cho phép thu thập dữ liệu chi tiết, đa chiều, tạo ra khối lượng dữ liệu lớn (big data) với độ chính xác cao. Điều này mở ra cơ hội khai thác thông tin sâu sắc từ mẫu phân tích, đồng thời đặt ra thách thức trong xử lý và phân tích dữ liệu.

Trong bối cảnh kỷ nguyên AI bùng nổ, chemometrics đang đứng trước cơ hội phát triển vượt bậc. Sự kết hợp giữa máy học (machine learning), học sâu (deep learning) và các thiết bị phân tích hiện đại hứa hẹn tạo ra các công cụ nhỏ gọn, nhanh chóng, thậm chí tích hợp trên smartphone hoặc thiết bị di động. Những công cụ này không chỉ tăng tốc độ phân tích mà còn mở rộng khả năng ứng dụng trong đời sống hàng ngày, từ kiểm tra chất lượng thực phẩm đến chẩn đoán y khoa. Ngoài ra, chemometrics còn có tiềm năng đóng góp vào các lĩnh vực như quản lý tài nguyên bền vững và giám sát môi trường thời gian thực, khẳng định vai trò không thể thiếu trong khoa học hiện đại.

1.3 MỘT SỐ PHƯƠNG PHÁP PHÂN TÍCH ĐA BIẾN

1.3.1 Hiệu chuẩn đơn biến và hiệu chuẩn đa biến

Hiệu chuẩn (calibration) là một quá trình điều chỉnh và xác định các thông số của một thiết bị hoặc mô hình sao cho nó phản ánh chính xác mối quan hệ giữa các biến số và kết quả mong muốn [48].

Hiệu chuẩn đơn biến là phương pháp phân tích của một thành phần đơn lẻ, vì nó dựa trên việc đo lường một giá trị hoặc dữ liệu duy nhất cho mỗi mẫu thử nghiệm [49]. Phương pháp này yêu cầu tín hiệu đo phải có tính chọn lọc đối với chất phân tích quan tâm, chính vì vậy đòi hỏi các thao tác phức tạp nhằm tách chất phân tích khỏi các tác nhân gây nhiễu có thể có mặt trong các mẫu thử [50].

Trong khi đó, hiệu chuẩn đa biến là thiết lập mối quan hệ giữa mẫu phân tích (có thể chứa nhiều thành phần khác nhau bao gồm các chất phân tích và các thành phần nhiễu) với tín hiệu đo được ở nhiều điều kiện khác nhau (số lượng có thể hàng nghìn

hoặc hàng triệu, ví dụ ở nhiều bước sóng khác nhau. Trong hiệu chuẩn đa biến, quá trình xác định nồng độ của một chất phân tích tương tự như trong hiệu chuẩn đơn biến nhưng việc xử lý dữ liệu bằng hồi quy tuyến hình bằng một mô hình đa biến phù hợp [49, 50]. Trong quá trình hiệu chuẩn đa biến, quá trình thiết lập mối quan hệ giữa các biến số và kết quả mong muốn đã xem xét đến các yếu tố ảnh hưởng của các thành phần khác, nền mẫu, thiết bị và môi trường chung quanh. Đây chính là những ưu thế nổi bật của hiệu chuẩn đa biến so với hiệu chuẩn đơn biến [38, 51].

1.3.2 Phương pháp bình phương tối thiểu cổ điển

Phương pháp bình phương tối thiểu cổ điển là phương pháp áp dụng hồi quy tuyến tính đa biến (MLR – Multiple Linear Regression) cho biểu thức của định luật Lambert-Beer. Trong phương trình (7), khi áp dụng vào trong mẫu thực tế thì luôn có sự sai khác giữa độ hấp thụ đo được với lý thuyết, do vậy phương trình (7) được viết lại:

$$\mathbf{A} = \mathbf{K} \cdot \mathbf{C} + \mathbf{E} \quad (9)$$

trong đó, $\mathbf{E}$ là ma trận sai số phổ hấp thụ. Phương pháp bình phương tối thiểu được sử dụng để xác định ma trận $\mathbf{K}$ sao cho $\mathbf{E}$ là nhỏ nhất. Giai đoạn này còn gọi là giai đoạn hiệu chuẩn. Công thức tính $\mathbf{K}$ như sau:

$$\mathbf{K} = \mathbf{A} \cdot \mathbf{C}^T \cdot (\mathbf{C} \mathbf{C}^T)^{-1} \quad (10)$$

Trong đó, $(\mathbf{C} \mathbf{C}^T)^{-1}$ là ma trận nghịch đảo của $(\mathbf{C} \mathbf{C}^T)$, $\mathbf{C}^T$ là ma trận chuyển vị của $\mathbf{C}$. Khi ma trận $\mathbf{K}$ đã tìm được, giai đoạn tiếp theo là xác định nồng độ của mẫu chưa biết được dự đoán từ phổ hấp thụ của nó:

$$\mathbf{C}_x = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \cdot \mathbf{A}_x \quad (11)$$

Trong phương pháp CLS số mẫu hiệu chuẩn để xây dựng mô hình là rất quan trọng. Vấn đề thứ 1, mẫu hiệu chuẩn phải đại diện và nồng độ phải bao trùm nồng độ của các chất có trong mẫu phân tích. Vấn đề thứ 2, cần phải biết chính xác tất cả các thành phần có trong mẫu phân tích để quá trình chuẩn bị mẫu hiệu chuẩn phải có các thành phần hoàn toàn giống như vậy. Đây là một công việc vô cùng khó khăn và không muốn nói là không thể. Chính điều này làm hạn chế phạm vi ứng dụng của phương pháp CLS. Một lĩnh vực có thể áp dụng được là phân tích các chất trong dược phẩm, ở đó mọi thành phần của các dạng bào chế có thể biết được nhiều nhất từ các hoạt chất đến các tá dược, tất cả các chất chuẩn đều có sẵn để chuẩn bị các hỗn hợp hiệu chuẩn. Vấn đề thứ 3 là trong biến đổi toán học, để ma trận $(\mathbf{C} \mathbf{C}^T)$ có nghịch đảo thì các hàng phải độc lập tuyến tính, có nghĩa là các nồng độ của các cấu tử trong các hỗn hợp hiệu chuẩn không tương quan nhau. Mặt khác, để nghịch đảo ma trận $(\mathbf{K}^T \mathbf{K})$ thì các cột của $\mathbf{A}$ cũng không tương quan với nhau, tức là nếu hai chất có phổ giống hệt nhau thì sẽ không xác định được.

Ưu điểm của phương pháp CLS là tính toán đơn giản, cho kết quả nhanh chóng, đáng tin cậy. Tuy nhiên, nếu các chất trong mẫu phân tích có phổ xen phủ nhau quá lớn và có những thành phần trong mẫu mà chưa biết thì sẽ gây ra sai số rất lớn cho việc áp dụng phương pháp CLS. Đây chính là nhược điểm lớn của phương pháp CLS, dẫn đến khó áp dụng cho các mẫu phức tạp như các mẫu có nguồn gốc tự nhiên, sinh học, công nghiệp hoặc môi trường [50-55].

1.3.2 Phương pháp bình phương tối thiểu nghịch đảo

Phương pháp bình phương tối thiểu nghịch đảo (ILS) dựa trên sự biến đổi từ định luật Lambert – Beer. Để đơn giản giải thích cho phương pháp bình phương tối thiểu nghịch đảo, chúng tôi lấy ví dụ thông qua 2 mẫu có 2 thành phần khác nhau: mẫu 1 gồm cả hai thành phần, mẫu 2 chứa một thành phần.

Theo định luật Lambert – Beer ta có:

$$A_1 = \varepsilon_{11}l.C_1 + \varepsilon_{12}l.C_2 \quad (12)$$

$$A_2 = \varepsilon_{22}l.C_2 \quad (13)$$

Rút C_2 thay vào phương trình (12):

$$A_1 = \varepsilon_{11}l.C_1 + \varepsilon_{12}l.\frac{A_2}{\varepsilon_{22}l} \quad (14)$$

Như vậy,

$$C_1 = \frac{A_1}{\varepsilon_{11}l} - \varepsilon_{12} \frac{A_2}{\varepsilon_{22}\varepsilon_{11}l} \quad (15)$$

Nếu ta đặt $P_{11} = \frac{1}{\varepsilon_{11}l}$ và $P_{12} = \varepsilon_{12} \frac{1}{\varepsilon_{22}\varepsilon_{11}l}$; nếu để ý rằng $\varepsilon_{11}, \varepsilon_{12}, \varepsilon_{22}$ là các hằng số chỉ phụ thuộc vào bản chất và bước sóng đo, như vậy các giá trị $P_{11}; P_{12}$ là các hằng số. Phương trình (15) được viết lại:

$$C_1 = P_{11}A_1 + P_{12}A_2 \quad (16)$$

Như vậy, phương pháp ILS thì các biến đóng vai trò ngược lại với phương pháp CLS với **A** là các biến độc lập và **C** là các biến phụ thuộc. Phương trình nghịch đảo của định luật Lambert-Beer viết dưới dạng ma trận được biểu diễn như sau:

$$\mathbf{C} = \mathbf{P} \mathbf{A} \quad (17)$$

Trong đó, **C** là ma trận nồng độ huấn luyện, **A** là ma trận độ hấp thụ, **P** là ma trận hệ số **P**. Trong giai đoạn hiệu chuẩn, **P** được xác định tương tự như phương pháp CLS:

$$\mathbf{P} = \mathbf{C} \mathbf{A}^T [\mathbf{A} \mathbf{A}^T]^{-1} \quad (18)$$

ở đây, $\mathbf{A}^T$ là ma trận chuyển vị của $\mathbf{A}$; $[\mathbf{A} \cdot \mathbf{A}^T]^{-1}$ là ma trận nghịch đảo của ma trận $\mathbf{A} \cdot \mathbf{A}^T$.

Sau khi xác định được P, mẫu phân tích được đo độ hấp thụ và nồng độ được xác định dựa vào phương trình (16):

$$\mathbf{C}_X = \mathbf{P} \mathbf{A}_X \quad (19)$$

Tương tự như phương pháp CLS, phương pháp ILS cũng có các phép biến đổi toán học tương đối đơn giản. Một số lưu ý từ phép biến đổi toán học cho thấy để tồn tại nghịch đảo $[\mathbf{A} \cdot \mathbf{A}^T]$ thì A phải có số hàng ít hơn hoặc bằng số cột tức là số mẫu đo phải nhiều hơn số bước sóng được chọn. Bên cạnh đó, nếu các bước sóng đã chọn mà phổ hấp thụ có sự tương quan hoặc xen phủ nhau thì ma trận $[\mathbf{A} \cdot \mathbf{A}^T]$ cũng không nghịch đảo được. Từ phương trình (18) cho thấy phương pháp bình phương tối thiểu nghịch đảo cho phép nghiên cứu các mẫu đa thành phần mà chỉ cần biết một hoặc một vài chất phân tích cần quan tâm trong khi nồng độ, phổ và đặc điểm hóa học của các thành phần còn lại không cần biết. Bằng cách này, phương pháp ILS đã khắc phục được nhược điểm chính của CLS là cần biết nồng độ của tất cả các thành phần mẫu trong giai đoạn hiệu chỉnh.

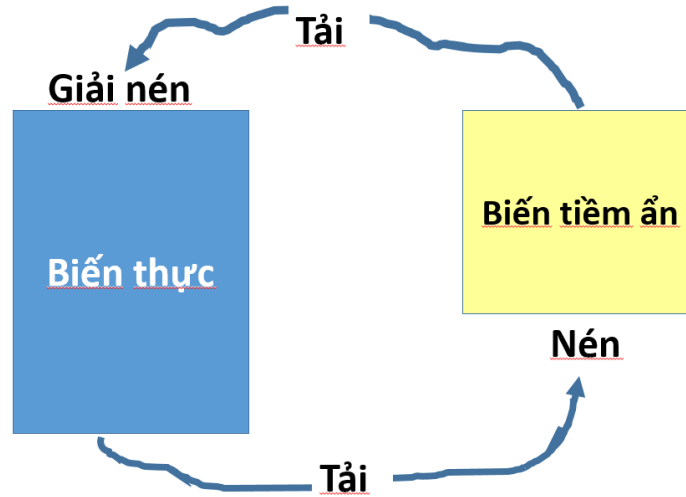
Ưu điểm chính của mô hình ILS là tính toán đơn giản, nhanh chóng, đặc biệt là ILS cho phép định lượng một chất phân tích trong mẫu mà có mặt các chất gây nhiễu khác, miễn là thành phần nhiễu này có đưa vào trong quá trình hiệu chuẩn nhưng không nhất thiết phải biết được nồng độ hoặc định danh hóa học của nó. Đây là ưu điểm mà nó khắc phục được cho mô hình CLS. Tuy nhiên, nhược điểm của phương pháp ILS là khó khăn trong việc lựa chọn bước sóng phù hợp để đo độ hấp thụ, công việc này có thể dẫn đến mất mát thông tin và độ nhạy tổng thể, và yêu cầu lựa chọn những bước sóng sao cho ít tương quan với nhau cũng không phải là một nhiệm vụ dễ dàng [51, 53-56].

Như vậy, phương pháp ILS có lợi thế hơn phương pháp CLS nếu trong mẫu có mặt các thành phần nhiễu nhưng không xác định được mà vẫn có mặt trong mô hình hiệu chuẩn thì vẫn có thể xác định được chính xác nồng độ của thành phần cần quan tâm. Khó khăn của ILS là phải lựa chọn được những bước sóng tốt nhất với số lượng bước sóng ít hơn số lượng mẫu thí nghiệm. Để giải quyết vấn đề này, các phương pháp giảm chiều dữ liệu được áp dụng như hồi quy thành phần chính, phương pháp bình phương tối thiểu riêng phần.

1.3.4 Phương pháp hồi quy thành phần chính

Hồi quy thành phần chính (PCR) là một phương pháp biến đổi giảm chiều dữ liệu, là sự kết hợp giữa phân tích thành phần chính (PCA) và hồi quy bình phương tối thiểu [50, 57-59]. Phương pháp giúp giảm số lượng biến đầu vào để xây dựng mô hình hiệu chuẩn. Để xây dựng mô hình PCR có thể phân chia thành hai giai đoạn: (1) Thực hiện

PCA để biến đổi tập dữ liệu phổ gốc A thành các thành phần chính (PCs) mang phần lớn phương sai của dữ liệu; (2) Thực hiện hồi quy tuyến tính giữa các thành phần chính và nồng độ mẫu C . Liên kết giữa các biến thực và PCs được thực hiện thông qua một tập hợp các vector gọi là vector tải (loading) được minh họa trong hình hình 1.3.



Hình 1.3. Sự biến đổi giữa biến thực (real variables) và biến tiềm ẩn (latent variable) bằng cách nén (compression) hoặc giải nén (decompression) thông qua các vector tải (loading) [50].

Sự biến đổi toán học trong phương pháp PCA cho đến hiện nay có một số cách để thực hiện như phương pháp phân rã giá trị suy biến (singula value decomposition – SVD), bình phương tối thiểu riêng phần lặp phi tuyến (non-linear iterative partial least-squares – NIPLS) [27, 60, 61]. Các cách thực hiện này có thể dựa trên biến đổi toán học khác nhau nhưng kết quả thu được là giống nhau, vì vậy có thể chọn bất kỳ phương pháp nào để thực hiện [50].

Thuật toán hồi qui thành phần chính cho dữ liệu phổ hấp thụ A (kích thước $p \times n$, với p mẫu và n bước sóng) và nồng độ C (kích thước $p \times m$, với p mẫu và m thành phần) theo phương pháp SVD được thực hiện với các bước sau:

Bước 1: chuẩn hóa dữ liệu bao gồm trung tâm hóa (mean-centering): trừ giá trị trung bình của mỗi cột trong A và C để đảm bảo dữ liệu có trung bình bằng 0. Chuẩn hóa: chia mỗi cột cho độ lệch chuẩn để các biến có cùng thang đo, điều này rất quan trọng khi các biến có đơn vị hoặc độ lớn khác nhau.

Bước 2: Phép biến đổi PCA được thực hiện bằng phân rã giá trị suy biến (SVD):

$$A = U\Sigma V^T \quad (20)$$

Trong đó: U (kích thước $p \times p$): ma trận các vector riêng trái; Σ (kích thước $p \times n$): ma trận đường chéo chứa các giá trị suy biến, sắp xếp theo thứ tự giảm dần; V (kích

thước $n \times n$): ma trận các vector riêng phải. Các giá trị suy biến trong Σ (kí hiệu $\sigma_1, \sigma_2, \dots$) biểu thị mức độ quan trọng của các hướng dữ liệu. Các cột của V là các vector tải (loading vector) tương ứng với các hướng của thành phần chính.

Bước 3: Trích xuất thành phần chính:

- Tính ma trận điểm (score):

$$T = XV = U\Sigma V^T V = U\Sigma \quad (21)$$

- Ma trận tải (loading): ma trận V chứa các vector tải, biểu thị cách các điểm gốc trong X đóng góp vào các thành phần chính.

Bước 4: Lựa chọn số thành phần chính:

Chọn số thành phần chính N dựa trên phần trăm phương sai được giải thích được tính bởi công thức:

$$\% \text{ phương sai giải thích} = \frac{\sigma_i^2}{\sum_i \sigma_i^2} \times 100\% \quad (22)$$

Hoặc theo phương pháp kiểm tra chéo (cross-validation) để đảm bảo hiệu suất hồi quy tối ưu.

Lấy N cột đầu tiên của T (kí hiệu T_N , kích thước $p \times N$) và N cột đầu tiên của V (kí hiệu V_N).

Bước 5: Hồi quy tuyến tính trên các thành phần chính: sử dụng ma trận điểm T_N làm biến độc lập mới để xây dựng hồi quy tuyến tính với C :

$$C = T_N \beta + E \quad (23)$$

Với E là ma trận sai số; β là ma trận hệ số hồi quy (kích thước $N \times m$) được xác định bằng phương pháp bình phương tối thiểu:

$$\beta = (T_N^T T_N)^{-1} T_N^T C \quad (24)$$

Bước 6: Dự đoán cho tập dữ liệu phổ mới $A_{\text{mới}}$:

- Chuẩn hóa dữ liệu mới

- Tính ma trận điểm mới:

$$T_{\text{mới}} = A_{\text{mới}} V_N \quad (25)$$

- Tính nồng độ trong mẫu mới:

$$C_{mới} = T_{mới} \beta \quad (26)$$

Như vậy, phương pháp PCA đã thu gọn dữ liệu hiệu chuẩn ban đầu thành một ma trận T mới nhưng không làm mất mát quá nhiều thông tin của dữ liệu phổ gốc, bởi các dữ liệu phổ gốc được chiếu vào không gian mới bởi một ma trận trọng số V . Đặc điểm này cho thấy, thông tin phổ gốc ban đầu được giữ lại đầy đủ đồng thời số chiều của dữ liệu được thu gọn lại, hai đặc tính này đều thỏa mãn cho hai mô hình hồi quy CLS và ILS. Chính vì vậy, phương pháp PCR tích hợp các ưu điểm của cả hai phương pháp CLS và ILS [53].

Tóm lại, có thể tổng quát một số ưu điểm chính của phương pháp PCR: (1) Không cần chọn bước sóng thủ công – có thể sử dụng toàn phổ. (2) Hiệu ứng trung bình hóa giúp mô hình ít bị nhiễu. (3) Cho phép hồi quy ngược, chỉ cần hiệu chuẩn trên thành phần quan tâm. (4) Áp dụng tốt cho hỗn hợp phức tạp. (5) Có thể nhận diện được mẫu chứa tạp không có trong tập hiệu chuẩn.

Nhược điểm của phương pháp PCR là việc lựa chọn bao nhiêu thành phần chính được giữ lại, số thành phần chính được giữ lại sẽ quyết định độ chính xác của mô hình. Nếu số thành phần chính giữ lại nhiều quá thì sẽ gây ra hiện tượng quá khớp (overfitting) nhưng nếu số thành phần chính giữ lại không đủ làm mất mát thông tin ban đầu. Tính toán chậm hơn so với hầu hết các phương pháp cổ điển; cần phải có khả năng phân tích sâu để có thể diễn giải các mối quan hệ giữa các thành phần chính và các biến đầu vào, cũng như hiểu được cách thức mà chúng ảnh hưởng đến kết quả của mô hình; Các thành phần chính chỉ giải thích phương sai tối đa với dữ liệu quang phổ, không liên quan đến thông tin nồng độ của các thành phần; Thông thường, cần một số lượng mẫu lớn để hiệu chuẩn chính xác; Việc thu thập các mẫu hiệu chuẩn có thể khó khăn; cần tránh sự tuyến tính nồng độ của các thành phần [62].

1.3.5 Phương pháp bình phương tối thiểu riêng phần

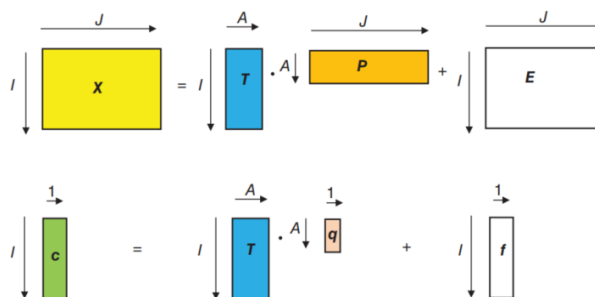
Mô hình PLS được xem là một kỹ thuật hồi quy chủ đạo trong phân tích dữ liệu đa biến, mặc dù việc sử dụng nó không phải lúc nào cũng hoàn toàn hợp lý và chính những người phát triển phương pháp này cũng ý thức rõ điều đó. Dù vậy, PLS đã đạt được nhiều thành công vượt bậc, đặc biệt trong các lĩnh vực như nghiên cứu mối liên hệ giữa cấu trúc và hoạt tính sinh học (QSAR), khoa học y sinh (metabolomics) hay khoa học đo lường đặc điểm tâm lý (psychometrics), nơi mà các yếu tố nền tảng thường không có ý nghĩa vật lý rõ ràng và mô hình tuyến tính khó được kỳ vọng. Trong các lĩnh vực như quang phổ hoặc sắc ký, nơi thường giả định tính cộng tính của tín hiệu, các phương pháp đơn giản hơn như hồi quy tuyến tính bội có thể đáp ứng tốt. Tuy nhiên, PLS vẫn giữ vai trò quan trọng khi các yếu tố cơ bản không thể diễn giải rõ ràng. Ví dụ, trong phép đo hàm lượng protein trong lúa mì bằng quang phổ hồng ngoại gần (NIR), PLS có thể xây dựng mô hình dự đoán từ một tập mẫu huấn luyện gồm các mẫu lúa mì có hàm

lượng protein đã biết. Nếu bộ dữ liệu huấn luyện được thiết kế tốt, mô hình PLS có thể hoạt động mạnh mẽ với các mẫu có đặc tính tương tự trong tương lai. Ngược lại, khi dữ liệu mới khác biệt đáng kể — chẳng hạn mô hình xác định vitamin C trong nước cam được xây dựng từ mẫu nước cam Tây Ban Nha nhưng áp dụng cho nước cam Brazil — kết quả dự đoán có thể không còn chính xác do sự khác biệt trong đặc tính quang phổ. Vì vậy, việc nhận thức rõ giới hạn của PLS là rất cần thiết, bởi phương pháp này không thể khắc phục những thiếu sót trong thiết kế thí nghiệm hoặc sai sót trong quá trình thực nghiệm. Những trường hợp ngoại lệ có cấu trúc dữ liệu khác biệt so với tập huấn luyện thuộc về một lớp phân tích khác mà đôi khi bị nhiều người bỏ qua [27].

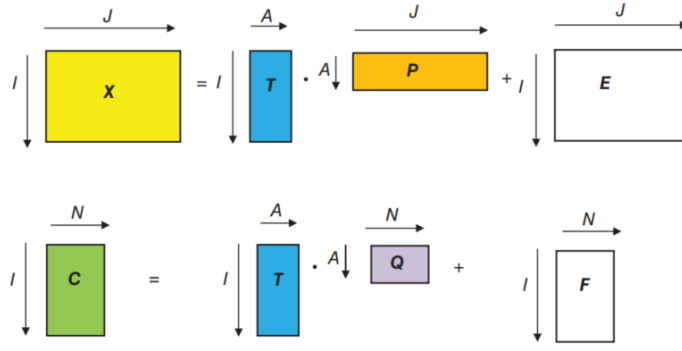
Một đặc điểm quan trọng của PLS là nó xem xét sai số trong cả dữ liệu phổ và ước tính nồng độ, trong khi các phương pháp như PCR giả định rằng nồng độ được xác định chính xác và mọi sai số chỉ tồn tại trong biến quang phổ. Trong thực tế, đặc biệt trong hóa học, các sai số có thể phát sinh ngay từ khâu chuẩn bị mẫu như cân, pha loãng hoặc đo lường, khiến cả biến độc lập cũng chứa sai số. Trái ngược với các phương pháp hồi quy cổ điển vốn yêu cầu xác định biến nào chịu sai số, PLS giả định rằng sai số được phân bố đồng đều trong cả hai khối dữ liệu phụ thuộc và độc lập, tạo nên điểm khác biệt quan trọng của phương pháp này so với các mô hình thống kê truyền thống [27].

Phương pháp PLS được phát triển và phổ biến bởi Word và được Mertens và Naes mở rộng [13, 63, 64]. PLS là một phương pháp giảm chiều, thu gọn dữ liệu bằng cách chiếu dữ liệu gốc lên không gian mới và tìm các biến tiềm ẩn phù hợp. PLS đã được phát triển để giải quyết các vấn đề hồi quy đa biến khi có sự tương quan và nhiễu ảnh hưởng đến dữ liệu [65, 66]. Cho đến nay, phương pháp PLS được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau, từ hóa học, kinh tế lượng cho đến nghiên cứu y sinh.

Có rất nhiều thuật toán PLS khác nhau, tuy nhiên kết quả các thuật toán đó mặc dù có cách biến đổi khác nhau nhưng kết quả thu được đều như nhau. Có hai phiên bản PLS gồm PLS1 và PLS2 [67-69]. PLS1 được giới thiệu bởi Word và Martens, PLS2 là một phần mở rộng của PLS1 được đề xuất giữa những năm 1980. Thực tế thì ít có sự khác biệt giữa ý tưởng thực hiện PLS1 và PLS2, thể hiện trong hình 1.4 và hình 1.5, ngoại trừ việc PLS2 sử dụng một ma trận nồng độ C thay vì các vector nồng độ cho từng hợp chất trong mỗi hỗn hợp.



Hình 1.4. Minh họa thuật toán của PLS1 [27]



Hình 1.5. Minh họa thuật toán của PLS2 [27]

Trong phần này, chúng tôi giới thiệu thuật toán PLS1 (và sau này gọi là PLS) được sử dụng phổ biến nhất, thuật toán NIPLS được giới thiệu bởi Wold. Giả sử có tập dữ liệu độ hấp thụ $\mathbf{A}_{(p,n)}$ ($\mathbf{A}$ là ma trận có kích thước p mẫu $\times$ n bước sóng) có mối quan hệ với tập giá trị nồng độ $\mathbf{c}_{(p,1)}$ ($\mathbf{c}$ là vector nồng độ của một chất có kích thước p mẫu $\times$ 1) thông qua mối quan hệ tuyến tính $\mathbf{c} = \boldsymbol{\alpha} + \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ với các tham số hồi quy chưa biết $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ và sai số $\boldsymbol{\varepsilon}$. Để thực hiện thuật toán PLS, đầu tiên các biến được căn giữa (center) dữ liệu: $\mathbf{A}_0 = \mathbf{A} - \mathbf{1}\bar{\mathbf{a}}^T$ và $\mathbf{c}_0 = \mathbf{c} - \mathbf{1}\bar{c}$. Đặt N là số thành phần biến tiềm ẩn được trích xuất. Khi $n = 1, 2, \dots, N$, kí hiệu T ở trên mũ trong các phương trình là biểu diễn phép toán chuyển vị, -1 là phép toán nghịch đảo của các ma trận, thuật toán được thực hiện như sau [26, 54, 70, 71]:

$$(1) \text{ Tính tải trọng (weight loading, } w\text{): mô hình: } \mathbf{A} = \mathbf{c}w_n^T + \mathbf{E}_A \quad (27)$$

$$w \text{ được tính: } \mathbf{w}_n = \mathbf{A}_{n-1}^T \mathbf{c}_{n-1} \quad (28)$$

Trọng số này xác định hướng trong không gian mở rộng bởi tối đa hiệp phương sai của $\mathbf{A}_{n-1}$ với $\mathbf{c}_{n-1}$. Chuẩn hóa LWs để có độ dài bằng 1: $\mathbf{w}_n \leftarrow \mathbf{w}_n / \|\mathbf{w}_n\|$

$$(2) \text{ Tính vector điểm } t_n: \text{ mô hình: } \mathbf{A} = \mathbf{t}_n \mathbf{w}_n^T + \mathbf{E}_A \quad (29)$$

$$t_n \text{ được tính: } \mathbf{t}_n = \mathbf{A}_{n-1} \mathbf{w}_n \quad (30)$$

Vector t_n , còn gọi là điểm số (score), đại diện cho biến tiềm ẩn phản ánh các đặc tính chính của dữ liệu từ ma trận $\mathbf{A}_{n-1}$.

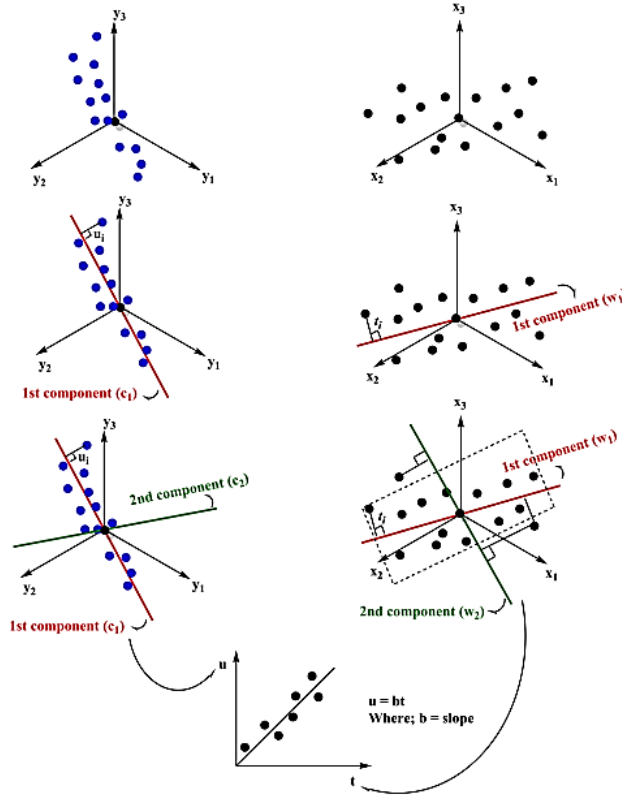
(3) Tìm hệ số tương quan (tải – loading) giữa vector điểm với dữ liệu ban đầu:

$$\text{mô hình: } \mathbf{A}_{n-1} = \mathbf{t}_n \mathbf{p}_n^T + \mathbf{E}_A \quad (31)$$

$$p_n \text{ được tính: } \mathbf{p}_n = \mathbf{A}_{n-1}^T \mathbf{t}_n (\mathbf{t}_n^T \mathbf{t}_n)^{-1} \quad (32)$$

Tính tương tự với hệ số tương quan của $\mathbf{c}$ với vector điểm $\mathbf{t}$ (q_n):

$$\mathbf{q}_n = \mathbf{c}_{n-1}^T \mathbf{t}_n (\mathbf{t}_n^T \mathbf{t}_n)^{-1} \quad (33)$$



Hình 1.6. Phân tích hồi quy bình phương tối thiểu riêng phần [72]

(4) Các giá trị $\mathbf{A}_n$ và $\mathbf{c}_n$ cho vòng lặp tiếp theo được tính bằng cách trừ đi đóng góp của t_n :

$$\mathbf{A}_n = \mathbf{A}_{n-1} - t_n \mathbf{p}_n^T \quad (34)$$

$$\mathbf{c}_n = \mathbf{c}_{n-1} - t_n \mathbf{q}_n \quad (35)$$

Việc trừ đi các phần đóng góp này giúp loại bỏ thành phần đã được giải thích trong dữ liệu, giảm bớt phần nhiễu và cải thiện độ chính xác của mô hình.

(5) Nếu $n < N$ quay lại thực hiện bước 1. Điều kiện dừng khi $n = N$ hoặc $c_n^T c_n < 10^{-6}$

Trong mô hình hồi quy này, các trọng số tải w xác định hướng của từng thành phần trong không gian của các biến độc lập sao cho tối đa hóa mối liên hệ với biến phụ thuộc. Điểm t là biến tiềm ẩn đại diện cho các yếu tố chính trong $\mathbf{A}$ và được tính dựa trên các trọng số w . Trọng số hồi quy p và q lần lượt mô tả cách mà từng thành phần t phản ánh các biến trong $\mathbf{A}$ và $\mathbf{c}$, giúp xác định mối quan hệ giữa các thành phần chính và biến phụ thuộc. Các bước thực hiện PLS được mô tả trực quan qua hình 1.6.

Sau khi huấn luyện được mô hình PLS, để xác định được nồng độ của cách chất có trong mẫu phân tích, thuật toán được thực hiện theo một trong hai phương pháp.

Phương pháp 1:

Bước 1: Căn giữa dữ liệu a bằng cách sử dụng dữ liệu hiệu chuẩn. Chuẩn hóa a nếu phổ trong dữ liệu hiệu chuẩn đã được chuẩn hóa. Đặt chỉ số $n = 1$.

Bước 2: Tính vector điểm: $t_n = w_n^T a$ (36)

trong đó w_n là vector trọng số của biến tiềm ẩn n .

Bước 3: Cập nhật vector phần dư cho nồng độ và dữ liệu:

$$c_n = c_{n-1} + q_n t_n \quad (37)$$

$$e_n = e_{n-1} + p_n t_n \quad (38)$$

trong đó q_n và p_n là các hệ số trọng số ước tính.

Bước 4: Tăng n lên 1 và thay e_n cho a , sau đó quay lại bước 2 và lặp lại quá trình cho đến khi $n = r$ (với r là số thành phần tối ưu).

Ghi chú: w_n , q_n và p_n là các vector trọng số được lấy từ quá trình hiệu chuẩn PLS1. c_0 là nồng độ trung bình của chất phân tích trong các mẫu hiệu chuẩn và $e_0 = a$ là dữ liệu ban đầu.

Phương pháp 2: Các vector tải trọng (w), điểm (t) và trọng số (p , q) được tính ở mỗi lần lặp lại cho đến khi thỏa mãn điều kiện dừng (hoặc chọn đến số biến tiềm ẩn tối ưu $n < N$) thu được ma trận W , T , P và vector q . Hệ số hồi quy trong mô hình tuyến tính được xác định bằng công thức

$$\beta = W(P^T W)^{-1} q \quad (39)$$

Trong đó W và P là các ma trận với r hàng chứa các vector w_n và p_n tương ứng cho từng thành phần, và q là vector của các trọng số ước tính q_n cho các thành phần trong mô hình PLS1 tối ưu (khi $n = r$).

Căn giữa a bằng cách sử dụng dữ liệu hiệu chuẩn (chuẩn hóa a nếu phổ trong dữ liệu hiệu chuẩn đã được chuẩn hóa).

Xây dựng giá trị dự đoán $\hat{c}$ từ dữ liệu a với công thức:

$$\hat{c} = a\beta + c_0 \quad (40)$$

trong đó c_0 là nồng độ trung bình của chất phân tích trong các mẫu hiệu chuẩn như trong phương pháp 1.

Hai phương pháp này cung cấp quy trình dự đoán nồng độ của chất phân tích dựa trên dữ liệu quang phổ và mô hình PLS đã hiệu chuẩn. Trong phương pháp 1, thuật toán đi qua từng thành phần chính một cách tuần tự, tính toán vector điểm và trọng số hồi quy, sau đó loại bỏ đóng góp của mỗi thành phần khỏi dữ liệu. Điều này tiếp tục cho đến khi đạt đến số thành phần tối ưu r . Phương pháp 2, sau khi hoàn thành hiệu chuẩn PLS, sẽ tổng hợp các hệ số hồi quy cuối cùng từ tất cả các thành phần đã chọn và sử dụng chúng trong một phương trình hồi quy đơn giản để dự đoán nồng độ dựa trên dữ liệu mới.

Phương pháp bình phương tối thiểu riêng phần (PLS) là công cụ mạnh mẽ trong phân tích quang phổ, nổi bật ở khả năng xử lý dữ liệu đa chiều, đặc biệt khi số biến lớn hơn số mẫu. PLS giúp khắc phục đa cộng tuyến, giảm số lượng biến đầu vào mà vẫn duy trì độ chính xác cao trong dự đoán, đồng thời có thể kết hợp linh hoạt với các phương pháp khác để tăng hiệu quả phân tích. PLS có nhiều ưu điểm như dễ tính toán và triển khai, thể hiện rõ mối quan hệ giữa các biến trong mô hình; chiếu dữ liệu gốc lên không gian nhỏ gọn hơn thông qua các biến tiềm ẩn nhưng vẫn giữ lại thông tin quan trọng; xác định và loại bỏ các biến không cần thiết giúp tối ưu thí nghiệm và giảm chi phí; trực quan hóa dữ liệu để phát hiện điểm ngoại lệ và mối quan hệ giữa các biến; hoạt động ổn định ngay cả khi dữ liệu bị nhiễu hoặc thiếu, và không yêu cầu dữ liệu tuân theo phân phối chuẩn [63, 67].

Tuy nhiên, phương pháp này cũng có những nhược điểm, như khó khăn trong việc giải thích ý nghĩa của các thành phần mới, sự phức tạp trong việc lựa chọn số lượng thành phần tối ưu, không thể kiểm tra ý nghĩa thống kê trừ khi chạy bootstrap, không rõ các đặc tính phân phối của các ước lượng [63]. Sự hiểu biết về những ưu và nhược điểm này là cần thiết để các nhà phân tích quang phổ lựa chọn phương pháp phù hợp cho từng tình huống cụ thể.

1.4 TỔNG QUAN VỀ ÁP DỤNG PHƯƠNG PHÁP PLS KẾT HỢP CÁC PHƯƠNG PHÁP PHỔ TRONG PHÂN TÍCH CÁC CHẤT

Phương pháp PLS là một kỹ thuật mạnh mẽ trong phân tích hóa học, được ứng dụng rộng rãi để xử lý các dữ liệu phổ phức tạp trong nhiều lĩnh vực, đặc biệt là trong phân tích dược phẩm và chất lượng thực phẩm. Kết hợp với các kỹ thuật phổ như Raman, UV-Vis, và NIR, PLS không chỉ giúp xác định nồng độ của các chất mà còn cải thiện độ chính xác và hiệu quả trong các nghiên cứu khoa học. Các nghiên cứu gần đây từ năm 2023 và 2024 đã chứng minh khả năng vượt trội của phương pháp này trong việc phân tích các chất phức tạp, đặc biệt là khi được kết hợp với các kỹ thuật phổ.

Một trong những nghiên cứu đáng chú ý trong năm 2024 của Satsuki Sato và Yasushi Numata [73] đã áp dụng phương pháp Raman kết hợp với PLS để phân tích các phân tử tương tự cấu trúc như rutin và quercetin trong các dung dịch ethanol. Các mẫu

Tartary buckwheat flour được phân tích để xác định nồng độ rutin và quercetin. Kết quả cho thấy phương pháp này đạt độ chính xác cao trong việc định lượng các chất này, với kết quả gần giống như phương pháp sắc ký lỏng hiệu suất cao (HPLC). Việc sử dụng PLS giúp phát triển một mô hình dự đoán hiệu quả, chứng minh khả năng ứng dụng mạnh mẽ của phương pháp trong phân tích các thành phần hóa học trong thực phẩm và dược phẩm.

Một nghiên cứu khác của Zain Ali và cộng sự [74] đã sử dụng phổ Raman và PLS để định lượng florfenicol, một loại kháng sinh phổ rộng, trong các mẫu dược phẩm. Các đỉnh phổ Raman quan trọng đã được xác định để liên kết với nồng độ API (active pharmaceutical ingredient) và các tá dược. Phương pháp này đã chứng minh được khả năng xác định chính xác các nồng độ khác nhau của florfenicol trong các mẫu thuốc, và việc áp dụng PLS giúp cải thiện độ chính xác của phương pháp phân tích này, tạo ra một công cụ đáng tin cậy cho việc kiểm soát chất lượng thuốc tại quy mô công nghiệp.

Một nghiên cứu nổi bật trong năm 2024 của Keita Yaginuma và cộng sự [75] đã giới thiệu phương pháp T-PLSR (T-shaped Partial Least Squares Regression) kết hợp với học chuyển giao (transfer learning - TL) để phát triển quy trình sản xuất thuốc mới. Phương pháp này được áp dụng trong việc mở rộng quy mô quá trình nén thuốc từ một mô phỏng ép viên có kinh nghiệm sang máy ép viên quay ít kinh nghiệm hơn. Kết quả nghiên cứu chỉ ra rằng T-PLSR+TL có hiệu suất dự đoán vượt trội hơn so với các mô hình T-PLSR truyền thống, giúp tối ưu hóa quy trình sản xuất thuốc và giảm thiểu sự thiếu hụt dữ liệu quy mô lớn trong phát triển sản phẩm mới.

Phương pháp PLS cũng đã được áp dụng để xác định đồng thời nhiều dược chất trong các hỗn hợp nhân tạo và các chế phẩm dược phẩm. Trong một nghiên cứu gần đây của Mahmood và cộng sự [76], PLS kết hợp với dữ liệu phổ UV-Vis đã được sử dụng để xác định đồng thời các thuốc như piroxicam, naproxen, diclofenac sodium và mefenamic acid trong các hỗn hợp và chế phẩm dược phẩm. Kết quả nghiên cứu cho thấy phương pháp này có thể xác định chính xác các dược chất mà không bị ảnh hưởng bởi các tá dược, làm nổi bật tính hiệu quả và đơn giản của phương pháp PLS trong kiểm tra chất lượng thuốc.

Trong nghiên cứu của Khanda F.M. Amin và cộng sự [77], phương pháp PLS đã được kết hợp với kỹ thuật phổ NIR để xác định các thành phần hóa học trong các mẫu thuốc khác nhau, bao gồm rabeprazole, lansoprazole, levofloxacin, amoxicillin, và paracetamol. Các mô hình PLS đã được phát triển để phân tích các thành phần này trong các mẫu hỗn hợp và chế phẩm dược phẩm, với kết quả cho thấy PLS là một phương pháp đáng tin cậy trong việc xác định các chất có trong các chế phẩm phức tạp.

Một nghiên cứu khác của Yong-Qi Zhong và cộng sự [78] đã sử dụng phương pháp PLS kết hợp với phổ NIR để định lượng năm thành phần hóa học chính trong

Arnebiae Radix (AR), một loại thảo dược có giá trị kinh tế và dược lý cao. Phương pháp này đã chứng minh khả năng nhanh chóng và chính xác trong việc xác định các thành phần trong AR, với kết quả cho thấy các mô hình PLS có độ chính xác cao và có thể được áp dụng trong việc đánh giá chất lượng AR trong các nghiên cứu kiểm tra chất lượng thực phẩm.

Trong nghiên cứu của Etil Guzelmeric và cộng sự [79], PLS đã được áp dụng để phân tích các thành phần đặc trưng trong sáp ong, một sản phẩm từ ong, sử dụng các hình ảnh HPTLC (High-Performance Thin-Layer Chromatography). Phương pháp này đã chứng minh khả năng sử dụng PLS kết hợp với HPTLC để xác định tỷ lệ các thành phần trong sáp ong, mang lại độ chính xác cao trong việc kiểm tra chất lượng sản phẩm sáp ong dùng làm thực phẩm bổ sung.

Nghiên cứu của Xuejia Zhao và cộng sự [80] đã phát triển một công nghệ mới dựa trên phổ Raman truyền qua (Transmission Raman Spectroscopy - TRS) kết hợp với PLS để xác định diclofenac sodium trong các viên thuốc thương mại. Phương pháp này đã chứng minh tính hiệu quả trong việc đo lường thành phần hoạt chất trong viên thuốc, đồng thời có thể được ứng dụng để kiểm tra chất lượng thuốc theo thời gian thực trong các dây chuyền sản xuất.

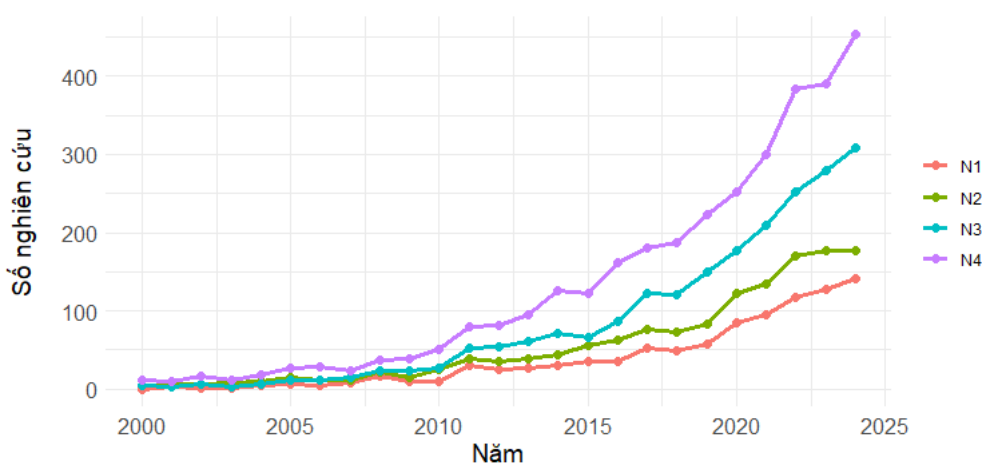
Cuối cùng, một nghiên cứu quan trọng của Yuwen Zhao và cộng sự [81] đã ứng dụng phương pháp SERS kết hợp với PLS và mạng nơ-ron nhân tạo (ANNs) để phân tích đồng thời các vi khuẩn *Escherichia coli*, *Staphylococcus aureus* và *Salmonella typhimurium*. Phương pháp này đã đạt được độ chính xác cao trong việc phát hiện và định lượng các vi khuẩn gây bệnh, cho thấy tiềm năng ứng dụng trong kiểm tra chất lượng vi sinh vật.

Từ các nghiên cứu trên, có thể thấy rằng phương pháp PLS kết hợp với các kỹ thuật phổ như Raman, UV-Vis, và NIR đang dần trở thành công cụ mạnh mẽ trong phân tích hóa học, đặc biệt là trong dược phẩm và chất lượng thực phẩm. Việc kết hợp các phương pháp này giúp giải quyết các vấn đề phức tạp liên quan đến định lượng và kiểm tra chất lượng, đồng thời mang lại các phương pháp nhanh chóng, chính xác và ít tốn kém hơn so với các phương pháp truyền thống như sắc ký lỏng hiệu suất cao (HPLC).

Để tìm hiểu tổng quan về việc áp dụng phương pháp PLS kết hợp phổ UV-Vis trong việc xác định hàm lượng các chất trong phân tích dược phẩm nói riêng và các đối tượng phân tích khác, chúng tôi sử dụng công cụ google Scholar để tìm kiếm các nghiên cứu được công bố trong khoảng thời gian từ 2000 đến 2024. Chúng tôi sử dụng tìm kiếm nâng cao để giới hạn lĩnh vực tìm kiếm phù hợp với nghiên cứu với các từ khóa, trong đó N1 = "Partial Least Squares" AND "UV-Vis spectroscopy" AND "quantification" AND "pharmaceutical"; N2 = "Partial Least Squares" AND "UV-Vis spectroscopy"

AND "pharmaceutical"; N3 = "Partial Least Squares" AND "UV-Vis spectroscopy" AND "quantification"; N4 = "Partial Least Squares" AND "UV-Vis spectroscopy".

Hình 1.7 cho thấy một rõ xu hướng tăng trong các công trình nghiên cứu ứng dụng PLS kết hợp với phổ UV-Vis từ năm 2000 cho đến năm 2024. Có thể thấy rằng, trước năm 2010 việc ứng dụng thuật toán PLS kết hợp với phổ UV-Vis chưa được chú trọng với số lượng công trình tương đối ít mỗi năm, dưới 50 bài mỗi năm. Bắt đầu từ 2010 đến 2018, số lượng các công trình công bố tăng lên đáng kể nhưng chưa thật sự mạnh mẽ. Từ năm 2018 cho đến nay, số lượng công bố hàng năm tăng nhanh, điều này có thể phản ánh sự ứng dụng rộng rãi hơn của PLS và UV-Vis trong các lĩnh vực khoa học phân tích bao gồm phân tích các thành phần trong dược phẩm, cũng như các lĩnh vực phân tích khác. Phương pháp PLS kết hợp phổ UV-Vis để định lượng các thành phần trong dược phẩm (đường N1) vẫn chiếm một trọng số tương đối ít so với tổng số ứng dụng của phương pháp PLS kết hợp phổ UV-Vis (142 bài so với 453 bài trong năm 2024). Điều này cho thấy khả năng ứng dụng một phương pháp thu thập dữ liệu đơn giản, chính xác với máy móc UV-Vis ngày càng hiện đại mà không cần phải xử lý mẫu phức tạp có thể là một trong những hướng tiềm năng cần tiếp tục nghiên cứu để giúp quá trình phân tích ngày càng đơn giản, rõ ràng những vẫn đảm bảo độ đúng, độ chính xác, độ tin cậy cao.



Hình 1.7. Biểu đồ mô tả số lượng các công trình nghiên cứu liên quan đến phương pháp PLS từ năm 2000 đến 2024

Bảng 1.1. Kết quả một số nghiên cứu phân tích định lượng bằng phương pháp quang phổ kết hợp PLS trong Dược phẩm

ST T	Thành phần	Phương pháp lấy dữ liệu	Phương pháp chemometrics	Kết quả			Năm, TLT K
				R ²	RMSE	Độ thu hồi	
1	paracetamol; tramadol HCl	Fourier Transform Infrared	PLS	0.9995 0.9998	0.06279 0.03785		2025, [82]
2	germacrone β-elemene bisdemethoxycurcumin demethoxycurcumin curcumin	(FT-NIR)	CARS-PLS	0.9708 0.8744 0.9511 0.9803 0.9567	0.2098 0.2065 0.0034 0.0066 0.0172		2025, [83]
3	Montelukast sodium (MLK), Levocetirizine dihydrochloride (LCZ)	UV-Vis	CLS PCR PLS GA-PLS	0,9932 0,9901	0,1872 và 0,1779	99,72– 100,86%	2024, [84]
4	Indomethacin	NIR	iPLS	0.9981			2024, [85]
5	Ciprofloxacin Lomefloxacin Enrofloxacin	UV-Vis	FA-PLS	0.9997 0.9995 0.9992	0.6103 0.9682 1.0840	98.18 - 101.83%	2024, [86]
6	ademetionine	ATR-FTIR	PLS	0.999			2024, [87]

7	Piroxicam Naproxen Diclofenac Sodium Mefenamic Acid	UV-Vis	PLS	0.9978 0.9932 0.9985 0.9970	0.5101 0.2978 0.5388 0.5822	100.45% 100.01% 101.93% 104.69%	
8	Insulin Aspart	Phổ Raman	PLS	0.969	7.21 IU/mL		2024, [88]
9	Levofloxacin (LEV) Metronidazole (MET) Rifampicin (RIF) Sulfamethoxazole (SUL)	UV-Vis	PLS	0.998 0.9984 0.9987 0.9986	0.1811 0.1804 0.1428 0.1716	95.67 - 103.30% 87.11 - 103.30% 97.10 - 98.94% 95.67 - 103.30%	2023, [89]
10	Ciprofloxacin: - cùng tá dược - tá dược khác nhau - Tập hợp gộp	FT-NIR	PLS, MCR- ALS		0.38% - 0.67% 0.47% - 1.76%		2023, [90]
11	Tetracaine Free Base (TC) Tetracaine Hydrochloride (TC·HCl) Indomethacin (IMC) Procaine (PC) và Procaine Hydrochloride (PC·HCl)	NMR	PLS	0.9986 0.9961 0.9974 0.9982	1.37% 2.01% 1.56% 1.82%		2023, [91]

12	Dexamethasone	FT-NIR	PLS, SPORT	0.9044	450 mg/kg	≈98-102%	2023, [92]
13	Acetylshikonin, β-Acetoxyisovalerylalkannin, Isobutylshikonin, Isovalerylshikonin, β,β-Dimethylacrylalkannin,	FT-NIR	PLS, SVM	0.9913 0.9855 0.9825 0.9916 0.9825			2023, [78]
14	Metformin	SERS (surface-enhanced Raman spectroscopy)	PCA PLS	0.97	0.42		2023, [93]
15	metformin	ATR-FTIR	PLS	0,997	0.447		2022, [94]
16	caffeine & fentanyl caffeine, fentanyl, mannitol, erythritol, heroin	UV-VIS Raman	PLSR	0.99 0.96	1.08% 1.63%		2021, [95]

17	Alprazolam	MIR, NIR, Raman	PLS, OPLS		1.118% 0.08% 0.059%		2020, [96]
18	Paracetamol và Piroxicam	UV-Vis	GA-PLS JK-PLSR	$R^2 > 0.999$	< 0.25 mg/L < 0.15 mg/L		2020, [97]
19	Icariin Salvianolic Puerarin	NIR	Si-PLS	0.9905 0.9781 0.9668	7.84% 9.48% 12.81%		2020, [98]
20	Doxorubicin Epirubicin Daunorubicin	Raman IR	PLSR	0.9972 0.9984 0.9971	0.0362 0.0244 0.0349		2019, [99]
21	propranolol montelukast haloperidol	Raman NIRR	PLS	0,95 đến 0,99.			2019, [100]

1.5 GIỚI HẠN PHÁT HIỆN TRONG PHÂN TÍCH ĐA BIẾN

Trong hóa phân tích định lượng, việc xác định các chỉ số để đánh giá độ tin cậy của phép đo như độ nhạy, giới hạn phát hiện, giới hạn định lượng ... là vô cùng quan trọng. Lý do cho sự quan trọng này bởi các chỉ số này xác định độ chính xác và độ nhạy của phương pháp phân tích [101]. Cho đến nay, LOD trong phân tích đa biến vẫn chưa có một hướng dẫn chính thức nào từ các tổ chức, hiệp hội quốc tế để xác định giá trị LOD cho các phép định lượng đa biến. Theo khái niệm, “LOD là lượng chất nhỏ nhất có thể phân biệt được một cách đáng tin cậy” [102]. Điều đó có nghĩa là LOD là lượng tối thiểu có thể phát hiện được với một xác suất xác định cho sai số loại I (dương tính giả) và sai số loại II (âm tính giả). Để xác định LOD có thể (1) dựa vào tính hiệu đo lặp lại mẫu trắng và các mẫu gần giới hạn phát hiện kết hợp kiểm định Neyman-Pearson xét đến hai loại sai số loại I và sai số loại II với một xác suất xác định hoặc (2) được tính từ việc xây dựng các mô hình hiệu chuẩn.

Trong phương pháp phân tích đơn biến, giá trị LOD có thể được tính bằng cách đo mẫu trắng lặp lại, giá trị LOD bằng tín hiệu mẫu trắng cộng 3 lần độ lệch chuẩn của mẫu trắng. Một cách tính khác được đưa ra từ việc xây dựng phương trình đường chuẩn, khi đó giá trị LOD bằng hệ số chặn cộng 3 lần độ lệch chuẩn của phần dư [103].

Trong phân tích đa biến, trong các mẫu khác nhau dù có cùng nồng độ chất phân tích nhưng thành phần mẫu khác nhau sẽ tạo ra phổ khác nhau, đây chính là sự phức tạp của các mẫu trong phân tích đa biến so với phân tích đơn biến [104]. Chính vì vậy, trong phân tích đa biến không có một mẫu trắng duy nhất mà mẫu trắng có thể thay đổi phụ thuộc vào thành phần, nồng độ của các chất còn lại. Vẫn còn các tranh cãi về cách xác định đúng đắn cho LOD trong phân tích đa biến. Cho đến nay đã có một số đề xuất về cách xác định LOD trong phân tích đa biến.

Lorber và cộng sự đã phát triển một phương pháp dựa trên khái niệm tín hiệu phân tích của chất phân tích tinh khiết (NAS – Net Analyte Signal). Ý tưởng cơ bản của NAS là tách biệt tín hiệu mà chỉ phản ánh sự hiện diện của chất phân tích từ những tín hiệu khác liên quan đến các thành phần không mong muốn hoặc các yếu tố nhiễu trong mẫu, hay nói cách khác NAS là phân tín hiệu đặc trưng cho chất phân tích. Phương pháp này đã được một số tác giả dùng để xác định LOD trong các phương pháp phân tích đa biến như CLS, ILS, PCR, PLS [105-108]. Tuy nhiên, cách tính này mắc nhược điểm chính là nó chỉ xem xét độ không đảm bảo đo trong các phép đo tín hiệu mà bỏ qua các nguồn sai số quan trọng khác là nồng độ và tín hiệu hiệu chuẩn [104, 109].

R. Boque và cộng sự đã đề xuất phương pháp để tính toán LOD đa biến dựa theo hướng dẫn của UIPAC liên quan đến khả năng phát hiện của các phương pháp phân tích, bao gồm dựa trên lý thuyết kiểm định giả thiết để xem xét cả sai số loại I (xác suất dương tính giả) và sai số loại II (xác suất âm tính giả) và tích hợp tất cả nguồn sai số ảnh hưởng

đến kết quả cuối cùng, bao gồm cả giai đoạn hiệu chuẩn và dự đoán. LOD bị ảnh hưởng bởi sự hiện diện của các chất phân tích khác trong mẫu do đó nó có giá trị khác nhau cho từng mẫu được kiểm tra, có nghĩa là LOD đa biến là một khoảng nồng độ thay vì một giá trị nồng độ đơn lẻ [104]. Một số nghiên cứu tương tự cũng được đề cập [110-112]. Các phương pháp này đều chỉ ra chỉ số ảnh hưởng (leverage) của mỗi mẫu ở nồng độ chất phân tích bằng không (nồng độ chất phân tích bằng 0 nhưng nồng độ các thành phần khác không bằng 0 và thay đổi) chỉ là một ước lượng và không có một quy trình được thiết lập để tính toán rõ ràng.

Một đề xuất khác về cách tính LOD đa biến đó là mở rộng khái niệm trong hiệu chuẩn đơn biến sang hiệu chuẩn đa biến. Đề xuất này được các tác giả Ortiz và cộng sự [113, 114] nghiên cứu và được áp dụng trong một số trường hợp. Mặc dù phương pháp “giả đơn biến” (psuedounivariable) có vẻ hợp lý nhưng nó không phù hợp với lý thuyết về lan truyền độ không đảm bảo đo, khi mà trong tập hiệu chuẩn và tập kiểm tra có các nền mẫu khác nhau [115].

Việc xác định giá trị giới hạn phát hiện trong phân tích đa biến cho đến nay vẫn chưa có một qui định hay hướng dẫn chính thức từ các tổ chức trên thế giới. Vì vậy, việc nghiên cứu để đưa ra cách tính giá trị LOD là cần thiết và quan trọng.

1.6 LỰA CHỌN BƯỚC SÓNG TRONG PHÂN TÍCH QUANG PHỔ

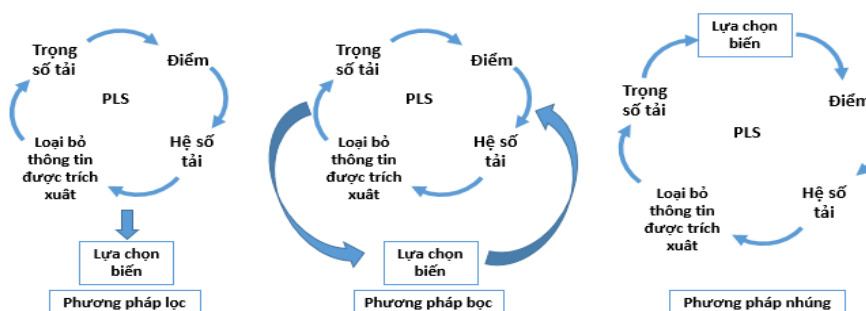
1.6.1 Sự cần thiết phải lựa chọn bước sóng

Kỹ thuật phân tích quang phổ hiện đại, dữ liệu đo độ hấp thụ của các mẫu phân tích trong một vùng ánh sáng quang học ở các bước sóng khác nhau dễ dàng thu thập được một cách nhanh chóng, chính xác trong một thời gian ngắn [116]. Thông tin mà dữ liệu độ hấp thụ mang lại rất nhiều và quan trọng, ví dụ đánh giá được khả năng hấp thụ mạnh yếu khác nhau do cấu trúc phân tử khác nhau, hoặc sự ảnh hưởng của các cấu tử hoặc dung môi ở nồng độ khác nhau trong mẫu hỗn hợp, hoặc thông tin nhiều của thiết bị, tá dược, cấu tử lạ... ở các bước sóng khác nhau. Các yếu tố ảnh hưởng này cũng tùy vào các vùng bước sóng khác nhau, ví dụ ở vùng tử ngoại xa sự hấp thụ chung của mẫu có thể có ảnh hưởng của dung môi sử dụng. Chính vì vậy, sự cần thiết phải đánh giá các yếu tố đến việc xây dựng các mô hình hồi quy tìm mối quan hệ giữa các chất cần phân tích với dữ liệu quang phổ thu được [117].

Thông thường, phương pháp PLS được huấn luyện với toàn bộ vùng phổ, bao gồm các biến không liên quan đến sự thay đổi của phản hồi (ví dụ: nồng độ); trong trường hợp này, có nguy cơ thu được một mô hình bị quá khớp (overfitting), tức là một mô hình có sai số nhỏ nhưng khả năng dự đoán kém. Tuy nhiên, việc huấn luyện PLS với một tập hợp các biến phổ được chọn lựa nên cho phép phân thông tin của phổ được trích xuất và mô hình hóa nhanh chóng, đồng thời loại bỏ các biến phổ khác không liên

quan hoặc không tương quan với biến phản hồi. Trong trường hợp này, nếu giữ lại quá ít biến, có thể dẫn đến mô hình bị underfitting, tức là một sự đánh giá quá cao về phương sai sai số [116, 118].

Mục tiêu của việc lựa chọn biến là thu được một tập hợp nhỏ các biến có khả năng tổng quát tốt hơn hoặc ít nhất là tương đương với tập hợp biến ban đầu. Những lợi ích chính của việc lựa chọn biến bao gồm cải thiện khả năng hiển thị dữ liệu và hiểu biết dữ liệu, giảm yêu cầu về đo lường và rút ngắn thời gian huấn luyện cũng như sử dụng, bằng cách khắc phục "lời nguyền của sự đa chiều" để cải thiện hiệu suất dự đoán (tăng cường độ tin cậy) [116, 119, 120].



Hình 1.8. Các mô hình lựa chọn bước sóng [121]

Nhiều nghiên cứu trên thế giới đã xây dựng các thuật toán cho phép các nhà phân tích có thể sử dụng để lựa chọn những khoảng bước sóng phù hợp mà sự ảnh hưởng của các yếu tố nhiễu là ít nhất. Nghiên cứu của Tahir Mehmood và cộng sự (2020) đã tổng quan được bao gồm 17 phương pháp lựa chọn bước sóng dựa trên mô hình PLS. Lựa chọn bước sóng trong PLS được phân thành ba nhóm chính: phương pháp lọc (filter methods), phương pháp bọc (wrapper method) và phương pháp nhúng (embedded method) được thể hiện qua hình 1.8. Phương pháp lọc được xây dựng qua hai bước, đầu tiên mô hình PLS được xây dựng, sau đó kết quả của thuật toán PLS được sử dụng để xác định tập con các bước sóng có đóng góp quan trọng. Phương pháp bọc (wrapper) sử dụng các thuật toán để tìm kiếm các tập con bước sóng để xây dựng mô hình PLS, quá trình được lặp lại nhiều lần cho đến khi đạt đến một ngưỡng tối ưu thì dừng lại và trích xuất các bước sóng được lựa chọn. Phương pháp nhúng lựa chọn các bước sóng dựa trên các thuật toán tích hợp trong thuật toán PLS [70, 121]. Trong một nghiên cứu khác, Mitchell Rogers và cộng sự đã mở rộng thêm các chiến lược lựa chọn bước sóng khác bao gồm phương pháp lựa chọn thủ công, phương pháp nối tiếp (concatenated) và phương pháp dựa trên khoảng (interval-based) [122]. Các mô hình máy học (machine learning), học sâu (deep learning) cũng được sử dụng để lựa chọn các biến tối ưu: thuật toán di truyền kết hợp support vector machine để lựa chọn các biến tối ưu cho nghiên cứu ảnh hưởng giữa gen và bệnh tật [123-126]. Mặc dù các mô hình lựa chọn biến này không được áp dụng cho việc lựa chọn các bước sóng trong phương pháp phân tích

quang phổ nhưng nó có thể là những gợi ý tuyệt vời cho các nhà phân tích quang phổ áp dụng.

Lựa chọn các bước sóng đặc trưng làm giảm bớt dữ liệu đầu vào cho quá trình tính toán, kết quả giải thích mối tương quan giữa các dữ liệu thu được với các biến phụ thuộc được tối ưu và chính xác hơn, kết quả dự đoán trên các mẫu thực tế cho kết quả tương đương hoặc tốt hơn so với việc sử dụng toàn bộ phổ [116, 127-131]. Tuy nhiên, xác định kỹ thuật lựa chọn bước sóng đặc trưng vẫn phụ thuộc vào ứng dụng cụ thể. Lựa chọn các tham số trong các thuật toán khác nhau cũng ảnh hưởng đến mức độ dự đoán khác nhau của mô hình, đặc biệt là những thuật toán có nhiều tham số tác động [122, 130, 132].

1.6.2 Tổng quan về phương pháp lựa chọn bước sóng trong PLS

Trong suốt nhiều thập kỷ qua, việc lựa chọn bước sóng kết hợp với PLS đã thu hút sự chú ý của nhiều nghiên cứu. Để có cái nhìn tổng quan về các công trình nghiên cứu đã được thực hiện (bao gồm cả các công trình liên quan), chúng tôi đã sử dụng một công cụ tìm kiếm phổ biến và được áp dụng rộng rãi: Google Scholar. Chúng tôi tìm kiếm các từ khóa xuất hiện trong tiêu đề, tóm tắt và nội dung các nghiên cứu, đồng thời loại trừ phần trích dẫn. Các chuỗi từ khóa tìm kiếm được chia thành ba nhóm chính:

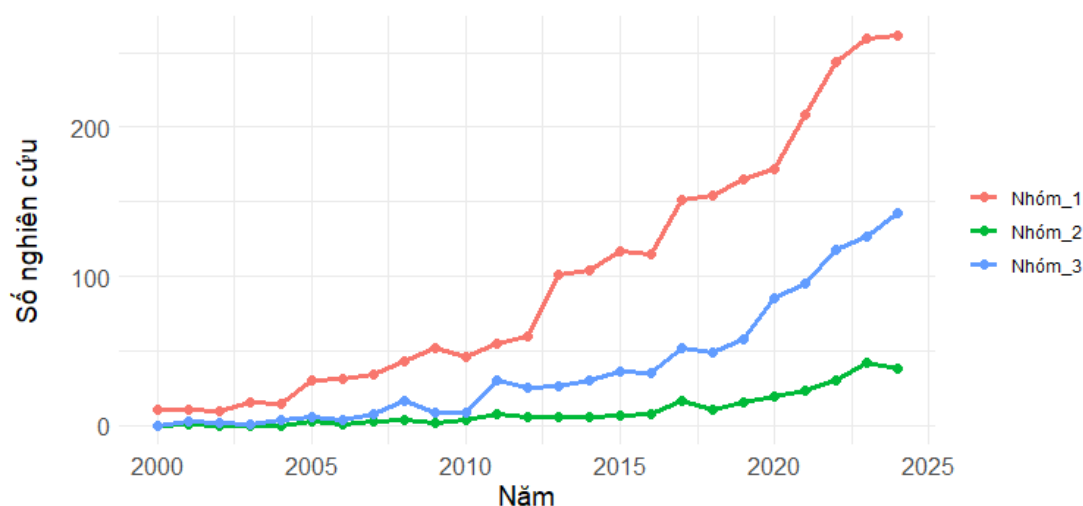
- Nhóm 1 bao gồm: "Partial Least Squares" AND "wavelength selection" OR "variable selection" AND "spectroscopy" AND "quantification" AND "pharmaceutical"
- Nhóm 2 bao gồm: "Partial Least Squares" AND "wavelength selection" OR "variable selection" AND "UV-Vis spectroscopy" AND "quantification" AND "pharmaceutical"
- Nhóm 3 bao gồm: "Partial Least Squares" AND "UV-Vis spectroscopy" AND "quantification" AND "pharmaceutical".

Bảng 1.2 Số lượng các nghiên cứu về lựa chọn bước sóng từ 2000 - 2024.

	Nhóm 1	Nhóm 2	Nhóm 3
2000	11	0	0
2001	11	1	3
2002	10	0	2
2003	15	0	1
2004	14	0	4
2005	30	3	6
2006	31	1	4
2007	34	3	8
2008	43	4	16
2009	52	2	9

2010	46	4	9
2011	55	8	30
2012	60	6	25
2013	101	6	26
2014	104	6	30
2015	117	7	36
2016	115	8	35
2017	151	16	52
2018	154	11	49
2019	165	15	58
2020	172	19	85
2021	208	23	95
2022	244	30	118
2023	259	42	127
2024	261	38	142

Bảng 1.2 liệt kê các công trình liên quan đến nghiên cứu lựa chọn bước sóng từ năm 2000 đến 2024. Biểu đồ hình 1.9 cho thấy sự gia tăng ứng dụng phương pháp lựa chọn bước sóng trong PLS trong các phương pháp quang phổ nói chung và phổ UV-Vis nói riêng trong phân tích dược phẩm, trong khoảng thời gian từ 2000 đến 2024. Tuy nhiên, trước năm 2010, số lượng nghiên cứu trong mỗi năm không đáng kể, đặc biệt là ứng dụng phương pháp lựa chọn bước sóng trong phổ UV-Vis cho phân tích dược phẩm, chỉ có dưới 10 công trình được tìm thấy trong một năm. Trong giai đoạn này, các thuật toán được phát triển chưa được ứng dụng nhiều hoặc các nhà nghiên cứu chưa chú trọng vào việc tối ưu các vùng bước sóng quan trọng để phát triển các nhóm phương pháp phân tích tối ưu. Từ năm 2010 cho đến 2020, số lượng các nghiên cứu bắt đầu tăng lên đáng kể trong nghiên cứu sử dụng các phương pháp lựa chọn biến sử dụng trong phân tích phổ, nhưng trong phổ UV-Vis (nhóm 2) vẫn chưa có sự gia tăng đáng kể. Bắt đầu từ năm 2020, cùng với sự phát triển của các mô hình học máy (machine learning), học sâu (deep learning) và sự cải tiến của công nghệ máy tính, việc xử lý các thuật toán với dữ liệu lớn trở nên dễ dàng hơn. Do đó, các nghiên cứu bắt đầu có sự phát triển vượt bậc so với những năm trước. Tuy nhiên, việc ứng dụng phương pháp lựa chọn bước sóng của phổ UV-Vis kết hợp PLS trong phân tích dược phẩm (nhóm 2) so với các phương pháp phổ khác như phổ hồng ngoại, phổ Raman... vẫn chiếm tỷ lệ tương đối ít, chỉ 38 công trình (nhóm 2) so với 261 công trình (nhóm 1) vào năm 2024. Điều này cho thấy, nghiên cứu tiếp tục về lựa chọn bước sóng cho phổ UV-Vis trong phân tích dược phẩm vẫn là một vấn đề cần được quan tâm để tối ưu hóa quá trình phân tích bằng phương pháp UV-Vis, một phương pháp đã được đánh giá là thiết bị đơn giản, rẻ tiền, tiện lợi nhưng vẫn đảm bảo được độ chính xác và tin cậy.



Hình 1.9. Biểu đồ số lượng các công trình nghiên cứu về lựa chọn bước sóng từ năm 2000 – 2024

Theo nghiên cứu tổng quan của Tahir Mehmood và cộng sự, nhóm phương pháp BVE-PLS và SR-PLS được đánh giá là những phương pháp có khả năng lựa chọn bước sóng hiệu quả. Trong khi đó, phương pháp GA là một trong những phương pháp có giải pháp tối ưu để chọn lọc biến. Vì vậy, chúng tôi đã tìm hiểu một số nghiên cứu cụ thể áp dụng ba phương pháp này trong phân tích quang phổ.

Năm 2024, nghiên cứu của Shengpeng Wang và cộng sự [133] sử dụng phương pháp quang phổ hồng ngoại gần (NIRS) kết hợp với phương pháp chemometrics và lựa chọn biến để phân biệt nguồn gốc của lá trà tươi. Các phương pháp chemometrics được sử dụng bao gồm phân tích phân biệt – bình phương tối thiểu riêng phần (PLS-DA), khoảng tương tác hiệu quả trong bình phương tối thiểu riêng phần (synergy interval PLS – siPLS), hồi quy cấu tử chính (PCA), và thuật toán di truyền (Genetic Algorithm – GA). Kết quả cho thấy phương pháp GA kết hợp siPLS sử dụng 37 điểm dữ liệu quang phổ cho kết quả tốt nhất ($R^2 = 0,9706$; RMSEP = 0,0772; RPD = 6,59). Mô hình có độ chính xác dự đoán là 96,67% đối với mẫu trong tập kiểm tra và 93,33% đối với mẫu ngoài. Phương pháp này nhanh chóng, chính xác và không xâm lấn, giúp giám sát và điều chỉnh việc buôn bán trái phép lá trà tươi.

Nghiên cứu của Noha S. Katamesh và cộng sự năm 2024 [84] khi xác định đồng thời Montelukast Sodium (MLK) và Levocetirizine dihydrochloride (LCZ) trong thuốc điều trị Covid-19 đã sử dụng các phương pháp phổ UV-Vis kết hợp với một số phương pháp bao gồm bình phương tối thiểu cổ điển (CLS), hồi quy thành phần chính (PCR), bình phương tối thiểu riêng phần (PLS) và bình phương tối thiểu riêng phần kết hợp thuật toán di truyền (GA-PLS). Kết quả cho thấy phương pháp GA-PLS đã giúp giảm kích thước ma trận phổ xuống 63% và 51% cho MLK và LCZ. Mô hình GA-PLS cho

độ chính xác cao với độ thu hồi từ 98-120%, giá trị RMSEC, RMSEP thấp, thân thiện với môi trường.

Năm 2020, M. M. Salim và cộng sự [134] đã xác định đồng thời Clopidogrel (CLO), Atorvastatin (ATOR), Aspirin (ASP) và acid Salicylic (SAL) bằng các phương pháp bình phương tối thiểu riêng phần (PLS) và bình phương tối thiểu riêng phần kết hợp thuật toán di truyền (GA-PLS). Phương pháp GA đã giảm ma trận dữ liệu độ hấp thụ xuống khoảng 25%, 15%, 21% và 18% tương ứng với CLO, ATOR, ASP và SAL. Phương pháp GA đã chọn các biến có thông tin quan trọng nhất và loại bỏ những biến không có thông tin, giúp giảm độ phức tạp của phương pháp mà không ảnh hưởng đến khả năng dự đoán. Những phương pháp này có thể được vận hành bằng một máy quang phổ đơn giản, giá rẻ, dễ dàng có sẵn và không cần các bước tách biệt trước đó; do đó, chúng giúp giảm thời gian phân tích. Khi cập nhật mô hình, chúng có thể được áp dụng thành công để xác định các thuốc được nghiên cứu trong các chế phẩm dược phẩm của chúng. Kết quả của các phương pháp đề xuất và các phương pháp đã báo cáo đã được so sánh thống kê, xác nhận rằng không có sự khác biệt đáng kể giữa chúng về độ chính xác và độ tin cậy.

Một nghiên cứu khác năm 2020, Ahmed Elsonbaty và cộng sự [135] đã sử dụng phổ UV kết hợp chemometric để định lượng đồng thời atenolol, ramipril, hydrochlorothiazide, simvastatin và aspirin. Phương pháp PLS và GA-PLS được sử dụng để giải quyết sự xen phủ phổ UV của các hỗn hợp này. Hiệu suất của các mô hình này đã được so sánh về độ chính xác và khả năng dự đoán bằng phương pháp kiểm tra chéo và đánh giá với tập hỗn hợp bên ngoài. Kết quả cho thấy phương pháp PLS có độ thu hồi tốt nhưng phương pháp GA-PLS có hiệu suất phân tích tốt hơn nhờ khả năng loại bỏ các biến dư thừa, tức là số lượng biến độ hấp thụ đã giảm xuống khoảng 19-28%. Các phương pháp phát triển đã cho phép xác định đáng tin cậy trong các hỗn hợp chuẩn bị trong phòng thí nghiệm và chế phẩm dược phẩm với kết quả so sánh được với phương pháp HPLC, vì vậy các phương pháp chemometric này trở thành công cụ phân tích giá trị và không thể thiếu trong kiểm tra quy trình và phân tích kiểm soát chất lượng dược phẩm.

Thuật toán di truyền kết hợp phương pháp PLS và ANN để xác định đồng thời Azilsartan Medozomil (AZM), Chlorthalidone (CHT) và Azilsatan (AZ) bằng phương pháp UV đã được Hany W. Darwish và cộng sự đã sử dụng [136]. Các tác giả đã so sánh giữa việc lựa chọn bước sóng và sử dụng toàn bộ bước sóng để xác định nồng độ của các thành phần trên trong hỗn hợp. Phương pháp GA-PLS đã lựa chọn 51, 32 và 27 bước sóng tương ứng của AZM, AZ và CHT trong tổng thể 111 bước sóng ban đầu. Kết quả của công trình này kết luận rằng các phương pháp GA-PLS, ANN và GA-ANN có thể được xếp vào nhóm phương pháp nhạy và chính xác. Những lợi thế này chứng tỏ tiềm năng của việc sử dụng các phương pháp được đề xuất trong nghiên cứu kiểm soát chất

lượng AZM, CHT và AZ tại các phòng thí nghiệm không có thiết bị sắc ký lỏng. Hơn nữa, kết quả của nghiên cứu này mang lại hy vọng sử dụng các phương pháp chemometric thông minh để phân tích các sản phẩm được phẩm bằng các thiết bị dễ sử dụng và rẻ tiền như máy quang phổ UV, và phổ của chất có sự trùng lặp rất cao.

Năm 2019, Farah Assadian và Ali Niazi [137] đã chỉ ra rằng thuật toán GA là phương pháp thích hợp để chọn bước sóng cho việc hiệu chuẩn bằng phương pháp hồi quy phần tử tối thiểu (PLS) cho các hỗn hợp ofloxacin (OFL) và riboflavin (B2) có phổ gần như giống hệt nhau mà không làm giảm khả năng dự đoán khi sử dụng phương pháp quang phổ huỳnh quang. Phổ huỳnh quang của mẫu hỗn hợp được đo trong khoảng 465 – 650 với 186 bước sóng, thuật toán GA đã giảm còn 44 (23,7%) bước sóng cho mỗi chất. Các khoảng tuyến tính lần lượt là 0,5-5,0 và 2,0-10,0 $\mu\text{g/mL}$ cho ofloxacin và riboflavin. Độ chính xác của phương pháp để xác định đồng thời OFL và B2 đã được đánh giá bằng sai số bình phương trung bình dự đoán (RMSEP) lần lượt là 0,0868 và 0,158 cho ofloxacin và riboflavin, và sai số chuẩn dự đoán tương đối (RSEP) lần lượt là 2,738 và 2,846 cho ofloxacin và riboflavin khi sử dụng mô hình GA-PLS. Quy trình này cho phép xác định đồng thời OFL và B2 trong mẫu nước tiểu và huyết thanh của con người với độ tin cậy cao của việc xác định.

Một báo cáo nổi bật năm 2022 của Pongpreda Kingwattanakul và cộng sự [138] khi xử lý hình ảnh siêu phổ để xác định hàm lượng lactone bằng phương pháp PCR và PLS. Phản xạ quang phổ của hình ảnh siêu phổ được ghi nhận trong khoảng bước sóng 355 – 1505 nm được tiền xử lý bằng phương pháp đạo hàm bậc nhất và bậc hai của bộ lọc Savitzky-Golay để loại bỏ nhiễu đi kèm với dữ liệu quang phổ thô. Để chọn bước sóng tối ưu cho hệ số hồi quy (RC), phương pháp chọn biến tiến (FVS) và loại bỏ biến ngược (BVE) đã được áp dụng trong nghiên cứu này. Sau khi chọn đặc trưng, hơn 88% các biến quang phổ đã được loại bỏ. Kết quả cho thấy phương pháp BVE cho hiệu suất tốt hơn so với các phương pháp khác với R^2 là 0.971, RMSEP là 1.050, RCV là 0.995 và RMSECV là 0.606. Kết quả cho thấy việc sử dụng hình ảnh hyperspectral kết hợp với mô hình PLSR có thể thực hiện một mô hình dự đoán đáng tin cậy và chính xác để xác định hàm lượng lactone.

Năm 2024, Tulio Merino và cộng sự [139] nghiên cứu một mô hình dựa trên các thuật toán học máy (ML) để ước tính các thông số chất lượng trong bột cá, sử dụng quang phổ hình ảnh hyperspectral (HSI). Các thuật toán chọn đặc trưng đã được áp dụng để tìm ra các dải bước sóng quan trọng nhất trong các đặc trưng quang phổ thu được bao gồm phương pháp chiếu các biến quan trọng (Variable Importance Projection - VIP), loại bỏ biến ngược (Backward Variable Elimination - BVE) kết hợp PLS. Cuối cùng, hiệu suất của cả hai mô hình đã được so sánh. Kết luận cho thấy phương pháp BVE-PLS đã chứng minh hiệu quả trong việc chọn các dải bước sóng quang phổ quan trọng nhất để ước tính các thông số chất lượng của bột cá.

Năm 2020, Ahmed M. Ibrahim và cộng sự [140] đã sử dụng các thuật toán lựa chọn biến để trích xuất các thông tin quan trọng từ dữ liệu phổ UV-Vis nhằm định lượng Terazosin (TZ) trong sự có mặt của Prazosin Hydrochloride (PZ). Ba mươi lăm mô hình học máy đã được đánh giá và hiệu suất của chúng đã được so sánh. Các giá trị của sai số bình phương trung bình (RMSE), hệ số xác định (R^2), và sai số trung bình tuyệt đối (MAE) đã được tính toán để đánh giá các mô hình đã phát triển. Các mô hình dự đoán trong nghiên cứu này có thể hữu ích cho các nhà nghiên cứu quan tâm đến việc xác định thành phần hoạt chất trong các dạng thuốc được phẩm có sự can thiệp của các chất khác bằng cách sử dụng quang phổ UV; do đó, phương pháp này đã được sử dụng để xác định TZ mà không bị ảnh hưởng bởi PZ.

Năm 2019, Joshi Ritu và cộng sự [141] đã xác định Benzen, một chất gây ung thư nhóm 1, có trong dầu ăn bằng phương pháp quang phổ hồng ngoại biến đổi Fourier (FTIR) kết hợp phân tích đa biến. Phương pháp chiếu các biến quan trọng (VIP) và tỉ lệ chọn lọc (SR) kết hợp PLS để lựa chọn bước sóng được áp dụng. Kết quả thu được các giá trị R^2 lần lượt là 0,95, 0,96 và 0,95 và các giá trị sai số dự đoán chuẩn (SEP) lần lượt là 38,5, 33,7 và 41,7 mg/L. Kỹ thuật FTIR được đề xuất kết hợp với phân tích đa biến và các phương pháp chọn biến có thể phát hiện dư lượng benzen trong dầu ăn với những ưu điểm là nhanh chóng và đơn giản, do đó có thể thay thế các phương pháp truyền thống được sử dụng cho cùng mục đích.

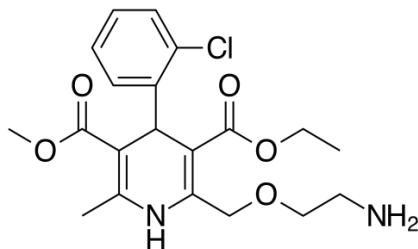
Năm 2016, B. Krakowska và cộng sự [142] đã sử dụng phương pháp HPLC sử dụng detector mảng diot kết hợp phương pháp phân biệt bình phương tối thiểu riêng phần (PLS-DA) và bốn phương pháp lựa chọn biến: loại bỏ biến không cung cấp thông tin, tầm quan trọng của biến trong phép chiếu, tỉ lệ chọn lọc và tương quan đa biến có ý nghĩa để phân biệt các mẫu thuốc Viagra. Mô hình PLS-DA tốt nhất được xây dựng trên một tập hợp các biến được chọn theo tầm quan trọng của chúng trong phương pháp chiếu. Đối với bộ kiểm tra độc lập, các ước tính trung bình với độ lệch chuẩn tương ứng của tỷ lệ phân loại đúng, độ nhạy, độ đặc hiệu và diện tích dưới đường cong lần lượt là $96,42\% \pm 2,04$, $98,69\% \pm 1,38$, $94,16\% \pm 3,52$ và $0,982 \pm 0,017$.

Kết quả tìm kiếm cụ thể bằng google scholar với nội dung lựa chọn bước sóng bằng các phương pháp GA, BVE, SR áp dụng cho phổ UV-Vis để phân tích đồng thời các chất trong Dược phẩm trong các năm gần đây rất ít, đặc biệt ở trong nước chưa tìm thấy được công trình nghiên cứu nào đã được công bố. Các phương pháp này hầu như chỉ được áp dụng lựa chọn bước sóng trong các phổ khác như phổ hồng ngoại, phổ Raman, hình ảnh siêu phổ... và lĩnh vực ứng dụng khác như nông nghiệp, thực phẩm, môi trường. Tính hiệu quả của việc lựa chọn các bước sóng được chỉ ra: có độ chính xác, độ tin cậy cao hơn, giảm kích thước chiều của dữ liệu, loại bỏ được ảnh hưởng của nhiễu. Trong lĩnh vực dược phẩm và phổ UV-Vis chưa nhận thấy nhiều công trình nghiên cứu được áp dụng.

1.7 GIỚI THIỆU AMLODIPIN BESYLAT, INDAPAMID, ATORVASTATIN CALCIUM, SPIRONOLACTON, FUROSEMID, RIFAMPICIN, ISONIAZID, PYRAZYNAMID, METRONIDAZOL, SPIRAMYCIN

1.7.1 Amlodipin Besylat

Công thức: $C_{20}H_{25}ClN_2O_5$ (viết tắt AML); $M = 408,88$; Các dạng được dùng: Amlodipin maleat: Am. $C_4H_4O_4$. Bột màu trắng; Amlodipin besylat (Am. Benzensulfonat): Am. $C_6H_5-SO_3H$



Hình 1.10. Công thức cấu tạo của Amlodipin

Tên khoa học: Ester 3-ethyl 5-methyl của acid 2-[(2-aminoethoxy) methyl]-4-2-(clorophenyl)-1,4-dihydro-6-methyl-3,5-pyridindicarboxylic [143].

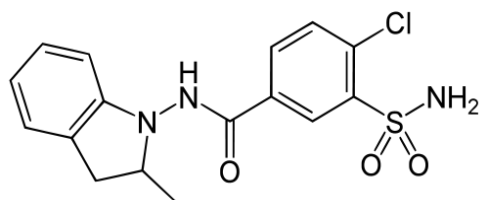
Tính chất: Thuộc nhóm chất không nitro; các tính chất lí-hoá chung của dẫn chất 1,4-dihydropyridin; dạng bột kết tinh màu trắng; dễ biến màu khi tiếp xúc ánh sáng; hấp thụ UV; dễ tan trong methanol, hơi tan trong ethanol khan, khó tan trong nước và 2-propanol [144].

Công dụng: Hạ huyết áp, giãn động mạch vành. Hấp thu tốt khi uống; Liều dùng: Người lớn, uống cho cả hai mục đích: 5-10 mg/lần/24 giờ; Chỉ định: Điều trị tăng huyết áp, đau thắt ngực. Chống chỉ định: Quá mẫn với dihydropyridin.

Tác dụng không mong muốn: Giãn mạch gây hạ huyết áp quá mức khi không có điều chỉnh liều, nhất là khi phải dùng thuốc kéo dài: hoa mắt, chóng mặt, đau đầu, phù ngoại vi ở mức độ thấp hơn [145].

1.7.2 Indapamid

Công thức: $C_{16}H_{16}ClN_3O_3S$ (viết tắt IND); $M = 365,8$



Hình 1.11. Công thức cấu tạo của Indapamid

Tên khoa học: 4-cloro-N-[(2RS)-2-methyl-2,3-dihydro-1H-indol-1-yl]-3-sulfamoylbenzamid, phải chứa từ 98,0% đến 102,0% $C_{16}H_{16}ClN_3O_3S$, tính theo chế phẩm khan. Tính chất: Bột màu trắng hoặc gần như trắng; thực tế không tan trong nước, tan trong ethanol (96%) [143].

Bảng 1.3. Một số công trình nghiên cứu về định lượng đồng thời Amlodipin và Indapamid trên thế giới

S T T	Phương pháp	Thành phần	Khoảng tuyến tính (ppm)	R^2	RSD (%)	Độ thu hồi Rev (%)	[TLTK]
1	UV-Vis (hiệu chỉnh độ hấp thụ)	AML	10 – 70	0,9979	0,55	100,99	[146]
		IND	2 – 16	0,9995	1,39	100,98	
2	UV-Vis (Phổ đạo hàm)	AML	10 – 70	0,9983	0,48	101,05	[146]
		IND	2 – 16	0,9973	0,39	101,06	
3	RP-HPLC	AML	5 – 30	0,9997	1,95	99,72	[147]
		IND	1,5 – 9	0,9998	1,07	100,28	
4	HPTLC	AML	-	0,9995	-	98,49 – 102,5	[148]
		IND	-	0,9977	-	99,2 – 102,1	

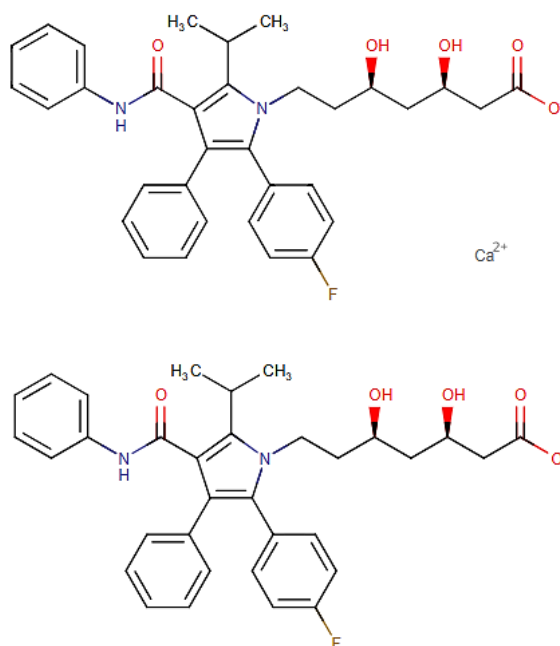
Dấu “-”: không có thông tin

Chỉ định: Indapamid dùng để điều trị bệnh tăng huyết áp vô căn. Hiệu lực điều trị tăng huyết áp của indapamid tương tự như thiazid. Thuốc có thể dùng một mình hoặc dùng phối hợp với các thuốc chống tăng huyết áp khác như các thuốc chẹn beta-adrenergic, thuốc ức chế enzym chuyển đổi angiotensin, thuốc chẹn kênh calci hoặc thuốc chẹn alpha-adrenergic. Indapamid có thể hiệp đồng hoặc tăng cường tác dụng các thuốc chống tăng huyết áp khác. Indapamid cũng được dùng điều trị phù và giữ muối do suy tim hoặc nguyên nhân khác. Trong suy tim trái cấp, nặng, trước tiên phải dùng các thuốc lợi niệu mạnh hơn như bumetanid hoặc furosemid.

Chống chỉ định: Không được dùng indapamid cho người bệnh mới bị tai biến mạch máu não, người vô niệu, người bệnh có tiền sử dị ứng với indapamid hoặc với các dẫn chất sulfonamid [145].

1.7.3 Atorvastatin calcium

- Công thức cấu tạo:



Hình 1.12 Công thức cấu tạo Atorvastatin calcium [149]

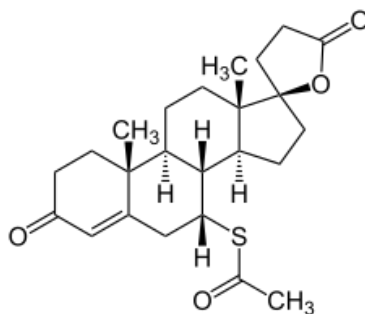
- Công thức: $(C_{33}H_{34}FN_2O_5)_2Ca \cdot 3H_2O$; Kí hiệu: ATO; Khối lượng phân tử: 1209,42; Tên khoa học: [R-(3R, 5R)]-7-[2-(4-fluorophenyl)-5-(1-methylethyl)-3-phenyl-4-(phenyl carbamoyl)-1H-pyrrol-1-yl] -3, 5-dihydroxy heptanoate trihydrate

Tính chất: chất bột tinh thể màu trắng, không hòa tan trong dung dịch nước có độ pH từ 4 trở xuống, nó rất ít hòa tan trong nước cất, dung dịch đệm photphat pH 7,4 và acetonitril. Nó hòa tan tự do trong methanol. Chỉ định: Ức chế men khử HMG-CoA, xúc tác bước đầu tiên trong quá trình tổng hợp cholesterol; hành động này làm giảm nồng độ cholesterol và lipoprotein tỷ trọng thấp (LDL), Có liên quan đến việc tăng nguy cơ mắc bệnh động mạch vành (CAD). Cũng làm tăng vừa phải nồng độ lipoprotein mật độ cao (HDL), liên quan đến việc giảm nguy cơ mắc bệnh CAD; Chống chỉ định: Mẫn cảm với Atorvastatin. Bệnh nhân có tiền sử bệnh gan, đau hoặc yếu cơ, bệnh thận, tiểu đường, rối loạn tuyến giáp [145].

Bảng 1.4. Một số công trình nghiên cứu xác định đồng thời AML và ATO trên thế giới

STT	Phương pháp	Thành phần	Khoảng tuyến tính (ppm)	R ²	RSD (%)	Độ thu hồi Rev (%)	[TLTK]
1	UV-Vis (Hiệu biên độ độ hấp thụ tỉ lệ)	AML					
		ATO	4-40	-	0,9122	100,22	[150]
			4-40	-	0,8968	100,12	
2	UV-Vis (Tỷ lệ hấp thụ)	AML	4-40	-	1,0124	100,6	[150]
		ATO	4-40	-	1,1045	100,86	
3	HPLC	AML	5-15	0,9999	2,56	100,39	[151]
		ATO	5-15	0,9999	2,02	100,18	
4	HPLC	AML	1-20	0,9993	1,65	100,85	[152]
		ATO	2-30	0,9994	1,75	99,9	
5	PLS2, iPLS, siPLS	ATO	0-32,0	> 0,99	1,1-1,22	96,8-102,3%	[153]
		AML	0-8,0		1,0 – 1,2	98,0-101,2%	
6	UV-Vis (Phổ tỉ đối)	AML	5-25	1	-	100,6-102	[154]
		ATO	10-50	0,9999	-	100,4-102,9	
7	LC-MS	ATO	0,2-20	>0,999	-	94,08-96,47	[155]
		AML	0,1-10		-	93,62-94,71	

1.7.4 Spironolacton

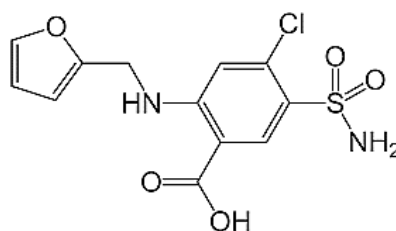


Hình 1.13. Công thức cấu tạo của Spironolacton [156]

Công thức: C₂₄H₃₂O₄S; Tên khoa học: γ -lacton của acid 7 α -acetylthio-17 β -hydroxy-3-oxo-pregn-4-en-21-carboxylic acetat. Khối lượng phân tử: 416,57

Tính chất: Bột kết tinh trắng hoặc trắng hơi vàng, mùi lợm giọng, vững bền ngoài không khí. Chảy ở khoảng 198°C - 207°C với sự phân hủy. Spironolacton thực tế không tan trong nước, dễ tan trong cloroform, tan trong ethanol [143]. **Chỉ định:** Phù do tăng quá mức Aldosteron (như phù vô căn, phù do xơ gan, hội chứng thận hư, suy tim sung huyết), thường phối hợp với các thuốc lợi tiểu khác. Tăng huyết áp: thường phối hợp với các thuốc lợi tiểu khác hoặc thuốc chống tăng huyết áp, tác dụng kém nếu dùng đơn độc. Suy tim sung huyết: dùng liều thấp, phối hợp với một thuốc ức chế enzym chuyển đổi Angiotensin. Tăng Aldosteron tiên phát: điều trị ngắn trước phẫu thuật tăng Aldosteron tiên phát hoặc điều trị duy trì lâu dài người bị Adenoma tuyến thượng thận tiết Aldosteron ít nhưng không phẫu thuật hoặc tăng Aldosteron vô căn. **Chống chỉ định:** Suy thận cấp, suy thận nặng, vô niệu, tăng kali huyết, giảm natri huyết, bệnh Addison's, mẫn cảm với Spironolacton [145].

1.7.5 Furosemid



Hình 1.14. Công thức cấu tạo của Furosemid [157]

Công thức: $C_{12}H_{11}ClN_2O_5S$; **Tên khoa học:** Acid 4-chloro-2-[(furan-2-ylmethyl)amino]-5-sulfamoylbenzoic; **Khối lượng phân tử:** 330,74

Tính chất: Bột kết tinh trắng hoặc gần như trắng, không mùi, không vị; không bền vững với ánh sáng, nhưng vững bền ngoài không khí; chảy ở khoảng 210°C với sự phân hủy; $pK_a = 3,9$. Furosemid thực tế không tan trong nước, tan trong aceton, khó tan trong ether, hơi tan trong ethanol, tan trong methanol và rất dễ tan trong dimethylformamid; tan trong các dung dịch kiềm loãng. Dưới tác dụng của ánh sáng, Furosemid biến màu dần [143, 144]. **Chỉ định:** Phù trong suy tim sung huyết, bệnh thận và xơ gan. Hỗ trợ điều trị phù phổi cấp. Đái ít do suy thận cấp hoặc mạn tính. Tăng huyết áp, đặc biệt khi do suy tim sung huyết hoặc do suy thận. Điều trị hỗ trợ cơn tăng huyết áp. Tăng calci huyết. **Chống chỉ định:** Mẫn cảm với Furosemid và các dẫn chất sulfonamid. Giảm thể tích máu, mất nước, hạ kali huyết nặng, hạ natri huyết nặng. Tình trạng tiền hôn mê gan,

hôn mê gan kèm xơ gan. Vô niệu hoặc suy thận do các thuốc gây độc đối với thận hoặc gan [145].

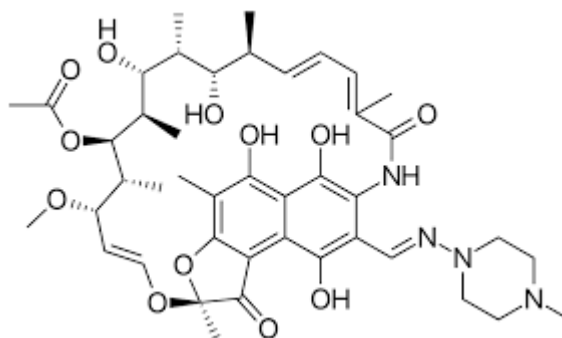
Bảng 1.5. Một số công trình nghiên cứu về định lượng đồng thời Spironolacton và Furosemid trên thế giới

STT	Phương pháp	Khoảng tuyến tính ($\mu\text{g/ml}$)	R^2	Độ thu hồi (%)	RSD, SD (%)	[TLTK]
1	UV-Vis (Vierordt)	2 - 10	0,98	102,02	RSD = 0,1	[158]
			0,99	101,97	RSD = 0,1	
2	RP-HPLC	40 - 160	0,9953	99,07 – 100,58	RSD = 1,19	[159]
			0,9977	98,05 – 100,17	RSD = 0,87	
3	UV-Vis (Phổ tỉ đối)	2 - 22	0,9999	100,41	SD = 0,980	[160]
		3 - 30		100,21	SD = 0,463	
4	PCR	2 - 12	0,986	92,72	SD = 5,53	[161]
		5 - 30	0,987	87,52	SD = 4,66	

1.7.6 Rifampicin

- Công thức: $\text{C}_{48}\text{H}_{58}\text{O}_{12}$; - Khối lượng phân tử: 822,94 g/mol; Tên khoa học: (3S, 4S, 5R, 9S, 15R, 16S, 17R, 18R, 19R, 20S, 21S, 22R, 23S)-2,4,12,16,20,22-hexahydroxy-5,9,17,19,21,23-hexamethyl-3-(propan-2-yl)-8,14,25-trioxo-3,4,5,6,9,15,16,17,18,19,20,21,22,23,24,25-hexadecahydro-1H-benzo[α]naphtho[2,1-d]furan-1-yl axetat.

- Công thức cấu tạo



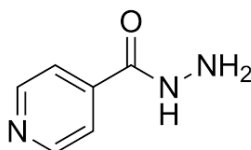
Hình 1.15. Công thức cấu tạo của Rifampicin

Tính chất: Bột kết tinh màu nâu đỏ hoặc đỏ nâu. Khó tan trong nước, khó tan trong acetone và ethanol 96 %, tan trong methanol. Chỉ định: điều trị bệnh lao (là một trong các thuốc chống lao hàng đầu). Là lựa chọn thay thế cho rifabutin trong điều trị nhiễm *M. avium* complex (MAC). Điều trị bệnh phong, dự phòng viêm màng não do *Haemophilus influenzae* type B và *Neisseria meningitidis*. Điều trị bệnh than. Điều trị nhiễm *Brucella*, nhiễm *Legionella pneumophila*, thay thế doxycycline trong điều trị nhiễm *E. phagocytophila*. Điều trị nhiễm khuẩn nặng do các chủng *Staphylococcus* và *Streptococcus*. Chống chỉ định: Mẫn cảm với rifampicin hoặc các dẫn chất rifamycin. Rối loạn chuyển hóa porphyrin ở những người nhạy cảm. Bệnh nhân đang dùng các thuốc kháng virus ritonavir-boosted). Phối hợp với các thuốc kháng virus nhóm ức chế Protease [145].

1.7.7 Isoniazid

Công thức: $C_6H_7N_3O$; Khối lượng phân tử: 137,14 g/mol; Tên khoa học: Isonicotinic hydrazid acid

Công thức cấu tạo



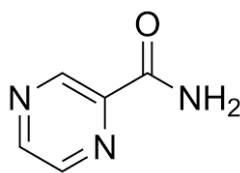
Hình 1.16. Công thức cấu tạo của Isoniazid

- Tính chất: Bột kết tinh trắng hay tinh thể không màu, không mùi. Dễ tan trong nước, hơi tan trong ethanol 96 %, khó tan trong chloroform, rất khó tan trong ether. Chỉ định: Isoniazid phối hợp với các thuốc chống lao khác theo các phác đồ chuẩn điều trị bệnh lao. Chỉ định dự phòng lao cho các nhóm người bệnh: những người trong gia đình có người bị lao và người thường xuyên tiếp xúc với người mới được chẩn đoán bệnh lao AFB(+) mà có test Mantoux dương tính và chưa tiêm phòng BCG, bất kể tuổi nào. Chống chỉ định: Người mẫn cảm với isoniazid. Người bị suy gan nặng, viêm gan nặng. Người bị viêm đa dây thần kinh và người bị động kinh [145].

1.7.8 Pyrazinamid

Công thức: $C_5H_5N_3O$; Khối lượng phân tử: 123,11 g/mol; Tên khoa học: Pyrazine-2-carboxamide

Công thức cấu tạo



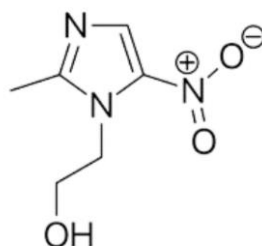
Hình 1.17. Công thức cấu tạo của Pyrazinamid

Tính chất: Bột kết tinh trắng hoặc gần như trắng, đa hình. Hơi tan trong nước, khó tan trong ethanol 96 % và methylen clorid. Chỉ định: Pyrazinamid là thuốc chống lao hàng 1, phối hợp với các thuốc chống lao khác điều trị các thể lao (bao gồm cả lao mới và lao kháng thuốc) do chủng *M. tuberculosis* đã biết hoặc dự đoán nhạy cảm với pyrazinamid. Chống chỉ định: Tổn thương gan nặng, bệnh gút cấp hoặc urê huyết cao, loạn chuyển hóa porphyrin, mẫn cảm với thuốc [145].

1.7.9 Metronidazol

Công thức: $C_6H_9N_3O_3$; Khối lượng phân tử: 171,2 ; Tên khoa học: 2-(2-methyl-5-nitro-1H-imidazol-1-yl)ethanol.

Công thức cấu tạo:



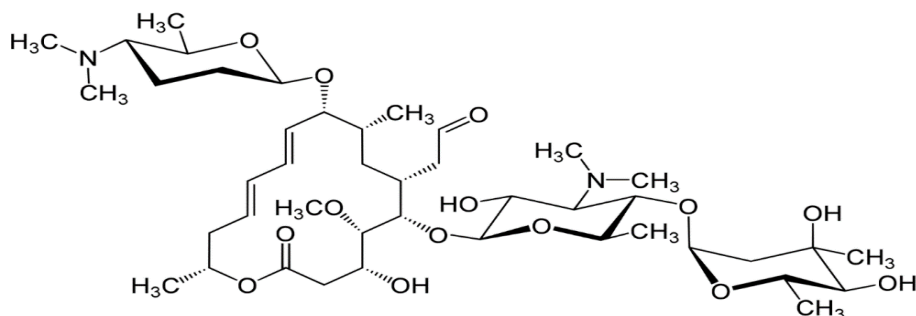
Hình 1.18. Công thức cấu tạo của Metronidazol [162]

Tính chất: Bột kết tinh màu trắng hoặc hơi vàng. Khó tan trong nước, trong aceton, ethanol 96 % và trong methylen clorid. Chỉ định: Điều trị các trường hợp nhiễm *Trichomonas vaginalis*, *Entamoeba histolytica* (thể cấp tính ở ruột và thể áp xe gan), *Dientamoeba fragilis* ở trẻ em, *Giardia lamblia* và *Dracunculus medinensis*. Trong khi điều trị bệnh nhiễm *Trichomonas*, cần điều trị cho cả nam giới. Điều trị nhiễm khuẩn nặng do vi khuẩn kỵ khí nhạy cảm như nhiễm khuẩn ổ bụng, nhiễm khuẩn phụ khoa, nhiễm khuẩn da và các cấu trúc da, nhiễm khuẩn hệ thần kinh trung ương, nhiễm khuẩn huyết và viêm màng trong tim. Phối hợp với uống neomycin, hoặc kanamycin để phòng ngừa khi phẫu thuật ở người phải phẫu thuật đại trực tràng và phẫu thuật phụ khoa. Viêm lợi hoại tử loét cấp, viêm lợi quanh thân răng và các nhiễm khuẩn răng khác do vi khuẩn kỵ khí. Bệnh Crohn thể hoạt động ở kết tràng, trực tràng. Viêm loét dạ dày - tá tràng do

Helicobacter pylori (phối hợp với 1 số thuốc khác) [145]. Chống chỉ định: Có tiền sử quá mẫn với metronidazol hoặc các dẫn chất nitro-imidazol khác.

1.7.10 Spiramycin

Công thức cấu tạo:



Hình 1.19. Công thức cấu tạo của Spiramycin [163]

Công thức: $C_{43}H_{74}N_2O_{14}$; Khối lượng phân tử: 843,1; Tên khoa học: (4R,5S,6S,7R,9R,10R,11E,13E,16R) – 6 - [[3,6 – dideoxy – 4 – O - (2,6-dideoxy – 3 – C – methyl – α – L – ribo - hexopyranosyl) – 3 - (dimethylamino) – β – D glucopyranosyl]oxy] – 4 – hydroxy – 5 – methoxy - 9,16 – dimethyl – 7 - (2 -oxoethyl) – 10 - [[2,3,4,6tetra-deoxy-4-(dimethylamino)-D-erythrohexopyranosyl]oxy] oxacyclohexadeca-11,13-dien-2-on.

Bảng 1.6. Một số công trình nghiên cứu về định lượng đồng thời Spironolacton và Furosemid trên thể giới

STT	Phương pháp	Thành phần	Khoảng tuyến tính ($\mu\text{g/ml}$)	R^2	Độ thu hồi (%)	RSD, SD (%)	[TLTK]
1	UV-Vis (Đường chuẩn)	MTZ	5-25	-	100,16	0,73	[164]
		SPI		-	100,09	0,7	
2	UV-Vis (Phổ đạo hàm)	MTZ	5-25	-	100,16	0,73	
		SPI		-	100,02	0,88	
3	UV-Vis (Phổ tỉ đối)	MTZ	5-25	-	101,1	0,9	
		SPI		-	100,49	1,26	

Tính chất: Là kết tinh màu trắng ánh vàng nhạt, hút ẩm nhẹ, ít tan trong nước, dễ tan trong ethanol 96%, aceton và methanol [144]. Chỉ định: Điều trị nhiễm khuẩn ở

đường hô hấp, da và sinh dục do vi khuẩn nhạy cảm. Điều trị dự phòng viêm màng não do Meningococcus (khi không có chống chỉ định với rifampicin). Dự phòng chứng nhiễm Toxoplasma bẩm sinh trong thời kỳ mang thai. Hóa dự phòng viêm khớp cấp tái phát ở người bệnh dị ứng với penicillin. Chống chỉ định: Người có tiền sử quá mẫn với spiramycin, erythromycin.

Tổng quan các tài liệu nghiên cứu cho thấy rằng các hoạt chất như Amlodipin, Indapamid, Atorvastatin, Spironolacton, Furosemid, Rifampicin, Isoniazid, Pyrazynamid, Metronidazol, Spiramycin đã được định lượng bằng nhiều phương pháp phân tích khác nhau như HPLC, LC-MS, HPTLC, các phương pháp phổ UV-Vis (phổ đạo hàm, phổ tỉ đối), hoặc một số mô hình hồi quy đa biến như PCR. Mặc dù các phương pháp này có hiệu quả nhất định, nhưng vẫn tồn tại nhiều hạn chế khi áp dụng cho các thuốc phối hợp chứa nhiều hoạt chất có phổ hấp thụ xen phủ nhau – một tình huống thường gặp trong thực tế sản xuất và kiểm nghiệm dược phẩm. Ngoài ra, một số kỹ thuật như LC-MS, HPLC yêu cầu thiết bị đắt tiền và quy trình phân tích phức tạp, khó triển khai đại trà trong điều kiện thực tế tại nhiều cơ sở kiểm nghiệm.

Đặc biệt, phương pháp bình phương tối thiểu riêng phần (PLS) vẫn chưa được ứng dụng đầy đủ và có hệ thống trong việc định lượng đồng thời các thành phần có phổ hấp thụ xen phủ, nhất là trong các thuốc phối hợp như AgimDogyl, Natrixam, Turbezid. Bên cạnh đó, các kỹ thuật lựa chọn bước sóng hiện đại như tỷ lệ chọn lọc (SR), loại biến ngược (BVE), và thuật toán di truyền (GA) cũng chưa được nghiên cứu sâu tại Việt Nam, dù chúng có tiềm năng nâng cao độ chính xác của mô hình PLS và giảm nhiều đáng kể bằng cách loại bỏ các bước sóng không mang thông tin.

Từ những lý do nêu trên, luận án lựa chọn hướng nghiên cứu ứng dụng mô hình PLS kết hợp với phổ UV-Vis để định lượng đồng thời các chất trong thuốc phối hợp, đồng thời khai thác các thuật toán lựa chọn bước sóng nhằm tối ưu hóa hiệu suất dự đoán và tăng độ tin cậy. Đây là một hướng tiếp cận hiện đại, có tính thực tiễn cao, góp phần bổ sung cho phương pháp phân tích dược phẩm trong bối cảnh yêu cầu kiểm soát chất lượng ngày càng khắt khe.

CHƯƠNG 2. NỘI DUNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

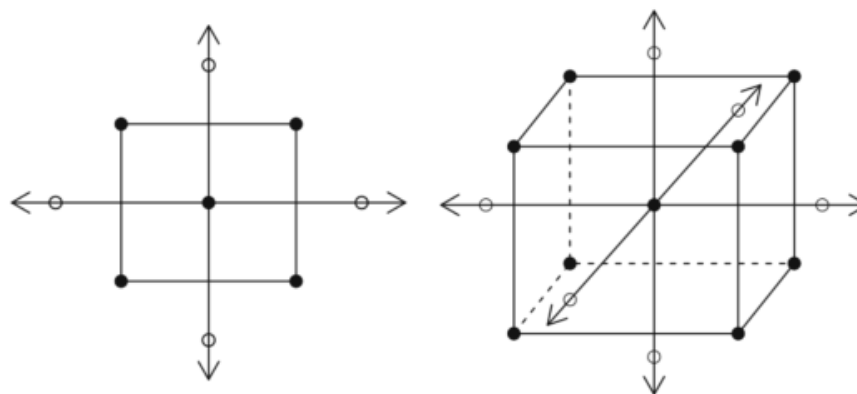
2.1. NỘI DUNG NGHIÊN CỨU

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 các hệ hai và ba thành phần có phổ hấp thụ xen phủ; trong đó tập hiệu chuẩn được thiết kế theo phương pháp thiết kế thí nghiệm (CCD), lựa chọn biến tiềm ẩn theo tiêu chuẩn PRESS, và tiến hành thẩm định theo các tiêu chí thống kê (RMSEC, RMSECV, RMSEP, R^2 , phân tích phần dư, độ thu hồi và độ lặp lại).
- Ứng dụng mô hình PLS–UV vào mẫu thực: áp dụng các mô hình đã xây dựng để định lượng đồng thời hoạt chất trong các thuốc thương mại (AgimDogyl, Zoamco-A, Spirosemid-40, Natrixam, Turbezid), so sánh kết quả với phương pháp chuẩn HPLC thông qua kiểm định thống kê (kiểm định t).
- So sánh hiệu năng giữa 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 trên dữ liệu phổ UV-Vis.
- Tối ưu hoá lựa chọn biến đầu vào cho PLS bằng các thuật toán lựa chọn bước sóng hiện đại (SR, BVE và GA) để rút gọn phổ hấp thụ, giảm số biến đầu vào nhưng vẫn duy trì hoặc cải thiện độ chính xác dự báo của mô hình PLS.
- Đề xuất và chuẩn hoá quy trình thực hành sử dụng PLS kết hợp phổ UV-Vis cho phân tích đa cấu tử trong dược phẩm, làm cơ sở cho triển khai ứng dụng trong thực tiễn phân tích QC/QA.

2.2. PHƯƠNG PHÁP NGHIÊN CỨU

2.2.1. Thiết kế thí nghiệm



Hình 2.1. Thiết kế thí nghiệm cấu trúc phức hợp với 2, 3 yếu tố.

Mục tiêu của thiết kế thí nghiệm là tối ưu hóa các phương pháp phân tích để áp dụng, như giảm số lần thử nghiệm cần thiết, giảm việc sử dụng hóa chất và giảm thiểu

công việc trong phòng thí nghiệm. Hơn nữa, thiết kế thí nghiệm cung cấp khuôn khổ tuần tự của một mô hình thống kê cho phép đánh giá các yếu tố quan trọng cùng với sự chấp nhận thống kê của các phản ứng đang được kiểm tra. Khía cạnh quan trọng nhất của thiết kế thí nghiệm là khả năng làm sáng tỏ các mối tương tác giữa các yếu tố ảnh hưởng trong quá trình phân tích [45, 165].

Trong toàn bộ các nghiên cứu này, thiết kế thí nghiệm dựa trên phương pháp thiết kế phức hợp trung tâm (central composite design – CCD) được sử dụng để tạo các tập thí nghiệm hiệu chuẩn và thẩm định trong các nghiên cứu, được mô tả trong hình 2.1. Bước đầu tiên là mã hóa các yếu tố, giả sử rằng điểm trung tâm cho mỗi yếu tố được mã hóa bằng 0, và thiết kế các điểm đối xứng quanh điểm đó, các thí nghiệm được bố trí như sau (bảng 2.1):

- Điểm tâm tương ứng với mức mã hóa bằng 0, là điểm trùng với trục tọa độ của mô hình.

Bảng 2.1. Mã hóa các điểm thiết kế thí nghiệm theo CCD

	Mô hình 2 yếu tố		Mô hình 3 yếu tố		
	A	B	A	B	C
Điểm tổ hợp các mức (Factorial points)	1	1	1	1	1
	1	-1	1	1	-1
	-1	1	1	-1	1
	-1	-1	1	-1	-1
			-1	1	1
			-1	1	-1
			-1	-1	1
			-1	-1	-1
Điểm sao (Star points)	α	0	α	0	0
	$-\alpha$	0	$-\alpha$	0	0
	0	α	0	α	0
	0	$-\alpha$	0	$-\alpha$	0
			0	0	α
			0	0	$-\alpha$
Điểm tâm	0	0	0	0	0

- Điểm tổ hợp các mức (Factorial points): Mỗi yếu tố tại những điểm này sẽ tương ứng với mã hóa -1 và $+1$, với phân này nếu mẫu có k yếu tố thì sẽ có 2^k điểm dữ liệu. Các điểm dữ liệu này nằm ở các đỉnh của hình vuông nếu mô hình có 2 yếu tố hoặc nằm ở các đỉnh của hình lập phương nếu mô hình có 3 yếu tố.

- Điểm sao (Star points): Mỗi điểm trong phần này là các điểm nằm ở một khoảng cách xác định α từ trung tâm thiết kế trên mỗi hướng dọc theo các trục được xác định bởi các yếu tố, tương ứng với các mẫu ở điểm này sẽ có $2k$ điểm dữ liệu. Giá trị α

phụ thuộc vào mỗi loại thiết kế, nếu thiết kế theo khả năng xoay (rotatability) thì $\alpha = \sqrt[4]{2^k}$ [27, 166-168].

2.2.2. Phương pháp thẩm định chéo

Trong học máy, thẩm định chéo (cross-validation (CV)) là một kỹ thuật đánh giá hiệu suất của mô hình học máy. Mục đích của CV là để kiểm tra khả năng tổng quát của mô hình bằng cách sử dụng dữ liệu huấn luyện và kiểm tra trên các phân vùng khác nhau của tập dữ liệu, tránh hiện tượng mô hình chỉ học tốt trên dữ liệu huấn luyện mà không thể áp dụng vào dữ liệu mới. Các phương pháp thẩm định chéo phổ biến bao gồm thẩm định chéo K lần (K-fold cross-validation) và thẩm định chéo loại bỏ từng phần (Leave-one-out cross-validation) [27, 169, 170].

Thẩm định chéo K lần chia tập dữ liệu thành K phần đều nhau. Mỗi phần được sử dụng một lần làm dữ liệu kiểm tra, trong khi các phần còn lại được sử dụng làm dữ liệu huấn luyện. Quá trình này được lặp lại K lần, mỗi lần với một phần khác nhau làm dữ liệu kiểm tra. Cuối cùng, hiệu suất của mô hình được tính bằng giá trị PRESS hoặc RMSE của các kết quả kiểm tra trên các phần thẩm định chéo K lần giúp giảm sự thiên lệch do việc chia tập huấn luyện và kiểm tra ngẫu nhiên, đặc biệt là khi dữ liệu có kích thước nhỏ. Nó đảm bảo rằng mọi phần của dữ liệu đều được sử dụng cả trong huấn luyện và kiểm tra, cung cấp đánh giá chính xác hơn về khả năng tổng quát của mô hình [171, 172].

Thẩm định chéo loại bỏ từng phần là trường hợp đặc biệt của thẩm định chéo K lần, trong đó số phần K bằng số lượng mẫu trong tập dữ liệu. Nói cách khác, mỗi lần lặp, một mẫu duy nhất được sử dụng làm dữ liệu kiểm tra, còn lại tất cả các mẫu khác được dùng làm dữ liệu huấn luyện. Thẩm định chéo loại bỏ từng phần giúp sử dụng tối đa dữ liệu huấn luyện, vì mỗi lần lặp mô hình chỉ bị loại bỏ một mẫu duy nhất. Thẩm định chéo loại bỏ từng phần thường được sử dụng khi có ít dữ liệu và khi mỗi mẫu đều có giá trị quan trọng. Tuy nhiên, thẩm định chéo loại bỏ từng phần có thể rất tốn kém về mặt tính toán khi dữ liệu lớn, vì nó yêu cầu huấn luyện mô hình N lần, với mỗi lần sử dụng một mẫu khác nhau làm dữ liệu kiểm tra [173].

Các bước tiến hành thẩm định chéo loại bỏ từng phần [169]:

- *Bước 1.* Một mẫu được loại bỏ khỏi tập hiệu chuẩn
- *Bước 2.* Mô hình PLS được xây dựng với các biến tiềm ẩn.
- *Bước 3.* Nồng độ của chất phân tích trong mẫu bị loại bỏ được tính bởi mô hình PLS đã xây dựng ở các biến tiềm ẩn khác nhau ($\hat{C}_i$).

- *Bước 4.* Bình phương của sai số giữa giá trị dự đoán và giá trị thực được tính:

$$(C_i - \hat{C}_i)^2$$
- *Bước 5.* Mẫu bị loại bỏ được đưa trở lại trong tập hiệu chuẩn, một mẫu khác bị loại bỏ ra.
- *Bước 6.* Lặp lại từ bước 1 cho đến bước 5 cho đến khi tất cả các mẫu đều được loại bỏ 1 lần.
- *Bước 7.* Giá trị tổng bình phương sai số phần dư dự đoán của mỗi biến tiềm ẩn được tính:

$$PRESS = \sum_{i=1}^n (C_i - \hat{C}_i)^2 \quad (41)$$

Hoặc căn bậc hai của trung bình bình phương sai số (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (C_i - \hat{C}_i)^2}{n}} \quad (42)$$

Trong đó, C_i và $\hat{C}_i$ là nồng độ thực và nồng độ tính được từ mô hình, n là số mẫu phân tích.

2.2.3. Lựa chọn số biến tiềm ẩn tối ưu cho mô hình PLS

Với mô hình PLS, để lựa chọn số biến tiềm ẩn phù hợp, quá trình thực hiện thuật toán PLS cho đến khi mô hình hội tụ, song song với quá trình như vậy, phương pháp thẩm định chéo loại bỏ từng phần cũng được áp dụng để tính giá trị PRESS ở mỗi biến tiềm ẩn được trích xuất. Dựa vào giá trị PRESS để xác định số biến tiềm ẩn tối ưu, có 3 trường hợp bao gồm dựa vào giá trị PRESS nhỏ nhất, dựa vào so sánh giá trị PRESS của các biến tiềm ẩn nhỏ so với PRESS của biến tiềm ẩn nhỏ nhất sao cho không có sự khác biệt theo đề xuất của Haaland và Thomas và dựa vào đánh giá giá trị PRESS tương tự như cách thứ 2 nhưng giá trị PRESS được tính trên tập kiểm tra, độc lập với tập hiệu chuẩn.

Các bước thực hiện theo đề xuất của tác giả Haaland và Thomas như sau:

Để lựa chọn mô hình tối ưu, chúng tôi đã tính toán PRESS cho các mô hình dựa vào quá trình thẩm định chéo (CV) với số lượng biến tiềm ẩn khác nhau. Mô hình mang lại giá trị PRESS tối thiểu được sử dụng làm mốc chuẩn, và được ký hiệu là h^* (h^* là giá trị biến tiềm ẩn mà ở đó có giá trị PRESS là nhỏ nhất). Tất cả các mô hình có số yếu tố ít hơn ($h < h^*$) sẽ được so sánh với mốc chuẩn này. Mục đích của sự so sánh này là tìm mô hình nhỏ nhất (số biến tiềm ẩn ít nhất) sao cho PRESS của mô hình này không lớn hơn đáng kể so với PRESS của mô hình có biến tiềm ẩn h^* .

Quá trình lựa chọn mô hình tối ưu diễn ra như sau:

Bước 1. Tính $F(h) = \frac{PRESS \text{ (mô hình với } h \text{ thành phần)}}{PRESS \text{ (mô hình với } h^* \text{ thành phần)}}$ với $h = 1, 2, \dots, h^*$. (43)

Bước 2. Chọn số lượng thành phần tối ưu h sao cho $F(h) < F_{(\alpha, m, m)}$, trong đó $F_{\alpha, m, m}$ là phân vị $(1 - \alpha)$ của phân phối F của Snedecor với m và m bậc tự do (với m là số mẫu hiệu chuẩn). Mô hình tối ưu được chọn là mô hình có giá trị PRESS đầu tiên có xác suất tỉ số F dưới 0,75 [54, 174].

2.2.4. Giới hạn phát hiện trong mô hình PLS

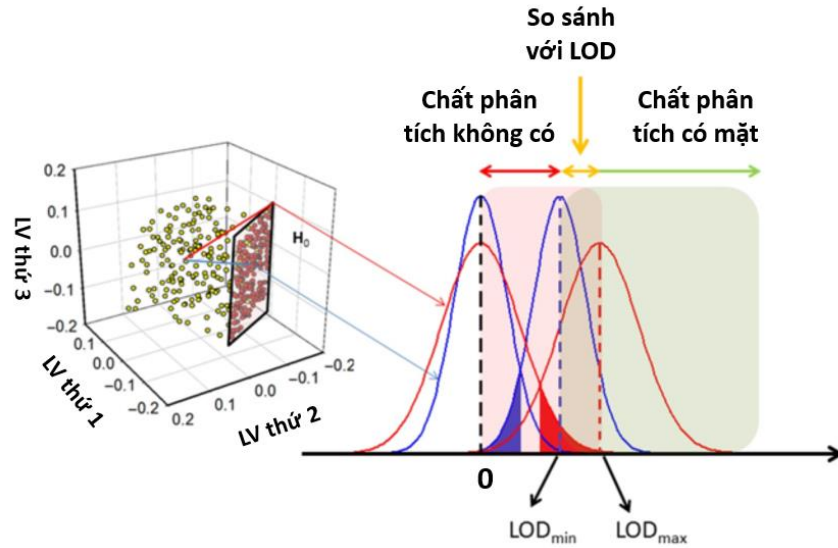
Theo tác giả Alejandro C. Olivieri, thay vì một giá trị LOD duy nhất, một dải giá trị LOD đã được cung cấp tùy thuộc vào sự biến động của các thành phần nền trong không gian hiệu chuẩn. [109, 175]. Khoảng giá trị LOD liên quan mật thiết đến khái niệm chỉ số ảnh hưởng (leverage), ký hiệu là h , đây là một tham số không có chiều, đo lường vị trí của một mẫu thử trong không gian hiệu chuẩn PLS, dựa trên khoảng cách Mahalanobis của mẫu so với trung tâm của không gian hiệu chuẩn. Chỉ số ảnh hưởng (leverage) h phản ánh mức độ mà một mẫu ảnh hưởng đến mô hình hiệu chuẩn, h cao cho thấy mẫu nằm xa trung tâm không gian hiệu chuẩn, làm tăng độ không chắc chắn trong dự đoán nồng độ, trong khi h thấp chỉ ra mẫu gần trung tâm, nơi dự đoán ổn định hơn. Khoảng LOD dựa trên bốn ý tưởng trong mô hình PLS:

(1) Mỗi mẫu thử có một giá trị LOD cụ thể. Trong hiệu chuẩn đơn biến, LOD thường là một giá trị cố định dựa trên độ nhạy và độ không chắc chắn của tín hiệu đo. Tuy nhiên, trong PLS, tín hiệu dụng cụ không đặc hiệu cho chất phân tích mà bị ảnh hưởng bởi các thành phần nền. Do đó, LOD không thể là một giá trị duy nhất mà thay đổi tùy theo thành phần nền của từng mẫu thử.

(2) Để tính giá trị LOD chính xác, các mẫu phân tích cần được phản ánh bởi tập hợp các mẫu hiệu chuẩn, tức là không gian hiệu chuẩn phải bao phủ được sự biến thiên của thành phần nền và tín hiệu dự kiến trong các mẫu phân tích. Điều này quan trọng vì nó đảm bảo rằng các đặc trưng của dữ liệu phân tích không vượt quá khả năng dự đoán của mô hình hiệu chuẩn.

(3) Các chỉ số ảnh hưởng (leverage) của các mẫu hiệu chuẩn có thể được ngoại suy về nồng độ chất phân tích bằng 0. Ý tưởng này đề xuất rằng bằng cách ngoại suy các chỉ số ảnh hưởng (leverage) của các mẫu hiệu chuẩn về điểm nồng độ chất phân tích bằng 0 (không có chất phân tích), ta có thể ước lượng tính chất của mô hình ở vùng thấp nhất của dải nồng độ. Điều này hữu ích để xác định LOD, vì LOD thường nằm gần ngưỡng không phát hiện được chất phân tích, và việc ngoại suy giúp dự đoán tính chất mô hình ở khu vực này mà không cần dữ liệu thực tế ở nồng độ 0.

(4) Khoảng giá trị LOD ($LOD_{min} - LOD_{max}$) có thể tính được dựa trên mô hình PLS. Giới hạn dưới và trên của khoảng LOD tương ứng với các mẫu hiệu chuẩn có độ nắm bắt ngoại suy thấp nhất và lớn nhất đối với nồng độ chất phân tích bằng không, được minh họa qua hình 2.2.



Hình 2.2. Biểu đồ minh họa giá trị LOD tối thiểu và tối đa, các quyết định liên quan đến sự có mặt và vắng mặt của chất phân tích trong các phạm vi nồng độ khác nhau. Khu vực màu xanh lam tương ứng với sai số loại I đối với LOD_{min} , khu vực màu đỏ ứng với sai số loại II đối với LOD_{max} (bản quyền của American Chemical Society – 2014) [175].

Để tính khoảng LOD, các bước được thực hiện như sau [109]:

Bước 1: Quang phổ hấp thụ của các mẫu hiệu chuẩn được đo, dẫn đến ma trận $A_{(J \text{ bước sóng} \times I \text{ mẫu})}$ và vector nồng độ $c_{(I \times 1)}$. Tiền xử lý dữ liệu bao gồm việc căn chỉnh trung bình dữ liệu A_m và c_m .

Bước 2: Triển khai thuật toán PLS để tìm số lượng biến tiềm ẩn tối ưu N , ma trận tải $P_{(J \times N)}$, ma trận trọng số tải $W_{(J \times N)}$, vector liên quan giữa ma trận điểm và nồng độ $v_{(N \times 1)}$, và ma trận điểm $T_{(I \times N)}$ [54].

Vector hệ số hồi quy b trong không gian bước sóng được tính theo công thức:

$$b = W(P^T W)^{-1} \cdot v \quad (44)$$

Bước 3. Độ nhạy được xác định theo công thức:

$$SEN = \frac{1}{\|b\|} \quad (45)$$

Trong đó, $\|b\|$ là chuẩn Euclid của một vector, $\|b\| = \sqrt{\sum_{j=1}^J b_j^2}$

Bước 4. Tính phương sai trong phổ

Từ bước hiệu chuẩn, phần dư phổ (spectral residuals) được tính:

$$A_{res} = P(W^T P)^{-1} W^T . A_m - A_m \quad (46)$$

Phương sai trong phổ:

$$var(A) = \sqrt{\frac{\sum_{j=1}^J \sum_{i=1}^I (A_{mji}^2)}{J \times I}} \quad (47)$$

Bước 5. Phương sai trong nồng độ (var(c)) được tính từ phần dư nồng độ (concentration residuals) trong quá trình thẩm định chéo (the cross-validation phase).

$$c_{res} = T . (v)^T - c_m \quad (48)$$

Tiếp theo, phương sai nồng độ được tính:

$$var(c) = \sqrt{\frac{\sum_{i=1}^I (c_{resi})^2}{I-1}} \quad (49)$$

Bước 6. Tính chỉ số ảnh hưởng (leverage) giả đơn biến cho một mẫu trắng:

$$h_{uni} = \frac{\bar{y}^2}{\sum_{i=1}^I y_{m_i}^2} \quad (50)$$

$\bar{y}$ là nồng độ trung bình trong tập hiệu chuẩn, y_m là giá trị nồng độ sau khi mean centering dữ liệu

Bước 7. Tính LOD_{min}:

Tính chỉ số ảnh hưởng (leverage) chiều tối thiểu

$$h_{omin} = h_{uni} + \frac{1}{I} \quad (51)$$

Và LOD_{min} được tính:

$$LOD_{min} = \sqrt{\frac{var(A)}{SEN^2} + \frac{h_{omin}.var(A)}{SEN^2} + var(c).h_{omin}} \quad (52)$$

Bước 8. Để tính LOD_{max}, đầu tiên cần tính chỉ số ảnh hưởng (leverage) của tất cả các mẫu hiệu chuẩn. Tính vector chỉ số ảnh hưởng (leverage) h cho mỗi mẫu

$$h_i = T(T^T T)^{-1}t \quad (53)$$

Trong đó, t là vector điểm của mỗi mẫu hiệu chuẩn có kích thước $(1 \times A)$. Tiếp theo, chỉ số ảnh hưởng (leverage) mẫu được tính:

$$h_{cal} = \sum_{i=1}^I (h_i)^2 \quad (54)$$

Tính chỉ số ảnh hưởng (leverage) của tất cả các mẫu hiệu chuẩn:

$$h_{ocal_i} = h_{cal} + h_{uni} \left(1 + \frac{c_{m_i}^2}{\bar{c}^2}\right) \quad (55)$$

Giá trị LOD_{max} được tính:

$$LOD_{max} = \sqrt{\frac{var(A)}{SEN^2} + \frac{h_{omax}.var(A)}{SEN^2} + var(c).h_{omax}} \quad (56)$$

Trong đó, $h_{max} = \max(h_{ocal_i}) + \frac{1}{I}$

Giới hạn phát hiện được chúng tôi xây dựng chương trình tính toán trên phần mềm R (phụ lục 1).

Khi khoảng LOD được xác định, nếu mẫu thử có nồng độ chất phân tích thấp hơn giá trị LOD_{min} tức là chất phân tích không được phát hiện, ngược lại, nếu nồng độ chất phân tích cao hơn giá trị LOD_{max} tức là chất phân tích có mặt. Trong trường hợp giá trị nồng độ nằm trong khoảng $LOD_{max} - LOD_{min}$, vấn đề được giải quyết bằng cách ước lượng một giá trị LOD riêng cho mẫu thử, bằng cách lấy giá trị chỉ số ảnh hưởng h của mẫu xem như gần bằng với chỉ số ảnh hưởng h_0 , tương ứng với thành phần nền của mẫu. Điều này tương đương với việc coi mẫu như một mẫu trắng vì trường hợp này nồng độ chất phân tích có khả năng rất thấp. Giá trị LOD thu được sau đó dùng để so sánh với nồng độ dự đoán, nếu thấp hơn thì quyết định là chất phân tích không có mặt, hoặc cao hơn thì quyết định chất phân tích có mặt.

Trên cơ sở tổng quan về các nghiên cứu liên quan đến tìm LOD trong phân tích đa biến và cách xác định LOD theo Alejandro C. Olivieri cho thấy rằng hiện nay vẫn chưa có sự thống nhất hoặc qui định chính thức về tính LOD trong phân các mô hình đa biến. Chúng tôi thấy rằng, sự giải thích, cách tính toán theo phương pháp của Alejandro C. Olivieri là hợp lý, chính vì vậy trong toàn bộ luận án giá trị LOD được tính theo phương pháp này.

2.2.5. Lựa chọn bước sóng bằng phương pháp tỉ lệ chọn lọc trong PLS

Tỉ lệ chọn lọc (Selectivity Ratio - SR) là một phương pháp lựa chọn biến trong hồi quy Partial Least Squares (PLS), được sử dụng để đo lường tầm quan trọng của từng biến đầu vào trong việc giải thích biến phản hồi. SR được tính toán bằng cách chia phần phương sai giải thích (explained variance) của từng bước sóng (hoặc biến) cho phần

phương sai không giải thích được (residual variance) của biến đó. Mục tiêu của SR là xác định những bước sóng hoặc biến có ảnh hưởng lớn nhất đến mô hình dự đoán và loại bỏ những biến không quan trọng [117, 176-178].

Các bước tiến hành lựa chọn biến bằng phương pháp SR-PLS như sau (hình 2.3) [178]:

Bước 1. Chuẩn bị dữ liệu gồm ma trận phổ hấp thụ A và vector nồng độ c.

Bước 2. Xây dựng mô hình PLS: Sử dụng phương pháp PLS để xây dựng mô hình giữa ma trận A và vector c. Xác định số biến tiềm ẩn (latent variables) dựa trên thẩm định chéo để tối ưu hóa mô hình.

Bước 3. Tính phương sai giải thích và không giải thích (residual variance) cho từng bước sóng: Sau khi có mô hình PLS, tính toán phần phương sai giải thích của từng bước sóng. Điều này bao gồm tính toán sự tương quan giữa các thành phần tiềm ẩn và biến phản hồi X. Để tính phương sai giải thích của biến X_j , ta cần tính toán phần tương quan giữa biến đầu vào X_j và thành phần tiềm ẩn T được sử dụng trong mô hình.

Công thức tổng quát:

$$V_{exp,j} = \frac{\sum_{i=1}^m (T_i P_j)^2}{\sum_{i=1}^m X_{ij}^2} \quad (57)$$

Trong đó: $V_{exp,j}$ là phương sai của biến X_j , T_i là giá trị của thành phần biến tiềm ẩn i; P_j là tải số (loading) của bước sóng X_j ; X_{ij} là giá trị của biến X_j tại mẫu i; m là số mẫu trong tập hiệu chuẩn. Phương sai giải thích là phần phương sai của biến X_j có thể được dự đoán bởi các thành phần tiềm ẩn của mô hình PLS.

Phương sai không giải thích được cho biến X_j là phần phương sai còn lại của biến đó sau khi đã giải thích phần phương sai thông qua các thành phần tiềm ẩn trong mô hình PLS. Đây là phần sai lệch giữa giá trị thực tế của biến X_j và giá trị mà mô hình dự đoán.

Công thức tính phương sai không giải thích được cho biến X_j :

$$V_{res,j} = \frac{\sum_{i=1}^m (X_{ij} - \hat{X}_{ij})^2}{\sum_{i=1}^m X_{ij}^2} \quad (58)$$

Trong đó, $V_{res,j}$ là phương sai không giải thích của biến X_j ; X_{ij} , $\hat{X}_{ij}$ là giá trị thực tế và dự đoán của biến X_j của mẫu i; m là số mẫu.

Bước 4. Tính SR

$$SR_j = \frac{V_{exp,j}}{V_{res,j}} \quad (59)$$

giá trị này càng cao càng phân biệt biến phổ giữa hai nhóm bước sóng.

Bước 5. Chọn các bước sóng quan trọng:

Việc lựa chọn ngưỡng của giá trị SR có ý nghĩa tốt cho mô hình PLS là rất quan trọng. Theo tác giả Olav M. Kvalheim [176, 178], đề xuất phương pháp xác định ngưỡng của SR có ý nghĩa dựa vào kiểm định Fisher (F-test). Để kết luận rằng biến có khả năng phân biệt cao, phương sai được giải thích trên thành phần TP (Target Projection) phải cao hơn đáng kể so với phương sai dư của một biến sau khi đã loại bỏ phương sai hệ thống được giải thích bởi thành phần TP. Điều này giống như việc đặt câu hỏi: đối với những biến nào mà việc giới thiệu thành phần TP giải thích đủ phương sai để nói rằng biến đó có khả năng phân biệt cao? Câu trả lời tất nhiên phụ thuộc vào kích thước mẫu N và mức xác suất α được chọn cho kiểm định F.

Để bác bỏ giả thuyết gốc rằng phương sai được giải thích và phương sai dư là như nhau, giá trị F tính toán, $F_{tính}$, được coi là bằng với chỉ số SR từ phương trình (3), phải lớn hơn giá trị F tới hạn của phân phối F, $F_{bảng}$:

$$F_{tính} = SR_j > F_{bảng} = F_{\alpha, n-2, n-3} \quad (60)$$

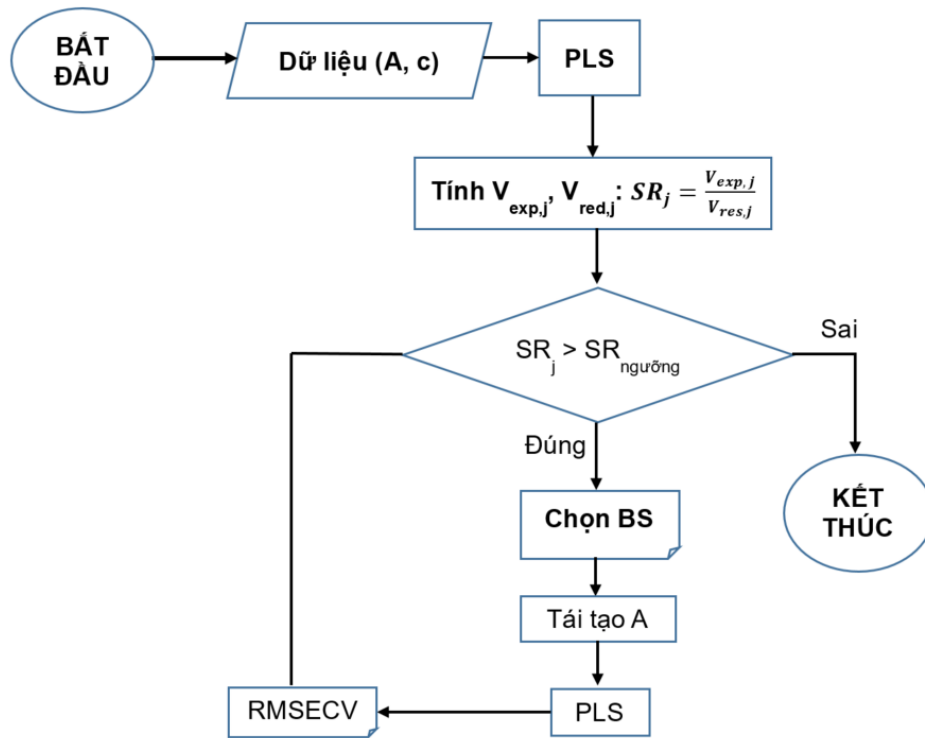
Trong đó, n là số mẫu, (n-2) và (n-3) tương ứng bậc tự do của tử số và mẫu số.

Trong nghiên cứu này, chúng tôi xác định ngưỡng tối ưu dựa trên xây dựng vòng lặp cho các giá trị SR khác nhau, từ thấp đến cao. Quá trình bắt đầu bằng việc lọc các biến có giá trị SR cao hơn ngưỡng đang xét. Những biến này được xem là có tiềm năng đóng góp đáng kể vào khả năng phân biệt trong mô hình PLS. Sau đó, một mô hình PLS được xây dựng bằng cách sử dụng các biến đã chọn, và số thành phần chính tối ưu (n_{comp}) được xác định thông qua phương pháp đã lựa chọn trước đó. Việc tính toán RMSECV được thực hiện để đánh giá hiệu quả dự đoán của mô hình tại ngưỡng SR đang xét. Giá trị RMSECV này phản ánh mức độ phù hợp của các biến được chọn với mục tiêu tối đa hóa độ chính xác của mô hình.

Trong mỗi vòng lặp, giá trị RMSECV nhỏ nhất được theo dõi và so sánh. Nếu RMSECV tại một ngưỡng SR cụ thể thấp hơn giá trị tối thiểu hiện tại, ngưỡng SR này được xem là tối ưu. Các bước sóng tương ứng với ngưỡng tối ưu cũng được lưu trữ để phục vụ phân tích và diễn giải sau cùng. Quá trình tiếp tục cho đến khi tất cả các ngưỡng SR được duyệt qua.

Kết quả cuối cùng bao gồm ngưỡng SR tối ưu, giá trị RMSECV nhỏ nhất và danh sách các bước sóng tương ứng. Phương pháp này cung cấp một cách tiếp cận định lượng để xác định ngưỡng SR, giúp cân bằng giữa việc lựa chọn các biến có ý nghĩa và giảm

thiểu tác động của các biến nhiễu. Kỹ thuật này có thể được mở rộng để áp dụng cho các dữ liệu phổ phức tạp, nơi mà việc lựa chọn biến đóng vai trò quan trọng trong việc xây dựng mô hình dự đoán đáng tin cậy. Việc loại bỏ các biến có giá trị SR thấp giúp giảm bớt số lượng biến đầu vào, cải thiện khả năng khái quát hóa của mô hình và đảm bảo rằng các biến được chọn thực sự đóng góp vào khả năng dự đoán của mô hình hồi quy.



Hình 2.3 Sơ đồ thuật toán SR-PLS

2.2.6. Lựa chọn bước sóng bằng phương pháp loại bỏ biến ngược trong PLS

Phương pháp loại bỏ biến ngược trong PLS (BVE-PLS) được áp dụng để lựa chọn các bước sóng quan trọng từ dữ liệu phổ UV-Vis, kết hợp với mô hình Hồi quy bình phương tối thiểu riêng phần (PLS) [116, 118, 120, 177, 179, 180]. Quy trình thực hiện gồm các bước sau (hình 2.4):

Bước 1. Chuẩn bị dữ liệu và thiết lập mô hình ban đầu

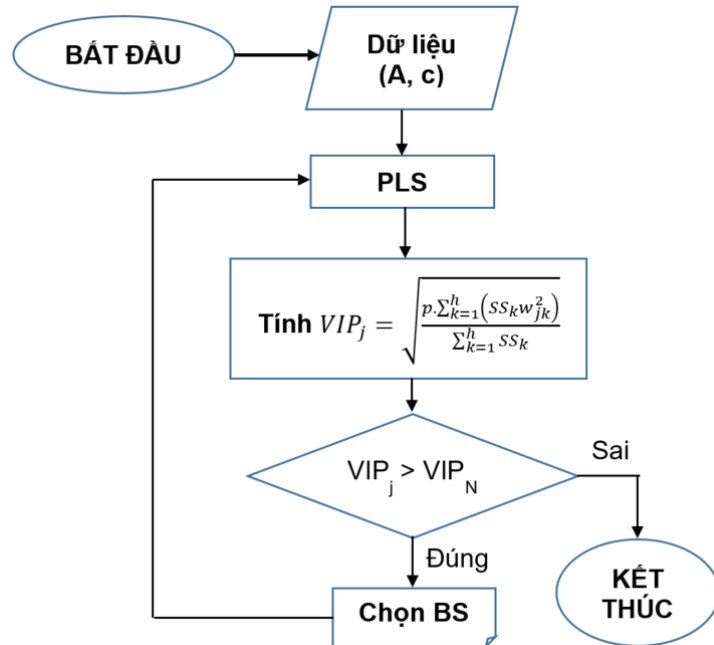
Dữ liệu đầu vào: Ma trận phổ X (kích thước $n \times p$, với n là số mẫu và p là số bước sóng) và vector nồng độ/chỉ số phân loại y .

Phân chia dữ liệu: Chia ngẫu nhiên dữ liệu thành tập huấn luyện (75%) và tập kiểm tra (25%).

Bước 2. Xây dựng mô hình PLS và xác định số thành phần tối ưu

Huấn luyện mô hình PLS: Sử dụng tập huấn luyện để xây dựng mô hình PLS với số thành phần tối đa 10.

Lựa chọn số thành phần tối ưu: Chọn số thành phần có giá trị PRESS (Predicted Residual Error Sum of Squares) nhỏ nhất qua kiểm định chéo (LOO-CV).



Hình 2.4 Sơ đồ thuật toán BVE-PLS

Bước 3. Tính tầm quan trọng của biến (VIP – Variable Important Project)

$$VIP_j = \sqrt{\frac{p \cdot \sum_{k=1}^h (SS_k w_{jk}^2)}{\sum_{k=1}^h SS_k}} \quad (61)$$

Trong đó, p: số biến (bước sóng) ban đầu; h: số thành phần chính PLS được chọn; SS_k : phương sai giải thích bởi thành phần thứ k; w_{jk} : trọng số của biến j ở thành phần k.

Ngưỡng của VIP: loại bỏ các bước sóng có $VIP < \text{ngưỡng}$, chỉ giữ lại biến có đóng góp đáng kể đến mô hình

Bước 4. Quy trình loại bỏ biến lặp lại:

- Khởi tạo: tất cả các bước sóng được xem xét ban đầu
- Lặp lại các bước sau:

1. Xây dựng mô hình PLS trên tập biến hiện tại
2. Tính VIP cho từng biến, xác định biến có $VIP < \text{ngưỡng}$

3. Loại bỏ 20% biến có VIP thấp nhất (hoặc theo tỉ lệ xác định)

4. Đánh giá hiệu suất của mô hình trên tập kiểm tra: tính RMSEP

5. Dừng khi: Số biến còn lại $\leq n_{\text{com}} + 1$ hoặc hiệu suất mô hình không cải thiện qua các lần lặp lại

Bước 5. Lựa chọn tập biến (bước sóng) tối ưu: tập biến (bước sóng) cuối cùng là tập có RMSEP thấp nhất

Ưu điểm của phương pháp BVE-PLS: loại bỏ các bước sóng không liên quan, tập trung vào vùng phổ quan trọng, cải thiện độ chính xác của mô hình PLS nhờ tối ưu hóa số thành phần PLS và biến đầu vào; Việc sử dụng phương pháp này trong các nghiên cứu hóa học phân tích giúp cải thiện mô hình dự đoán đồng thời giảm số lượng biến giúp tiết kiệm chi phí và thời gian phân tích. Tuy nhiên, do phương pháp BVE khi thực hiện loại bỏ biến ngược tức là những biến (bước sóng) bị loại bỏ thì không được đưa lại vào trong mô hình tiếp theo, chính điều này có thể làm loại bỏ những biến mang thông tin quan trọng.

2.2.7. Lựa chọn bước sóng bằng phương pháp thuật toán di truyền trong PLS

Thuật toán di truyền là các thuật toán nhằm mục đích tìm kiếm giải pháp tối ưu dựa trên khái niệm về chọn lọc tự nhiên và di truyền học. Chính vì vậy, các khái niệm trong thuật toán di truyền rất gần với các khái niệm trong sinh học như quần thể, cá thể (nhầm sắc thể), gen, chọn lọc, lai ghép, đột biến [70, 181-184].

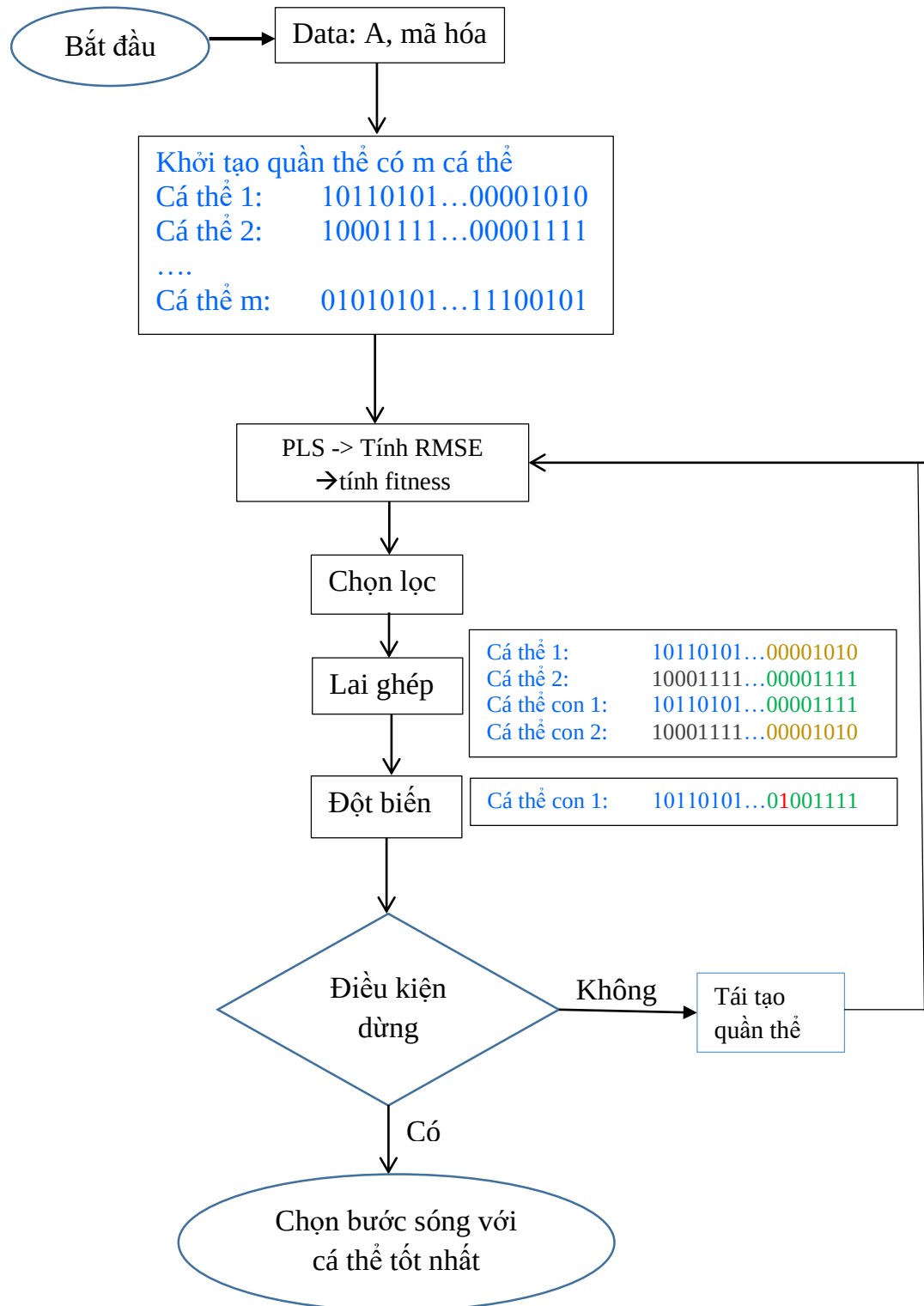
Bước 1. Khởi tạo một quần thể gồm n cá thể được chọn ngẫu nhiên từ tập dữ liệu bước sóng trong tập hiệu chuẩn ban đầu, mỗi cá thể là một tập hợp ngẫu nhiên có các bước sóng khác nhau, mỗi bước sóng được gọi là 1 gen. Để thực hiện thuật toán GA, các gen được mã hóa thành các bit nhị phân 0 – 1, trong đó 0 tương ứng với bước sóng không được chọn, 1 tương ứng với bước sóng được chọn. Số lượng gen trong mỗi cá thể bằng số lượng bước sóng trong tập dữ liệu ban đầu.

Bước 2. Tính giá trị RMSE của mỗi cá thể bởi hàm mục tiêu PLS. Giá trị thích nghi (fitness) được tính như sau:

$$fitness = \frac{1}{RMSE} \quad (62)$$

Bước 3. Chọn lọc: những cá thể có giá trị thích nghi lớn là những cá thể chứa những bước sóng tối ưu nhất, những cá thể này có khả năng giao phối cao hơn để tạo ra các cá thể con phù hợp.

Bước 4. Lai ghép: Thực hiện lai giữa các cá thể bố mẹ, có nhiều kiểu lai khác nhau. Từ mỗi cặp “bố-mẹ”, hai cá thể mới (con) được tạo ra bằng cách gán ngẫu nhiên các gen từ một trong hai bố mẹ cho cá thể con.



Hình 2.5. Sơ đồ thuật toán GA-PLS

Bước 5. Đột biến: Đột biến trên một số cá thể con ở một vị trí nào đó trên 1 cá thể (1 nhiễm sắc thể) với một tỉ lệ đột biến phù hợp (thông thường từ 2-5%), cách thực hiện đột biến là thay thế 0 → 1 hoặc 1 → 0.

Bước 6. Tính giá trị thích nghi của các cá thể con sinh ra. Sau đó lặp lại bước 3 cho đến khi thỏa mãn điều kiện dừng

Điều kiện dừng: khi giá trị thích nghi trong quần thể hầu như không thay đổi sau n lần lặp lại, hoặc dừng sau khi thực hiện vòng lặp n lần.

Các bước sóng tương ứng với cá thể có giá trị thích nghi tốt nhất được lựa chọn để xây dựng các mô hình tiếp theo. Hình 2.3 minh họa sơ đồ vòng lặp của thuật toán di truyền.

2.3 THỰC NGHIỆM

2.3.1. Thiết bị, hóa chất

2.3.1.1. Thiết bị

Máy quang phổ hai chùm tia UV – Vis UH5300 (Hitachi – Nhật), khoảng quét bước sóng 190 - 1100 nm, kết nối với máy tính để chứa dữ liệu phổ thu được. Cuvet thạch anh hãng Labomed – Mỹ: Kích thước 12,5 x 12,5 x 45 mm.

Cân phân tích 4 chữ số thập phân Ohaus Mỹ; máy cất nước 2 lần Bibby Scientific A4000D – Anh; bể rung siêu âm Elma 18L – Đức; tủ sấy Memmert – Đức. Micropipet Boeco – Đức các loại: 10 – 100 μ L; 20 – 200 μ L; 100 – 1000 μ L.

2.3.1.2. Hóa chất

Dùng môi các loại: Methanol (Fisher), ống chuẩn HCl 0,1N (Công ty TNHH Hóa chất và vật tư Khoa học kỹ thuật Thanh phố Hồ Chí Minh), nước cất hai lần.

Các chất chuẩn:

- Chất chuẩn Amlodipin Besylat (100,3%- tính trên nguyên trạng) và Indapamid (97,7% - tính trên nguyên trạng) được mua từ Viện Kiểm nghiệm thuốc Thành phố Hồ Chí Minh;

- Chất chuẩn Amlodipine besilate (100,3%) và Atorvastatin calcium (94,6%) được mua từ Viện Kiểm nghiệm thuốc Thành phố Hồ Chí Minh;

- Chất chuẩn Spironolacton (99,6%) và Furosemid (94,6%) được mua từ Viện Kiểm nghiệm thuốc Thành phố Hồ Chí Minh;

- Chất chuẩn Rifampicin (96,0%) , Isoniazid (99,63%), Pyrazinamid (100,00%) được mua từ Viện kiểm nghiệm thuốc trung ương, Hà Nội;

- Chất chuẩn Metronidazol (99,9%), Spiramycin (4220 IU/mg) được mua từ Viện Kiểm nghiệm thuốc Thành phố Hồ Chí Minh;

Các thuốc đa thành phần:

- Viên nén AgimDogyl chứa 125 mg Metronidazol và 750.000 IU Spiramycin được sản xuất bởi công ty cổ phần dược phẩm Agimmexpharm, số lô No. 211222, hạn dùng ngày 30/12/2025;

- Viên nén Natrixam 1,5mg/5mg chứa 1,5 mg Indapamid và 5 mg Amlodipin. Được sản xuất bởi Công ty TNHH Servier (Việt Nam), số lô: 6075868, ngày sản xuất: 12/2022; Hạn dùng: 12/2024;

- Viên nén Zoamco-A với thành phần mỗi viên chứa 5 mg Amlodipin và 10 mg Atorvastatin được sản xuất bởi Công ty cổ phần Pymepharco, số lô 201220, ngày sản xuất 16/12/2020, hạn dùng 16/12/2023;

- Viên nén Spiromide-40 với mỗi viên chứa 50 mg spironolactone và 40 mg Furosemid được sản xuất bởi công ty Hapharco J.S.C, số lô C0033, ngày sản xuất 6/12/2021, hạn dùng 5/12/2023;

- Viên nén Turbezid với mỗi viên chứa 150 mg Rifampicin, 75 mg Isoniazid và 400 mg Pyrazynamid được sản xuất bởi công ty Dược Nam Hà, số lô 23121, ngày sản xuất 20/8/2023, hạn dùng 20/8/2027.

2.3.2. Chuẩn bị mẫu

2.3.2.1 Chuẩn bị mẫu chuẩn

Chuẩn bị mẫu chuẩn: mẫu chuẩn được chuẩn bị bằng cách cân một lượng chính xác 0,0500 (g) chất chuẩn bằng cân phân tích điện tử 4 chữ số thập phân cho vào bình định mức 50 mL thêm MeOH vừa đủ, siêu âm trong 5 phút cho chất chuẩn hòa tan hoàn toàn, ta được các dung dịch chuẩn gốc có nồng độ 1000 $\mu\text{g/mL}$, sau đó bảo quản trong tủ lạnh [129].

Các mẫu chuẩn trong tập hiệu chuẩn và tập kiểm tra được pha trực tiếp từ các dung dịch chuẩn gốc theo các tỉ lệ được tính toán theo phần thiết kế thí nghiệm. Các dung dịch chuẩn trước khi đem đi pha được lấy ra khỏi tủ lạnh để cân bằng đến nhiệt độ phòng (25°C).

2.3.2.2 Chuẩn bị mẫu thuốc

Mẫu thuốc: cân chính xác 20 viên, tính khối lượng trung bình của mỗi viên. Nghiền mịn thuốc, cân một lượng chính xác m g bột thuốc (tương ứng với hàm lượng phù hợp với mỗi loại thuốc) cho vào bình định mức 50 mL, thêm vừa đủ dung môi để hòa tan thuốc, siêu âm 5 phút, lọc lấy phần dịch lọc bằng màng lọc 0,45 μm . Hút lượng chính xác V mL dịch lọc vào bình định mức 10 mL, thêm dung môi vừa đủ, lấy dung dịch quét phổ hấp thụ UV-Vis [129].

Hàm lượng hoạt chất X trong một viên thuốc được tính bởi công thức sau:

$$m_{X/viên} = \frac{C \times 10}{1000} \times \frac{50}{V} \times \frac{m_{TB\ 1\ viên}}{m_{cân}} \quad (63)$$

Trong đó, C là nồng độ của chất X, µg/mL; V là thể tích dịch lọc, mL; $m_{TB\ 1\ viên}$ là khối lượng trung bình của 1 viên thuốc và $m_{cân}$ khối lượng thuốc bột được cân.

Bảng 2.2. Các thông số chuẩn bị mẫu thuốc

Thuốc	$m_{20\ viên\ (g)}$	$m_{TB\ 1\ viên\ (g)}$	$m_{cân\ (g)}$	V (µL)
AgimDogyl	12,4227	0,6211	0,6211	60
Spiromid	4,7964	0,2398	0,1199	200
Zoamco-A	5,6274	0,2814	0,2814	1000
Natrixam	8,4331	0,4216	0,6325	1000
Turbezid	18,9724	0,9486	0,4743	60

2.3.3 Đánh giá độ đúng

Độ đúng của phương pháp là khái niệm chỉ mức độ gần nhau giữa giá trị trung bình của kết quả thử nghiệm và giá trị thực hoặc giá trị được chấp nhận là đúng. Độ đúng được thực hiện bởi một trong các cách, hoặc so sánh kết quả thử nghiệm với kết quả thực hiện bởi một phương pháp đối chiếu hoặc sử dụng mẫu đã biết nồng độ (mẫu kiểm tra hoặc mẫu chuẩn được chứng nhận hoặc phương pháp đánh giá độ thu hồi

Đối với phương pháp so sánh với phương pháp chuẩn, mẫu phân tích được thực hiện bằng phương pháp khảo sát và phương pháp đối chiếu của các tổ chức có uy tín. Tính kết quả trung bình và độ lặp lại của hai phương pháp. So sánh kết quả độ lặp lại bằng tiêu chuẩn F, so sánh hai giá trị trung bình bằng tiêu chuẩn t.

Đối với phương pháp sử dụng vật liệu so sánh được cấp chứng chỉ (mẫu chuẩn), mẫu chuẩn được xác định lại nồng độ bằng phương pháp khảo sát, sau đó nồng độ được so sánh với nồng độ của mẫu chuẩn đã được công bố.

Đối với phương pháp xác định độ thu hồi, cách thực hiện bằng cách thêm một lượng chất chuẩn xác định vào mẫu thử hoặc mẫu trắng, phân tích các mẫu thêm chuẩn đó, tính độ thu hồi theo công thức:

$$Rev(\%) = \frac{C_t \times 100}{C_t^0} \quad (64)$$

Trong đó, C_t^0 là nồng độ chuẩn thêm theo lý thuyết, C_t là nồng độ chất phân tích tính được.

Chất chuẩn thêm vào theo 3 mức nồng độ: mức thấp (80-85%), mức trung bình (100%) và mức cao (115 – 120%) so với nồng độ xác định.

Tiêu chí đánh giá: Độ thu hồi được chấp nhận ở các mức nồng độ khác nhau theo qui định của AOAC, trong đó với ngưỡng mức nồng độ 10 ppm độ thu hồi chấp nhận nằm trong khoảng 80 – 110%, ngưỡng mức nồng độ 100 ppm độ thu hồi chấp nhận trong khoảng 90 – 107% [185, 186].

2.3.4. Đánh giá độ lặp lại

Độ lặp lại được thực hiện tương tự như phần chuẩn bị mẫu để định lượng nhưng được thực hiện lặp lại 6 lần và được thực hiện trong 2 ngày khác nhau, trong đó ngày thứ 2 được thực hiện bởi một người khác. Nồng độ của các thành phần trong các mẫu lặp lại được qui về tương ứng với khối lượng cân dự kiến.

$$C = C_t \cdot \frac{m_t}{m_{tt}} \quad (65)$$

Trong đó, C là nồng độ được qui về khối lượng cân dự kiến; C_t là nồng độ tính được thực tế; m_t là khối lượng cân của mẫu theo dự kiến; m_{tt} là khối lượng cân mẫu thực tế.

Độ chính xác được đánh giá thông qua độ lệch chuẩn (SD) [187, 188]:

$$SD = \sqrt{\sum_{i=1}^n (C_i - \bar{C})^2 / (n - 1)} \quad (66)$$

Và độ lệch chuẩn tương đối (RSD) hoặc hệ số biến thiên (Coefficient of variation – CV):

$$RSD = \frac{100 \times SD}{\bar{C}} \quad (67)$$

Trong đó, C_i và $\bar{C}$ là nồng độ tính được lần thứ i và nồng độ trung bình của các lần thử nghiệm.

Tiêu chí đánh giá: dựa vào giá trị RSD tính được, so sánh với độ lặp lại tối đa chấp nhận tại các nồng độ khác nhau theo qui định của AOAC, trong đó với ngưỡng nồng độ 10 ppm giá trị RSD là 7,3; ngưỡng nồng độ 100 ppm giá trị RSD là 5,3 [185, 186].

2.4. PHƯƠNG PHÁP ĐÁNH GIÁ THỐNG KÊ

2.4.1. Khoảng tin cậy

$$\mu = \bar{x} \pm \frac{t \cdot s}{\sqrt{n}} \quad (68)$$

Trong đó, t là chuẩn Student, phụ thuộc vào độ tin cậy và số dữ liệu thực nghiệm

2.4.2. So sánh độ lặp lại giữa hai dãy thí nghiệm khác nhau (chuẩn F – Fisher) [188]

Bước 1. Tính phương sai của hai dãy thí nghiệm theo công thức

$$S_n^2 = \sum_{i=1}^n (C_i - \bar{C})^2 / (n - 1) \quad (69)$$

Bước 2. Tính F_{tn} : $F_{tn} = \frac{s_1^2}{s_2^2}$; trong đó: $s_1^2 > s_2^2$

Bước 3. So sánh F_{tn} với $F_{(\alpha, n_1, n_2)}$, trong đó, $F_{(\alpha, n_1, n_2)}$ là giá trị F tra bảng với xác suất α và bậc tự do $n_1 - 1$ (tử số), $n_2 - 1$ (mẫu số).

+ Nếu $F_{tn} < F_{(\alpha, n_1, n_2)}$: độ lệch chuẩn của 2 dãy số liệu không khác nhau

+ Nếu $F_{tn} > F_{(\alpha, n_1, n_2)}$: độ lệch chuẩn của 2 dãy số liệu khác nhau đáng kể.

2.4.3. So sánh giá trị trung bình của 2 dãy thí nghiệm [188]

Trường hợp 1, nếu độ lệch chuẩn của 2 dãy thí nghiệm không khác nhau, tính độ lệch chuẩn chung:

$$s_d = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ và } \varepsilon = t. s_d \quad (70)$$

Trong đó t được tra từ bảng phân phối chuẩn t với mức tin cậy α và bậc tự do $n_1 + n_2 - 2$.

So sánh ε với $|\bar{x}_1 - \bar{x}_2|$.

Trường hợp 2, nếu độ lệch chuẩn của 2 dãy thí nghiệm khác nhau đáng kể:

- Tính: $v_1 = \frac{s_1^2}{n_1}$ và $v_2 = \frac{s_2^2}{n_2}$

- Tính giá trị T: $T = \frac{v_1 t_1 + v_2 t_2}{\sqrt{v_1 + v_2}}$; trong đó t_1, t_2 được tra từ bảng.

- So sánh giá trị $|\bar{x}_1 - \bar{x}_2|$ với T

2.4.4. Đánh giá phân phối chuẩn của tập dữ liệu

Kiểm định Shapiro-Wilk là một trong những phương pháp phổ biến để kiểm tra tính chuẩn của dữ liệu. Dưới đây là các bước thực hiện kiểm định phân phối chuẩn Shapiro-Wilk :

Bước 1. Thu thập dữ liệu: một bộ dữ liệu mẫu $(x_1, x_2, \dots, x_n)$ với n ($3 < n < 5000$, nhưng hiệu quả nhất với $n < 50$) quan sát thực tế. Dữ liệu này phải là một mẫu ngẫu nhiên, và các quan sát trong mẫu phải độc lập với nhau.

Bước 2. Xác định giả thuyết: Giả thuyết không (H_0): Dữ liệu có phân phối chuẩn (phân phối chuẩn $N(\mu, \sigma^2)$); Giả thuyết thay thế (H_1): Dữ liệu không có phân phối chuẩn.

Bước 3. Tính các giá trị cần thiết:

Xếp thứ tự các giá trị trong mẫu: Sắp xếp các giá trị $(x_1, x_2, \dots, x_n)$ theo thứ tự tăng dần

Giá trị của thống kê Shapiro-Wilk:

$$W = \frac{(\sum_{i=1}^n a_i x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (71)$$

Trong đó, a_i là Hằng số Shapiro-Wilk được lấy từ bảng (phụ lục 4)

Giá trị W càng gần 1, dữ liệu càng phân phối chuẩn [189, 190].

Tiêu chí đánh giá: So sánh giá trị W với giá trị giới hạn từ bảng phân phối W (phụ lục 2) hoặc so sánh giá trị p với mức ý nghĩa α ; Nếu $p > \alpha$, không có đủ chứng cứ để bác bỏ giả thuyết không (H_0), nghĩa là dữ liệu có phân phối chuẩn; Nếu giá trị $p < \alpha$, bác bỏ giả thuyết không (H_0) và kết luận rằng dữ liệu không phân phối chuẩn.

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

Nghiên cứu trong luận án được thực hiện trên 5 loại thuốc khác nhau bao gồm Spirosemid-40 (SPR : FUR = 5:4), Zoamco-A (AML : ATO = 1:2), Natrixam (IND : AML = 3:10), AgimDogyl (MTZ : SRM = 2:3), Turbezid (RIF : INZ : PYR = 6 : 3 : 16). Thiết kế các mẫu trong tập hiệu chỉnh sao theo CCD sao cho mẫu tâm có tỉ lệ tương ứng với tỉ lệ của các thuốc, các mẫu ở các điểm giai thừa và điểm trục có tổng độ hấp thụ của các thành phần ở trong khoảng bước sóng khảo sát không được vượt quá ngưỡng đo của máy (trong khoảng 2,5 – 3). Kết quả các mẫu trong tập hiệu chuẩn được thể hiện qua bảng 3.1.

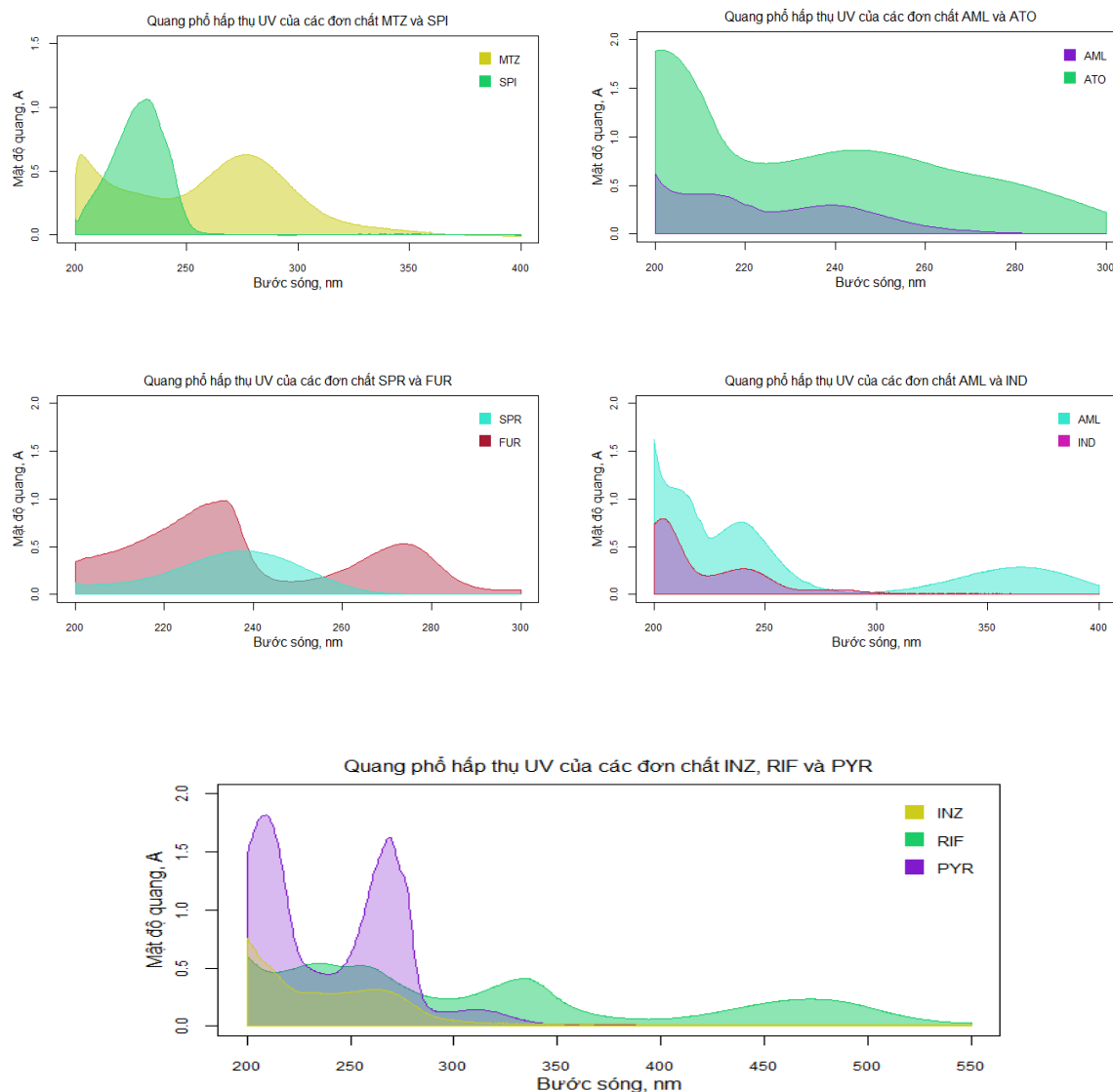
Bảng 3.1. Nồng độ của các chất trong tập hiệu chuẩn của các mẫu thuốc

STT	Spirosemid		Zoamco-A		Natrixam		AmgimDogyl		Turbezid		
	SPR	FUR	AML	ATO	IND	AML	MTZ	SPI*	RIF	INZ	PYR
T1	6,0	4,0	8,0	16,0	3,0	10,0	10,0	60,0	7,5	7,5	20,0
T2	14,0	4,0	12,0	16,0	6,0	10,0	20,0	60,0	10,5	7,5	20,0
T3	6,0	12,0	8,0	24,0	3,0	40,0	10,0	120,0	7,5	10,5	20,0
T4	14,0	12,0	12,0	24,0	6,0	40,0	20,0	120,0	10,5	10,5	20,0
T5	4,4	8,0	7,2	20,0	2,4	25,0	8,0	90,0	7,5	7,5	28,0
T6	15,6	8,0	12,8	20,0	6,6	25,0	22,0	90,0	10,5	7,5	28,0
T7	10,0	2,4	10,0	14,4	4,5	4,0	15,0	48,0	7,5	10,5	28,0
T8	10,0	13,6	10,0	25,5	4,5	46,0	15,0	132,0	10,5	10,5	28,0
T9	10,0	8,0	10,0	20,0	4,5	25,0	15,0	90,0	6,5	9,0	24,0
T10					4,5	15,0			11,5	9,0	24,0
T11					4,5	30,0			9,0	6,5	24,0
T12									9,0	11,5	24,0
T13									9,0	9,0	17,3
T14									9,0	9,0	30,7
T15									9,0	9,0	24,0

Lưu ý: nồng độ của các thành phần có đơn vị $\mu\text{g/mL}$; riêng (*) đơn vị IU/mL.

Chúng tôi đã khảo sát các phổ đơn chất của các chất có trong các mẫu phân tích. Các dung dịch chuẩn đơn của các chất được chuẩn bị có nồng độ tương ứng với nồng độ của chất ở mẫu tâm của thiết kế thí nghiệm (màu đỏ). Hình 3.1 thể hiện phổ hấp thụ của các đơn chất trong các thuốc lần lượt là AgimDogyl (gồm Metronidazol và Spiramycin), Zoamco-A (Amlodipin và Atorvastatin), Spiromide-40 (Spironolacton và Furosemid), Natrixam 1,5/5 (Amlodipin và Indapamid) và Turbezid (Isoniazid, Rifampicin và Pyrazinamid). Hầu hết phổ các chất đã khảo sát đều có sự chồng phổ nhau rất lớn. Đối với Metronidazol và Spiramycin, từ khoảng bước sóng 200 -250 nm cả hai chất MTZ và SPI đều hấp thụ ánh sáng trong khoảng bước sóng này. Đối với AML và ATO đều hấp thụ ánh sáng trong khoảng bước sóng ánh sáng khảo sát, từ 200 – 300 nm. Điều tương tự cũng xảy ra đối với các cặp chất còn lại trong các mẫu thuốc.

Sự chồng lấn phổ này làm ảnh hưởng đến phổ chung của hỗn hợp khi thực hiện phân tích đồng thời bằng phương pháp phân tích phổ hấp thụ UV-Vis. Các hiệu ứng tương tác lưỡng cực, khả năng tạo liên kết Hydro, hiệu ứng tương tác giữa các phân tử... làm ảnh hưởng đến khả năng hấp thụ của các phân tử riêng lẻ, chính các tác động này làm cho tính cộng tính không còn được đảm bảo, làm cho quá trình phân tích theo phương pháp đơn biến có sai số lớn, không chính xác.



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

3.2 KẾT QUẢ PHƯƠNG PHÁP PLS KẾT HỢP PHỔ HẤP THỤ UV-VIS

3.2.1 Kết quả thuật toán PLS

Phương pháp PLS là một quá trình biến đổi các phép toán nhằm khái quát hóa các dữ liệu gốc lên trên một không gian mới. Dữ liệu thu thập được bao gồm tập dữ liệu ma trận nồng độ và ma trận độ hấp thụ quang phổ UV-Vis được gọi là dữ liệu thực, các dữ liệu được thu thập nhờ quá trình biến đổi toán học được gọi là biến tiềm ẩn. Các biến

này được gọi là tiềm ẩn bởi vì chúng không thể đo trực tiếp mà phải nhờ các phép toán mới có được chúng.

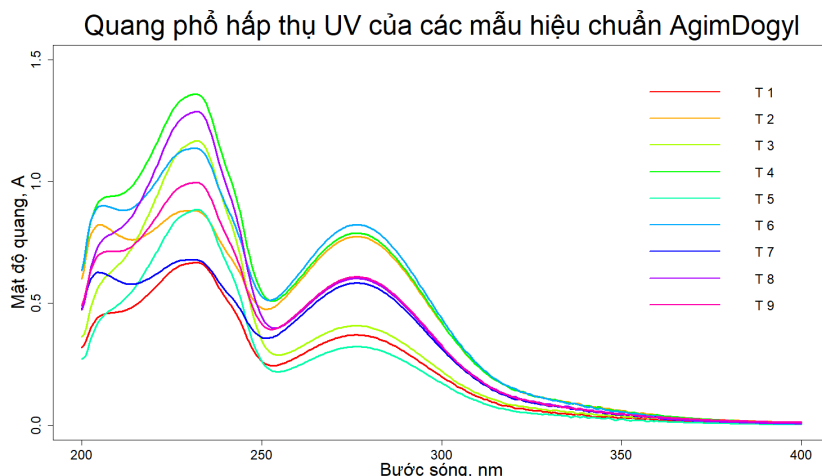
Mối liên hệ giữa các biến thực và biến tiềm ẩn được thực hiện thông qua một tập hợp các vector được gọi là loadings (vector tải), cung cấp sự kết nối giữa dữ liệu thực và tiềm ẩn hay nói một cách khác, vector tải là các công cụ cho phép nén thông tin, thu gọn thông tin gốc bằng cách chiếu thông tin lên không gian mà chúng đại diện hoặc giải nén các biến tiềm ẩn để tái tạo lại thông tin gốc. Tương ứng với mỗi vector tải là một giá trị điểm (score), đây là giá trị tương ứng với biến tiềm ẩn trên không gian mới, đại diện cho tập dữ liệu mới phục vụ cho quá trình xây dựng mô hình hiệu chuẩn, hồi quy.

Các tham số sau khi thực hiện thuật toán PLS bao gồm: ma trận P , W , T và vector q và ma trận chiếu $P_{proj} = W.(P^T.W)^{-1}$ biểu diễn mối quan hệ giữa dữ liệu gốc và dữ liệu điểm T trong không gian mới, P_{proj} khi nhân với dữ liệu gốc ta được ma trận điểm T (bảng 3.2). Các tham số quan trọng bao gồm ma trận dữ liệu điểm T , ma trận chiếu P_{proj} và ma trận hệ số tương quan sẽ được làm rõ trong các phần tiếp theo.

Bảng 3.2. Các tham số được trích xuất trong mô hình PLS

Tham số	Tên đầy đủ	Ý nghĩa
P	Hệ số tải của phổ	Là ma trận hệ số hồi qui liên kết T với dữ liệu gốc, tức là dùng để tái tạo dữ liệu phổ từ điểm số T .
W	Hệ số trọng số tải	Là ma trận trọng số liên kết giữa không gian biến độc lập (phổ) và không gian biến tiềm ẩn, phản ánh mức độ đóng góp của từng biến phổ vào thành phần tiềm ẩn, giúp chọn hướng tối ưu trong không gian biến tiềm ẩn
T	Điểm số phổ	Là ma trận điểm số đại diện cho các mẫu trong không gian biến ẩn, là sự chiếu dữ liệu phổ xuống các thành phần tiềm ẩn. T chứa thông tin chính của phổ đã giảm chiều.
U	Điểm số nồng độ	Là ma trận điểm số của biến phụ thuộc (nồng độ), đại diện cho thông tin nồng độ được chiếu xuống không gian ẩn, giúp kết nối giữa phổ và nồng độ. U tối ưu hóa sự tương quan với T trong PLS
Q	tải của nồng độ	Là ma trận hệ số hồi qui liên kết giữa U với dữ liệu nồng độ, giúp tái tạo nồng độ từ U .
β	Hệ số hồi quy	Dự đoán nồng độ C từ ma trận phổ
P_{proj}	ma trận chiếu	Là ma trận chiếu được sử dụng để ánh xạ dữ liệu mới (phổ mẫu thử) vào không gian điểm số T của mô hình huấn luyện để tính điểm số T mới, từ đó tính nồng độ dự đoán.

Chúng tôi áp dụng trên dữ liệu của tập thuốc AgimDogyl để tính toán và làm rõ các thông số trên. Để tạo mô hình hiệu chuẩn PLS chúng tôi chuẩn bị tập hiệu chuẩn gồm 9 mẫu chứa 2 thành phần là MTZ và SPI, bảng 3.1. Phổ hấp thụ UV-Vis của các mẫu hiệu chuẩn được quét trong khoảng bước sóng từ 200 – 400 nm, khoảng quét là 0,5 nm; mỗi mẫu được pha lặp lại 3 lần. Như vậy, dữ liệu để xây dựng tập hiệu chuẩn gồm tập nồng độ **C** với kích thước ma trận 2 thành phần $\times$ 27 mẫu và tập dữ liệu phổ hấp thụ **A** với kích thước ma trận 401 bước sóng $\times$ 27 mẫu. Hình 4.2 thể hiện phổ hấp thụ của các mẫu trong tập hiệu chuẩn trong khoảng bước sóng từ 200 – 400 nm.



Hình 3.2. Phổ hấp thụ UV của hỗn hợp MTZ và SPI trong các mẫu hiệu chuẩn

Với thuật toán đã được trình bày trong mục 1.3.4, kết quả từ tập dữ liệu phổ hấp thụ A và nồng độ C đã phân tích thành ma trận điểm T (kích thước 27×25 , trong đó 25 là số thành phần chính tối đa mà mô hình PLS đã thu gọn), ma trận tải P (kích thước 401×25), ma trận tải trọng W (kích thước 401×25), vector hệ số hồi qui giữa nồng độ và ma trận điểm q (kích thước) và hệ số hồi qui β giữa nồng độ và phổ hấp thụ (kích thước 401×25).

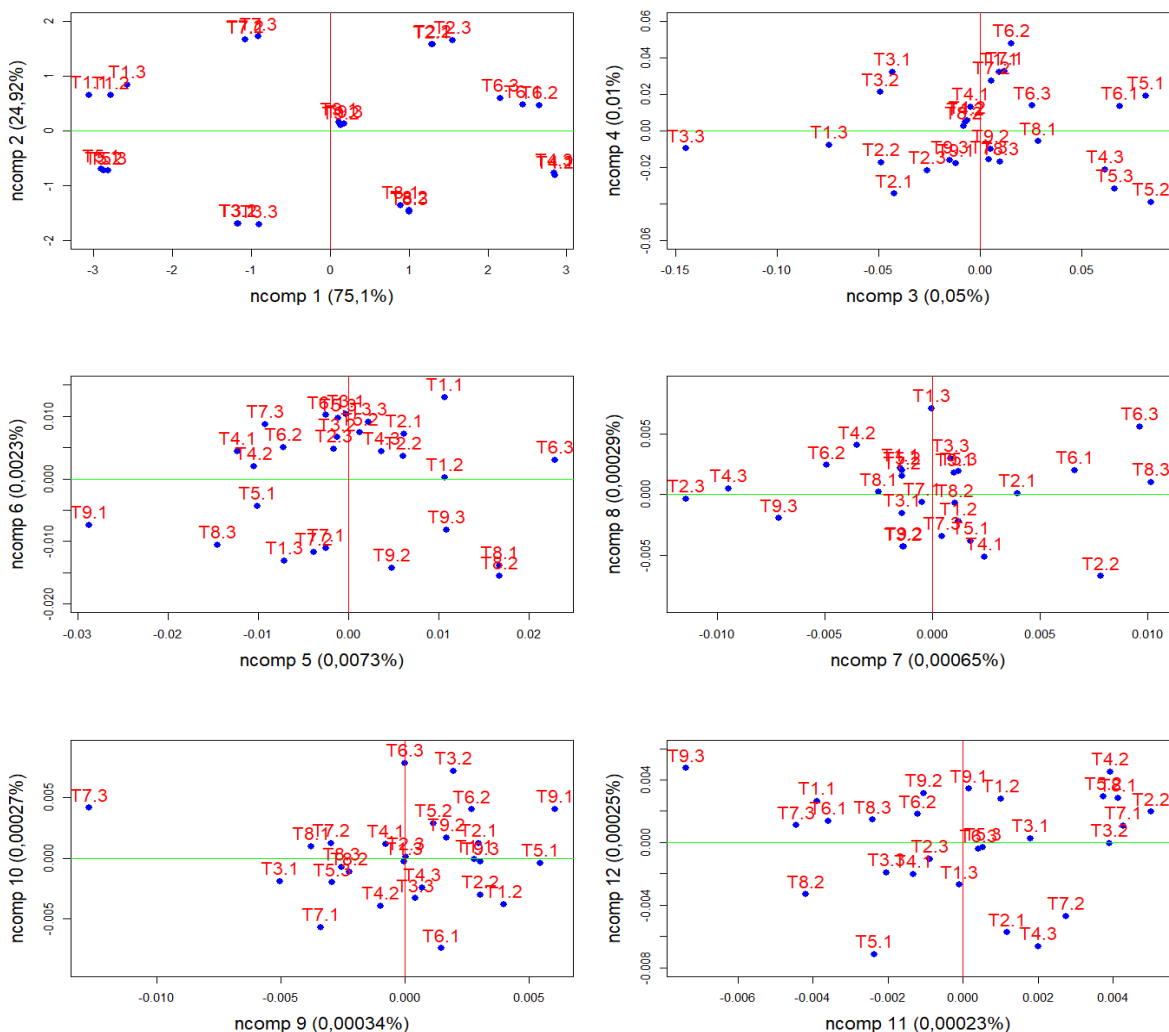
3.2.1.1 Biểu đồ điểm trong mô hình PLS

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. Mô hình PLS khái quát hai tập dữ liệu phổ hấp thụ A và vector nồng độ của MTZ hoặc SPI thành hai ma trận điểm, một đại diện cho phổ hấp thụ A và một đại diện cho nồng độ MTZ hoặc SPI. Hình 3.3 mô tả biểu đồ điểm PLS của MTZ và SPI với 12 thành phần chính đầu tiên.

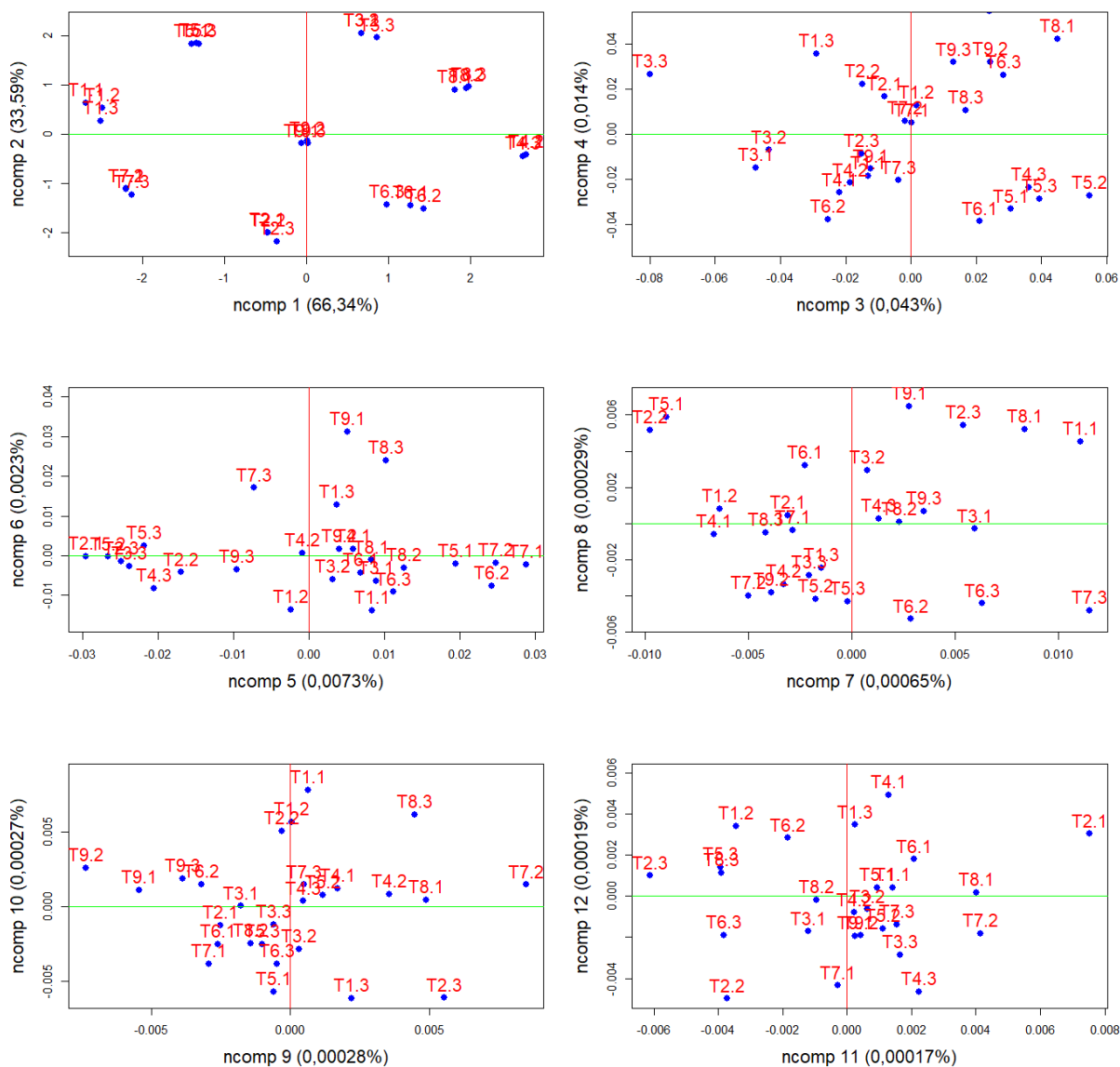
Kết quả thu được từ hai biểu đồ điểm trong hình 3.3 cung cấp cái nhìn sâu sắc về cấu trúc và phân phối của dữ liệu trong không gian tọa độ mới được xác định bởi các

thành phần chính của mô hình PLS. Một số đặc điểm nổi bật trong kết quả của hình 3.3 như sau:

- Đối với biểu đồ điểm của MTZ và SPI phân bố trong hệ xác định bởi thành phần chính thứ 1 và 2: cho thấy phân phối của các mẫu trong không gian của thành phần chính thứ nhất và thứ hai, hai thành phần này giải thích phần lớn phương sai của dữ liệu. Biểu đồ này cho thấy các cụm mẫu dữ liệu rõ ràng, tương ứng với các mẫu chuẩn khác nhau. Các mẫu trong cùng một cụm nằm gần nhau, từ các cụm điểm T1 (T1.1, T1.2, T1.3) cho đến T9 (T9.1, T9.2, T9.3), cho thấy độ lặp lại tốt giữa các lần đo của cùng một mẫu chuẩn. Điều này chỉ ra rằng mô hình PLS đã thành công trong việc phân biệt các mẫu dựa trên đặc tính quang phổ của chúng trong khoảng bước sóng đã đo. Bên cạnh đó, sự phân tách giữa các cụm phản ánh sự khác biệt đáng kể về tín hiệu quang phổ của các mẫu hiệu chuẩn. Điều này phù hợp với các nồng độ khác nhau của MTZ hoặc SPI trong tập hiệu chuẩn dẫn đến tín hiệu quang phổ khác nhau. Sự phân bố của các điểm T trong không gian mới với 2 trục của thành phần chính thứ nhất và thứ hai rất phù hợp với các điểm nồng độ được thiết kế cho tập hiệu chuẩn ban đầu bằng thiết kế thí nghiệm CCD (hình 2.1).



Hình 3.3a. Biểu đồ điểm của 12 biến tiềm ẩn đầu tiên của MTZ

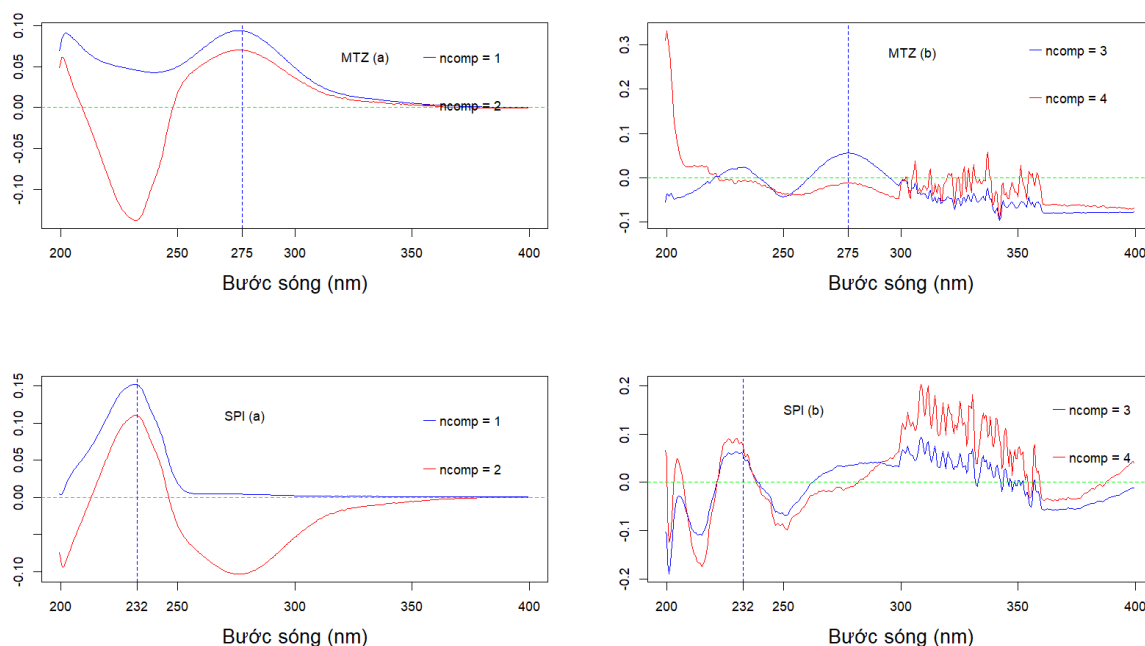


Hình 3.4b. Biểu đồ điểm của 12 biến tiềm ẩn đầu tiên của SPI

Đối với các hình còn lại, phương sai giải thích cho mỗi điểm số trên các trục ứng với mỗi thành phần chính còn lại không đáng kể và giảm dần từ thành phần chính thứ 3 trở đi. Có sự phân tán rất rõ ràng giữa các cụm điểm T1 (T1.1, T1.2, T1.3) cho đến các cụm điểm T9 (T9.1, T9.2, T9.3) ở trong không gian của các thành phần chính thứ 3 trở đi. Điều này giải thích mô hình PLS đã bắt đầu phân biệt được các yếu tố nhiễu trong mỗi mẫu đo ở các lần lặp lại khác nhau, bao gồm nhiễu do sai lệch khi pha mẫu, do tính ổn định của thiết bị, hoặc do các yếu tố môi trường xung quanh... Sự đóng góp ít hơn vào việc giải thích phương sai dữ liệu của các thành phần chính thứ 3 trở đi gợi ý cho thấy mô hình PLS đã trực quan tốt các dữ liệu gốc ban đầu. Khoảng cách giữa các cụm ở trên các thành phần chính khác nhỏ hơn chỉ ra một số mức độ giao thoa tín hiệu. Do đó, các nghiên cứu tiếp theo có thể tập trung vào việc tối ưu hóa điều kiện đo để cải thiện hơn nữa khả năng phân biệt của mô hình.

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 quy của mô hình PLS

Trong mô hình hồi quy PLS (Partial Least Squares), ma trận chiếu đóng vai trò rất quan trọng để biểu diễn mối quan hệ giữa dữ liệu gốc (dữ liệu quang phổ) và các điểm trong không gian tiềm ẩn của PLS. Mỗi thành phần trong ma trận chiếu đại diện cho sự đóng góp của các bước sóng cụ thể vào việc giải thích biến thiên trong dữ liệu ban đầu.



Hình 3.5. 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.

Hình 3.5 thể hiện phổ của các vector chiếu dữ liệu quang phổ lên không gian tiềm ẩn của MTZ và SPI. Phổ của thành phần thứ nhất và thứ 2 của MTZ và SPI (hình 3.5 – MTZ(a) và SPI (a)) cho thấy hai đường cong tương đối phẳng. Thành phần thứ nhất có phổ đồng dạng với phổ của các đơn chất MTZ và SPI, như vậy mô hình PLS có thể diễn giải được các đặc trưng của các chất riêng lẻ trong hỗn hợp, tương ứng với khoảng từ bước sóng 200 – 320 nm đối với MTZ và 200 – 250 đối với SPI trùng với sự hấp thụ ánh sáng của MTZ và SPI trong khoảng bước sóng tương ứng, cực đại của các vector chiếu trùng với cực đại hấp thụ của MTZ và SPI đơn chất. Thành phần thứ 2 bắt đầu xuất hiện các giá trị âm ở các vùng phổ mà MTZ và SPI hấp thụ mạnh, đối với MTZ giá trị âm trong khoảng từ 210 – 250 nm (là khoảng chứa hấp thụ cực đại của SPI) trong khi đối với SPI là từ 250 – 320 nm (Khoảng chứa hấp thụ cực đại của MTZ). Điều này chỉ ra rằng thành phần thứ 2 bắt đầu xuất hiện sự phản ánh không hoàn toàn tương quan với tập nồng độ và giá trị âm xuất hiện vì dữ liệu đã được chiếu vào một không gian mới mà ở đó mô hình đang phản ánh những yếu tố phụ hoặc những biến thể của dữ liệu mà thành phần thứ nhất không thể giải thích được.

Bắt đầu từ thành phần thứ 3 trở đi, các đường cong bắt đầu xuất hiện các dao động mạnh hơn, đặc biệt là vùng ở bước sóng từ 300 – 400 nm, trong khi đó đối với thành phần thứ nhất và hai gần như không ảnh hưởng. Ở các thành phần lớn hơn, sự dao động này xuất hiện hầu như trên toàn bộ các khoảng bước sóng 200 – 400 nm (phụ lục 4). Điều này chứng minh rằng, bắt đầu từ thành phần thứ 3, mô hình nắm bắt các thông tin khác mà thành phần chính thứ 1 và thứ 2 chưa giải thích được. Sự dao động mạnh ở các thành phần này cho thấy sự không chắc chắn trong việc giải thích cho các dữ liệu gốc ban đầu, có thể ám chỉ sự hiện diện của nhiễu hoặc những biến ngẫu nhiên trong dữ liệu gốc. Vấn đề này cũng phù hợp với vấn đề đã được đề cập trong biểu đồ điểm ở phần 4.2.1.



Hình 3.6. Phổ của các hệ số tải (P), hệ số trọng số tải (W), hệ số chiếu (P_{proj}) và hệ số hồi quy của MTZ và SPI

Hình 3.6 là biểu đồ biểu thị các vector của hệ số hồi quy, tải của phổ, trọng số rút trích của phổ, vector chiếu của phổ lên không gian mới tương ứng với các thành phần chính của mô hình PLS được xây dựng cho MTZ và SPI. Khi số biến tiềm ẩn tăng dần, các đường hệ số tải (loadings), hệ số trọng số tải (loading.weights), hệ số chiếu (P_{proj}) gần như bằng 0 cho thấy mối tương quan giữa dữ liệu gốc ban đầu với các biến tiềm ẩn rất bé. Bên cạnh đó, khi số biến tiềm ẩn tăng, hệ số hồi quy tăng đồng thời dao động mạnh (phụ lục 5). Các hệ số hồi quy cho biết mức độ và hướng tác động của từng biến độc lập trong ma trận phổ hấp thụ A đối với biến phụ thuộc nồng độ C. Hệ số lớn cho thấy tác động mạnh hơn của biến đó lên nồng độ, đồng nghĩa với sự thay đổi nhỏ về sự hấp thụ của phổ cũng tác động đến sự thay đổi rất lớn của nồng độ. Vì vậy, việc chọn được số biến tiềm ẩn phù hợp cho mô hình PLS là rất quan trọng, số biến tiềm ẩn phù hợp sẽ giúp mô hình PLS tránh các hiện tượng quá khớp (overfitting) hoặc chưa phù hợp (underfitting), giảm được sai số do ảnh hưởng của các yếu tố thực nghiệm cũng như do mẫu phân tích đến kết quả định lượng.

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

Việc trực quan các tham số trong mô hình PLS chỉ giúp chúng ta tổng quan được cách thức mà thuật toán PLS đã xây dựng để tạo ra các mối quan hệ trong các dữ liệu gốc ban đầu. Để xây dựng mô hình hồi quy, cần phải lựa chọn được số thành phần chính phù hợp để khi xây dựng mô hình giúp giải quyết được việc dự đoán nồng độ của một chất có trong mẫu phân tích. Trong trường hợp lựa chọn số biến tiềm ẩn ít dẫn đến hiện tượng chưa phù hợp (underfitting), tức là mô hình quá đơn giản, nó có thể đúng trong trường hợp dự đoán các mẫu hiệu chuẩn, mẫu chuẩn; mô hình không tận dụng tối đa thông tin có trong dữ liệu. Trường hợp lựa chọn nhiều số biến tiềm ẩn, mô hình sẽ cho hiện tượng quá khớp (overfitting) với tập dữ liệu ban đầu, hiện tượng này làm cho kết quả dự đoán trên tập hiệu chuẩn rất chính xác, mô hình đạt sai số gần như bằng 0 trên tập huấn luyện; Mô hình có số lượng tham số quá lớn hoặc cấu trúc quá phức tạp so với dữ liệu. Cả hai trường hợp này đều cho sai số lớn khi áp dụng trên tập kiểm tra hoặc trên mẫu thực. Điều đó cho thấy lựa chọn số thành phần chính phù hợp cho một quy trình phân tích bằng phương pháp PLS là rất quan trọng.

Điều này làm cho khi đánh giá trên mô hình thì kết quả mang lại rất chính xác nhưng khi dự đoán cho mẫu thực tế thì kết quả sai khác là rất lớn. Chính vì vậy, phương pháp thẩm định chéo loại bỏ một lần được lựa chọn để tìm số biến tiềm ẩn phù hợp dựa trên kết hợp với quá trình đánh giá thống kê.

Trong phần này, nghiên cứu áp dụng ba cách để thực hiện lựa chọn số thành phần chính phù hợp bao gồm dựa trên giá trị PRESS nhỏ nhất, dựa trên phương pháp đánh giá thông kê do Halaand và Thomas đề xuất và theo cách đánh giá dựa trên một tập mẫu độc lập nằm trong không gian mẫu hiệu chuẩn của mô hình. Các chương trình tính toán được xây dựng trên phần mềm R và code của chương trình được trình bày trong phần phụ lục 6.

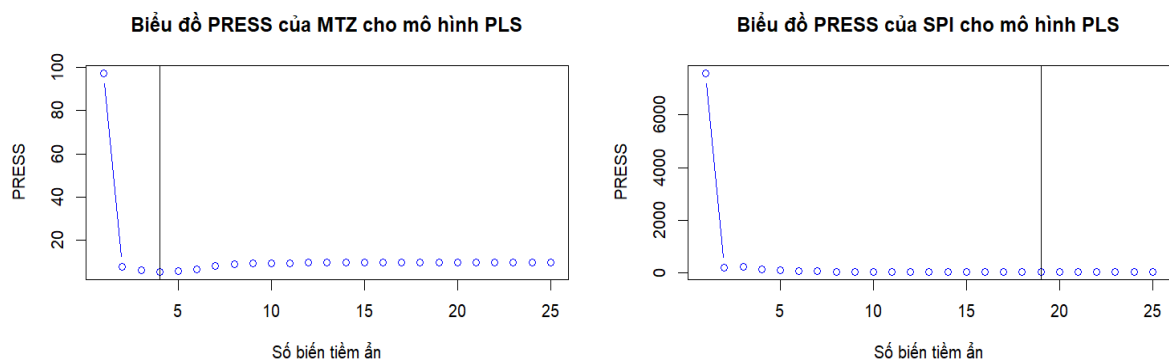
3.2.2.1 Chọn số biến tiềm ẩn dựa vào giá trị PRESS nhỏ nhất

Giá trị PRESS (Prediction Error Sum of Squares) là tổng bình phương của sai số dự đoán được thực hiện trong quá trình thẩm định chéo. Các giá trị này được trích xuất từ mô hình PLS thực hiện trên phần mềm R, với phương pháp thẩm định chéo loại bỏ từng phần. Kết quả các giá trị PRESS và RMSECV được thể hiện trong bảng 3.3.

Kết quả cho thấy, mô hình PLS đã thu gọn dữ liệu ban đầu thành 25 biến tiềm ẩn. Kết quả kiểm định chéo loại bỏ 1 lần ở các biến tiềm ẩn từ 1 đến 25 cho thấy, giá trị PRESS giảm dần từ 1 đến 4 đối với MTZ và từ 1 đến 19 đối với SPI. Giá trị PRESS là tổng bình phương sự khác biệt giữa giá trị dự đoán được và giá trị nồng độ ban đầu. Giá trị càng nhỏ thì mô hình càng chính xác. Kết quả trong trường hợp này số biến tiềm ẩn được chọn lần lượt là 4 và 19 tương ứng với MTZ và SPI. Hình 4.6 thể hiện trực quan các giá trị PRESS của MTZ và SPI.

Bảng 3.3. Các tham số thống kê của mô hình PLS cho MTZ và SPI ở các biến tiềm ẩn khác nhau

Số biến tiềm ẩn	METRONIDAZOL				SPIRAMYCIN			
	PRESS	RMSECV (µg/mL)	Phần trăm phương sai giải thích tích lũy (%)		PRESS	RMSECV (µg/mL)	Phần trăm phương sai giải thích tích lũy (%)	
			A	C,MTZ			A	C,SPI
1	97,17	1,90	75,01	86,22	7569,51	16,74	66,34	70,87
2	7,86	0,54	99,93	98,94	215,27	2,82	99,93	99,19
3	6,08	0,47	99,98	99,25	250,03	3,04	99,97	99,34
4	5,36	0,45	99,99	99,45	146,58	2,33	99,99	99,64
5	5,69	0,46	99,99	99,56	115,90	2,07	99,99	99,74
6	6,34	0,48	100,00	99,68	87,98	1,81	100,00	99,86
7	8,07	0,55	100,00	99,89	68,83	1,60	100,00	99,95
8	8,82	0,57	100,00	99,96	57,50	1,46	100,00	99,99
9	9,12	0,58	100,00	99,99	53,53	1,41	100,00	100
10	9,34	0,59	100,00	100,00	51,70	1,38	100,00	100
11	9,45	0,59	100,00	100,00	51,42	1,38	100,00	100
12	9,49	0,59	100,00	100,00	50,87	1,37	100,00	100
13	9,52	0,59	100,00	100,00	50,90	1,37	100,00	100
14	9,54	0,59	100,00	100,00	50,91	1,37	100,00	100
15	9,55	0,59	100,00	100,00	50,87	1,37	100,00	100
16	9,55	0,59	100,00	100,00	50,85 ₁	1,37	100,00	100
17	9,55	0,59	100,00	100,00	50,85 ₂	1,37	100,00	100
18	9,55	0,59	100,00	100,00	50,85 ₀	1,37	100,00	100
19	9,55	0,59	100,00	100,00	50,85₀	1,37	100,00	100
20	9,55	0,59	100,00	100,00	50,85	1,37	100,00	100
21	9,55	0,59	100,00	100,00	50,85	1,37	100,00	100
22	9,55	0,59	100,00	100,00	50,85	1,37	100,00	100
23	9,55	0,59	100,00	100,00	50,85	1,37	100,00	100
24	9,55	0,59	100,00	100,00	50,85	1,37	100,00	100
25	9,55	0,59	100,00	100,00	50,85	1,37	100,00	100



Hình 3.7. Biểu đồ giá trị PRESS trong mô hình PLS của MTZ và SPI

Dựa vào bảng 3.3 và hình 3.7 cho chúng ta thấy rằng, với MTZ ở các giá trị biến tiềm ẩn nhỏ hơn 4 sự sai khác giữa các giá trị PRESS hoặc là RMSECV không khác nhau nhiều, ở các biến tiềm ẩn 2 và 3 lớn hơn không nhiều so với ở biến tiềm ẩn 4. Cũng tương tự với SPI cho các biến tiềm ẩn từ 6 cho đến 19.

3.2.2.2 Chọn số biến tiềm ẩn dựa vào kiểm định thống kê

Bảng 3.4. Kết quả phân tích thống kê kiểm tra chéo loại bỏ một phần theo Haaland và Thomas cho tập hiệu chuẩn AgimDogyl

Biến tiềm ẩn	MTZ				SPI		
	PRESS	Fcalc	p_value	PRESS	Fcalc	p_value	
1	97,17	18,13	1,00	7569,51	148,86	1,00	
2	7,86	1,47	0,84	215,27	4,23	0,999	
3	6,08	1,13	0,63	250,03	4,92	0,999	
4	5,36	1,00	0,50	146,58	2,88	0,996	
5				115,90	2,29	0,982	
6				87,98	1,73	0,920	
7				68,82	1,35	0,782	
8				57,50	1,13	0,624	
9				53,53	1,05	0,553	
10				51,70	1,02	0,517	
11				51,41	1,01	0,511	
12				50,87	1,00	0,500	
13				50,90	1,00	0,501	
14				50,91	1,00	0,501	
15				50,87	1,00	0,500	
16				50,85	1,00	0,500	
17				50,85	1,00	0,500	
18				50,85	1,00	0,500	
19				50,85	1,00	0,500	

Kết quả lựa chọn số biến tiềm ẩn dựa vào giá trị nhỏ nhất của PRESS (hoặc của RMSECV) cho thấy rằng số lượng biến tiềm ẩn có thể quá lớn. Điều này dẫn đến mô hình trở nên quá khớp với dữ liệu huấn luyện nhưng có thể dự đoán không đúng với các

mẫu không có mặt trong tập huấn luyện. Một cách lựa chọn khác theo đề xuất của Haaland và Thomas, chúng tôi dựa trên kết quả của mô hình PLS để tính các giá trị F_{calc} được tính khi so sánh giữa giá trị PRESS ở các biến tiềm ẩn nhỏ hơn so với giá trị PRESS ở biến tiềm ẩn có giá trị nhỏ nhất. 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, tương ứng với giá trị xác suất là 0,627 và 0,624 (bảng 3.4).

3.2.2.3 Lựa chọn số biến tiềm ẩn dựa trên đánh giá tập kiểm tra

Một cách chọn khác được chúng tôi đề xuất đó là dựa vào kết quả của dự đoán của mô hình PLS đã xây dựng trên một tập nồng độ được xây dựng (tập kiểm tra) độc lập với tập hiệu chuẩn, nằm trong không gian của tập hiệu chuẩn ở các biến tiềm ẩn khác nhau. Các giá trị PRESS được tính ở 4 biến tiềm ẩn đầu tiên đối với MTZ và 19 biến tiềm ẩn đầu tiên đối với SPI. Bảng 3.5 cho kết quả thống kê khi đánh giá lựa chọn số thành phần chính dựa trên phương pháp Haaland và Thomas.

Bảng 3.5. Kết quả phân tích thống kê giá trị PRESS trên tập hiệu chuẩn AgimDogyl theo Haaland và Thomas

Biến tiềm ẩn	MTZ			SPI		
	PRESS	F_{calc}	p_value	PRESS	F_{calc}	p_value
1	21,87	7,54	1,00	1707,03	25,60	1,0
2	2,90	1,00	0,50	201,31	3,02	0,99
3	8,99			97,92	1,47	0,84
4	4,23			193,63	2,90	1,00
5				192,65	2,89	1,00
6				190,20	2,85	1,00
7				67,02	1,01	0,51
8				66,68	1,00	0,50
9				81,43		
10				79,40		
11				81,04		
12				80,63		
13				80,89		
14				80,71		
15				80,68		
16				80,66		
17				80,67		
18				80,67		
19				80,67		

Trong trường hợp này, tương tự, các giá trị phần trăm xác suất p trong phân phối Fisher được tính với bậc tự do bằng số mẫu trong tập tham định là 27 cho cả tử và mẫu (mẫu tham định bao gồm 27 mẫu). Kết quả ở số biến tiềm ẩn bằng 2 và 7 tương ứng với MTZ và SPI cho kết quả giá trị p nhỏ hơn 0,75 được chọn để xây dựng mô hình PLS.

Như vậy, với 3 phương pháp đề xuất sẽ có 3 sự lựa chọn biến tiềm ẩn khác nhau, đối với MTZ số biến tiềm ẩn được đề xuất lần lượt là 4, 3 và 2; đối với SPI số biến tiềm ẩn được đề xuất lần lượt là 19, 8 và 7.

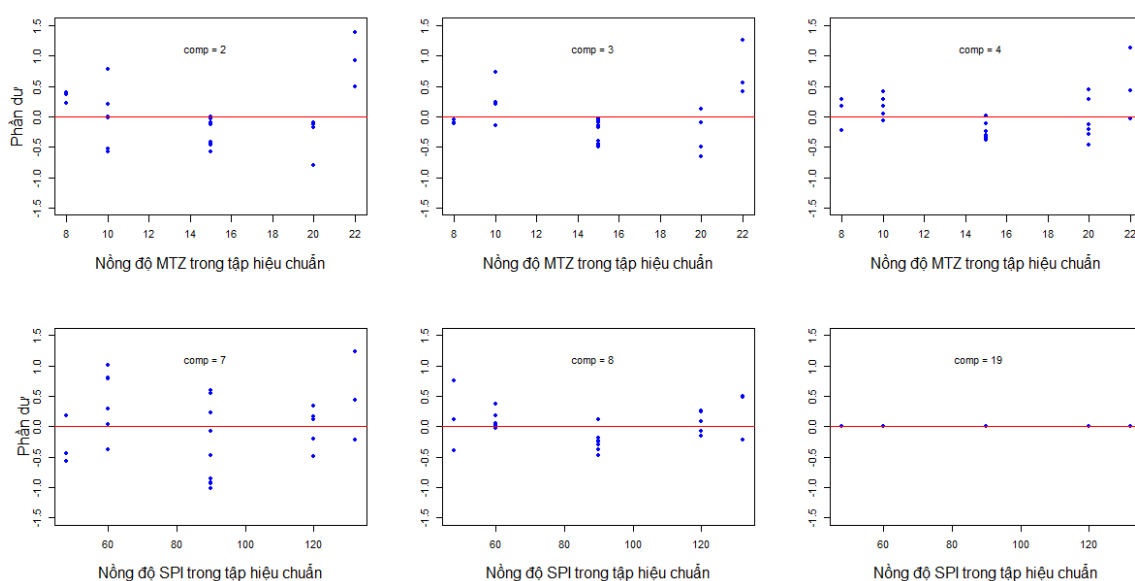
3.2.3 Kết quả đánh giá mô hình PLS để xác định đồng thời MTZ và SPI

Bảng 3.6. Kết quả các thông số thống kê của mô hình PLS cho MTZ và SPI ở các biến tiềm ẩn được chọn

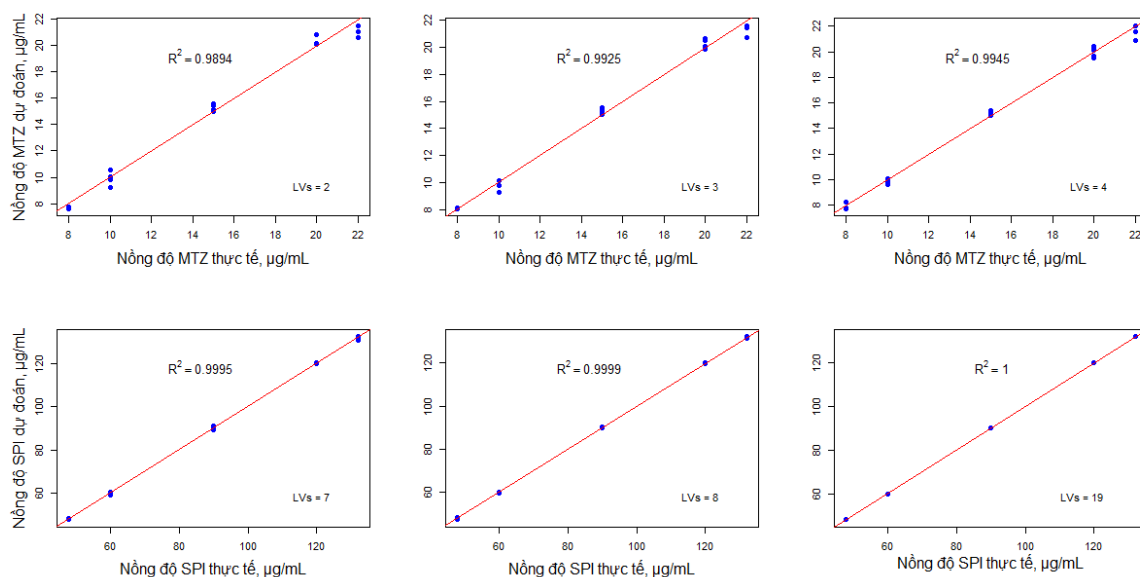
Mẫu	Nồng độ tập hiệu chuẩn		Phần dư của MTZ			Phần dư của SPI		
	MTZ (µg/mL)	SPI (IU/mL)	n = 2	n = 3	n = 4	n = 7	n = 8	n = 19
T1.1	10,0	60,0	0,78	0,73	0,42	1,02	0,38	0,00
T1.2	10,0	60,0	0,20	0,23	0,18	0,30	0,18	0,00
T1.3	10,0	60,0	-0,53	-0,14	-0,06	-0,38	-0,03	0,00
T2.1	20,0	60,0	-0,10	0,12	0,45	0,04	-0,03	0,00
T2.2	20,0	60,0	-0,13	0,12	0,29	0,79	0,05	0,00
T2.3	20,0	60,0	-0,80	-0,66	-0,46	0,81	0,03	0,00
T3.1	10,0	120,0	0,00	0,23	-0,07	-0,20	-0,16	0,00
T3.2	10,0	120,0	-0,02	0,24	0,04	0,34	-0,08	0,00
T3.3	10,0	120,0	-0,57	0,20	0,29	-0,49	-0,08	0,00
T4.1	20,0	120,0	-0,12	-0,09	-0,21	0,16	0,24	0,00
T4.2	20,0	120,0	-0,13	-0,09	-0,13	-0,20	0,27	0,00
T4.3	20,0	120,0	-0,17	-0,50	-0,29	0,12	0,08	0,00
T5.1	8,0	90,0	0,39	-0,05	-0,23	0,59	-0,25	0,00
T5.2	8,0	90,0	0,36	-0,09	0,28	-0,48	0,12	0,00
T5.3	8,0	90,0	0,23	-0,12	0,18	-0,85	-0,23	0,00
T6.1	22,0	90,0	0,92	0,55	0,43	0,23	-0,23	0,00
T6.2	22,0	90,0	0,50	0,41	-0,04	-0,94	-0,19	0,00
T6.3	22,0	90,0	1,39	1,26	1,13	-0,92	-0,29	0,00
T7.1	15,0	48,0	0,00	-0,07	-0,38	-0,44	-0,39	0,00
T7.2	15,0	48,0	-0,01	-0,04	-0,30	0,19	0,76	0,00
T7.3	15,0	48,0	-0,45	-0,47	-0,32	-0,57	0,12	0,00
T8.1	15,0	132,0	-0,03	-0,18	-0,12	1,23	0,48	0,00
T8.2	15,0	132,0	-0,13	-0,09	-0,11	-0,21	-0,22	0,00
T8.3	15,0	132,0	-0,10	-0,15	0,01	0,43	0,50	0,00
T9.1	15,0	90,0	-0,47	-0,40	-0,24	0,55	-0,37	0,00
T9.2	15,0	90,0	-0,42	-0,45	-0,36	-1,02	-0,48	0,00
T9.3	15,0	90,0	-0,58	-0,50	-0,35	-0,08	-0,18	0,00
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 > 0,05		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)			1,019	1,139	1,074	3,186	3,289	1,500

Tập nồng độ hiệu chuẩn được chuẩn bị với 9 mẫu độc lập ở các mức tỉ lệ nồng độ khác nhau, quét phổ hấp thụ UV-Vis trong khoảng 200 – 400 nm với khoảng quét 0,5 nm, mỗi mẫu được pha lặp lại 3 lần. Bảng 3.6 mô tả các kết quả đạt được khi xây dựng mô hình PLS cho các thành phần MTZ và SPI có trong tập hiệu chuẩn.

Kết quả bảng 3.6 cho thấy rằng, giá trị RMSECV đều có giá trị nhỏ, với MTZ lần lượt là 0,5396 (3,6%), 0,4745 (3,2%) và 0,347 (2,3%) tương ứng với số biến tiềm ẩn là 2, 3, 4; với SPI lần lượt là 1,597 (1,8%), 1,459 (1,6%), 1,372 (1,5%) tương ứng với số biến tiềm ẩn là 7, 8, 19. Theo Olivieri, phần trăm giữa giá trị RMSE so với nồng độ trung bình trong tập hiệu chuẩn (là 15 $\mu\text{g/mL}$ đối với MTZ và 90 IU/mL đối với SPI) nhỏ hơn 2% là rất tốt, 2-5% là tốt; 5-10% là chấp nhận được và lớn hơn 10% là không tốt [50]. Kết quả cho thấy với các thành phần chính đã chọn để đánh giá kết quả mô hình xây dựng là rất tốt đối với SPI hoặc tốt đối với MTZ. Mặc dù độ lệch chuẩn cung cấp một chỉ số về độ phân tán của một tập hợp kết quả xung quanh giá trị trung bình, nó không chỉ ra hình dạng của phân phối, vì vậy cần phải có phân tích phần dư để kiểm tra phân phối chuẩn [188]. Kết quả phân tích phần dư bằng biểu đồ hình 3.8 thể hiện các giá trị phần dư đều phân bố quanh điểm 0. Tuy nhiên, kết quả phân tích phần dư bằng kiểm định Shapiro - Wilk cho thấy rằng, một tập hiệu chuẩn tốt khi phần dư đánh giá trên tập hiệu chuẩn phải đạt phân phối chuẩn tức là $p > 0,05$ (Điều này cho thấy rằng phân phối các phép đo có hình dạng gần như đối xứng quanh giá trị trung bình, với phần lớn các phép đo tập trung vào trung tâm.); do đó với MTZ số biến tiềm ẩn bằng 2 là hợp lý, trong khi đó SPI số biến tiềm ẩn là 7. Điều này cũng phù hợp với nguyên tắc cơ bản của phương pháp PLS: để giảm các yếu tố tác động lên mô hình và tránh hiện tượng quá khớp (overfitting), số biến tiềm ẩn càng nhỏ càng tốt.



Hình 3.8. Biểu đồ thể hiện sự phân tán của phần dư quanh điểm 0 của MTZ và SPI.



Hình 3.9. Đồ thị biểu diễn mức độ chính xác giữa nồng độ dự đoán bởi mô hình PLS và nồng độ thực tế của MTZ và SPI

Kết quả đánh giá mức độ chính xác khả năng dự đoán của mô hình PLS đã xây dựng cho MTZ và SPI trên tập hiệu chuẩn đều có hệ số xác định R^2 rất tốt, bằng 0,9894, 0,9925, 0,9945 tương ứng với các biến tiềm ẩn 2, 3, 4 cho MTZ; bằng 0,9995, 0,9999 và 1 tương ứng với số biến tiềm ẩn 7, 8, 19 cho SPI (hình 3.8). Điều này có thể giải thích rằng, với số thành phần chính càng lớn thì mức độ diễn giải cho tập huấn luyện càng khớp giữa dữ liệu phổ và nồng độ, đây chính là hiện tượng quá khớp của mô hình.

Bảng 3.7. Kết quả định lượng MTZ và SPI đối với các mẫu trong tập kiểm tra

Mẫu	Nồng độ		Độ thu hồi (%)	
	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ	SPI
V1	12,5	75,0	98,6	99,6
V2	17,5	75,0	100,3	99,5
V3	12,5	105,0	99,1	98,9
V4	17,5	105,0	99,9	101,7
V5	11,5	90,0	99,0	101,7
V6	18,5	90,0	102,2	99,9
V7	15,0	69,0	101,3	97,1
V8	15,0	111,0	103,9	97,5
V9	15,0	90,0	104,1	98,1
Trung bình (%)			100,9	99,3
RSD (%)			2,061	1,658
RMSEP ($\mu\text{g/mL}$)			0,328	1,575

Kết luận: Đã xây dựng thành công mô hình PLS dựa trên các tập nồng độ xây dựng trong tập hiệu chuẩn, đánh giá mô hình cho thấy số biến tiềm ẩn phù hợp cho MTZ và SPI lần lượt là 2 và 7. Các giá trị thống kê đánh giá phù hợp để sử dụng mô hình PLS dự đoán nồng độ MTZ và SPI trong các mẫu phân tích.

Với mô hình PLS đã xây dựng với số biến tiềm ẩn đã chọn lần lượt là 2 và 7 cho MTZ và SPI, chúng tôi tiến hành xác định nồng độ của MTZ và SPI của các mẫu trong tập thẩm định, kết quả thể hiện trong bảng 3.7. Độ thu hồi của MTZ trong các mẫu nằm trong khoảng 98,6 – 104,1%, giá trị RSD = 2,06%; độ thu hồi của SPI nằm trong khoảng 97,1 – 101,7% với RSD = 1,658%. Kết quả này hoàn toàn phù hợp với quy định của AOAC.

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

a. Đánh giá độ đúng bằng phương pháp thêm chuẩn

Độ đúng được đánh giá bằng phương pháp thêm chuẩn ở ba mức nồng độ thấp – trung bình – cao trong khoảng nồng độ làm việc, kết quả đánh giá bằng cách tính độ thu hồi. Thuốc AgimDogyl có hàm lượng có tỷ lệ tương ứng là 15 µg/mL – 90 IU/mL tương ứng với MTZ và SPI, vì vậy, nồng độ thêm chuẩn lần lượt cho MTZ là 12 – 15 – 18 µg/mL, cho SPI là 72 – 90 – 108 IU/mL. Kết quả độ thu hồi được thể hiện qua bảng 3.8

Bảng 3.8. Kết quả đánh giá độ thu hồi của MTZ và SPI

Mẫu	Thêm chuẩn		Độ thu hồi (%)	
	MTZ (µg/mL)	SPI (IU/mL)	MTZ	SPI
1	12,0	72,0	96,6	100,4
2	12,0	72,0	91,8	100,7
3	12,0	72,0	90,6	99,6
Trung bình			93,0	100,2
RSD			3,08	1,05
4	15,0	90,0	98,6	103,6
5	15,0	90,0	96,9	101,6
6	15,0	90,0	100,2	99,1
Trung bình			98,6	101,4
RSD			1,50	2,08
7	18,0	108,0	96,6	105,4
8	18,0	108,0	95,1	105,1
9	18,0	108,0	97,2	104,3
Trung bình			96,3	104,9
RSD			1,01	0,71

Đối với MTZ, độ thu hồi nằm trong khoảng 90,3% đến 100,2%; trong khi đó, đối với SPI có độ thu hồi nhỏ nhất là 98,3% và lớn nhất là 106,1%. Theo quy định của AOAC, mức nồng độ của MTZ xác định từ 12 – 18 µg/mL, mức nồng độ của SPI xác định từ 72 IU/mL (16 µg/mL) đến 108 IU/mL (24 µg/mL) nhỏ hơn 100 µg/mL nên khoảng độ thu hồi cho phép là 90 – 107%. Như vậy, độ thu hồi cho MTZ và SPI đều nằm trong khoảng giới hạn cho phép của AOAC, do đó, kết quả đánh giá độ đúng đạt yêu cầu.

b. So sánh với phương pháp đối chiếu

Mẫu thuốc AgimDogyl được gửi tại trung tâm Kiểm nghiệm thuốc, Mỹ phẩm, Thực phẩm Thành phố Huế để xác định hàm lượng MTZ bằng phương pháp HPLC và SPI bằng phương pháp vi sinh. Kết quả được thể hiện trong bảng 3.9 cho thấy, giá trị $F_{tính}$ của MTZ và SPI đều nhỏ hơn $F_{bảng}$ chứng tỏ phương pháp xác định hàm lượng MTZ và SPI bằng phổ hấp thụ UV-Vis kết hợp phương pháp bình phương tối thiểu riêng phần có độ lặp lại không có sự khác biệt có ý nghĩa thống kê ($p = 0,05$) với phương pháp tham chiếu. Kết quả so sánh giá trị trung bình giữa hai phương pháp có giá trị $t_{tính} < t_{bảng}$ ($p = 0,05$) cho thấy hai giá trị trung bình không khác nhau có ý nghĩa thống kê.

Bảng 3.9. Kết quả đánh giá độ đúng thông qua so sánh với phương pháp tham chiếu

Mẫu	MTZ (mg/viên)		SPI (IU/viên)	
	PLS	HPLC	PLS	Vi sinh
1	122,7	124,54	742416,7	774172
2	125,3	123,25	764333,3	768611
3	125,4	122,84	756250	765896
$F_{tính}$	2,977		6,902	
$F_{(2,2-0,05)}$	15,439		15,439	
$t_{tính}$	0,904		2,224	
$t_{(0,05 - 4)}$	2,776		2,776	

Như vậy, thông qua đánh giá độ thu hồi và so sánh với phương pháp tham chiếu cho thấy phương pháp định lượng MTZ và SPI bằng phổ hấp thụ UV-Vis kết hợp phương pháp bình phương tối thiểu riêng phần đạt chỉ tiêu về độ đúng.

3.2.4.2 Đánh giá độ lặp lại

Độ lặp lại được thực hiện trong hai ngày, mỗi ngày được làm lặp lại 6 lần giống nhau bắt đầu từ cân mẫu. Cách tiến hành: Cân chính xác 20 viên thuốc AgimDogyl có khối lượng 12,422 gam sau đó cạo sạch vỏ, nghiền mịn; Cân chính xác 0,6211 gam bột thuốc cho vào bình định mức 50 mL, thêm khoảng 30 – 40 mL MeOH, siêu âm trong 5 phút, thêm MeOH cho vừa đủ và lắc nhẹ. Để yên dung dịch 10 phút và lọc qua mạng lọc 0,45 µm để lấy khoảng 5 mL dịch lọc. Lấy 0,06 mL dịch lọc cho vào bình định mức 10 mL, pha loãng bằng dung môi HCl 0,1 mol/L vừa đủ. Lắc nhẹ và thực hiện quét phổ

UV trong khoảng bước sóng từ 200 – 400 nm, khoảng quét 0,5 nm. Mỗi mẫu được pha 3 lần lặp lại để quét phổ, kết quả tính dựa trên mô hình PLS đã xây dựng được thể hiện trong bảng 3.10

Kết quả cho thấy, đối với MTZ đánh giá độ lặp lại ngày 1 có nồng độ trung bình tính được là 14,96 µg/mL, độ lệch chuẩn tương đối (RSD) là 1,303%; Đánh giá độ lặp lại ngày thứ 2 có giá trị trung bình là 14,98 µg/mL, giá trị RSD (%) là 1,275%. Độ lặp lại ngày thứ 1 của SPI có giá trị trung bình là 88,86 IU/mL (19,75 µg/mL) với độ lệch chuẩn tương đối là 1,511%, trong khi đó đánh giá độ lặp lại ngày thứ hai đối với SPI có giá trị trung bình là 88,06 IU/mL (19,57 µg/mL) và RSD (%) bằng 0,907%. Kết quả các giá trị RSD đều nhỏ hơn mức quy ước của AOAC ở mức nồng độ dưới 100 µg/mL là 5,3, điều này cho thấy kết quả phân tích cho độ lặp lại tốt.

Nghiên cứu cũng đã so sánh kết quả độ lặp lại hai ngày dựa vào chuẩn Fisher (F-test) và Student (t-test). Đối với MTZ, $F_{tính} = 1,181 < F_{bảng}$ với bậc tự do cho cả tử và mẫu là 5 và 5 ở mức $\alpha = 0,05$; $t_{tính} = 0,344 < t_{bảng}(\alpha = 0,05, 10)$ cho thấy độ lệch chuẩn của hai dãy số liệu không khác nhau đáng kể và các giá trị trung bình không khác nhau có ý nghĩa thống kê ($p > 0,05$). Đối với SPI, $F_{tính} = 7,518 > F_{bảng(5,5)}(\alpha = 0,05) = 5,820$; $t_{tính} = 1,540 < t_{bảng}(\alpha = 0,05, 10)$ cho thấy độ lệch chuẩn của 2 dãy số liệu khác nhau đáng kể nhưng các giá trị trung bình không khác nhau có ý nghĩa thống kê ($p > 0,05$).

Bảng 3.10. Kết quả đánh giá độ lặp lại định lượng đồng thời MTZ và SPI trong viên nén AgimDogyl

	Độ lặp lại (ngày 1)		Độ lặp lại (ngày 2)	
	MTZ (µg/mL)	SPI (IU/mL)	MTZ (µg/mL)	SPI (IU/mL)
Lần 1	15,08	88,07	14,83	88,35
Lần 2	14,9	88,7	15,18	88,65
Lần 3	14,7	87,92	14,97	87,49
Lần 4	15,04	88,65	15,01	87,61
Lần 5	14,99	88,6	15,05	88,13
Lần 6	15,03	91,23	14,86	88,1
TB	14,96	88,86	14,98	88,06
RSD (%)	1,303	1,511	1,275	0,907

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 thể hiện trong bảng 3.11, hàm lượng MTZ $124,38 \pm 1,26$ mg và SPI là 735971 ± 3370 IU. Theo qui định của Dược điển Việt Nam V, hàm lượng MTZ nằm trong khoảng từ 95,0% đến 110,0% so với lượng ghi trên nhãn, hàm lượng SPI từ 90,0% đến 120,0% so với lượng ghi trên nhãn. Kết quả cho thấy, hàm lượng của MTZ và SPI đều nằm trong giới hạn theo quy định của Dược điển Việt Nam V.

Bảng 3.11. Kết quả định lượng MTZ và SPI trong viên nén AgimDogyl bằng phương pháp PLS

Mẫu	Hàm lượng trong 1 viên		%HLGTN	
	MTZ (mg)	SPI (IU)	MTZ (%)	SPI (%)
Q1	123,77	731333	99,0	97,5
Q2	122,33	737139	97,9	98,3
Q3	124,37	736639	99,5	98,2
Trung bình	123,49 ± 1,26	735972 ± 3370	99,5 ± 0,99	98,1 ± 0,45

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

Trong nghiên cứu này, chúng tôi tiến hành xác định đồng thời SPR và FUR trong viên nén Spirosemid, AML và ATO trong viên nén Zoamco-A, IND và AML trong viên nén Natrixam 1,5/5. Các tập hiệu chuẩn được thiết kế theo thiết kế thí nghiệm CCD được trình bày trong bảng 3.1. Các mẫu trong tập hiệu chuẩn Spirosemid được đo phổ UV-Vis trong khoảng 200 – 400 nm, bước nhảy 0,5 nm; tập hiệu chuẩn Zoamco-A đo trong khoảng 230 – 260 nm, bước nhảy 0,2 nm; tập hiệu chuẩn Natrixam được đo trong khoảng 200 – 400 nm, bước nhảy 0,2 nm. Các dữ liệu phổ hấp thụ và nồng độ của tập hiệu chuẩn sau đó được dùng để xây dựng mô hình PLS với quá trình thẩm định chéo loại bỏ từng phần để xác định các tham số thống kê cho mô hình. Kết quả được thể hiện trong bảng 3.12.

Về lựa chọn số biến tiềm ẩn, đối với mô hình PLS cho tập hiệu chuẩn thuốc Spirosemi số biến tiềm ẩn được chọn cho SPR và FUR lần lượt là 2 và 2, số biến tiềm ẩn của AML và ATO cho tập hiệu chuẩn thuốc Zoamco-A là 3 và 2, 4 và 2 là số biến tiềm ẩn của IND và AML cho mô hình PLS khi xây dựng cho tập hiệu chuẩn thuốc Natrixam.

Về đánh giá các giá trị RMSE, đối với tập hiệu chuẩn thuốc Spirosemid phần trăm giá trị RMSE đối với nồng độ trung bình của tập hiệu chuẩn của SPR là 1,9%, 2,6%; của FUR là 2,7%, 3,5% tương ứng với RMSECV, RMSEC; Đối với tập hiệu chuẩn của thuốc Zoamco-A, phần trăm tương ứng của AML là 1,3%, 0,6%; của ATO là 2,4%, 1,6%; Đối với tập hiệu chuẩn của thuốc Natrixam 1,5/5, phần trăm tương ứng của IND là 2,9%, 2,4%; của AML là 1,5%, 1,4%. Kết quả cho thấy, đối với khả năng dự đoán trên tập dữ liệu chưa thấy trước đó bằng kỹ thuật thẩm định chéo (RMSECV) cho thấy mô hình PLS đã xây dựng cho kết quả rất tốt (<2%) hoặc tốt (2-5%). Chỉ số RMSEC là giá trị đánh giá độ chính xác của mô hình trong việc dự đoán các giá trị trên cùng một tập dữ liệu mà mô hình đã được huấn luyện, với số biến tiềm ẩn đã chọn tương ứng với

các thành phần trong các mẫu hiệu chuẩn đều cho giá trị RMSEC nhỏ (<2% hoặc trong khoảng 2-5%).

Bảng 3.12. Các thông số thống kê đánh giá các tập hiệu chuẩn Spirosemid, Zoamco-A và Natrixam bằng phương pháp PLS

Mẫu	Nồng độ tập hiệu chuẩn tính từ mô hình PLS ($\mu\text{g/mL}$)					
	Spirosemid		Zoamco-A		Natrixam	
	SPR	FUR	AML	ATO	IND	AML
T1	5,86	4,32	7,94	16,24	2,86	9,6
T2	13,70	3,80	11,99	16,03	6,05	10,13
T3	5,89	11,96	8,07	23,46	3,09	39,84
T4	14,15	11,89	12,05	23,83	6,08	40,21
T5	10,14	7,94	9,96	20,26	4,46	25,22
T6	4,57	7,73	7,16	20,35	2,33	25,15
T7	15,64	8,16	12,73	20,28	6,32	24,4
T8	10,19	2,44	10,1	13,91	4,59	4,20
T9	9,86	13,77	9,99	25,55	4,43	45,97
T10					4,58	14,76
T11					4,50	30,0
Khoảng bước sóng (nm)	200 - 400		230-260		200 - 400	
Khoảng quét ($\Delta\lambda$)	0,5		0,2		0,2	
n	2	2	3	2	4	2
RMSECV ($\mu\text{g/mL}$)	0,190	0,212	0,129	0,489	0,128	0,385
RMSEC ($\mu\text{g/mL}$)	0,260	0,280	0,057	0,314	0,109	0,354
Shapiro-W	0,8842	0,97416	0,9062	0,8676	0,9093	0,8871
Wilk test p(>0,05)	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} ($\mu\text{g/mL}$)	0,375	0,272	1,00	4,69	0,269	1,263
LOD _{max} ($\mu\text{g/mL}$)	0,420	0,320	1,06	4,83	0,338	1,708

Về đánh giá phân tích phần dư bằng kiểm định Shapiro – Wilk, các giá trị p thu được đều lớn hơn 0,05 cho thấy các giá trị phần dư phân bố đều quanh điểm 0 và phân phối chuẩn. Chứng tỏ mô hình PLS đã xây dựng cho các tập hiệu chuẩn là hợp lý, có độ tin cậy cao, mô hình PLS đã nắm bắt được các thông tin quan trọng trong dữ liệu, đã mô phỏng tốt dữ liệu và không có sự thiên lệch hệ thống nào trong phần dư.

Về đánh giá sự phù hợp giữa giá trị dự đoán và giá trị thực thông qua hệ số xác định R² cho thấy các giá trị R² cho các thành phần SPR, FUR trong mẫu hiệu chuẩn Spirosemid; AML, ATO trong mẫu hiệu chuẩn Zoamco-A; IND, AML trong mẫu hiệu chuẩn Natrixam đều lớn hơn 0,99. Kết quả cho thấy khả năng giải thích của mô hình PLS đã xây dựng đối với tập nồng độ ban đầu là rất tốt.

Đánh giá giới hạn phát hiện cho mô hình phân tích đa biến cho thấy rằng các mô hình PLS kết hợp phổ hấp thụ UV –Vis đều cho giá trị giới hạn phát hiện nhỏ.

Áp dụng mô hình PLS kết hợp phổ hấp thụ UV-Vis đã xây dựng để định lượng các thành phần trong các mẫu thử định được chuẩn bị trong phòng thí nghiệm. Các giá trị phần trăm độ thu hồi, độ lệch chuẩn tương đối, RMSEP đã được tính và được trình bày cụ thể trong bảng 3.13. Độ thu hồi của SPR và FUR nằm trong khoảng 92,8-106,3%, 97,0-108,0%; độ thu hồi của AML và ATO lần lượt là 97,3-104,5%, 98,0-105,2%; và của IND và AML là 96,9-103,9%, 97,3-106,3%. Khoảng độ thu hồi này đều nằm trong khoảng độ thu hồi theo hướng dẫn của AOAC là 90-110%. Các giá trị RMSEP đo lường sai số của mô hình trên dữ liệu ngoài mẫu hiệu chuẩn, các giá trị này đều có phần trăm so với giá trị nồng độ trung bình trong tập kiểm tra là nhỏ hơn 2% hoặc trong khoảng 2-5%. Như vậy, với việc đánh giá trên tập kiểm tra cho thấy các mô hình PLS đã xây dựng dựa trên dữ liệu phổ hấp thụ UV-Vis cho kết quả tốt, có thể sử dụng để định lượng trên các mẫu thực tế.

Bảng 3.13. Kết quả đánh giá độ thu hồi và RMSEP trên tập kiểm tra của Spirosemid, Zoamco-A và Natrxam

Mẫu	Độ thu hồi được tính theo mô hình PLS (%)					
	Spirosemid		Zoamco-A		Natrxam	
	SPR	FUR	AML	ATO	IND	AML
V1	99,3	97,0	102,2	105,2	101,0	102,8
V2	102,8	102,8	100,1	99,9	99,1	105,1
V3	101,5	104,6	99,5	98,0	99,4	105,9
V4	103,8	106,0	102,3	104,5	99,7	106,3
V5	101,5	100,5	104,5	99,7	103,9	102,5
V6	92,8	104,5	97,3	99,7	103,5	103,7
V7	106,3	108,0	103,4	100,9	103,7	100,8
V8	101,1	100,2	97,3	99,9	102,3	97,3
V9	103,0	103,8	99,0	103,1	96,9	100,3
Trung bình	101,3	103,1	100,6	101,2	101,1	102,7
RSD(%)	3,70	3,25	2,59	2,44	2,43	2,82
RMSEP (µg/mL)	0,389	0,379	0,259	0,547	0,123	1,153

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

Độ đúng được đánh giá thông qua độ thu hồi bằng cách thêm chuẩn ở 3 mức nồng độ tương ứng với mức thấp – trung bình – cao. Theo quy định của AOAC, với mức nồng độ dưới 100 ppm độ thu hồi cho phép nằm trong khoảng 90 – 110%. Kết quả từ bảng 3.14 cho thấy, 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 Natruxam độ thu hồi của IND là 96,7 – 104,7% và của AML là 101,1 – 104,6%.

Bảng 3.14. Kết quả đánh giá độ thu hồi các thuốc Spirosemid, Zoamco-A và Natruxam

Mẫu	Spirosemid				Zoamco-A				Natruxam			
	C _{SPR} ^T (µg/mL)	Rev (%)	C _{FUR} ^T (µg/mL)	Rev (%)	C _{AML} ^T (µg/mL)	Rev (%)	C _{ATO} ^T (µg/mL)	Rev (%)	C _{IND} ^T (µg/mL)	Rev (%)	C _{AML} ^T (µg/mL)	Rev (%)
1	3,0	106,7	2,4	103,8	8,0	105,6	16,0	105,1	3,6	101,4	12,0	101,2
2	3,0	107,0	2,4	104,2	8,0	108,1	16,0	106,4	3,6	100,6	12,0	101,4
3	3,0	103,0	2,4	104,2	8,0	108,0	16,0	107,2	3,6	104,7	12,0	104,6
Trung bình		105,6		104,1		107,3		106,2		102,2		102,4
RSD(%)		2,11		0,22		1,31		1,01		2,15		1,85
4	5,0	106,2	4,0	105,3	10,0	103,3	20,0	103,1	4,5	100,7	15,0	104,2
5	5,0	107,8	4,0	105,0	10,0	102,8	20,0	104,6	4,5	100,7	15,0	104,1
6	5,0	105,2	4,0	104,0	10,0	105,4	20,0	105,5	4,5	99,6	15,0	103,3
Trung bình		106,4		104,8		103,8		104,4		100,3		103,9
RSD(%)		1,23		0,65		1,33		1,16		0,64		0,498
7	7,0	102,3	5,6	101,4	12,0	101,6	24,0	99,1	5,4	96,7	18,0	103,3
8	7,0	99,6	5,6	104,1	12,0	102,5	24,0	101,2	5,4	97,6	18,0	102,8
9	7,0	98,4	5,6	103,9	12,0	102,3	24,0	103,1	5,4	96,7	18,0	103,2
Trung bình		100,1		103,1		102,1		101,2		97,0		103,1
RSD(%)		2,00		1,46		0,48		1,98		0,548		0,228

Bảng 3.15. Kết quả đánh giá độ lặp lại các thuốc Spirosemid, Zoamco-A và Natruxam

	Spirosemid		Zoamco-A		Natruxam	
	SPR	FUR	AML	ATO	IND	AML
Lần 1	9,57	7,57	10,67	21,26	4,44	14,29
Lần 2	10,23	8,03	10,31	20,08	4,38	14,42
Lần 3	9,79	7,72	10,38	21,02	4,36	14,47
Lần 4	9,59	7,65	10,27	20,24	4,36	14,41
Lần 5	9,62	7,7	10,57	21,26	4,28	14,55
Lần 6	9,77	7,89	10,42	20,99	4,3	14,53
Trung bình	9,76	7,76	10,44	20,81	4,35	14,45
SD	0,25	0,17	0,15	0,52	0,057	0,095
RSD (%)	2,54	2,18	1,48	2,49	1,32	0,655
$\frac{1}{2}$ RSD _H	5,66	5,85	5,66	5,1	6,38	5,32

Đơn vị: µg/mL

Đánh giá độ lặp lại bằng cách chuẩn bị mẫu được trình bày trong mục 3.3.2. Đối với thuốc Spironosemid: khối lượng thuốc được cân $m = 0,1199$ g, thể tích mẫu thuốc lấy để pha loãng là $200 \mu\text{L}$; Đối với thuốc Zoamco-A: $m_{\text{cân}} = 0,2814$ g, $V_{\text{mẫu}} = 1000 \mu\text{L}$; Đối với thuốc Natrixam: $m_{\text{cân}} = 0,6325$ g, $V_{\text{mẫu}} = 1000 \mu\text{L}$. Tiến hành lặp lại 6 lần từ cân mẫu, kết quả thể hiện trong bảng 3.15 cho thấy độ lặp lại có giá trị RSD nhỏ hơn $\frac{1}{2}$ RSD tính theo phương trình Horwitz, điều này cho thấy phương pháp có độ lặp lại tốt, ổn định trong quá trình thực hiện.

c. Kết quả định lượng

Bảng 3.16. Kết quả định lượng (mg/viên) các thành phần trong viên nén Spirosemid; Zoamco-A và Natrixam 1,5/5

Mẫu	Spirosemid				Zoamco-A				Natrixam			
	SPR		FUR		AML		ATO		IND (mg)		AML (mg)	
	PLS	HPLC	PLS	HPLC	PLS	HPLC	PLS	HPLC	PLS	HPLC	PLS	HPLC
1	51,14	49,75	40,13	39,65	5,34	5,2	10,63	10,49	1,45	1,49	4,8	4,73
2	48,96	49,15	38,58	39,97	5,16	5,24	10,04	10,54	1,49	1,49	4,88	4,8
3	48,84	50,02	39,45	40,25	5,19	5,17	10,51	10,46	1,47	1,51	4,82	4,75
Trung bình	49,65	49,64	39,39	39,96	5,23	5,2	10,39	10,5	1,47	1,5	4,83	4,76
SD	1,29	0,44	0,78	0,30	0,10	0,04	0,31	0,04	0,02	0,01	0,04	0,04
RSD(%)	2,6	0,9	1,97	0,75	1,84	0,68	3,00	0,38	1,36	0,77	0,86	0,76
$F_{\text{tính}}$	8,4		6,67		7,54		59,531		3,00		1,33	
$t_{\text{tính}}$	0,01		1,19		0,45		0,569		2,00		2,31	

Định lượng đồng thời các chất trong các mẫu thuốc với 3 lần lặp lại. Kết quả thể hiện trong bảng 3.16. Đối với viên nén Spirosemid: SPR: $49,65 \pm 3,21$ mg/viên; FUR: $39,39 \pm 1,93$ mg/viên; Đối với viên nén Zoamco-A: AML: $5,23 \pm 0,24$ mg/viên; ATO: $10,39 \pm 0,77$ mg/viên; Đối với viên nén Natrixam 1,5/5: IND: $1,47 \pm 0,05$ mg/viên; AML: $4,83 \pm 0,10$ 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 và trình bày trong bảng 3.17. 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 V.

Bảng 3.17. Kết quả hàm lượng phần trăm so với hàm lượng ghi trên nhãn của các thuốc Spirosemid, Zoamco-A và Natrixam

Mẫu	Spirosemid		Zoamco-A		Natrixam	
	SPR	FUR	AML	ATO	IND	AML
	(%)	(%)	(%)	(%)	(%)	(%)
1	102,3	100,3	106,8	106,3	96,7	96
2	97,9	96,5	103,2	100,4	99,3	97,6
3	97,7	98,6	103,8	105,1	98	96,4
Trung bình	99,3	98,5	104,6	103,9	98	96,7
DDVN V	90-110%					

Các mẫu thuốc cũng được gửi tại Trung tâm Kiểm nghiệm thuốc, Mỹ phẩm, Thực phẩm Thành phố Huế bằng phương pháp sắc ký lỏng hiệu năng cao. 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 ($t_{\text{tính}} < t_{\text{bảng}} = 2,775$).

3.2.6.2 Đối với thuốc có 3 thành phần

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

Bảng 3.18. Các thông số thống kê đánh giá trên tập hiệu chuẩn Turbezid bằng phương pháp PLS

Mẫu	Nồng độ tập hiệu chuẩn ($\mu\text{g/mL}$)			Nồng độ tính từ mô hình PLS ($\mu\text{g/mL}$)		
	RIF	INZ	PYR	RIF	INZ	PYR
T1	7,5	7,5	20,0	7,53	7,48	20,18
T2	10,5	7,5	20,0	10,66	7,34	20,44
T3	7,5	10,5	20,0	7,38	10,39	19,77
T4	10,5	10,5	20,0	10,5	10,89	19,98
T5	7,5	7,5	28,0	7,62	7,71	27,38
T6	10,5	7,5	28,0	10,51	7,6	28,14
T7	7,5	10,5	28,0	7,5	10,58	28,4
T8	10,5	10,5	28,0	10,48	10,63	28,17
T9	6,5	9,0	24,0	6,54	9,01	24,34
T10	11,5	9,0	24,0	11,39	8,83	23,69
T11	9,0	6,3	24,0	9,03	6,66	24,06
T12	9,0	11,5	24,0	9,01	11,25	23,93
T13	9,0	9,0	17,3	8,9	8,97	17,19
T14	9,0	9,0	30,7	8,98	8,72	30,13
T15	9,0	9,0	24	9,0	8,98	24,23
Khoảng bước sóng (nm)	220-440	220 - 425	220 - 300	220-440	220 - 425	220 - 300
Khoảng quét ($\Delta\lambda$)	0,5					
Số biến tiềm ẩn				3	4	3
RMSECV ($\mu\text{g/mL}$)				0,136	0,227	0,367
RMSEC ($\mu\text{g/mL}$)				0,107	0,184	0,317
Shapiro-Wilk test	W			0,9825	0,9816	0,9452
	p(>0,05)			0,888	0,8661	0,1254
R^2				0,9944	0,9836	0,9929
LOD_{min} ($\mu\text{g/mL}$)				0,43	0,87	1,30
LOD_{max} ($\mu\text{g/mL}$)				0,45	0,97	1,35

Thiết kế thí nghiệm cho tập hiệu chuẩn để định lượng đồng thời RIF, INZ và PYR trong mẫu Turbezid được trình bày trong bảng 3.18. Phổ hấp thụ UV-Vis của các mẫu hiệu chuẩn được đo trong khoảng 200 – 550 nm, khoảng quét là 0,5 nm. Để xác

định khoảng bước sóng để định lượng tốt mỗi thành phần trong mẫu Turbezid, chúng tôi đã thực hiện một chương trình lựa chọn khoảng bước sóng sao cho khả năng dự đoán của mô hình PLS áp dụng trên tập hiệu chuẩn là nhỏ nhất, chương trình được viết trên phần mềm R và được trình bày trong phần phụ lục 7. Kết quả cho thấy, với RIF khoảng bước sóng được chọn từ 220 – 440 nm, với INZ là 220 – 425 nm và PYR là 220 – 300 nm. Mô hình PLS đã được áp dụng để đánh giá khả năng hoạt động của mô hình PLS, kết quả được trình bày trong bảng 3.18.

Số biến tiềm ẩn được chọn lần lượt là 3, 4, 3 tương ứng với RIF, INZ và PYR. Các giá trị phần trăm của RMSECV và RMSEC của RIF lần lượt là 1,5%, 1,2%; của INZ là 2,5%, 2,0%; và của PYR là 1,5% và 1,3%. Kết quả cho thấy các %RMSE đều nhỏ hơn 2% (đòi với RIF và PYR) hoặc nằm trong khoảng từ 2-5% (đối với INZ). Như vậy mô hình PLS đã xây dựng cho kết quả tốt hoặc rất tốt.

Phân tích phần dư cho thấy các giá trị phần dư phân bố đều quanh điểm 0 và có phân phối chuẩn khi kiểm định Shapiro – Wilk ($p > 0,05$). Hệ số xác định xấp xỉ bằng 1 cho thấy khả năng dự đoán của mô hình là chính xác. Các giá trị giới hạn phát hiện thấp cho phép mô hình PLS kết hợp phổ UV-Vis có thể phát hiện và định lượng được các chất có nồng độ thấp.

Bảng 3.19. Kết quả định lượng RIF, INZ và PYR trong tập kiểm tra bằng phương pháp PLS

Mẫu	Nồng độ tập thẩm định ($\mu\text{g/mL}$)			Độ thu hồi (%)		
	RIF	INZ	PYR	RIF	INZ	PYR
V1	8,3	8,3	22,0	96,9	98,1	101,7
V2	9,8	8,3	22,0	99,9	96,0	98,8
V3	8,3	9,8	22,0	99,6	103,1	99,2
V4	9,8	9,8	22,0	100,2	98,4	97,3
V5	8,3	8,3	26,0	100,4	105,9	102,2
V6	9,8	8,3	26,0	101,9	100,7	101,4
V7	8,3	9,8	26,0	94,7	101,3	102,7
V8	9,8	9,8	26,0	99,5	98,9	102,0
V9	7,8	9,0	24,0	101,8	101,3	98,9
V10	10,3	9,0	24,0	105,0	102,9	99,8
V11	9,0	7,8	24,0	99,1	104,4	100,4
V12	9,0	10,3	24,0	100,7	103,3	98,8
V13	9,0	9,0	20,5	98,7	103,9	100,8
V14	9,0	9,0	27,5	102,1	101,1	100,1
V15	9,0	9,0	24,0	100,0	102,8	99,6
Trung bình (%)				100,0	101,5	100,2
RSD (%)				2,35	2,66	1,54
RMSEP ($\mu\text{g/mL}$)				0,209	0,271	0,384

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. Kết quả được thể hiện trong bảng 3.19, độ tìm lại 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.

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

Đánh giá độ đúng bằng cách thêm chuẩn của RIF, INZ và PYR ở các mức thấp tương ứng 7,7 – 7,7 – 20,4 µg/mL; mức trung bình: 9,0 – 9,0 – 24,0 µg/mL; mức cao: 10,4 – 10,4 – 27,6 µg/mL. 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.

Bảng 3.20. Kết quả đánh giá độ thu hồi RIF, INZ và PYR bằng phương pháp PLS

Mẫu	C _{RIF} ^T	Rev (%)	C _{INZ} ^T	Rev (%)	C _{PYR} ^T	Rev (%)
1	7,7	100,7	7,7	97,5	20,4	101,8
2	7,7	98,9	7,7	107,3	20,4	100,1
3	7,7	97,6	7,7	100	20,4	96,4
Trung bình		99,1		101,6		99,4
RSD (%)		1,57		5,01		2,78
4	9,0	100,3	9,0	95,4	24,0	95,9
5	9,0	100,4	9,0	101,2	24,0	98,0
6	9,0	100,3	9,0	108,2	24,0	100,1
Trung bình		100,3		101,6		98
RSD (%)		0,06		6,31		2,14
7	10,4	100,0	10,4	104,2	27,6	98,6
8	10,4	102,4	10,4	99,8	27,6	100,1
9	10,4	100,9	10,4	106,6	27,6	102,1
Trung bình		101,1		103,5		100,3
RSD (%)		1,2		3,33		1,75

Nồng độ được tính theo đơn vị µg/mL

Đánh giá độ lặp lại được thực hiện giữa hai ngày, trong đó khối lượng thuốc được lấy bằng 0,4743 g, thể tích dịch lọc được lấy là 60 µL để pha mẫu định lượng. Do phổ của INZ có trong mẫu thuốc đóng góp nhỏ vào phổ chung của hỗn hợp nên khi tiến hành định lượng INZ chúng tôi thêm chuẩn INZ vào trong mẫu thuốc và tiến hành đo mẫu. Kết quả đánh giá độ lặp lại được thể hiện qua bảng 3.21.

Giá trị trung bình của RIF, INZ và PYR trong đánh giá độ lặp lại lần lượt là 9,13, 4,57 và 24,53 $\mu\text{g/mL}$; trong đánh giá độ chính xác trung gian lần lượt là 9,10, 4,60 và 24,41 $\mu\text{g/mL}$. Kết quả đánh giá độ lặp tốt, giá trị RSD (%) đều nhỏ hơn giá trị RSD tính theo hàm Horwitz. So sánh độ lặp lại được thực hiện giữa 2 ngày khác nhau cho thấy $F_{\text{tính}}$ cho cả 3 chất đều nhỏ hơn $F_{\text{bảng}}$, chứng tỏ độ lặp lại giữa 2 lần thực hiện không có sự khác biệt có ý nghĩa thống kê. 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 thống kê với $t_{\text{tính}} < t_{\text{bảng}}$. Như vậy phương pháp phân tích đáp ứng yêu cầu về độ chính xác.

Bảng 3.21. Kết quả đánh giá độ lặp lại xác định RIF, INZ và PYR trong viên nén Turbezid bằng phương pháp PLS

Mẫu	Độ lặp lại (ngày 1)			Độ lặp lại (ngày 2)		
	RIF ($\mu\text{g/mL}$)	INZ ($\mu\text{g/mL}$)	PYR ($\mu\text{g/mL}$)	RIF ($\mu\text{g/mL}$)	INZ ($\mu\text{g/mL}$)	PYR ($\mu\text{g/mL}$)
R1	9,17	4,3	24,05	8,94	4,6	24,75
R2	8,97	4,63	24,66	9,23	4,66	25,21
R3	9,24	4,64	25,22	9,26	4,71	25,3
R4	9,28	4,65	25,3	8,9	4,37	23,26
R5	9,09	4,63	23,46	9,28	4,65	23,5
R6	9,04	4,56	24,48	9,01	4,6	24,46
Mean	9,13	4,57	24,53	9,1	4,6	24,41
RSD (%)	1,31	3,06	2,85	1,87	2,61	3,52
$\frac{1}{2}\text{RSD}_H$	5,57	6,36	4,96	5,57	6,36	4,96
$F_{\text{tính}}$				2,078	1,287	1,502
$F_{(5,5)(0,05)}$				5,820	5,820	5,820
$t_{\text{tính}}$				0,331	0,408	0,254
$t_{(0,05; 10)}$				2,228	2,228	2,228
p				0,748	0,692	0,805

c. Kết quả định lượng

Các mẫu thuốc Turbezid được chuẩn bị và đo độ hấp thụ UV-Vis, kết quả xác định hàm lượng của RIF, INZ và PYR trong viên nén ba thành phần được trình bày trong bảng 3.22. Khoảng tin cậy 95% của kết quả định lượng của RIF: $154,11 \pm 6,36$ mg/viên; INZ: $77,39 \pm 2,74$ mg/viên; PYR: $407,83 \pm 10,40$ 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 với phương pháp HPLC cho thấy, độ lặp lại khi xác định RIF không có sự khác biệt giữa 2 phương pháp ($F_{\text{tính}} = 7,734 < F_{\text{bảng}} = 15,439$), trong khi đó việc xác định INZ và PYR có độ lặp lại khác với với phương pháp HPLC. 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.

Bảng 3.22. Kết quả định lượng các thành phần trong viên turbezid (mg/viên) bằng phương pháp PLS

Mẫu	RIF		INZ		PYR	
	PLS	HPLC	PLS	HPLC	PLS	HPLC
1	151,17	156,48	76,17	76,86	412,67	402,86
2	155,33	154,95	78,33	76,62	405,33	403,95
3	155,83	154,83	77,67	76,39	405,5	402,69
Trung bình	154,11	155,42	77,39	76,62	407,83	403,17
RSD	1,66	0,59	1,43	0,31	1,03	0,17
%HLGTN	102,7	103,6	103,2	102,2	102	100,8
$F_{tính}$	7,734		22,182		37,550	
$F_{(2,2-0,05)}$	15,439		15,439		15,439	
$t_{tính}$	0,835		1,174		1,904	
$t_{(4-0,05)}$	2,776		4,303		4,303	

3.2.7 Kết luận phương pháp PLS

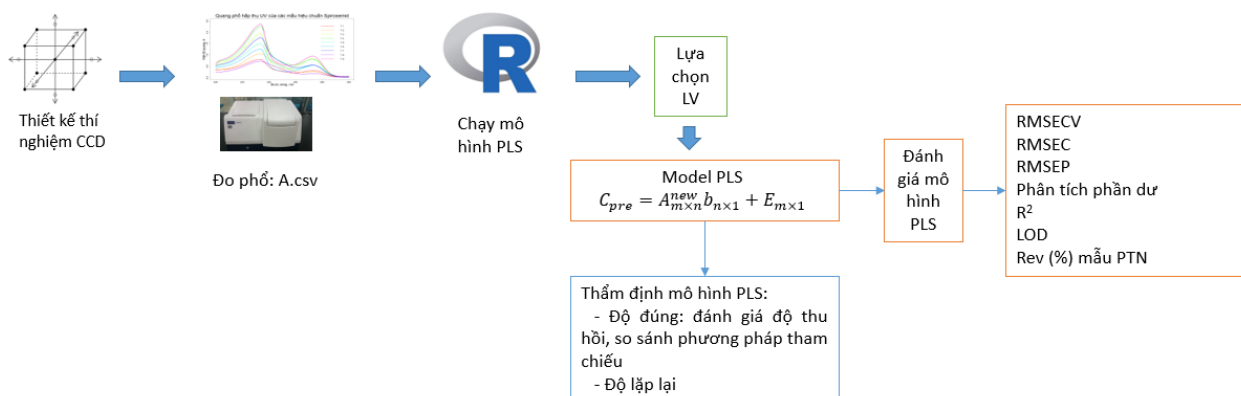
Phương pháp phân tích đa biến là xu hướng tất yếu của sự phát triển ứng dụng các phương pháp phân tích công cụ trong việc định tính cũng như định lượng. Nếu phân tích đơn biến chỉ thu thập được thông tin chỉ duy nhất một thành phần phân tích cần quan tâm và bỏ qua sự tương tác, ảnh hưởng của các yếu tố khác thì phân tích đa biến cho phép đánh giá được sự tác động qua lại giữa các yếu tố đó. Một phép phân tích đa biến cho phép tổng hợp được sự ảnh hưởng lẫn nhau khi nồng độ các thành phần trong mẫu phân tích thay đổi, các ảnh hưởng đó bao gồm sự tương tác lưỡng cực, liên kết Hidro giữa các phân tử, sự ảnh hưởng của lực Van der Waals giữa các phân tử; sự ảnh hưởng của quá trình không ổn định của thiết bị; ảnh hưởng của quá trình thực nghiệm... Phương pháp bình phương tối thiểu riêng phần là một trong những phương pháp phân tích đa biến được ứng dụng phổ biến. Trong nghiên cứu này, chúng tôi đã diễn giải được cách thực hiện của mô hình PLS cũng như ý nghĩa của các tham số. Phương pháp PLS đã thu gọn được số lượng các biến số một cách đơn giản nhất, giúp cho việc tính toán được dễ dàng, nhanh hơn. Việc trích xuất số biến tiềm ẩn là một bước rất quan trọng trong quá trình xây dựng mô hình PLS. Trong nghiên cứu này, chúng tôi đã thử nghiệm 3 cách để lựa chọn số thành phần chính, bao gồm dựa vào giá trị PRESS nhỏ nhất, dựa vào đánh giá thống kê để lựa chọn số biến tiềm ẩn là ít nhất nhưng giá trị PRESS không có sự khác biệt có ý nghĩa thống kê với giá trị PRESS nhỏ nhất và dựa vào đánh giá trên một tập thí nghiệm độc lập. Sự khác biệt giữa việc chọn số biến tiềm ẩn cho trường hợp thứ 2 và thứ 3 không khác nhau nhiều, trong nghiên cứu này chúng tôi lựa chọn trường hợp thứ 3 để lựa chọn số biến tiềm ẩn. Tuy nhiên, đối với trường hợp này thì số thí nghiệm sẽ tăng lên nhất nhiều, vì vậy trong trường hợp các mẫu có nhiều thành phần cần cân nhắc lựa chọn cho phù hợp.

Chúng tôi cũng đã xây dựng thành công 5 mô hình PLS để định lượng đồng thời các mẫu thuốc có 2 hoặc 3 thành phần. Đối với tập hiệu chuẩn, cần đánh giá độ tin cậy,

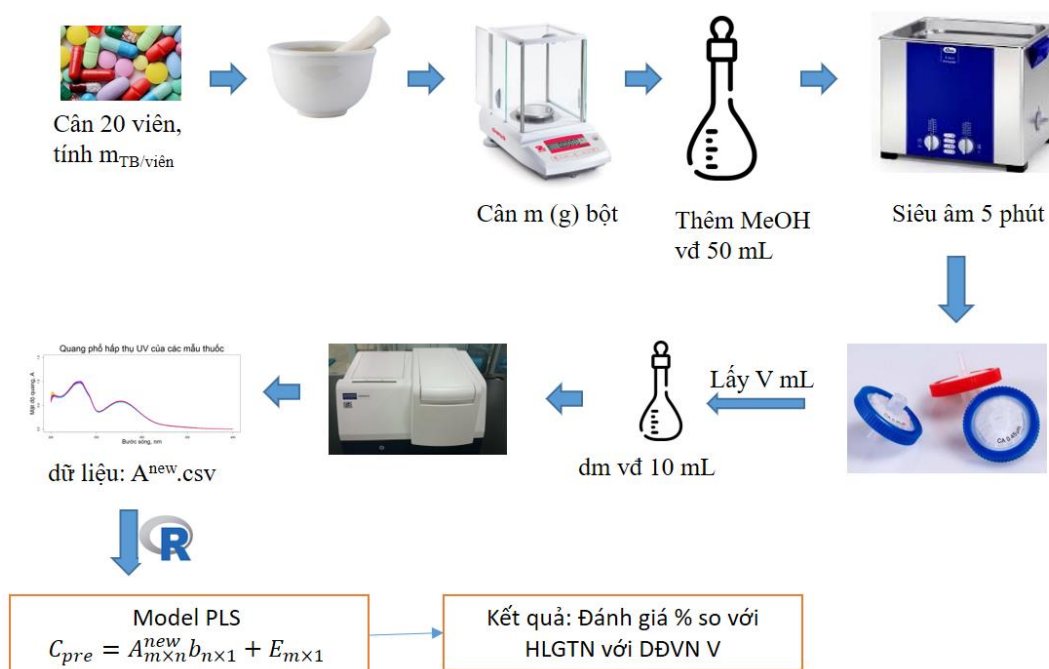
chính xác của phương pháp thông qua các tham số RMSECV, RMSEC, phân tích phần dư, xem xét hệ số xác định R^2 để chứng minh mô hình đã giải thích được bao nhiêu phần trăm của dữ liệu ban đầu. Đối với tập kiểm tra, đánh giá độ thu hồi của các mẫu được chuẩn bị trong phòng thí nghiệm ở các mức nồng độ và tỉ lệ nồng độ khác nhau. Để chứng minh qui trình phân tích đã xây dựng có phù hợp để phân tích các mẫu thực tế cần phải được chứng minh qua đánh giá độ đúng, độ chính xác của phương pháp. Kết quả cho thấy, các qui trình phân tích đã xây dựng xác định đúng kết quả hàm lượng của các mẫu thuốc đã nghiên cứu, phần trăm hàm lượng so với hàm lượng ghi trên nhãn của các nhà sản xuất đều nằm trong khoảng qui định của Dược điển Việt Nam V, so sánh kết quả với kết quả được xác định bằng các phương pháp tham chiếu cho thấy không có sự khác nhau có ý nghĩa thống kê.

Trên cơ sở đó, 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**



Như vậy, phương pháp PLS kết hợp với dữ liệu phổ UV-Vis đã thành công trong việc xác định các chất có phổ trùng nhau, có khả năng có ảnh hưởng của tá dược mà không cần phải xử lý. Sự thành công của phương pháp cho phép có thể được ứng dụng vào trong việc kiểm soát hàm lượng thuốc cho các trung tâm kiểm nghiệm, phòng thí nghiệm và các cơ sở sản xuất thuốc. Đồng thời tiếp tục nghiên cứu để có thể xây dựng thêm nhiều qui trình phân tích thuốc đa thành phần, áp dụng trong lĩnh vực phân tích Dược phẩm và có thể mở rộng ra cho các lĩnh vực khác.

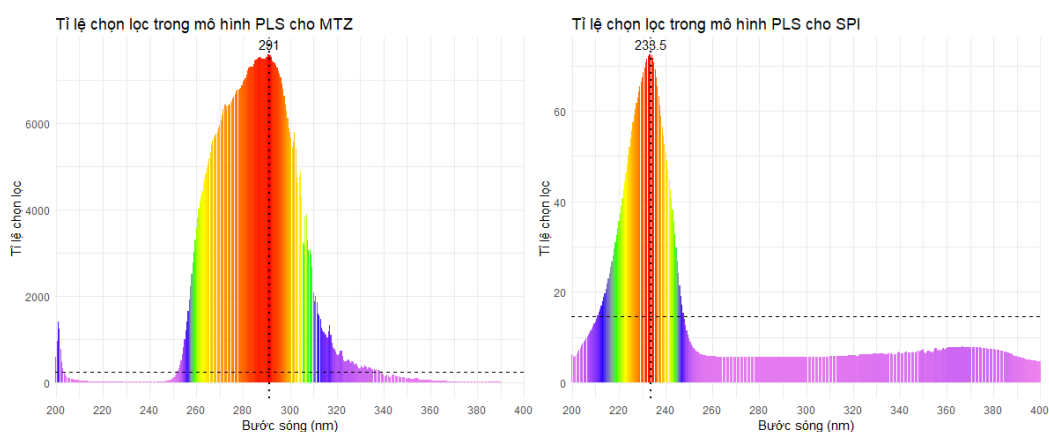
3.3 LỰA CHỌN BƯỚC SÓNG TRONG PLS

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

Theo kết quả phương pháp PLS áp dụng cho toàn phổ đối với MTZ và SPI, số thành phần chính tối ưu được chọn tương ứng là 2 và 7 để xây dựng mô hình hồi qui. Theo ý nghĩa của phương pháp SR, giá trị SR là tỷ lệ giữa phương sai giải thích và phương sai phần dư cho từng bước sóng, đại diện cho mức độ quan trọng của mỗi bước sóng trong mô hình PLS. Giá trị SR càng lớn thì mức độ của bước sóng đó càng quan trọng, ít bị ảnh hưởng của các thành phần khác trong mô hình; Giá trị SR càng nhỏ thì ở bước sóng đó bị ảnh hưởng bởi thông tin của các thành phần khác, kết quả dự đoán của mô hình gây sai số lớn. Mặt khác, giá trị SR tương tự như giá trị độ hấp thụ phổ UV, giá trị càng lớn biểu thị giữa giá trị phổ hấp thụ ở những bước sóng đó có mối liên hệ chặt chẽ với biến nồng độ được dự đoán.

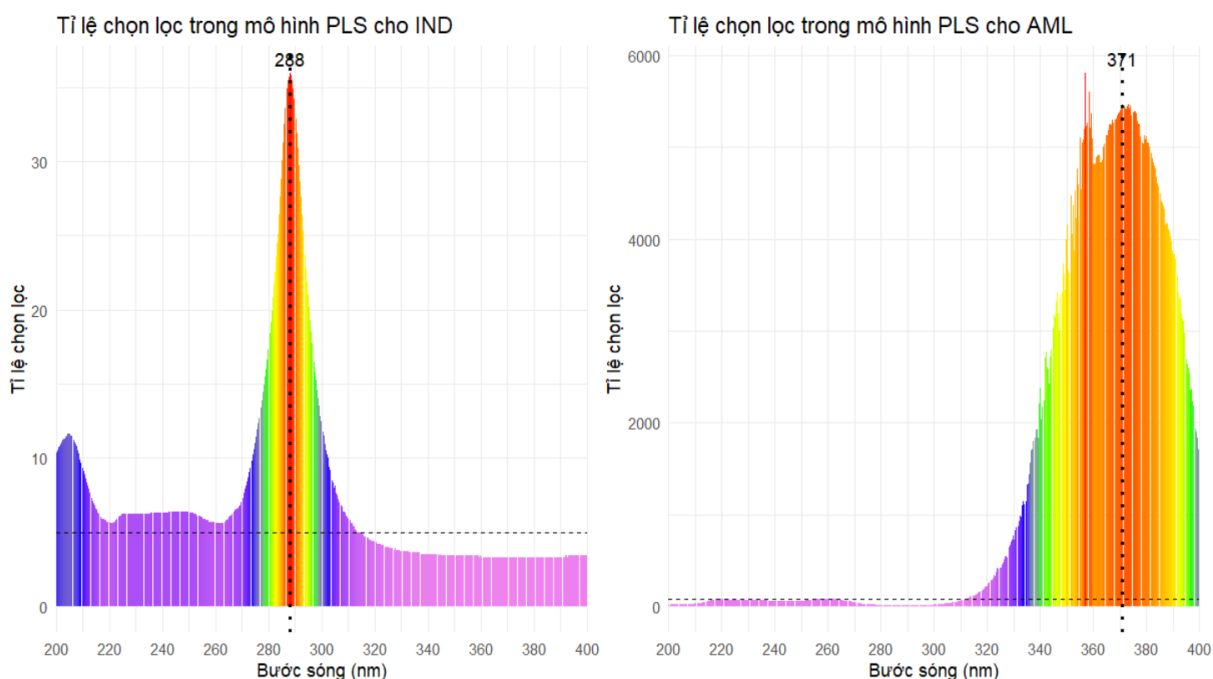
Theo kết quả khảo sát phân tính chất hấp thụ ánh sáng trong vùng UV từ 200 – 400 nm cho thấy, khoảng bước sóng hấp thụ mạnh của MTZ là khoảng 200 – 320 nm, phổ có đỉnh hấp thụ cực đại ở bước sóng 278 nm. Trong khi đó, SPI có khoảng bước sóng hấp thụ từ 200 - 250 nm với đỉnh cực đại hấp thụ ở bước sóng 245 nm. Sự xem phủ giữa hai phổ ở khoảng bước sóng 200 – 250 nm cho thấy sẽ có sự ảnh hưởng lẫn nhau trong khoảng bước sóng này khi xét đến hỗn hợp hai chất này.



Hình 3.10 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 tính được cho MTZ nhỏ nhất bằng 5,4 và lớn nhất là 7605, rất rõ ràng chúng ta có thể thấy ở khoảng bước sóng từ 250 đến 350 nm, giá trị SR có sự khác biệt rất lớn đối với vùng dưới 250 nm và trên 350 nm. Đối với vùng bước sóng 250 – 315, giá trị SR lớn hơn 1000 trong khi đó ở khoảng 205 – 250 và từ bước sóng lớn hơn 350 giá trị SR nhỏ hơn 10 (phụ lục 8). Đối với SPI, giá trị SR tính được nhỏ nhất bằng 4,6 và lớn nhất là 72,5 tại bước sóng 233,5 nm. Giá trị SR trong khoảng bước sóng 210 đến 247 nm lớn hơn rất nhiều đối với các bước sóng còn lại.

Đánh giá tương tự với IND và AML trong viên nén Natrixam 1,5/5. Theo mô hình PLS, số biến tiềm ẩn tối ưu cho mô hình là 3 đối với IND và 4 đối với AML. Phổ UV-Vis được đo trong khoảng 200 – 400 nm với khoảng quét là 0,2 nm. Kết quả đánh giá giá trị SR tại các bước sóng khác nhau của mỗi chất IND và AML được thể hiện trong hình 3.11.



Hình 3.11. Kết quả giá trị SR cho IND và AML trong mô hình PLS đã xây dựng

Đỉnh SR của IND và AML nằm trong vùng có độ hấp thụ UV-Vis mạnh của IND và AML. 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.

Điều này cho thấy khi xác định nồng độ MTZ và SPI trong hỗn hợp, các vùng phổ từ bước sóng 250 – 315 nm (đối với MTZ) và từ bước sóng 210 đến 247 nm (đối với SPI) là các bước sóng quan trọng; Tương tự với IND và AML, kết quả SR cho thấy từ bước sóng 200 – 320 nm đối với IND và từ 320- 400 nm đối với AML là các bước sóng quan trọng, có ý nghĩa trong việc xây dựng mô hình hồi qui, ở các bước sóng khác

bị ảnh hưởng của nhiễu (bao gồm nhiễu do nhiều nguồn khác nhau như thiết bị, hóa chất, phương pháp ...) hoặc do của sự xen phủ phổ giữa MTZ và SPI hoặc giữa IND và AML. Vùng bước sóng này gần như trùng khớp với vùng hấp thụ mạnh các thành phần trong mẫu đơn chất. Với kết quả giá trị SR trong mô hình PLS có phân loại khá rõ ràng đối với mỗi thành phần có trong mẫu hỗn hợp.

Từ kết quả tính toán SR cho hai hỗn hợp SPI – MTZ và IND – AML, chúng ta có thể rút ra những kết luận quan trọng về khả năng phân biệt của các chất này trong mô hình PLS. Sự khác biệt trong SR của các thành phần trong hỗn hợp tại các bước sóng khác nhau có thể cung cấp thông tin giá trị trong việc thiết kế các phương pháp phân tích UV cho các chất này. Mô hình PLS kết hợp với phân tích SR giúp tối ưu hóa việc lựa chọn bước sóng và làm cho quá trình phân tích trở nên hiệu quả hơn.

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

Phương pháp lựa chọn ngưỡng SR theo phương pháp đề xuất, chúng tôi tạo bước nhảy cho các giá trị SR trong mô hình này bằng 10. Như vậy, đối với MTZ có giá trị SR nhỏ nhất là 5,4 và lớn nhất 7605 có 760 khoảng SR được xem xét; đối với SPI có giá trị SR nhỏ nhất là 4,6 và lớn nhất là 72,6 có 7 khoảng được xem xét. Kết quả cho thấy, ngưỡng SR cho MTZ và SPI để lựa chọn bước sóng phù hợp lần lượt là 245,4 và 14,6 tương ứng với đường màu đen ngang trong hình 3.10.

Đối với hỗn hợp IND và AML, giá trị nhỏ nhất và lớn nhất của IND lần lượt là 3,3 và 36, của AML lần lượt là 6,9 và 5814,9. Ngưỡng tương ứng của IND và AML để lựa chọn bước sóng quan trọng là 5 và 76,9. Kết quả lựa chọn bước sóng được trình bày trong bảng 3.23

Bảng 3.23. Kết quả lựa chọn bước sóng theo phương pháp SR cho MTZ và SPI trong mô hình PLS

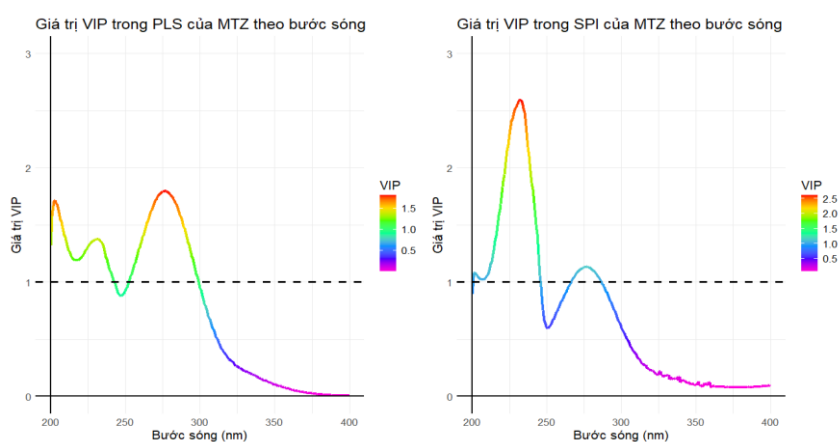
Thuốc	Thành phần	Số bước sóng được chọn	số bước sóng ban đầu	tỉ lệ (%)
AgimDogyl	MTZ	181 (252 – 338,5 và 200 – 203 nm)	401	45,1%
	SPI	74 (211 – 247,5 nm)	401	18,4%
Natrixam 1,5/5	IND	569 (200 – 313,6 nm)	1001	56,8%
	AML	514 (219 – 222 và 300,6 – 400 nm)	1001	51,3%

Như vậy, phương pháp SR đã loại bỏ được một số lượng bước sóng khá lớn có khả năng ảnh hưởng đến khả năng tính toán của mô hình PLS. Đối với hỗn hợp MTZ và SPI, số lượng bước sóng được chọn của MTZ là 182/401 bước sóng, chiếm 45,1% so với tổng số lượng bước sóng ban đầu, trong khi đó, số lượng bước sóng được giữ lại với SPI là 74 bước sóng, chiếm 18,4%. Đối với hỗn hợp IND và AML, số lượng bước sóng được giữ lại tương ứng chiếm lần lượt là 56,8% và 51,3%.

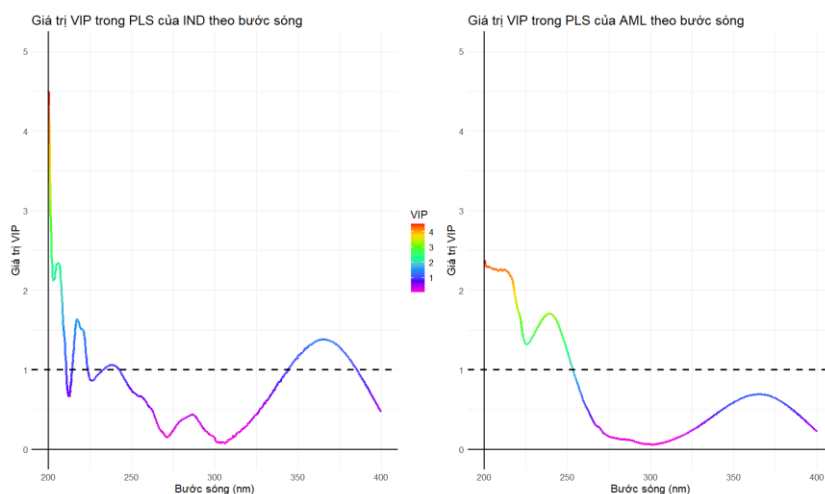
3.3.2 Lựa chọn bước sóng bằng phương pháp loại bỏ biến ngược kết hợp bình phương tối thiểu riêng phần (BVE-PLS)

3.3.2.1 Kết quả tính giá trị VIP

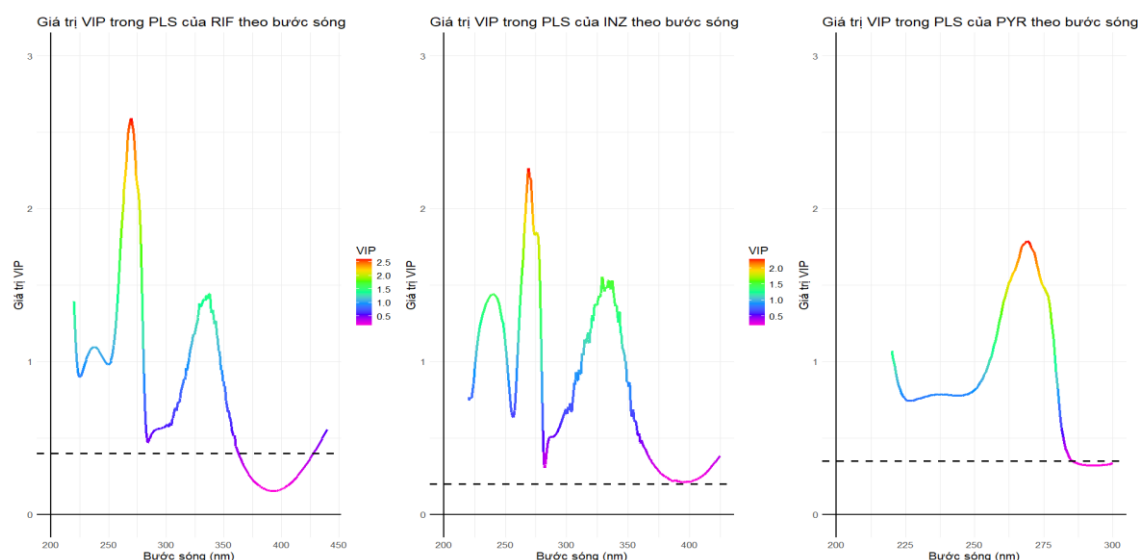
Trong phương pháp loại bỏ biến ngược, chúng tôi áp dụng để tính cho ba mẫu thuốc AgimDogyl, Natrixam 1,5/5 và Turbezid. Phương pháp lựa chọn biến ngược dựa trên kết quả của mô hình PLS đã xây dựng cho các tập hiệu chuẩn trên, tại các biến tiềm ẩn được lựa chọn tiến hành tính các giá trị VIP cho mỗi bước sóng. Với mẫu thuốc AgimDogyl có số biến tiềm ẩn đã được chọn cho MTZ và SPI lần lượt là 2 và 7, mẫu thuốc Natrixam 1,5/5 số biến tiềm ẩn được chọn cho IND và AML lần lượt là 4 và 2, mẫu thuốc Turbezid có số biến tiềm ẩn được chọn cho RIF, INZ và PYR lần lượt là 3, 4 và 3. Các giá trị VIP được tính và thể hiện trong hình 3.12; hình 3.13; hình 3.14.



Hình 3.12. 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 3.13. Các giá trị VIP được tính cho IND và AML trong mô hình PLS tương ứng với số biến tiềm ẩn là 4 và 2.



Hình 3.14. 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

Tương tự trong phương pháp SR-PLS, ngưỡng để lựa chọn bước sóng cho phương pháp BVE-PLS là một giá trị rất quan trọng. Tùy vào mỗi tập mẫu mà ngưỡng để lựa chọn biến có giá trị khác nhau. Kết quả lựa chọn bước sóng bằng phương pháp BVE-PLS được thể hiện trong bảng 3.24

Kết quả lựa chọn bước sóng bằng phương pháp BVE-PLS đối với các tập dữ liệu UV-Vis cho các mẫu thuốc AgimDogyl và Natrixam 1,5/5 đã loại bỏ được một số lượng lớn các bước sóng trong dữ liệu phổ gốc ban đầu. Đối với tập dữ liệu trong mẫu thuốc AgimDogyl, số lượng bước sóng được chọn đối với MTZ là 180 bước sóng bao gồm các bước sóng từ 200 – 241,5 nm; 252 – 299 nm chiếm 44,9% so với tổng số lượng bước sóng ban đầu (401 bước sóng), các bước sóng được chọn cùng nằm trong vùng hấp thụ mạnh của MTZ. Đối với SPI, số lượng bước sóng được chọn là 90 (chiếm 22,4%), bao gồm các bước sóng từ 200,5 – 245,5 nm và 265 – 288 nm. Vùng có VIP mạnh chủ yếu tập trung trong khoảng 200,5 – 245,5 nm tương ứng với vùng hấp thụ mạnh của SPI. Đối với tập dữ liệu trong mẫu thuốc Natrixam 1,5/5, số bước sóng được giữ lại cho IND và AML lần lượt là 37,0% và 26,7% so với tổng số 1001 bước sóng ban đầu. Kết quả này cho thấy rằng, thuật toán BVE dựa trên việc tính toán giá trị VIP để lựa chọn các bước sóng phù hợp với các khoảng hấp thụ tương ứng với các chất có trong mẫu hỗn hợp, có thể loại bỏ được các yếu tố ảnh hưởng cũng như các vùng quang phổ không có đóng góp nhiều vào trong phổ hỗn hợp, góp phần giảm được sự ảnh hưởng qua lại giữa các chất trong phổ chung của các mẫu hiệu chuẩn.

Bảng 3.24. 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 nm)	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%

Đối với mẫu Turbezid, chỉ có RIF là đã loại bỏ một số lượng đáng kể số bước sóng so với lượng bước sóng thu thập dữ liệu ban đầu, tuy nhiên với INZ và PYR thì lượng bước sóng bị loại bỏ không đáng kể. Điều này có thể giải thích là do trong quá trình thực hiện để xác định hàm lượng các chất trong mẫu bằng phương pháp PLS đã loại bỏ một số bước sóng ảnh hưởng đến hiệu quả thực hiện của mô hình PLS, chính vì vậy các bước sóng được giữ lại để huấn luyện cho mô hình BVE-PLS hầu như đều có ý nghĩa quan trọng cho mô hình, chính vì vậy việc áp dụng BVE-PLS lúc này không loại bỏ đáng kể các bước sóng.

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

Bảng 3.25. Các thông số cho GA-PLS để lựa chọn bước sóng

Tham số	Giá trị
Số bit trong mỗi cá thể của quần thể (Size)	Số bước sóng
Kích thước quần thể (Popsiz)	100
Số thế hệ của thuật toán (Iters)	50
Mô hình hồi qui	PLS
Xác suất đột biến (mutationChance)	1/(n+1)
Số cá thể giữ lại mà không bị biến đổi trong mỗi thế hệ (elitism)	Popsiz/5
Tỷ lệ giữa số gen mang giá trị 0 và 1 khi khởi tạo quần thể (zeroToOneRatio = GA.Threshold)	10

Thuật toán di truyền được triển khai là một thuật toán di truyền nhị phân. Nó được sử dụng để tối ưu hóa một hàm đánh giá (fitness function) bằng cách tiến hóa một

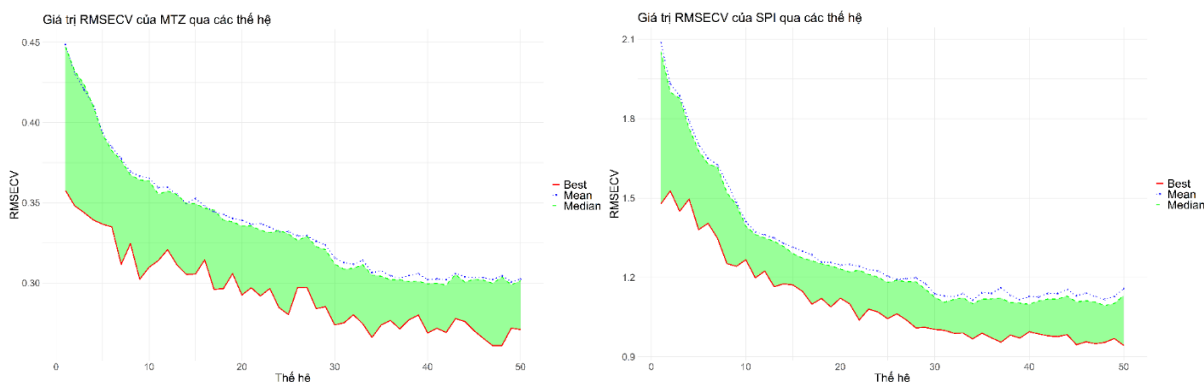
quần thể nhiễm sắc thể (chromosome) mã hóa nhị phân (chuỗi bit 0 và 1). Thuật toán di truyền được áp dụng cho tập dữ liệu phân tích thuốc AgimDogyl, Natrixam 1,5/5 và Turbezid. Các thông số khởi tạo cho thuật toán được trình bày trong bảng 3.25.

Thuật toán di truyền kết hợp phương pháp bình phương tối thiểu riêng phần được khởi tạo bằng các cá thể có số bit bằng với số bước sóng để xây dựng mô hình PLS, cụ thể, với tập dữ liệu AgimDogyl là 401, Natrixam là 1001, Turbezid tương ứng với mỗi chất RIF, INZ và PYR lần lượt là 441, 411 và 161. Số cá thể được khởi tạo đầu tiên là 100 cá thể được tạo ngẫu nhiên từ các dữ liệu gốc ban đầu và có tỉ lệ giữa bước sóng được chọn và không được chọn bằng 10. Với mỗi cá thể (tương ứng với một tập dữ liệu) sử dụng mô hình PLS để huấn luyện và đánh giá giá trị RMSE tương ứng với mỗi thành phần, các giá trị RMSE nhỏ nhất của trong mỗi thế hệ được ghi lại trong bảng 3.26. Thuật toán GA-PLS được lặp lại 50 lần với xác suất đột biến qua mỗi thế hệ là $\frac{1}{n+1}$, số cá thể giữ lại mà không bị biến đổi là 20. Các hình 3.15; hình 3.16; hình 3.17 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. Điều này thể hiện rằng với khoảng 50 thế hệ di truyền thì hầu như không có xu hướng thay đổi giá trị RMSE hơn nữa, có nghĩa rằng các bước sóng đã được chọn ở các thế hệ cuối là những bước sóng có thông tin hữu ích nhất để xây dựng mô hình PLS dựa trên các tập huấn luyện đã được xây dựng. Tần suất các bước sóng được chọn qua 50 thế hệ được trình bày trong phụ lục 9.

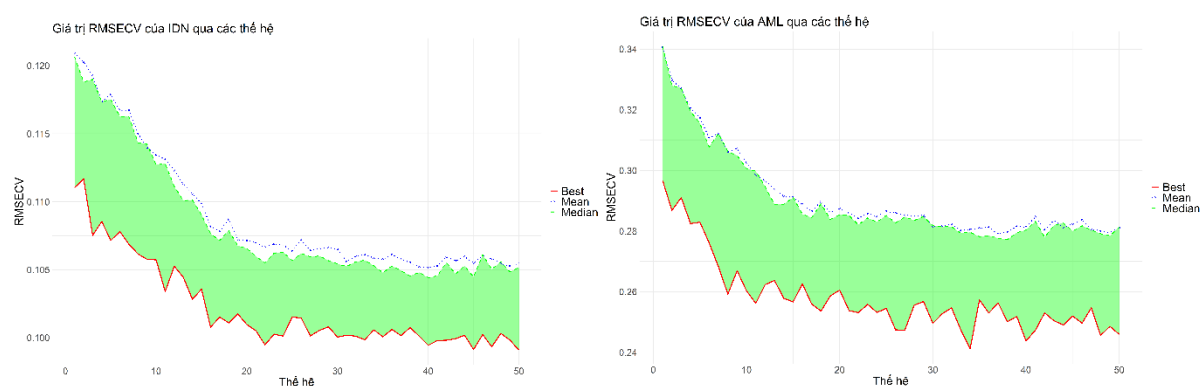
Bảng 3.26. Giá trị RMSE ($\mu\text{g/mL}$) tốt nhất nhất qua 50 thế hệ của thuật toán di truyền

Thế hệ	AgimDogyl		Natrixam 1,5/5		Turbezid		
	MTZ	SPI	IDN	AML	RIF	IND	PYR
1	0,358	1,477	0,111	0,297	0,096	0,129	0,128
2	0,348	1,527	0,112	0,287	0,096	0,132	0,116
3	0,344	1,450	0,108	0,291	0,097	0,127	0,121
4	0,339	1,496	0,108	0,283	0,097	0,126	0,119
5	0,337	1,380	0,107	0,283	0,090	0,124	0,116
6	0,335	1,404	0,108	0,276	0,093	0,114	0,111
7	0,312	1,348	0,107	0,268	0,089	0,117	0,109
8	0,325	1,252	0,106	0,259	0,091	0,115	0,108
9	0,302	1,242	0,106	0,267	0,090	0,110	0,103
10	0,310	1,267	0,106	0,260	0,086	0,107	0,102
11	0,314	1,198	0,103	0,256	0,087	0,103	0,099
12	0,321	1,224	0,105	0,262	0,085	0,101	0,095
13	0,311	1,165	0,104	0,264	0,085	0,102	0,096
14	0,305	1,176	0,103	0,258	0,087	0,093	0,097

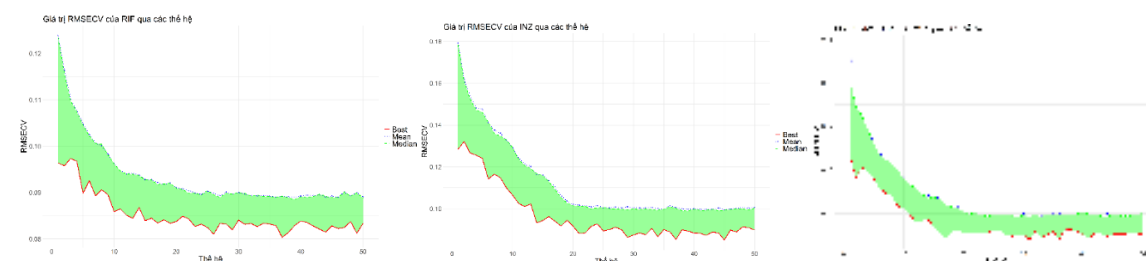
Thế hệ	AgimDogyl		Natrixam 1,5/5		Turbezid		
	MTZ	SPI	IDN	AML	RIF	IND	PYR
15	0,306	1,171	0,104	0,257	0,084	0,094	0,093
16	0,314	1,147	0,101	0,263	0,085	0,096	0,091
17	0,296	1,099	0,102	0,256	0,083	0,094	0,089
18	0,296	1,121	0,101	0,254	0,084	0,092	0,091
19	0,306	1,088	0,102	0,259	0,083	0,094	0,093
20	0,293	1,122	0,101	0,261	0,084	0,092	0,091
21	0,297	1,100	0,101	0,254	0,085	0,088	0,091
22	0,292	1,039	0,100	0,253	0,084	0,088	0,091
23	0,297	1,080	0,100	0,256	0,083	0,092	0,091
24	0,285	1,069	0,100	0,253	0,083	0,093	0,091
25	0,280	1,044	0,102	0,255	0,082	0,089	0,092
26	0,297	1,062	0,101	0,247	0,081	0,090	0,092
27	0,297	1,039	0,100	0,247	0,083	0,091	0,090
28	0,284	1,009	0,101	0,256	0,083	0,090	0,090
29	0,285	1,011	0,101	0,257	0,082	0,086	0,090
30	0,274	1,003	0,100	0,250	0,084	0,087	0,090
31	0,275	1,000	0,100	0,253	0,083	0,088	0,088
32	0,280	0,989	0,100	0,255	0,083	0,088	0,092
33	0,275	0,989	0,100	0,247	0,083	0,091	0,089
34	0,266	0,967	0,101	0,241	0,083	0,087	0,091
35	0,274	0,989	0,100	0,257	0,083	0,090	0,092
36	0,277	0,972	0,101	0,253	0,083	0,089	0,088
37	0,271	0,955	0,100	0,256	0,080	0,085	0,093
38	0,277	0,982	0,101	0,250	0,081	0,090	0,093
39	0,280	0,971	0,100	0,252	0,083	0,089	0,092
40	0,269	0,994	0,099	0,244	0,084	0,088	0,090
41	0,272	0,986	0,100	0,247	0,084	0,088	0,089
42	0,269	0,979	0,100	0,253	0,083	0,087	0,091
43	0,278	0,976	0,100	0,250	0,082	0,089	0,091
44	0,276	0,984	0,100	0,249	0,082	0,088	0,089
45	0,270	0,945	0,099	0,252	0,083	0,085	0,091
46	0,265	0,956	0,100	0,250	0,082	0,090	0,091
47	0,261	0,949	0,099	0,255	0,083	0,089	0,091
48	0,261	0,954	0,100	0,246	0,084	0,091	0,091
49	0,272	0,969	0,100	0,249	0,081	0,091	0,091
50	0,271	0,943	0,099	0,246	0,083	0,090	0,091



Hình 3.15. Giá trị RMSECV qua 50 thể hệ di truyền của MTZ và SPI



Hình 3.16. Giá trị RMSECV qua 50 thể hệ di truyền của IND và MTZ



Hình 3.17. Giá trị RMSECV qua 50 thể hệ di truyền của RIF, INZ và PYR

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

Thuật toán di truyền kết hợp PLS là thuật toán dựa trên việc lấy ngẫu nhiên các bước sóng khác nhau và đánh giá hiệu quả hoạt động của mô hình thông qua đánh giá sai số trên các tập dữ liệu thu được. Việc lựa chọn không phải dựa trên một ngưỡng mà dựa trên việc tạo ra tổ hợp các mẫu dữ liệu mới với các bước sóng được lấy ngẫu nhiên từ các vị trí khác nhau. Kết quả cho thấy các bước sóng được chọn ở cá thể tối ưu của thể hệ cuối cùng không phải là liên tục, không nằm cạnh nhau hoặc không phải luôn nằm trong khoảng hấp thụ tối ưu tương ứng với độ hấp thụ mạnh của các đơn chất mà các bước sóng được chọn là tổ hợp các bước sóng sao cho khả năng đánh giá bằng phương pháp thẩm định chéo mà sai số là nhỏ nhất. Các bước sóng được chọn là các bước sóng có tần suất xuất hiện nhiều nhất ở giai đoạn cuối cùng và được mô tả trong hình của thể hệ 50 trong phần phụ lục.

Bảng 3.27. Số lượng và tỉ lệ các bước sóng được chọn của các tập dữ liệu AgimDogyl, Natrixam và Turbezid bằng phương pháp GA-PLS

Mẫu thuốc	Thành phần	Số bước sóng ban đầu	Số bước sóng được chọn	Tỉ lệ
AgimDogyl	MTZ	401	49	12,2%
	SPI	401	34	8,5%
Natrixam 1,5/5	IND	1001	105	10,5%
	AML	1001	105	10,5%
Turbezid	RIF	441	42	9,5%
	INZ	411	39	9,5%
	PYR	161	14	8,7%

Kết quả từ bảng 3.27 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

Kết quả đánh giá các phương pháp SR-PLS, BVE-PLS và GA-PLS cho thấy sự khác biệt về hiệu suất khi tính toán nồng độ MTZ và SPI, được thể hiện trong bảng 3.28. Phương pháp SR-PLS vượt trội hơn cả về độ chính xác và khả năng phù hợp với dữ liệu. Cụ thể, SR-PLS có RMSECV và RMSEC thấp nhất, với các giá trị lần lượt là 0,500 và 0,445 cho MTZ, và 1,125 và 0,628 cho SPI, cho thấy sự chính xác cao trong việc dự đoán nồng độ. Bên cạnh đó, SR-PLS cũng có R^2 cao (0,9910 cho MTZ và 0,9995 cho SPI), chứng tỏ mô hình có độ phù hợp rất tốt với dữ liệu thực tế. Hơn nữa, với LOD_{min} thấp nhất (0,94 cho MTZ và 4,97 cho SPI), SR-PLS có khả năng phát hiện nồng độ thấp nhất trong các phương pháp.

BVE-PLS mặc dù không đạt được mức độ chính xác cao như SR-PLS, nhưng vẫn có hiệu suất tốt với RMSECV và RMSEC ở mức chấp nhận được (0.529 và 0.474 cho MTZ, 1.398 và 0.868 cho SPI). Phương pháp này cũng cho thấy R^2 cao (0.9898 cho MTZ và 0.9995 cho SPI), nhưng thấp hơn SR-PLS.

Bảng 3.28. Kết quả đánh giá các tham số thông kê đánh giá trên tập hiệu chuẩn bởi các mô hình lựa chọn bước sóng cho tập hiệu chuẩn AgimDogyl

Mẫu	Nồng độ tập hiệu chuẩn		Nồng độ tính được từ mô hình					
			Phương pháp SR-PLS		Phương pháp BVE-PLS		Phương pháp GA-PLS	
	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)
T1	10,0	60,0	9,75	59,67	9,84	59,83	9,85	59,24
T2	20,0	60,0	20,26	59,41	20,33	59,67	20,34	60,27
T3	10,0	120,0	9,87	120,02	10,17	119,84	10,18	120,26
T4	20,0	120,0	20,20	119,94	20,15	119,94	20,14	119,96
T5	8,0	90,0	8,00	90,11	7,70	90,28	7,69	90,65
T6	22,0	90,0	21,14	90,39	21,08	90,28	21,06	90,13
T7	15,0	48,0	15,16	48,39	15,16	47,99	15,16	47,97
T8	15,0	132,0	15,13	131,68	15,08	131,7	15,09	131,30
T9	15,0	90,0	15,50	90,40	15,48	90,48	15,49	90,22
n			2	5	2	5	2	4
RMSECV ($\mu\text{g/mL}$)			0,500	1,125	0,529	0,889	0,538	1,159
RMSEC ($\mu\text{g/mL}$)			0,445	0,628	0,474	0,642	0,482	0,868
Shapiro-Wilk	W		0,8985	0,9769	0,925	0,961	0,926	0,951
	p(>0,05)		0,243	0,786	0,053	0,401	0,055	0,223
R ²			0,9910	0,9995	0,9898	0,9995	0,9895	0,9990
LOD _{min} ($\mu\text{g/mL}$)			0,94	4,97	0,96	14,80	0,94	8,74
LOD _{max} ($\mu\text{g/mL}$)			1,09	5,53	1,05	16,60	1,04	9,94

Trong khi đó, GA-PLS có hiệu suất thấp hơn đáng kể, với RMSECV và RMSEC cao nhất (0.889 và 0.482 cho MTZ, 1.159 và 2.184 cho SPI). Mặc dù giá trị R² của GA-PLS vẫn đạt được mức khá (0.9895 cho MTZ và 0.9990 cho SPI), nhưng khả năng phát hiện nồng độ thấp của phương pháp này (LOD_{min} cao nhất) kém hơn so với SR-PLS và BVE-PLS.

Tóm lại, SR-PLS là phương pháp ưu việt nhất trong việc tính toán nồng độ MTZ và SPI, với độ chính xác cao, khả năng phát hiện tốt và độ phù hợp cao. BVE-PLS và GA-PLS cũng cho kết quả khả quan

Bảng kết quả 3.29 đánh giá hiệu quả hoạt động của ba phương pháp SR-PLS, BVE-PLS và GA-PLS trong việc lựa chọn bước sóng trên tập nồng độ kiểm tra của thuốc AgimDogyl. Các chỉ số chính được xem xét bao gồm độ thu hồi (%), RSD, và RMSEP. Những chỉ số này giúp đánh giá độ chính xác, độ ổn định và độ sai lệch của các mô hình khi dự đoán nồng độ MTZ và SPI từ các mẫu thử nghiệm.

Độ thu hồi của các phương pháp có sự khác biệt không đáng 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. Các giá trị trung bình đều gần với 100% cho thấy các phương pháp lựa chọn bước sóng cho kết quả dự đoán nồng độ tốt đối với MTZ và SPI trong các mẫu có nồng độ khác nhau được chuẩn bị trong phòng thí nghiệm.

Bảng 3.29. Kết quả đánh giá hiệu của hoạt động của các mô hình lựa chọn bước sóng trên tập nồng độ kiểm tra của thuốc AgimDogyl

Mẫu	Độ thu hồi (%)							
	Nồng độ tập kiểm tra		Phương pháp SR-PLS		Phương pháp BVE-PLS		Phương pháp GA-PLS	
	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ	SPI	MTZ	SPI	MTZ	SPI
V1	12,5	75,0	100,4	99,9	98,9	99,5	98,7	98,8
V2	17,5	75,0	103,0	99,7	100,6	99,3	100,5	99,0
V3	12,5	105,0	101,7	99,3	99,4	99,2	99,3	97,6
V4	17,5	105,0	102,5	102,1	100,1	101,8	100,1	100,6
V5	11,5	90,0	102,8	100,2	99,1	99,8	99,2	103,0
V6	18,5	90,0	104,9	99,6	102,4	99,4	102,4	99,6
V7	15,0	69,0	104,1	97,4	101,5	97,1	101,4	99,1
V8	15,0	111,0	103,1	99,2	103,8	97,9	103,9	95,6
V9	15,0	90,0	104,8	98,5	104,1	98,6	104,1	98,5
Trung bình			103,0	99,5	101,1	99,2	101,1	99,1
RSD (%)			1,41	1,28	1,89	1,33	2,00	2,03
RMSEP ($\mu\text{g/mL}$)			0,531	1,14	0,330	1,40	0,336	2,18

RMSEP đo độ sai lệch giữa giá trị dự đoán và giá trị thực tế, với giá trị càng thấp càng cho thấy mô hình càng chính xác. 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 đúng

Kết quả từ bảng 3.30 đánh giá độ đúng khi sử dụng các phương pháp SR-PLS, BVE-PLS và GA-PLS để xác định nồng độ của MTZ và SPI ở ba mức nồng độ (thấp, trung bình, và cao) cho thấy sự chênh lệch đáng kể về hiệu quả của từng phương pháp.

Bảng 3.30. Kết quả đánh giá độ thu hồi của MTZ và SPI bằng các phương pháp lựa chọn bước sóng SR-PLS, BVE-PLS, GA-PLS

Mẫu	Nồng độ thêm chuẩn		Độ thu hồi (%)					
			Phương pháp SR-PLS		Phương pháp BVE-PLS		Phương pháp GA-PLS	
	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ	SPI	MTZ	SPI	MTZ	SPI
1	12,0	72,0	99,3	101,5	96,9	97,9	96,6	104,1
2	12,0	72,0	96,2	101,4	92,2	99,1	91,9	102,3
3	12,0	72,0	93,5	102,5	91,3	100,9	90,5	101,5
Trung bình			96,3	101,8	93,5	99,3	93,0	102,6
RSD (%)			3,01	0,6	3,22	1,52	3,44	1,3
4	15,0	90,0	98,9	103,5	98,7	96,1	98,3	104,2
5	15,0	90,0	98,1	103,0	97,3	98,3	96,7	103,8
6	15,0	90,0	100,9	101,0	100,5	100,6	100,1	101,8
Trung bình			99,3	102,5	98,8	98,3	98,4	103,3
RSD (%)			1,45	1,29	1,62	2,29	1,73	1,24
7	18,0	108,0	100,2	106,0	97	97,6	96,6	102,4
8	18,0	108,0	97,3	103,8	95,2	94,5	94,9	103,4
9	18,0	108,0	100,4	104,6	97,4	96,3	97,1	103,5
Trung bình			99,3	104,8	96,5	96,1	96,2	103,1
RSD (%)			1,75	1,06	1,21	1,62	1,2	0,59

Phương pháp SR-PLS có độ thu hồi trung bình khá ổn định với MTZ đạt 96,3% và SPI đạt 101,8% ở mức nồng độ thấp, 99,3% và 102,5% ở mức nồng độ trung bình, 99,3% và 104,8% ở mức cao. Các kết quả này đều nằm trong khoảng cho phép của AOAC (90-107%) và chỉ số RSD khá thấp, lần lượt là 3,01%, 1,45% và 1,75% đối với MTZ và 0,60%; 1,29%; 1,06% đối với SPI, đều nhỏ hơn 5,3% với ngưỡng nồng độ dưới 100 $\mu\text{g/mL}$ (AOAC), cho thấy độ ổn định và chính xác của phương pháp SR-PLS trong việc xác định nồng độ MTZ và SPI.

Phương pháp BVE-PLS và phương pháp GA-PLS khi xác định MTZ đều có độ thu hồi thấp hơn và có RSD hầu hết cao hơn phương pháp SR-PLS cho thấy các bước sóng lựa chọn bởi phương pháp SR-PLS cho các thông tin hiệu quả hơn trong việc dự đoán kết quả nồng độ MTZ trong các mẫu đánh giá độ đúng. Đối với SPI, độ thu hồi khi áp dụng phương pháp BVE-PLS cho kết quả thấp hơn cả 2 phương pháp còn lại, trong khi đó cả phương pháp SRPLS và GAPLS cho kết quả tương đương với nhau.

Như vậy, mặc dù cả 3 phương pháp đều cho kết quả đánh giá độ thu hồi và có giá trị RSD nằm trong khoảng cho phép của AOAC nhưng khi so sánh giữa các phương pháp cho thấy phương pháp SR-PLS cho độ thu hồi gần với 100% hơn với hai phương pháp còn lại.

* Đánh giá độ chính xác

Bảng 3.31. Kết quả đánh giá độ lặp lại bằng các phương pháp SR-PLS, BVE-PLS và GA-PLS

Mẫu	Độ lặp lại (ngày 1)						Độ lặp lại (ngày 2)					
	SR-PLS		BVE-PLS		GA-PLS		SR-PLS		BVE-PLS		GA-PLS	
	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)	MTZ ($\mu\text{g/mL}$)	SPI (IU/mL)
Lần 1	14,35	94,58	15,17	93,58	14,96	92,13	14,22	94,51	14,93	93,83	14,71	91,90
Lần 2	14,26	94,17	14,97	93,06	14,79	93,92	14,46	96,59	15,29	94,95	15,06	91,26
Lần 3	14,40	93,50	14,79	92,36	14,61	92,16	14,25	94,34	15,07	93,32	14,85	91,72
Lần 4	14,83	94,10	15,13	93,39	14,96	93,27	14,40	94,34	15,10	93,65	14,89	91,76
Lần 5	14,84	93,09	15,06	92,06	14,92	91,95	14,38	94,57	15,14	93,60	14,93	91,85
Lần 6	14,82	96,24	15,10	95,02	14,95	96,00	14,41	94,20	14,95	93,08	14,76	91,88
TB ($\mu\text{g/mL}$)	14,58	94,28	15,04	93,25	14,87	93,24	14,35	94,76	15,08	93,74	14,87	91,73
SD ($\mu\text{g/mL}$)	0,27	1,10	0,14	1,05	0,14	1,56	0,10	0,91	0,13	0,65	0,13	0,24
RSD (%)	1,85	1,17	0,93	1,13	0,95	1,67	0,70	0,96	0,86	0,69	0,84	0,26
$F_{\text{tính}}$	8,17	1,46	1,10	2,61	1,27	42,21						
$t_{\text{tính}}$	1,94	0,82	0,55	0,98	0,02	2,35						

Đánh giá độ chính xác bằng kết quả đánh giá độ lặp lại giữa 2 ngày khác nhau. Kết quả được thể hiện trong bảng 3.31. Kết quả từ bảng cho thấy giá trị RSD đều nhỏ hơn 5,3%, như vậy kết quả giữa 6 lần lặp lại khi sử dụng các phương pháp lựa chọn bước sóng để tính nồng độ của MTZ và SPI trong các mẫu phân tích cho kết quả lặp lại tốt.

Chúng tôi cũng so sánh kết quả tính toán giữa các phương pháp trong kết quả xác định MTZ và SPI giữa độ lặp lại ngày thứ 1 và ngày thứ 2. Kết quả cho thấy, phương pháp SR-PLS khi tính nồng độ MTZ giữa hai kết quả có sự khác biệt đáng kể về độ lệch chuẩn giữa hai dãy số liệu ($F_{\text{tính}} = 8,17 > F_{(5,5)}(\alpha = 0,05) = 0,582$), kết quả tương tự khi sử dụng phương pháp GA-PLS để xác định nồng độ của SPI trong hai dãy số liệu, cho thấy có sự khác biệt đáng kể về độ lệch chuẩn giữa hai dãy số liệu thu được ($F_{\text{tính}} = 42,21 > F_{(5,5)}(\alpha = 0,05) = 0,582$). Tuy nhiên, khi so sánh kết quả trung bình giữa các dãy số liệu giữa kết quả độ lặp lại và độ chính xác trung gian bằng chuẩn t, kết quả cho thấy các giá trị $t_{\text{tính}}$ đều nhỏ hơn $t_{\text{bảng}}$, 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

Bảng 3.32. Kết quả xác định hàm lượng MTZ (mg/viên) và phần trăm so với hàm lượng ghi trên nhãn bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC

Mẫu	SR-PLS		BVE-PLS		GA-PLS		HPLC	
	m	%	m	%	m	%	m	%
Q1	121,67	97,3	124,45	99,6	123,05	98,4	124,54	99,6
Q2	123,67	98,9	123,03	98,4	121,64	97,3	123,25	98,6
Q3	120,83	96,7	125,14	100,1	123,33	98,7	122,84	98,3
Trung bình	122,06	97,6	124,21	99,4	122,7	98,2	123,5	98,8
RSD (%)	1,195		0,866		0,738		0,718	

Bảng 3.33. Kết quả xác định hàm lượng SPI (IU/viên) và phần trăm so với hàm lượng ghi trên nhãn bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và vi sinh.

Mẫu	SR-PLS		BVE-PLS		GA-PLS		Phương pháp vi sinh	
	IU	%	IU	%	IU	%	IU	%
Q1	776583	103,5	770527,8	102,7	770027,8	102,7	774172	103,2
Q2	777750	103,7	771388,9	102,9	779194,4	103,9	768611	102,5
Q3	785083	104,7	781888,9	104,3	780555,6	104,1	765896	102,1
Trung bình	779806	104	774601,9	103,3	776592,6	103,5	769559,7	102,6
RSD (%)	0,591		0,817		0,737		0,548	

Kết quả sử dụng các phương pháp SR-PLS, BVE-PLS, GA-PLS để xác định hàm lượng MTZ và SPI trong các mẫu thuốc AgimDogyl được trình bày trong các bảng 3.32 và bảng 3.33. Kết quả định lượng MTZ với khoảng tin cậy ($\alpha = 0,05$): $122,06 \pm 3,62$ mg; $124,21 \pm 2,67$ mg; $123,54 \pm 2,20$ mg tương ứng với các phương pháp SR-PLS; BVE-PLS; GA-PLS. Kết quả định lượng SPI với khoảng tin cậy ($\alpha = 0,05$): 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 quy 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 được trình bày trong bảng 3.34. Đối với MTZ, kết quả $F_{\text{tính}} = 2,20 < F_{\text{bảng}} = 4,07$, $p = 0,166 > 0,05$ 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 $F_{\text{tính}} = 1,98 < F_{\text{bảng}} = 4,07$; $p = 0,196 > 0,05$ cho thấy giá trị trung bình kết quả phân tích SPI của các nhóm không khác nhau đáng kể. Kết quả này khẳng định các phương pháp đã xây dựng có độ đúng tốt.

Bảng 3.34. Kết quả phân tích ANOVA kết quả định lượng MTZ và SPI giữa các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC

MTZ							
Nguồn biến thiên	Tổng bình phương	Bậc tự do	Trung bình bình phương	F _{tính}	p	F _{bảng}	
Khác biệt giữ các nhóm	8,07	3	2,69	2,20	0,166	4,07	
Khác biệt trong từng nhóm	9,79	8	1,22				
Tổng số	17,86	11					
SPI							
Nguồn biến thiên	Tổng bình phương	Bậc tự do	Trung bình bình phương	F _{tính}	p	F _{bảng}	
Khác biệt giữ các nhóm	165915268,2	3	55305089,4	1,98	0,196	4,07	
Khác biệt trong từng nhóm	223650993,4	8	27056374,2				
Tổng số	389566261,6	11					

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

Các mẫu trong tập hiệu chuẩn được pha với các nồng độ khác nhau và dữ liệu phổ hấp thụ được đo trong khoảng từ 200 – 400 nm với khoảng quét 0,2 nm. Các phương pháp SR-PLS, BVE-PLS và GA-PLS đã được huấn luyện dựa trên các dữ liệu thu được, các kết quả được trình bày trong bảng 3.35. Về nhận xét chung, các tham số thống kê để đánh giá độ chính xác, độ tin cậy của các mô hình đều đạt theo các yêu cầu chung. Khi so sánh giá trị RMSECV và RMSEC, đây là tham số đánh giá độ chính xác của mô hình, cho thấy phương pháp GA có các giá trị RMSECV và RMSEC là nhỏ nhất cho cả IND và AML, trong khi đó sự chênh lệch của hai giá trị này đối với phương pháp SR-PLS và BVE-PLS là không đáng kể. Đo lường mức độ mà mô hình giải thích sự biến đổi trong dữ liệu bằng hệ số xác định R^2 cho thấy các giá trị đạt được đề tiệm cận đến 1, chứng tỏ mức độ giải thích tốt của các mô hình đã xây dựng.

Phân tích phần dư bằng kiểm định Shapiro-Wilk cho thấy các giá trị $p > 0,05$, chứng tỏ phần dư đều phân phối chuẩn, điều này có nghĩa mô hình đã học được đúng sự biến động của dữ liệu mà không có sự sai lệch hệ thống, sai số dự đoán của mô hình được phân bố ngẫu nhiên và không có xu hướng rõ ràng chứng tỏ khả năng dự đoán của các mô hình đã xây dựng đáng tin cậy.

Các giá trị LOD thấp cũng cho thấy các phương pháp tính toán SR-PLS, BVE-PLS, GA-PLS dựa trên phổ hấp thụ UV-Vis có thể phát hiện được các thành phần trong mẫu phân tích ở nồng độ thấp.

Bảng 3.35. Kết quả đánh giá các tham số thông kê đánh giá trên tập hiệu chuẩn bởi các mô hình lựa chọn bước sóng cho tập hiệu chuẩn Natruxam 1,5/5

Mẫu	Nồng độ tập hiệu chuẩn (µg/mL)		Nồng độ tính được từ mô hình (µg/mL)					
			Phương pháp SR-PLS		Phương pháp BVE-PLS		Phương pháp GA-PLS	
			IND	AML	IND	AML	IND	AML
T1	3,0	10,0	2,90	9,44	2,82	9,35	2,89	9,59
T2	6,0	10,0	5,98	9,69	5,92	9,98	6,03	10,14
T3	3,0	40,0	3,03	39,12	2,90	39,74	3,04	39,85
T4	6,0	40,0	6,13	40,33	6,07	39,94	6,10	40,17
T5	4,5	25,0	4,44	25,24	4,45	25,10	4,46	25,21
T6	2,4	25,0	2,35	25,65	2,43	25,18	2,36	25,14
T7	6,6	25,0	6,35	24,44	6,33	24,36	6,34	24,41
T8	4,6	4,0	4,59	4,10	4,60	4,34	4,57	4,20
T9	4,5	46,0	4,39	46,12	4,49	46,25	4,42	46,01
T10	4,5	15,0	4,58	15,37	4,72	15,29	4,56	14,77
T11	4,5	30,0	4,49	29,84	4,47	30,02	4,50	30,00
Số biến tiềm ẩn			4	3	4	4	4	4
RMSECV (µg/mL)			0,142	0,381	0,138	0,385	0,118	0,335
RMSEC (µg/mL)			0,120	0,337	0,122	0,324	0,104	0,297
Shapiro-Wilk	W		0,9557	0,9606	0,9836	0,9221	0,9067	0,9594
	p(>0,05)		0,084	0,128	0,988	0,207	0,2226	0,1155
R ²			0,9906	0,9986	0,9874	0,9992	0,9915	0,9994
LOD _{min} (µg/mL)			0,24	0,53	0,28	0,56	1,17	0,45
LOD _{max} (µg/mL)			0,30	0,73	0,34	0,72	1,30	0,66

Đánh giá hiệu quả hoạt động của các mô hình SR-PLS, BVE-PLS, GA-PLS bằng cách đánh giá độ thu hồi và RMSEP dựa trên các mẫu được chuẩn bị trong phòng thí nghiệm với các mức nồng độ và tỉ lệ khác nhau, các nồng độ này đều nằm trong không gian của tập nồng độ hiệu chuẩn. Kết quả được trình bày trong bảng 3.36. Kết quả cho thấy, phương pháp SR-PLS và GA-PLS cho kết quả dự đoán IND với giá trị trung bình tìm lại gần bằng 100% và giá trị RMSEP nhỏ hơn phương pháp BVE-PLS; cả ba phương pháp đều cho kết quả dự đoán nồng độ AML tương đương nhau, sự khác biệt không đáng kể giữa các giá trị trung bình phần trăm độ thu hồi và RMSEP.

Bảng 3.36. Kết quả đánh giá hiệu của hoạt động của các mô hình lựa chọn bước sóng trên tập nồng độ kiểm tra của thuốc Natruxam 1,5/5

Mẫu	Nồng độ tập kiểm tra ($\mu\text{g/mL}$)		Độ thu hồi (%)					
			Phương pháp SR-PLS		Phương pháp BVE-PLS		Phương pháp GA-PLS	
	IND	AML	IND	AML	IND	AML	IND	AML
V1	3,5	15,0	100,7	101,2	103,4	100,8	100,5	103
V2	5,5	15,0	99,4	103,6	103,1	107,3	98,3	105,7
V3	3,5	35,0	98,7	105,3	104,9	107,1	98,7	106,1
V4	5,5	35,0	99,4	105,4	102,7	107,1	98,8	106,4
V5	4,5	25,0	102,8	101,6	105,3	101,9	103,9	102,3
V6	3,1	25,0	101,4	102,8	105,6	102,8	103,3	103,7
V7	5,9	25,0	102,8	100,1	104,4	100,7	103,7	100,7
V8	4,5	11,0	101,8	93,9	104,5	96,2	101,2	97,3
V9	4,5	39,0	95,7	99,3	100,9	101,2	95,1	100,4
Trung bình			100,3	101,4	103,9	102,7	100,5	102,7
RSD(%)			2,29	3,51	1,42	3,52	2,95	2,90
RMSEP ($\mu\text{g/mL}$)			0,107	1,004	0,183	1,312	0,137	1,176

b. Đánh giá phương pháp phân tích

*** Đánh giá độ đúng**

Đánh giá độ đúng ở ba mức nồng độ khác nhau, với IND cả ba mức nồng độ đều nhỏ hơn 10,0 $\mu\text{g/mL}$, với mức nồng độ này theo qui định của AOAC độ thu hồi yêu cầu trong khoảng 80-110% và giá trị RSD < 7,3; Kết quả từ bảng 3.37 cho thấy phương pháp BVE-PLS có giá trị độ thu hồi lần thứ 3 ở mức thấp là 111,4%, kết quả này cho thấy phương pháp BVE-PLS lựa chọn các bước sóng có thể chưa phải là tối ưu để có thể xác định chính xác nồng độ IND trong các mẫu phân tích; Đối với các phương pháp còn lại, độ thu hồi và giá trị RSD đều thỏa mãn theo qui định. Tương tự với AML, với mức nồng độ xác định, độ thu hồi theo qui định là 90-107% và RSD < 5,3; ở mức nồng độ thấp, phương pháp BVE-PLS có giá trị độ thu hồi là 107,8 > 107% cho thấy độ thu hồi này chưa phù hợp. Vì vậy, đối với phương pháp BVE-PLS khi xác định nồng độ IND và AML trong mẫu chưa đạt yêu cầu về độ đúng.

Bảng 3.37. Kết quả đánh giá độ thu hồi xác định nồng độ IND và AML bằng các phương pháp SR-PLS, BVE-PLS và GA-PLS

Mẫu	Nồng độ thêm chuẩn ($\mu\text{g/mL}$)		Độ thu hồi (%)					
			Phương pháp SR-PLS		Phương pháp BVE-PLS		Phương pháp GA-PLS	
	IND	AML	IND	AML	IND	AML	IND	AML
1	3,6	12,0	100,8	98,6	107,8	102,4	100,0	102,4
2	3,6	12,0	99,7	98,3	107,5	102,8	98,9	102,7
3	3,6	12,0	104,2	102,7	111,4	107,9	103,3	106,1
Trung bình			101,6	99,9	108,9	104,4	100,7	103,7
RSD (%)			2,28	2,51	1,99	2,93	2,29	1,97
4	4,5	15,0	99,8	102,9	100	102	100,2	103,8
5	4,5	15,0	100,0	103,1	99,8	101,9	100,9	103,7
6	4,5	15,0	98,9	103,0	102,7	104,8	98,9	103,9
Trung bình			99,6	103,0	100,8	102,9	100,0	103,8
RSD (%)			0,59	0,13	1,59	1,63	1,02	0,07
7	5,4	18,0	95,2	100,3	103,7	106,1	95,2	104,0
8	5,4	18,0	96,8	100,1	103,7	105,1	96,5	103,4
9	5,4	18,0	95,2	100,1	103,3	105,4	95,4	103,8
Trung bình			95,7	100,2	103,6	105,5	95,7	103,7
RSD (%)			1,01	0,12	0,21	0,49	0,73	0,30

* Đánh giá độ chính xác

Bảng 3.38. Kết quả đánh giá độ lặp lại mẫu Natruxam bằng phương pháp SR-PLS, BVE-PLS, GA-PLS

	SR-PLS		BVE-PLS		GA-PLS	
	IND	AML	IND	AML	IND	AML
	($\mu\text{g/mL}$)	($\mu\text{g/mL}$)	($\mu\text{g/mL}$)	($\mu\text{g/mL}$)	($\mu\text{g/mL}$)	($\mu\text{g/mL}$)
Lần 1	4,45	14,45	4,62	15,48	4,39	14,40
Lần 2	4,40	14,37	4,60	15,53	4,37	14,54
Lần 3	4,37	14,44	4,58	15,49	4,33	14,59
Lần 4	4,38	14,44	4,56	15,49	4,32	14,52
Lần 5	4,27	14,50	4,52	15,53	4,21	14,67
Lần 6	4,30	14,40	4,54	15,54	4,25	14,63
Trung bình	4,36	14,43	4,57	15,51	4,31	14,56
SD	0,07	0,05	0,037	0,026	0,069	0,095
RSD (%)	1,52	0,309	0,819	0,168	1,61	0,655

Kết quả đánh giá độ lặp lại 6 lần bằng các phương pháp SR-PLS, BVE-PLS và GA-PLS được thể hiện trong bảng 3.38. Kết quả độ lệch chuẩn tương đối RSD được so sánh với giá trị RSD tính bằng phương trình Horwitz. Đối với IND, mức nồng độ chất phân tích được xem xét là 4,5 $\mu\text{g/mL}$ có giá trị RSD theo phương trình Horwitz là 6,38. Như vậy, các giá trị RSD cho IND của các phương pháp SR-PLS, BVE-PLS và GA-

PLS lần lượt là 1,52; 0,819; 1,61 đều nhỏ hơn giá trị 6,38 chứng tỏ các phương pháp đều có độ lặp lại tốt. Tương tự với AML, giá trị Horwitz tính được ở mức nồng độ 15 µg/mL là 5,32; giá trị này đều lớn hơn các giá trị RSD của các phương pháp được tính. Như vậy, các phương pháp SR-PLS, BVE-PLS, GA-PLS đạt tiêu chuẩn về đánh giá độ lặp lại.

c. Kết quả định lượng

Dựa trên kết quả thực nghiệm trong bảng 3.39 và 3.40 cho thấy rằng, hàm lượng các hoạt chất AML và IND so với hàm lượng ghi trên nhãn của mẫu thuốc Natrixam 1,5mg/5mg nằm trong giới hạn cho phép (90,0% - 110,0%). Các mẫu thuốc đều đạt yêu cầu về hàm lượng theo tiêu chuẩn của Dược điển Việt Nam V.

Bảng 3.39. Kết quả xác định hàm lượng IND và giá trị phần trăm so với hàm lượng ghi trên nhãn trong viên nén Natrixam 1,5/5 bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC

Mẫu	SR-PLS		BVE-PLS		GA-PLS		HPLC	
	m	%	m	%	m	%	m	%
	(mg)	HLGTN	(mg)	HLGTN	(mg)	HLGTN	(mg)	HLGTN
Q1	1,43	95,3	1,46	97,3	1,45	96,7	1,49	99,3
Q2	1,46	97,3	1,49	99,3	1,48	98,7	1,49	99,3
Q3	1,46	97,3	1,48	98,7	1,47	98,0	1,51	100,7
TB	1,45	96,63	1,48	98,4	1,47	97,8	1,50	100,0
F _{tính}	2,25		1,75		1,75			
t _{tính}	3,88		1,81		2,71			

Bảng 3.40. Kết quả xác định hàm lượng AML và giá trị phần trăm so với hàm lượng ghi trên nhãn trong viên nén Natrixam 1,5/5 bằng các phương pháp SR-PLS, BVE-PLS, GA-PLS và HPLC

Mẫu	SR-PLS		BVE-PLS		GA-PLS		HPLC	
	m	%	m	%	m	%	m	%
	(mg)	HLGTN	(mg)	HLGTN	(mg)	HLGTN	(mg)	HLGTN
Q1	4,87	97,4	5,19	103,8	4,81	96,2	4,73	94,6
Q2	4,93	98,6	5,31	106,2	4,87	97,4	4,80	96
Q3	4,90	98,0	5,26	105,2	4,82	96,4	4,75	95
TB	4,90	98,0	5,25	105,1	4,83	96,7	4,76	95,2
F _{tính}	1,44		2,80		1,26			
t _{tính}	5,17		12,17		2,63			

Khi so sánh với phương pháp HPLC kết quả từ bảng 3.39 và bảng 3.40 cho thấy đối với phương pháp SR-PLS giá trị F_{tính} của IND và AML lần lượt là 2,25 và 1,44 đều nhỏ hơn $F_{(2,2)(\alpha=0,05)} = 15,44$ cho thấy độ lệch chuẩn của hai dãy số liệu không khác nhau đáng kể, nhưng các giá trị t_{tính} có được lần lượt là 3,88 và 5,17 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. Đối với phương

pháp BVE-PLS, giá trị trung bình của AML có sự khác biệt đáng kể khi so sánh với phương pháp HPLC. Trong khi đó, độ lệch chuẩn giữa hai dãy số liệu và giá trị trung bình được tính bằng phương pháp GA-PLS không khác với phương pháp HPLC khi sử dụng kiểm định F và t để so sánh.

Như vậy, với phương pháp lựa chọn bước sóng để xác định hàm lượng IND và AML trong viên nén Natrixam 1,5/5 phương pháp GA-PLS cho kết quả phù hợp nhất từ phần trăm hàm lượng ghi trên nhãn và so sánh với phương pháp tham chiếu. Đối với phương pháp SR-PLS và BVE-PLS cho kết quả phân tích chưa chính xác.

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

Bảng 3.41 Kết quả đánh giá các tham số thống kê đánh giá trên tập hiệu chuẩn bởi các mô hình BVE-PLS, GA-PLS cho tập hiệu chuẩn Turbezid

Mẫu	Nồng độ tập hiệu chuẩn ($\mu\text{g/mL}$)			Nồng độ tính được từ mô hình ($\mu\text{g/mL}$)					
	RIF	INZ	PYR	Phương pháp BVE-PLS			Phương pháp GA-PLS		
				RIF	INZ	PYR	RIF	INZ	PYR
T1	7,5	7,5	20,0	7,59	7,48	20,18	7,57	7,47	20,17
T2	10,5	7,5	20,0	10,69	7,34	20,45	10,68	7,33	20,07
T3	7,5	10,5	20,0	7,45	10,39	19,77	7,51	10,49	20,15
T4	10,5	10,5	20,0	10,51	10,88	19,98	10,55	10,67	20,11
T5	7,5	7,5	28,0	7,60	7,70	27,38	7,57	7,60	27,24
T6	10,5	7,5	28,0	10,52	7,60	28,14	10,51	7,56	27,75
T7	7,5	10,5	28,0	7,51	10,58	28,4	7,53	10,48	28,77
T8	10,5	10,5	28,0	10,50	10,63	28,17	10,49	10,54	28,32
T9	6,5	9,0	24,0	6,49	9,01	24,34	6,50	9,00	24,52
T10	11,5	9,0	24,0	11,37	8,84	23,69	11,41	8,97	23,43
T11	9,0	6,5	24,0	8,95	6,66	24,06	8,92	6,61	23,67
T12	9,0	11,5	24,0	9,03	11,25	23,93	9,02	11,43	24,28
T13	9,0	9,0	17,5	8,87	8,98	17,19	8,87	9,00	17,25
T14	9,0	9,0	30,5	8,98	8,72	30,14	8,98	8,89	30,11
T15	9,0	9,0	24,0	8,97	8,98	24,23	8,94	9,01	24,19
Số biến tiềm ẩn				3	4	3	3	4	2
RMSECV ($\mu\text{g/mL}$)				0,141	0,223	0,382	0,111	0,126	0,461
RMSEC ($\mu\text{g/mL}$)				0,109	0,179	0,317	0,096	0,088	0,413
Shapiro-Wilk	W			0,9681	0,9824	0,9457	0,9540	0,9911	0,9741
	p(>0,05)			0,489	0,885	0,130	0,216	0,995	0,659
R ²				0,9941	0,9842	0,9929	0,9955	0,9962	0,9879
LOD _{min} ($\mu\text{g/mL}$)				0,39	0,49	1,79	1,68	0,63	2,52
LOD _{max} ($\mu\text{g/mL}$)				0,41	0,52	1,85	1,74	0,66	2,59

Xây dựng các mô hình BVE-PLS, GA-PLS dựa trên dữ liệu thập hiệu chuẩn Turbezid đã xây dựng. Các tham số thống kê được trình bày trong bảng 3.41. Các giá trị RMSECV, RMSEC nhỏ hơn 5% so với nồng độ trung bình của tập hiệu chuẩn của mỗi thành phần phản ánh các mô hình đã xây dựng chính xác, tin cậy. Phân tích phần dư

bằng kiểm định Shapiro-Wilk có giá trị p đều lớn hơn 0,05 cho thấy các phần dư có phân phối chuẩn, ngẫu nhiên quanh điểm 0. Giá trị R^2 đều tiệm cận 1 cho thấy các mô hình đã giải thích tốt các nồng độ trong tập hiệu chuẩn. Trong phương pháp GA-PLS, số lượng bước sóng được giữ lại để xây dựng mô hình hiệu chuẩn nhỏ hơn 10% so với tổng số lượng bước sóng ban đầu, có thể vì lí do này mà trong mô hình GA-PLS có các giá trị LOD đều lớn hơn giá trị LOD của phương pháp BVE-PLS, số lượng bước sóng ít có thể làm giảm thông tin để xây dựng mô hình nên có thể làm giảm độ nhạy của phương pháp GA-PLS.

Bảng 3.42. Kết quả xác định trên tập kiểm tra bằng phương pháp BVE-PLS và GA-PLS đối với hỗn hợp RIF, INZ và PYR

Mẫu	Nồng độ tập kiểm tra ($\mu\text{g/mL}$)			Độ thu hồi (%)					
				Phương pháp BVE-PLS			Phương pháp GA-PLS		
	RIF	INZ	PYR	RIF	INZ	PYR	RIF	INZ	PYR
V1	8,3	8,3	22,0	98,8	97,5	101,8	98,8	99,0	101,8
V2	9,8	8,3	22,0	101,7	95,3	99,0	101,5	98,2	98,1
V3	8,3	9,8	22,0	101,5	102,7	99,2	101,8	103,3	100,5
V4	9,8	9,8	22,0	102,0	97,9	97,4	101,8	100,1	98,1
V5	8,3	8,3	26,0	102,3	105,3	102,3	102,0	105,5	102,6
V6	9,8	8,3	26,0	103,4	100,2	101,4	103,3	100,8	100,8
V7	8,3	9,8	26,0	96,5	100,9	102,8	96,7	101,6	103,7
V8	9,8	9,8	26,0	101,5	98,4	102,1	101,4	100,3	102,4
V9	7,8	9,0	24,0	103,4	100,9	98,9	104,1	101,1	99,4
V10	10,3	9,0	24,0	106,5	102,5	99,8	106,5	102,9	99,2
V11	9,0	7,8	24,0	100,5	103,8	100,5	100,6	104,2	100,0
V12	9,0	10,3	24,0	102,7	103	98,9	103,0	103,1	100,0
V13	9,0	9,0	20,5	100,1	103,5	100,8	100,7	104,2	101,4
V14	9,0	9,0	27,5	103,8	100,8	100,2	103,6	102,6	100,4
V15	9,0	9,0	24,0	101,6	102,3	99,6	101,8	103,0	99,8
Trung bình				101,7	101,0	100,3	101,8	102,0	100,5
RSD(%)				2,26	2,81	1,59	2,27	2,06	1,61
RMSEP ($\mu\text{g/mL}$)				0,265	0,257	0,387	0,266	0,258	0,424

Các mẫu kiểm tra được chuẩn bị từ các mẫu chuẩn trong phòng thí nghiệm với các nồng độ khác nhau và có tỉ lệ khác nhau giữa các mẫu. Kết quả độ thu hồi trung bình trên tập kiểm tra đều xấp xỉ 100%, các giá trị RMSEP nhỏ cho phép có thể xác định nồng độ của các chất trong mẫu thực tế.

b. Đánh giá phương pháp phân tích

* Đánh giá độ đúng

Kết quả đánh giá độ đúng của các phương pháp BVE-PLS và GA-PLS được trình bày trong bảng 3.43. Đối với mức nồng độ thấp và trung bình, 2 thành phần RIF và INZ có mức độ độ dưới 10 $\mu\text{g/mL}$ giá trị RSD ngưỡng cho phép là 7,3% và độ thu hồi nằm

trong khoảng 80 – 110%. Các kết quả về độ thu hồi và RSD từ bảng cho thấy phù hợp với quy định, vì vậy đánh giá độ đúng bằng cách đánh giá độ thu hồi là đạt yêu cầu.

Bảng 3.43. Kết quả đánh giá độ thu hồi cho mẫu hỗn hợp Turbezid

Mẫu	Nồng độ thêm chuẩn			Độ thu hồi (%)					
	(µg/mL)			Phương pháp BVE-PLS			Phương pháp GA-PLS		
	RIF	INZ	PYR	RIF	INZ	PYR	RIF	INZ	PYR
1	7,7	7,7	20,4	102,6	96,9	101,9	102,7	99	101,7
2	7,7	7,7	20,4	99,8	107	100,1	100,2	105,3	100,6
3	7,7	7,7	20,4	99,4	99,4	96,5	99,5	101	96,5
Trung bình				100,6	101,1	99,5	100,8	101,8	99,6
RSD (%)				1,73	5,2	2,76	1,47	2,81	2,38
4	9,0	9,0	24,0	102,4	94,8	96	102,7	98,3	95,9
5	9,0	9,0	24,0	102	100,7	98,1	102,2	102	98,3
6	9,0	9,0	24,0	101,2	107,8	100	101,5	106,8	100,7
Trung bình				101,9	101,1	98	102,1	102,4	98,3
RSD (%)				0,6	6,44	2,04	0,81	4,54	2,6
7	10,4	10,4	27,6	101,1	103,9	98,6	100,7	103,7	99,5
8	10,4	10,4	27,6	104,2	99,3	100,2	104,2	101,7	100,6
9	10,4	10,4	27,6	102,2	106,3	102,1	102,1	106,9	103,2
Trung bình				102,5	103,2	100,3	102,3	104,1	101,1
RSD (%)				1,53	3,45	1,75	1,54	2,27	1,64

* Đánh giá độ chính xác

Đánh giá kết quả độ lặp lại dựa trên dữ liệu đo được từ 6 mẫu thuốc bắt đầu từ cân mẫu. Kết quả được trình bày trong bảng 3.44.

Đối với phương pháp BVE-PLS, giá trị RSD tính được khi xác định nồng độ RIF, INZ và PYR giữa hai lần khác nhau đều cho kết quả tốt, giá trị RSD đều nhỏ hơn giá trị RSD tính theo phương trình Horwitz cho RIF, INZ, PYR lần lượt là 5,75; 6,38 và 4,96. Đối với phương pháp GA-PLS, giá trị RSD tính được cho RIF và PYR cho độ lặp lại nhỏ hơn giá trị RSD tính theo hàm Horwitz, tuy nhiên, đối với INZ có giá trị RSD tính được cho đánh giá độ lặp lại ngày thứ 1 và ngày thứ 2 lần lượt là 9,94 và 8,39 đều lớn hơn giá trị RSD tính theo hàm Horwitz. Như vậy, đối với phương pháp GA-PLS khi đánh giá độ lặp lại cho INZ chưa đạt yêu cầu.

Bên cạnh đó, khi so sánh giá trị trung bình giữa hai kết quả của hai phương pháp BVE-PLS và GA-PLS, các giá trị F và t đã được tính; Kết quả cho thấy, giá trị trung bình của RIF và INZ giữa hai phương pháp có sự khác biệt có ý nghĩa thống kê ($p < 0,05$).

Bảng 3.44. Kết quả đánh giá độ lặp lại và độ chính xác trung gian của thuốc Turbezid bằng các phương pháp BVE-PLS, GA-PLS

	Độ lặp lại (ngày 1)						Độ lặp lại (ngày 2)					
	BVE-PLS			GA-PLS			BVE-PLS			GA-PLS		
	RIF	INZ	PYR	RIF	INZ	PYR	RIF	INZ	PYR	RIF	INZ	PYR
Lần 1	9,42	4,2	24,09	9,19	2,93	23,72	9,19	4,5	24,79	8,95	3,11	24,47
Lần 2	9,26	4,51	24,7	9,04	3,07	24,35	9,44	4,54	25,24	9,20	3,20	24,89
Lần 3	9,47	4,53	25,26	9,23	2,97	24,89	9,48	4,61	25,33	9,26	3,09	24,97
Lần 4	9,51	4,54	25,33	9,27	3,01	24,99	9,16	4,26	23,3	8,95	3,04	22,97
Lần 5	9,32	4,52	23,49	9,11	3,00	23,15	9,51	4,54	23,54	9,28	3,13	23,19
Lần 6	9,17	4,49	24,52	9,07	3,74	24,33	9,14	4,54	24,49	8,96	3,76	24,3
Mean	9,36	4,47	24,57	9,15	3,12	24,23	9,32	4,5	24,45	9,10	3,22	24,13
SD	0,13	0,13	0,7	0,09	0,31	0,7	0,17	0,12	0,86	0,16	0,27	0,86
RSD	1,39	2,91	2,85	0,98	9,94	2,89	1,82	2,67	3,52	1,76	8,39	3,56
F _{tính}	2,02	5,50	1,00				1,14	4,85	1,00			
t _{tính}	3,16	9,86	0,81				2,26	10,59	0,64			
t _(0,05-10)	2,23						2,23					

Đơn vị nồng độ: µg/mL

c. Kết quả định lượng

Bảng 3.45. Kết quả xác định hàm lượng RIF, INZ, PYR (mg/viên) trong viên nén Turbezid bằng các phương pháp BVE-PLS, GA-PLS và HPLC

Mẫu	BVE-PLS			GA-PLS			HPLC		
	RIF	INZ	PYR	RIF	INZ	PYR	RIF	INZ	PYR
Q1	154,67	74,5	413,17	151,5	50,09	399,67	156,48	76,86	402,86
Q2	159,67	76,33	406	153,42	53,42	395,92	154,95	76,62	403,95
Q3	160,00	75,83	406,17	155,92	51,67	393,67	154,83	76,39	402,69
Trung bình	158,11	75,55	408,45	153,61	51,73	396,42	155,42	76,62	403,17
t _{tính}	1,493	1,902	2,205	1,304	25,63	3,761			
t _(0,05-4)	2,776	4,303	4,303	2,776	4,303	4,303			

Kết quả các định nồng độ của RIF, INZ và PYR trong 3 mẫu Q1, Q2, Q3 bằng các phương pháp BVE-PLS, GA-PLS và HPLC được trình bày trong bảng 3.45. Kết quả được so sánh với phương pháp HPLC được đo tại Trung tâm Kiểm nghiệm Dược phẩm, Mỹ phẩm, Thực phẩm Thành phố Huế cho thấy phương pháp GA-PLS khi xác định hàm lượng RIF và INZ có sự khác biệt đáng kể ($t_{\text{tính}} > t_{\text{bảng}}$), trong khi đó phương pháp BVE-PLS có kết quả tương đồng với kết quả đo được bằng phương pháp

HPLC. Khoảng tin cậy, tính theo BVE-PLS, của RIF là $158,11 \pm 7,42$ mg/viên; INZ là $75,55 \pm 2,36$ mg/viên; PYR = $408,45 \pm 10,15$ mg/viên ($\alpha = 0,05$)

Bảng 3.46. Kết quả so sánh với hàm lượng ghi trên nhãn của thuốc Turbezid bằng phương pháp BVE-PLS và GA-PLS

Mẫu	BVE-PLS			GA-PLS		
	RIF (%)	INZ (%)	PYR (%)	RIF (%)	INZ (%)	PYR (%)
Q1	103,1	99,3	103,3	101	66,8	99,9
Q2	106,4	101,8	101,5	102,3	71,2	99
Q3	106,7	101,1	101,5	103,9	68,9	98,4
Trung bình	105,4	100,7	102,1	102,4	69	99,1

Đánh giá hàm lượng tính được so với hàm lượng ghi trên nhãn của nhà sản xuất cho thấy, phương pháp BVE-PLS có phần trăm hàm lượng ghi trên nhãn đều nằm trong khoảng 90 – 110% theo quy định của Dược điển Việt Nam V; kết quả xác định bằng phương pháp GA-PLS cho kết quả nhỏ hơn 90% cho thấy phương pháp GA-PLS khi xác định hàm lượng INZ là chưa chính xác.

3.3.5 Kết luận phương pháp lựa chọn bước sóng trong PLS

Với việc tìm hiểu về cơ sở lý thuyết, thuật toán và ứng dụng trong một số thí nghiệm thực nghiệm, chúng tôi đi đến một số kết luận như sau:

- Đề tài đã khái quát được vị trí, vai trò và những ứng dụng cụ thể của các phương pháp lựa chọn bước sóng trong phương pháp phổ nói chung và phổ UV-Vis nói riêng.

- Đề tài đã làm rõ các thuật toán SR-PLS, BVE-PLS và GAPLS áp dụng trong việc lựa chọn bước sóng cho phổ UV-Vis.

- Đối với phương pháp SR-PLS, cùng với việc triển khai thuật toán trên phần mềm R đã tính được các giá trị SR cho mỗi bước sóng cho các biến tiềm ẩn tối ưu của phương pháp PLS, từ đó lựa chọn được ngưỡng tối ưu cho SR để lựa chọn các bước sóng phù hợp cho mỗi tập thí nghiệm. Kết quả xác định hàm lượng của MTZ và SPI không có sự khác biệt đáng kể so với phương pháp tham chiếu, nhưng kết quả xác định INZ và AML có sự khác biệt đáng kể mặc dù kết quả so với hàm lượng ghi trên nhãn đều nằm trong khoảng cho phép của Dược điển Việt Nam.

- Đối với phương pháp BVE-PLS, các giá trị VIP đã được tính và thuật toán lựa chọn bước sóng phù hợp đã được xác định. Áp dụng trên các mẫu thực tế cho thấy kết

quả đối với mẫu hỗn hợp MTZ và SPI; IND và AML; RIF, INZ và PYR cho kết quả phù hợp với phương pháp tham chiếu, các giá trị so với hàm lượng ghi trên nhãn đều nằm trong khoảng cho phép của Dược điển Việt Nam

- Đối với phương pháp GA-PLS, các tham số cho thuật toán di truyền đã được giải thích và lựa chọn để thực hiện. Với các mẫu hai thành phần, phương pháp đã lựa chọn được các bước sóng có thông tin tối ưu cho việc dự đoán kết quả trong các mẫu thực tế. Ngược lại, với việc trích xuất quá ít bước sóng cho các thành phần trong mẫu ba thành phần dẫn đến mô hình không tổng quát được hết các thông tin làm cho kết quả áp dụng trên các mẫu trong phòng thí nghiệm thì phù hợp nhưng với các mẫu định lượng thực tế thì có sự khác biệt đáng kể, đặc biệt với INZ có nồng độ bé trong hỗn hợp và mức độ đóng góp vào phổ chung nhỏ.

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

Để so sánh hiệu quả hoạt động của các mô hình phân tích đa biến CLS và PCR với phương pháp PLS, nghiên cứu này thực hiện so sánh các giá trị thống kê trong các tập hiệu chuẩn và tập kiểm tra. Kết quả được phân tích theo các chỉ số bao gồm: Các giá trị RMSECV/RMSEC/RMSEP: đánh giá sai số hiệu chuẩn và dự đoán, giá trị RMSECV là sai số cho kiểm định chéo trên tập huấn luyện, RMSEC là sai số khi áp dụng phương pháp trên bộ dữ liệu huấn luyện, RMSEP là sai số dự đoán trên tập kiểm tra, là chỉ số chính để đánh giá hiệu suất hoạt động của mô hình. Các giá trị RMSE càng nhỏ thể hiện mô hình càng chính xác; R^2 (hệ số xác định): đo lường độ tương quan giữa giá trị thực tế và dự đoán của mô hình; Độ thu hồi: đo lường khả năng của mô hình trong việc dự đoán đúng các thành phần trong mẫu. Độ thu hồi càng gần 100% cho thấy mô hình có khả năng dự đoán tốt các thành phần có trong mẫu; RSD (độ lệch chuẩn tương đối): chỉ ra biến động của các giá trị độ thu hồi, càng thấp cho thấy tính ổn định của mô hình; Kiểm tra khả năng phân phối chuẩn của phần dư thông qua kiểm định Shapiro-Wilk, nếu giá trị p nhỏ hơn 0,05 thì mô hình có phần dư không phân phối chuẩn, cho thấy mô hình không chính xác. Các giá trị thống kê được mô tả trong bảng 3.47.

Xét về hiệu suất trên tập huấn luyện, các chỉ số RMSECV và RMSEC cho thấy khả năng khớp của mô hình với dữ liệu huấn luyện:

- Trong thuốc Zoamco-A: với AML, PLS có RMSECV là 1,3% và RMSEC là 0,6%, trong khi PCR có RMSECV cao hơn (1,9%) và RMSEC là 1,4%, còn CLS không có RMSECV nhưng RMSEC là 2,2%. Điều này cho thấy PLS vượt trội trong việc khớp dữ liệu huấn luyện, với sai số thấp hơn đáng kể so với cả PCR và CLS, có thể do khả năng tối ưu hóa hiệp phương sai giữa biến độc lập và phụ thuộc. Đối với thành phần ATO trong cùng loại thuốc, PLS và PCR có hiệu suất tương đương với RMSECV lần

lượt là 2,4% và 2,3%, RMSEC là 1,6% và 1,5%, trong khi CLS có RMSEC là 1,6%, cho thấy hiệu suất tương đương với PLS và PCR.

Bảng 3.47. Các chỉ số đánh giá mô hình và phân phối dữ liệu phân tích phần dư của các phương pháp PLS, PCR và CLS

Thuốc	Thành phần	Phương pháp	RMSECV (%)	RMSEC (%)	Shapiro-Wilk		R ²	RMSEP (%)	Độ thu hồi (%)	RSD (%)
					W	p				
Zoamco_A	AML	PLS	1,3	0,6	0,91	0,290	0,9991	2,6	100,6	2,6
		PCR	1,9	1,4	0,95	0,709	0,9939	1,5	99,9	1,6
		CLS		2,2	0,91	0,307	0,9904	2,1	99,6	2,4
	ATO	PLS	2,4	1,6	0,87	0,116	0,9929	2,7	101,2	2,4
		PCR	2,3	1,5	0,90	0,196	0,9929	2,7	101,2	2,4
		CLS		1,6	0,87	0,100	0,9927	2,9	101,2	2,6
Natrixam	AML	PLS	1,5	1,4	0,89	0,128	0,9992	4,6	102,7	2,8
		PCR	1,5	1,4	0,97	0,910	0,9991	4,6	102,6	2,8
		CLS		5	0,85	0,016	0,9903	5,8	101,3	4,6
	IND	PLS	2,9	2,4	0,91	0,240	0,9902	2,7	101,1	2,4
		PCR	2,6	2,3	0,91	0,120	0,9910	3,5	102,4	2,6
		CLS		15,5	0,86	0,022	0,5980	13,1	108,9	12,6
Spirosemid-40	FUR	PLS	2,7	3,5	0,97	0,928	0,9977	4,7	103,1	3,3
		PCR	4,3	2,1	0,99	0,390	0,9979	4,8	103,9	2,9
		CLS		2,8	0,89	0,007	0,9969	4,3	103,1	2,5
	SPR	PLS	1,9	2,6	0,88	0,175	0,9980	3,9	101,3	3,7
		PCR	3,1	1,6	0,91	0,850	0,9981	3,4	99,2	4,3
		CLS		2,1	0,93	0,053	0,9975	4,3	100,9	4,1
AgimDogyl	MTZ	PLS	3,6	3,2	0,93	0,061	0,9894	2,2	100,9	0,3
		PCR	3,3	2,9	0,92	0,048	0,9915	3,7	103,3	1,3
		CLS		3,3	0,93	0,090	0,9888	2,2	101,2	1,8
	SPI	PLS	1,8	0,3	0,98	0,789	0,9995	1,8	99,3	1,6
		PCR	2,4	1,7	0,90	0,016	0,9984	3,5	98,2	2,7
		CLS		2,9	0,92	0,034	0,9914	2,8	97,3	1,2
Turbezyd	RIF	PLS	1,5	1,2	0,98	0,888	0,9944	2,3	100,0	2,4
		PCR	1,4	1,1	0,98	0,888	0,9849	2,4	99,6	2,4
		CLS		1,6	0,98	0,902	0,9910	2,4	101,1	2,1
	INZ	PLS	2,5	2	0,98	0,866	0,9836	3,0	101,5	2,7
		PCR	2,4	1,6	0,98	0,866	0,9892	2,0	100,2	2,1
		CLS		3,3	0,96	0,289	0,9660	18,1	117,9	2,1
	PYR	PLS	1,5	1,3	0,95	0,125	0,9929	1,6	100,2	1,5
		PCR	1,6	1,3	0,95	0,125	0,9929	1,6	100,4	1,6
		CLS		1,5	0,93	0,043	0,9914	1,6	100,0	1,7

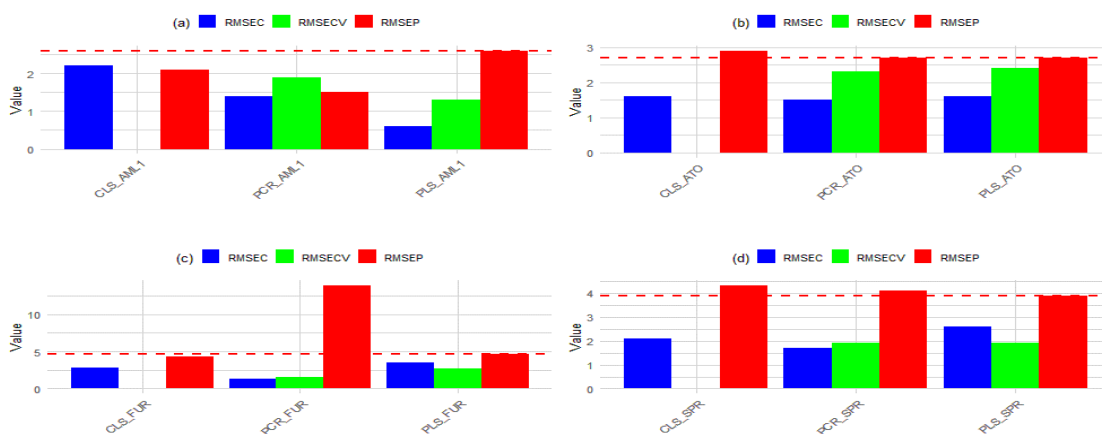
- Trong thuốc Natrixam: với AML, PLS và PCR có RMSECV và RMSEC đều là 1,5% và 1,4%, trong khi CLS có RMSEC cao hơn đáng kể (5,0%), cho thấy CLS kém hiệu quả hơn trong việc khớp dữ liệu huấn luyện. Tương tự, với thành phần IND, PLS có RMSECV là 2,9% và RMSEC là 2,4%, PCR cải thiện nhẹ với RMSECV là 2,6% và RMSEC là 2,3%, nhưng CLS có RMSEC rất cao (15,5%), cho thấy khả năng khớp rất kém.

- Trong thuốc Spirosemid-40: với FUR, PLS có RMSECV là 2,7% và RMSEC là 3,5%, PCR có RMSECV cao hơn (4,3%) nhưng RMSEC thấp hơn (2,1%), còn CLS có RMSEC là 2,8%, cho thấy sự dao động trong hiệu suất. Đối với SPR, PLS có RMSECV là 1,9% và RMSEC là 2,6%, PCR có RMSECV cao hơn (3,1%) và RMSEC thấp hơn (1,6%), còn CLS có RMSEC là 2,1%, cho thấy PLS và CLS tương đối ổn định hơn PCR.

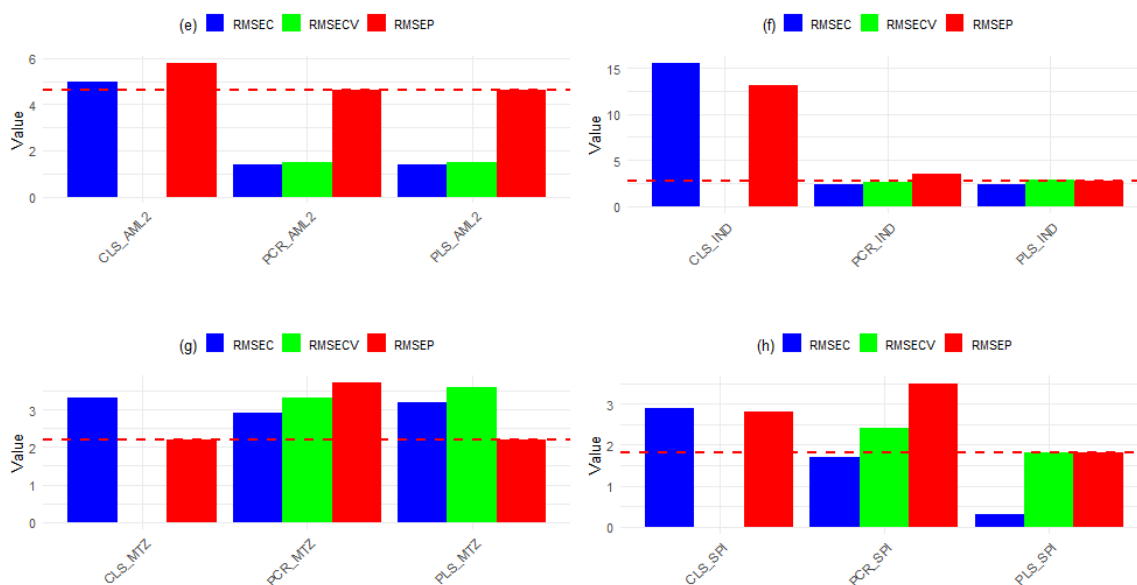
- Trong thuốc AgimDogyl: với MTZ, PLS có RMSECV là 3,6% và RMSEC là 3,2%, PCR cải thiện nhẹ với RMSECV là 3,3% và RMSEC là 2,9%, CLS có RMSEC là 3,3%, cho thấy hiệu suất tương đương. Đối với SPI, PLS có RMSECV là 1,8% và RMSEC rất thấp (0,3%), trong khi PCR có RMSECV là 2,4% và RMSEC là 1,7%, CLS có RMSEC là 2,9%, cho thấy PLS vượt trội hơn.

- Trong thuốc Turbezid: với RIF, PLS có RMSECV là 1,5% và RMSEC là 1,2%, PCR cải thiện nhẹ với RMSECV là 1,4% và RMSEC là 1,1%, CLS có RMSEC là 1,6%, cho thấy hiệu suất tương đương. Đối với INZ, PLS có RMSECV là 2,5% và RMSEC là 2,0%, PCR cải thiện với RMSECV là 2,4% và RMSEC là 1,6%, CLS có RMSEC là 3,3%, cho thấy CLS kém hơn. Với PYR, PLS và PCR có RMSECV và RMSEC tương đương (1,5%–1,6% và 1,3%), CLS có RMSEC là 1,5%, cho thấy hiệu suất tương đương.

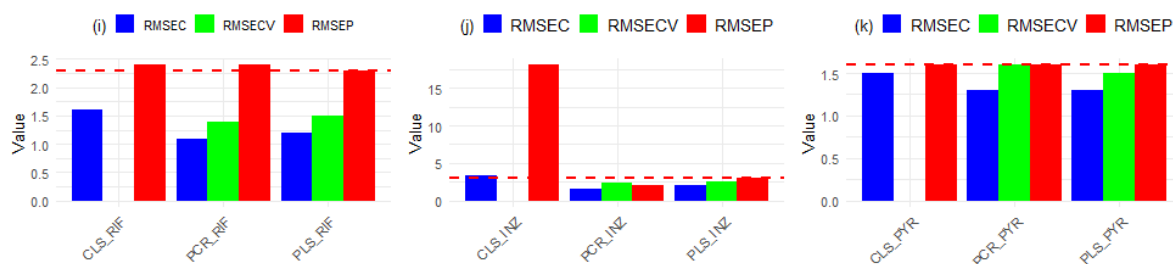
Nhìn chung, trên tập huấn luyện, PLS thường có hiệu suất tốt hơn PCR và CLS, đặc biệt với các thành phần như AML và SPI, trong khi CLS có xu hướng kém hơn, đặc biệt với IND và AML của Natrixam. So sánh các giá trị RMSE được minh họa trên các hình 3.18; hình 3.19; hình 3.20.



Hình 3.18. So sánh các giá trị RMSE trong các phương pháp PLS, CLS và PLS cho các tập thuốc Zoamco_A ((a) AML; (b) ATO) và Spirosemid-40 ((c) FUR; (d) SPR)



Hình 3.19. So sánh các giá trị RMSE trong các phương pháp PLS, CLS và PLS cho các tập thuốc Natrixam 1,5/5 ((e) AML; (f) IND) và AgimDogyl ((g) MTZ; (h) SPI)



Hình 3.20. So sánh các giá trị RMSE trong các phương pháp PLS, CLS và PLS cho các tập thuốc Turbezid ((i) RIF, (j) INZ và (k) PYR)

Về khả năng dự đoán trên tập kiểm tra, RMSEP là chỉ số quan trọng để đánh giá hiệu suất thực tế.

- Trong thuốc Zoamco-A: với AML, PLS có RMSEP là 2,6%, PCR giảm xuống 1,5%, và CLS là 2,1%, cho thấy PCR vượt trội trong dự đoán, có thể do khả năng tổng quát hóa tốt hơn, trong khi CLS vẫn tốt hơn PLS một chút. Với ATO, PLS và PCR có RMSEP tương đương (2,7%), nhưng CLS tăng lên 2,9%, cho thấy CLS kém hơn một chút.

- Trong thuốc Natrixam: với AML, PLS và PCR có RMSEP tương đương (4,6%), nhưng CLS tăng lên 5,8%, cho thấy CLS kém hiệu quả hơn đáng kể, có thể do không xử lý tốt nhiễu trong dữ liệu thực tế. Với IND, PLS có RMSEP là 2,7%, PCR tăng lên 3,5%, và CLS có RMSEP rất cao (13,1%), cho thấy CLS hoàn toàn không phù hợp trong trường hợp này, có thể do yêu cầu phổ thuần khiết không được đáp ứng.

- Trong thuốc Spirosemid-40: với FUR, PLS có RMSEP là 4,7%, PCR tăng nhẹ lên 4,8%, và CLS giảm xuống 4,3%, cho thấy CLS có hiệu suất tốt hơn một chút. Với

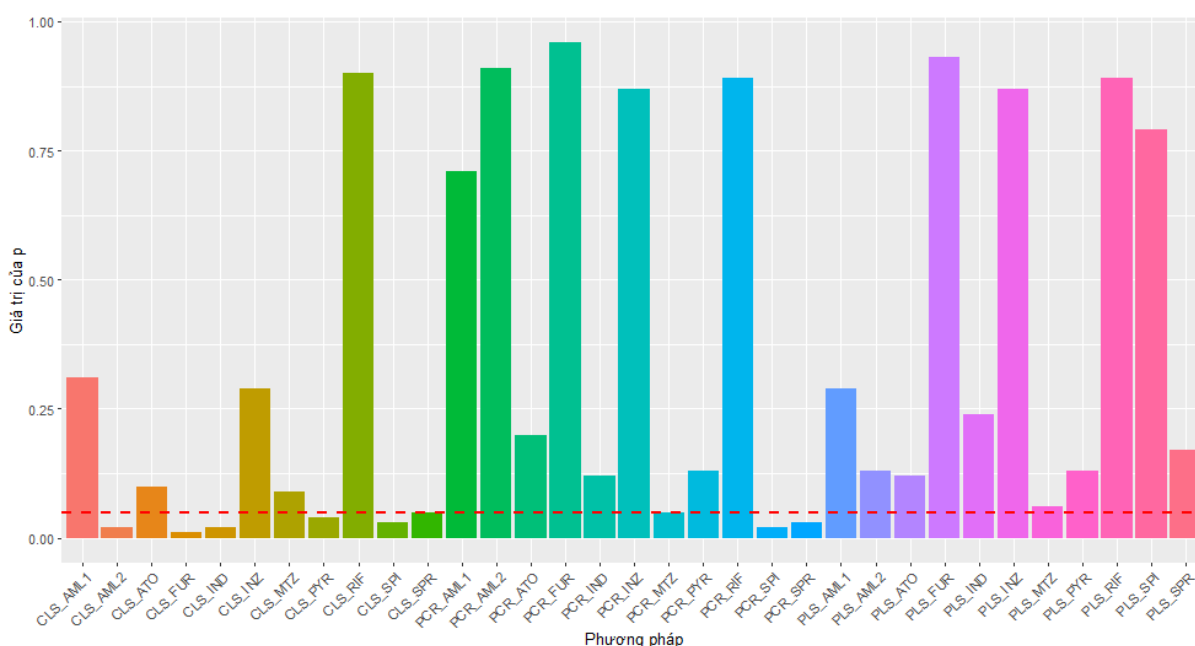
SPR, PLS có RMSEP là 3,9%, PCR giảm xuống 3,4%, CLS tăng lên 4,3%, cho thấy PCR vượt trội hơn.

- Trong thuốc AgimDogyl: với MTZ, PLS có RMSEP là 2,2%, PCR tăng lên 3,7%, CLS giữ ở 2,2%, cho thấy PLS và CLS tốt hơn PCR. Với SPI, PLS có RMSEP là 1,8%, PCR tăng lên 3,5%, CLS là 2,8%, cho thấy PLS vượt trội nhất.

- Trong thuốc Turbezid: với RIF, PLS có RMSEP là 2,3%, PCR và CLS đều là 2,4%, cho thấy hiệu suất tương đương. Với INZ, PLS có RMSEP là 3,0%, PCR giảm xuống 2,0%, nhưng CLS tăng mạnh lên 18,1%, cho thấy CLS thất bại hoàn toàn, có thể do nhiễu hoặc dữ liệu phức tạp. Với PYR, cả ba phương pháp có RMSEP tương đương (1,6%), cho thấy hiệu suất ngang nhau.

Nhìn chung, PLS thường có RMSEP ổn định hơn CLS, đặc biệt với các thành phần như AML và INZ của Natrixam, trong khi PCR đôi khi vượt trội (như AML của Zoamco_A và SPR), nhưng CLS có xu hướng kém hơn đáng kể trong nhiều trường hợp (như IND và INZ).

Chỉ số R^2 phản ánh khả năng giải thích phương sai của mô hình. Hầu hết các trường hợp của các phương pháp PLS, PCR và CLS giá trị R^2 đều cho giá trị xấp xỉ 0,99 hoặc lớn hơn. Riêng đối với INZ, phương pháp CLS có giá trị bằng 0,9660; đặc biệt đối với IND, giá trị R^2 chỉ bằng 0,5980, cho thấy CLS không giải thích tốt dữ liệu trong trường hợp này. Nhìn chung, PLS thường có R^2 cao hơn CLS, đặc biệt với IND và INZ, trong khi PCR có thể vượt trội trong một số trường hợp (như MTZ và INZ), nhưng CLS thường kém hơn đáng kể.



Hình 3.21. Giá trị p trong kiểm tra phân phối chuẩn của phần dư trong tập hiệu chuẩn của các phương pháp CLS, PCR và PLS bằng kiểm định Shapiro-Wilk

Về đánh giá phân tích phần dư giữa kết quả dự đoán nồng độ của mô hình và nồng độ trong tập hiệu chuẩn dựa trên kiểm định Shapiro-Wilk. Ngưỡng để khẳng định tập dữ liệu phân phối chuẩn là dựa vào giá trị $p > 0,05$. Kết quả được thể hiện trong hình 3.17. Phương pháp CLS có phân tích phần dư không tốt, ở các tập thí nghiệm AML và IND trong mẫu Natrifax, FUR trong mẫu Spirosemid, SPI trong mẫu AgimDogyl, PYR trong mẫu Turbezid đều có giá trị $p < 0,05$ chứng tỏ phần dư trong các mẫu này không tuân theo phân phối chuẩn, cho thấy mô hình không chính xác. Khi đánh giá phân phối chuẩn cho các tập dữ liệu phần dư được dự đoán bởi phương pháp PCR, có 3 tập hiệu chuẩn có phần dư chưa đạt phân phối chuẩn, có giá trị $p < 0,05$, bao gồm tập hiệu chuẩn MTZ, SPI trong mẫu AgimDogyl và PYR trong mẫu Turbezid. Điều này cho thấy phương pháp PCR có hiệu suất dự báo nhỏ hơn so với phương pháp PLS, mặc dù sự khác biệt này hầu như không đáng kể.

Về độ thu hồi và RSD, các chỉ số này cho thấy tính chính xác và ổn định của mô hình.

- Trong thuốc Zoamco_A: với AML, PLS có độ thu hồi 100,6% và RSD 2,6%, PCR có 99,9% và 1,6%, CLS có 99,6% và 2,4%, cho thấy PCR ổn định hơn. Với ATO, cả ba phương pháp có độ thu hồi 101,2%, nhưng CLS có RSD cao hơn (2,6%), cho thấy kém ổn định.

- Trong thuốc Natrifax: với AML, PLS có độ thu hồi 102,7% và RSD 2,8%, PCR gần tương đương với 102,6% và 2,8%, nhưng CLS có 101,3% và RSD cao (4,6%), cho thấy kém ổn định. Với IND, PLS có độ thu hồi 101,1% và RSD 2,4%, PCR là 102,4% và 2,6%, CLS là 108,9% và 12,6%, cho thấy CLS có sai lệch lớn.

- Trong thuốc Spirosemid-40: với FUR, PLS có độ thu hồi 103,1% và RSD 3,3%, PCR là 103,9% và 2,9%, CLS là 103,1% và 2,5%, cho thấy CLS ổn định hơn. Với SPR, PLS có độ thu hồi 101,3% và RSD 3,7%, PCR là 99,2% và 4,3%, CLS là 100,9% và 4,1%, cho thấy PLS chính xác hơn.

- Trong thuốc AgimDogyl: với MTZ, PLS có độ thu hồi 100,9% và RSD rất thấp (0,3%), PCR là 103,3% và 1,3%, CLS là 101,2% và 1,8%, cho thấy PLS ổn định nhất. Với SPI, PLS có độ thu hồi 99,3% và RSD 1,6%, PCR là 98,2% và 2,7%, CLS là 97,3% và 1,2%, cho thấy CLS ổn định hơn nhưng kém chính xác.

- Trong thuốc Turbezid: với RIF, PLS có độ thu hồi 100,0% và RSD 2,4%, PCR là 99,6% và 2,4%, CLS là 101,1% và 2,1%, cho thấy hiệu suất tương đương. Với INZ, PLS có độ thu hồi 101,5% và RSD 2,7%, PCR là 100,2% và 2,1%, CLS là 117,9% và 2,1%, cho thấy CLS có sai lệch lớn. Với PYR, cả ba phương pháp có độ thu hồi quanh 100,0%–100,4% và RSD 1,5%–1,7%, cho thấy hiệu suất tương đương.

Nhìn chung, PLS thường có độ thu hồi và RSD ổn định hơn CLS, đặc biệt với IND và INZ, trong khi PCR có thể tốt hơn trong một số trường hợp (như AML của Zoamco_A).

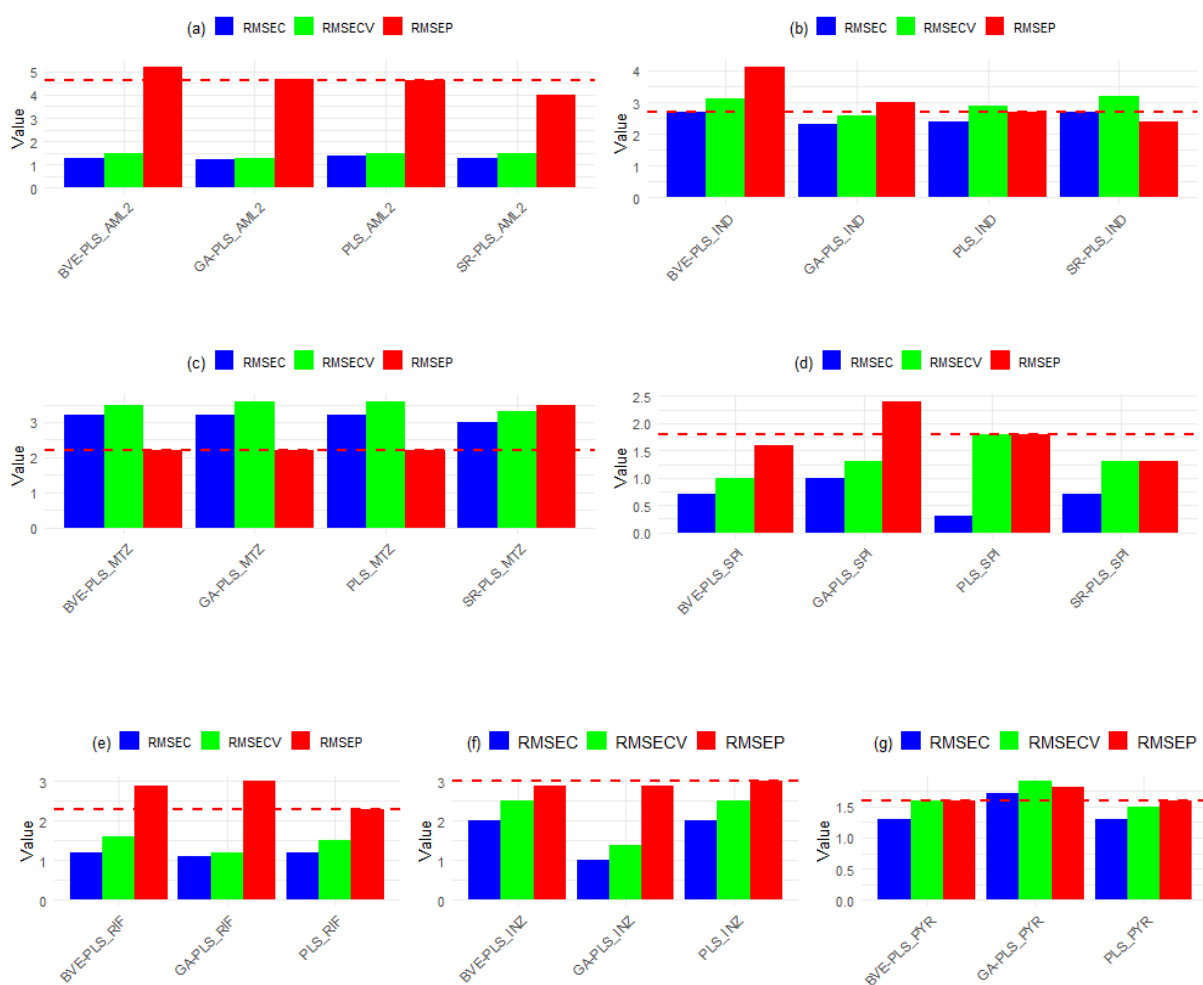
Tóm lại, so sánh giữa CLS, PCR và PLS cho thấy PLS thường vượt trội hơn CLS về RMSEP, R^2 , và độ thu hồi, đặc biệt với các thành phần như AML và IND của Natrixam, cũng như INZ của Turbezid, nơi CLS có sai số dự đoán và sai lệch rất lớn. PCR có hiệu suất dao động, đôi khi vượt trội hơn PLS về RMSEP (như AML của Zoamco_A và SPR), nhưng thường kém hơn về R^2 và độ thu hồi (như SPI). CLS có xu hướng kém hiệu quả nhất, đặc biệt trong dữ liệu phức tạp, do yêu cầu phổ thuần khiết không được đáp ứng. PLS là phương pháp đáng tin cậy nhất trong hầu hết các trường hợp, nhờ khả năng xử lý nhiễu và đa cộng tuyến tốt hơn, trong khi PCR có thể là lựa chọn thay thế trong một số trường hợp cụ thể.

3.4.2 Phương pháp PLS và phương pháp lựa chọn bước sóng

Bảng 3.48. Các chỉ số đánh giá mô hình và phân phối dữ liệu phân tích phần dư của các phương pháp lựa chọn bước sóng và phương pháp PLS

Thuốc	Thành phần	Phương pháp	RMSECV (%)	RMSEC (%)	Shapiro-Wilk		R^2	RMSEP (%)	Độ thu hồi	RSD (%)
					W	p				
Natrixam	AML	PLS	1,5	1,4	0,89	0,128	0,9992	4,6	102,7	2,8
		SR-PLS	1,5	1,3	0,96	0,128	0,9986	4,0	101,4	3,5
		BVE-PLS	1,5	1,3	0,92	0,207	0,9992	5,2	102,7	3,5
		GA-PLS	1,3	1,2	0,96	0,116	0,9994	4,7	102,7	2,9
	IND	PLS	2,9	2,4	0,91	0,240	0,9902	2,7	101,1	2,4
		SR-PLS	3,2	2,7	0,96	0,084	0,9906	2,4	100,3	2,3
		BVE-PLS	3,1	2,7	0,98	0,988	0,9874	4,1	103,9	1,4
		GA-PLS	2,6	2,3	0,91	0,223	0,9915	3,0	100,5	3,0
AgimDogyl	MTZ	PLS	3,6	3,2	0,93	0,061	0,9894	2,2	100,9	0,3
		SR-PLS	3,3	3,0	0,90	0,243	0,9910	3,5	103,0	1,4
		BVE-PLS	3,5	3,2	0,93	0,053	0,9898	2,2	101,1	1,9
		GA-PLS	3,6	3,2	0,93	0,055	0,9895	2,2	101,1	2,0
	SPI	PLS	1,8	0,3	0,98	0,789	0,9995	1,8	99,3	1,6
		SR-PLS	1,3	0,7	0,98	0,786	0,9995	1,3	99,5	1,3
		BVE-PLS	1,0	0,7	0,96	0,401	0,9995	1,6	99,2	1,3
		GA-PLS	1,3	1,0	0,95	0,223	0,9990	2,4	99,1	2,0
Turbezid	RIF	PLS	1,5	1,2	0,98	0,888	0,9944	2,3	100,0	2,4
		BVE-PLS	1,6	1,2	0,97	0,489	0,9941	2,9	101,7	2,3
		GA-PLS	1,2	1,1	0,95	0,216	0,9955	3,0	101,8	2,3
	INZ	PLS	2,5	2,0	0,98	0,866	0,9836	3,0	101,5	2,7
		BVE-PLS	2,5	2,0	0,98	0,885	0,9842	2,9	101,0	2,8
		GA-PLS	1,4	1,0	0,99	0,995	0,9962	2,9	102,0	2,1
	PYR	PLS	1,5	1,3	0,95	0,125	0,9929	1,6	100,2	1,5
		BVE-PLS	1,6	1,3	0,95	0,130	0,9929	1,6	100,3	1,6
		GA-PLS	1,9	1,7	0,97	0,659	0,9879	1,8	100,5	1,6

Phân tích và so sánh hiệu suất của các phương pháp lựa chọn bước sóng (SR, BVE, và GA) so với phương pháp PLS dựa trên các chỉ số thống kê được cung cấp trong bảng 3.48, là một nhiệm vụ quan trọng để đánh giá khả năng dự đoán và độ tin cậy của các mô hình trong phân tích dược phẩm. Các chỉ số bao gồm RMSECV, RMSEC, kiểm định Shapiro-Wilk (W và p-value), R^2 , RMSEP, độ thu hồi, và độ lệch chuẩn tương đối (RSD). Dữ liệu được thu thập từ các loại thuốc như Natrixam, AgimDogyl, Turbezid, với các thành phần hoạt chất như AML, IND, MTZ, SPI, RIF, INZ, và PYR, cho thấy sự đa dạng trong việc áp dụng các phương pháp hồi quy. Mục tiêu của phân tích này là đánh giá xem các phương pháp lựa chọn bước sóng có cải thiện hiệu suất so với PLS nguyên bản hay không, dựa trên các chỉ số chính như RMSEP và R^2 , đồng thời xem xét tính ổn định (RSD) và độ thu hồi.



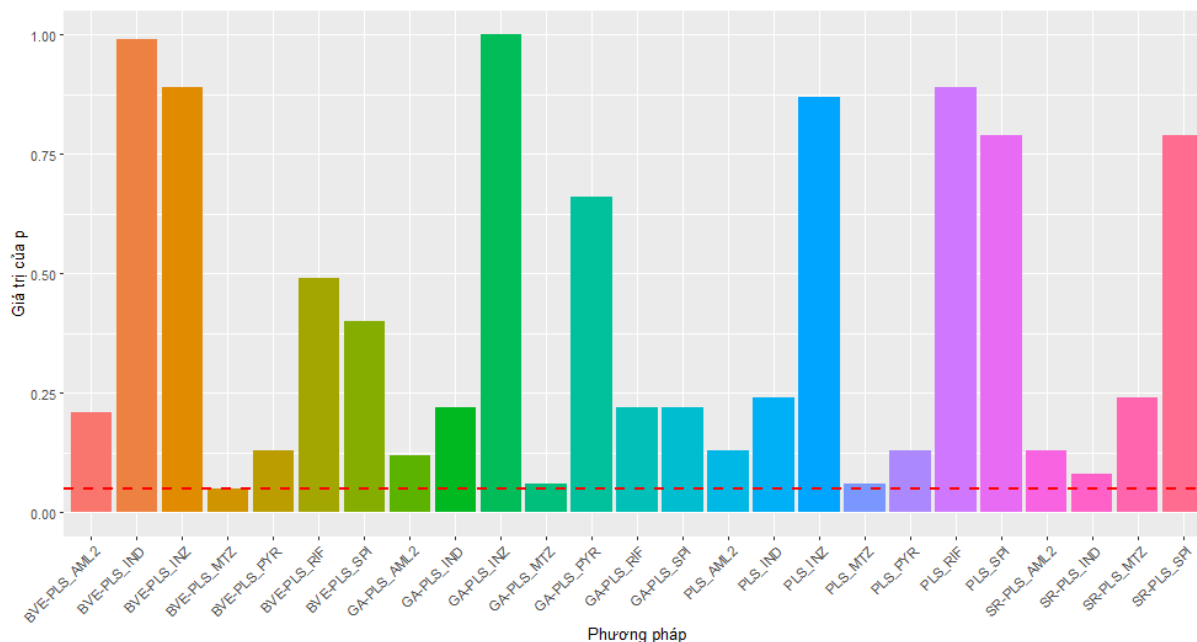
Hình 3.22. So sánh các giá trị RMSE trong các phương pháp SR-PLS, BVE-PLS, GA-PLS và PLS cho các tập thuốc Natrixam (a, b); AgimDogyl (c,d) và Turbezid (e, f, g)

Khi xem xét hiệu suất trên tập huấn luyện, các chỉ số RMSECV và RMSEC cung cấp cái nhìn ban đầu về khả năng khớp của mô hình với dữ liệu huấn luyện, được minh họa qua hình 3.21. Đối với thuốc Natrixam với thành phần AML, phương pháp PLS có RMSECV là 1,5% và RMSEC là 1,4%, trong khi các phương pháp SR-PLS, BVE-PLS,

và GA-PLS lần lượt đạt 1,5%, 1,5%, và 1,3% cho RMSECV, cùng với RMSEC lần lượt là 1,3%, 1,3%, và 1,2%. Điều này cho thấy GA-PLS có xu hướng cải thiện nhẹ RMSECV và RMSEC so với PLS, với sự giảm từ 1,5% xuống 1,3% cho RMSECV và từ 1,4% xuống 1,2% cho RMSEC, cho thấy khả năng tối ưu hóa tốt hơn nhờ thuật toán di truyền chọn các bước sóng quan trọng. Tuy nhiên, sự khác biệt này không quá lớn, và SR-PLS cùng BVE-PLS duy trì giá trị tương đương với PLS, cho thấy các phương pháp lựa chọn bước sóng không nhất thiết cải thiện đáng kể khả năng khớp trên tập huấn luyện. Đối với thành phần IND trong Natrixam, PLS có RMSECV là 2,9% và RMSEC là 2,4%, trong khi SR-PLS cải thiện nhẹ với RMSECV là 3,2% và RMSEC là 2,7%, nhưng BVE-PLS và GA-PLS cho thấy sự dao động với RMSECV lần lượt là 3,1% và 2,6%, RMSEC là 2,7% và 2,3%. Sự giảm xuống 2,6% của GA-PLS so với 2,9% của PLS cho thấy GA có thể loại bỏ các biến nhiễu hiệu quả hơn, trong khi BVE-PLS có sự tăng nhẹ RMSEC, có thể do quá trình loại bỏ biến không tối ưu. Với AgimDogyl (MTZ), PLS có RMSECV và RMSEC lần lượt là 3,6% và 3,2%, trong khi các phương pháp SR-PLS, BVE-PLS, và GA-PLS dao động quanh mức này với các giá trị từ 3,3% đến 3,6% cho RMSECV và 3,0% đến 3,2% cho RMSEC, cho thấy hiệu suất tương đương. Đối với SPI, PLS có RMSEC rất thấp (0,3%), trong khi SR-PLS và BVE-PLS giảm thêm xuống 0,7%, nhưng GA-PLS tăng lên 1,0%, cho thấy sự không nhất quán trong hiệu quả của GA trên thành phần này. Nhìn chung, các phương pháp lựa chọn bước sóng không mang lại sự cải thiện đáng kể trên tập huấn luyện so với PLS, với GA-PLS đôi khi nổi bật nhờ khả năng chọn biến tối ưu, nhưng SR-PLS và BVE-PLS thường duy trì hiệu suất tương đương.

Về khả năng dự đoán trên tập kiểm tra, chỉ số RMSEP là yếu tố quan trọng nhất để đánh giá hiệu suất thực tế của mô hình. Với Natrixam (AML), PLS có RMSEP là 4,6%, trong khi SR-PLS giảm xuống 4,0%, BVE-PLS tăng lên 5,2%, và GA-PLS là 4,7%. Sự giảm của SR-PLS so với PLS cho thấy phương pháp này có thể loại bỏ các bước sóng không liên quan, cải thiện độ chính xác dự đoán, trong khi BVE-PLS có sự gia tăng đáng kể, có thể do quá trình loại bỏ biến làm mất thông tin quan trọng. Đối với IND, PLS có RMSEP là 2,7%, SR-PLS giảm xuống 2,4%, nhưng BVE-PLS tăng lên 4,1% và GA-PLS là 3,0%, cho thấy SR-PLS hiệu quả nhất trong việc giảm sai số dự đoán, trong khi BVE-PLS có hiệu suất kém nhất, có thể do loại bỏ không chính xác các biến quan trọng. Với AgimDogyl (MTZ), PLS có RMSEP là 2,2%, trong khi SR-PLS tăng lên 3,5%, BVE-PLS và GA-PLS duy trì ở 2,2%, cho thấy SR-PLS không hiệu quả trong trường hợp này, trong khi BVE-PLS và GA-PLS giữ ổn định. Đối với SPI, PLS có RMSEP là 1,8%, SR-PLS giảm xuống 1,3%, BVE-PLS là 1,6%, và GA-PLS tăng lên 2,4%, cho thấy SR-PLS vượt trội trong việc cải thiện dự đoán, trong khi GA-PLS lại có hiệu suất kém hơn. Với Turbezid (RIF), PLS có RMSEP là 2,3%, BVE-PLS tăng lên 2,9%, và GA-PLS là 3,0%, cho thấy cả hai phương pháp lựa chọn đều không cải thiện so với PLS. Đối với INZ, PLS có RMSEP là 3,0%, BVE-PLS giảm nhẹ xuống 2,9%, và GA-PLS cũng là 2,9%, trong khi với PYR, PLS có RMSEP là 1,6%, các phương pháp

BVE-PLS và GA-PLS dao động quanh mức 1,6% đến 1,8%. Nhìn chung, SR-PLS thường cho thấy khả năng giảm RMSEP tốt nhất trong một số trường hợp (ví dụ: AML, IND, SPI), nhưng hiệu quả không nhất quán qua các loại thuốc, trong khi BVE-PLS và GA-PLS có xu hướng dao động, đôi khi tăng sai số dự đoán so với PLS.



Hình 3.23. So sánh giá trị p giữa các phương pháp lựa chọn bước sóng với phương pháp PLS

Chỉ số R^2 phản ánh khả năng giải thích phương sai của mô hình, và sự so sánh giữa PLS và các phương pháp lựa chọn bước sóng cho thấy một số biến thiên. Với Natrixam (AML), PLS có R^2 là 0,9992, trong khi SR-PLS là 0,9986, BVE-PLS là 0,9992, và GA-PLS là 0,9994, cho thấy GA-PLS có xu hướng cải thiện nhẹ khả năng giải thích so với PLS. Đối với IND, PLS có R^2 là 0,9902, SR-PLS tăng lên 0,9906, BVE-PLS giảm xuống 0,9874, và GA-PLS là 0,9915, cho thấy GA-PLS và SR-PLS có hiệu suất tốt hơn, trong khi BVE-PLS kém hơn. Với AgimDogyl (MTZ), PLS có R^2 là 0,9894, và các phương pháp SR-PLS, BVE-PLS, GA-PLS dao động từ 0,9895 đến 0,9910, cho thấy sự tương đương. Đối với SPI, PLS có R^2 rất cao (0,9995), và cả SR-PLS, BVE-PLS, GA-PLS duy trì ở mức 0,9990 đến 0,9995, cho thấy hiệu suất tương đương. Với Turbezid (RIF), PLS có R^2 là 0,9944, BVE-PLS là 0,9941, và GA-PLS là 0,9955, cho thấy GA-PLS có cải thiện nhẹ. Đối với INZ, PLS có R^2 là 0,9836, BVE-PLS tăng lên 0,9842, và GA-PLS đạt 0,9962, cho thấy GA-PLS vượt trội. Với PYR, PLS có R^2 là 0,9929, và các phương pháp BVE-PLS và GA-PLS dao động quanh mức này. Nhìn chung, GA-PLS thường cải thiện R^2 tốt nhất trong một số trường hợp (như AML, INZ), nhưng hiệu quả không đồng đều, trong khi SR-PLS và BVE-PLS đôi khi duy trì hoặc giảm nhẹ so với PLS.

Độ thu hồi và RSD cung cấp thông tin về tính chính xác và ổn định của mô hình. Với Natrixam (AML), PLS có độ thu hồi 102,7% và RSD 2,8%, trong khi SR-PLS có 101,4% và 3,5%, BVE-PLS 102,7% và 3,5%, GA-PLS 102,7% và 2,9%, cho thấy SR-PLS giảm độ thu hồi nhưng tăng RSD, trong khi GA-PLS duy trì ổn định. Đối với IND, PLS có độ thu hồi 101,1% và RSD 2,4%, SR-PLS có 100,3% và 2,3%, BVE-PLS 103,9% và 1,4%, GA-PLS 100,5% và 3,0%, cho thấy BVE-PLS cải thiện RSD đáng kể, nhưng độ thu hồi cao hơn có thể là dấu hiệu quá khớp. Với AgimDogyl (MTZ), PLS có độ thu hồi 100,9% và RSD 0,3%, trong khi SR-PLS có 103,0% và 1,4%, BVE-PLS 101,1% và 1,9%, GA-PLS 101,1% và 2,0%, cho thấy SR-PLS tăng độ thu hồi nhưng giảm ổn định. Đối với SPI, PLS có độ thu hồi 99,3% và RSD 1,6%, SR-PLS có 99,5% và 1,3%, BVE-PLS 99,2% và 1,3%, GA-PLS 99,1% và 2,0%, cho thấy SR-PLS và BVE-PLS cải thiện RSD, trong khi GA-PLS kém hơn. Với Turbezid (RIF), PLS có độ thu hồi 100,0% và RSD 2,4%, BVE-PLS 101,7% và 2,3%, GA-PLS 101,8% và 2,3%, cho thấy hiệu suất tương đương. Đối với INZ, PLS có độ thu hồi 101,5% và RSD 2,7%, BVE-PLS 101,0% và 2,8%, GA-PLS 102,0% và 2,1%, cho thấy GA-PLS ổn định hơn. Với PYR, PLS có độ thu hồi 100,2% và RSD 1,5%, các phương pháp khác dao động quanh mức này. Nhìn chung, SR-PLS và BVE-PLS có xu hướng cải thiện RSD trong một số trường hợp, nhưng GA-PLS không nhất quán, và độ thu hồi đôi khi tăng quá mức, có thể dẫn đến sai lệch.

Tóm lại, so sánh giữa PLS và các phương pháp lựa chọn bước sóng cho thấy hiệu quả không đồng đều. SR-PLS thường cải thiện RMSEP tốt nhất trong một số trường hợp (như AML, IND, SPI), nhờ khả năng chọn biến dựa trên tỷ lệ chọn, nhưng hiệu quả không ổn định qua các loại thuốc. BVE-PLS có xu hướng tăng RMSEP (như AML, IND), có thể do loại bỏ không chính xác các biến quan trọng, nhưng lại cải thiện RSD trong một số trường hợp (như IND). GA-PLS cho thấy tiềm năng giảm RMSECV và RMSEC (như AML, IND), nhờ thuật toán di truyền tối ưu hóa, nhưng hiệu suất dự đoán (RMSEP) và RSD không nhất quán, có thể do phụ thuộc vào thiết lập tham số. PLS vẫn là phương pháp cơ bản hiệu quả, đặc biệt với độ thu hồi cao và R^2 tốt, nhưng các phương pháp lựa chọn bước sóng có thể hữu ích trong các trường hợp cụ thể, tùy thuộc vào đặc điểm dữ liệu và yêu cầu tối ưu hóa.

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 được.

(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 thỏa 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 Trâm, 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>

TÀI LIỆU THAM KHẢO

- [1] WHO. "1 in 10 medical products in developing countries is substandard or falsified." <https://www.who.int/news/item/28-11-2017-1-in-10-medical-products-in-developing-countries-is-substandard-or-falsified> (accessed April 5, 2025).
- [2] F. El-Dahiyat, K. M. Fafelelbom, A. A. Jairoun, and S. S. Al-Hemyari, "Combatting substandard and falsified medicines: public awareness and identification of counterfeit medications," (in English), *Frontiers in Public Health*, vol. 9, p. 754279, 2021.
- [3] M.K. Olson, "Regulation of Safety, Efficacy, and Quality," (in English), *Encyclopedia of Health Economics*, pp. 240-248, 2014, doi: <https://doi.org/10.1016/B978-0-12-375678-7.01202-5>.
- [4] V. D'Atri *et al.*, "Trends in Pharmaceutical Analysis: The Evolving Role of Liquid Chromatography," (in English), *Analytical Chemistry*, 2025.
- [5] I. Siddique, "High-performance liquid chromatography: comprehensive techniques and cutting-edge innovations," (in English), *European Journal of Advances in Engineering and Technology*, vol. 10, no. 9, pp. 66-70, 2023.
- [6] D. Gupta *et al.*, "Simultaneous spectrophotometric determination of drug components from their dosage formulations," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 270, p. 120819, 2022.
- [7] N. K. Karmakar, S. Pandey, R. K. Pandey, and S. S. Shukla, "Solvatochromism: a tool for solvent discretion for UV-Vis spectroscopic studies," (in English), *Applied Spectroscopy Reviews*, vol. 56, no. 6, pp. 513-529, 2021.
- [8] C. Vogt, C. S. Wondergem, and B. M. Weckhuysen, "Ultraviolet-visible (UV-vis) spectroscopy," in *Springer handbook of advanced catalyst characterization*: Springer, 2023, pp. 237-264.
- [9] K. Attala and A. Elsonbaty, "Advanced eco-friendly UV spectrophotometric approach for resolving overlapped spectral signals of antihypertensive agents in their binary and tertiary pharmaceutical dosage form," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 258, p. 119855, 2021.
- [10] M. Ebrahimi, M. R. Sohrabi, F. Motiee, and M. Davallo, "Rapid simultaneous spectrophotometric determination of acetaminophen, phenylephrine, and guaifenesin in a cold syrup formulation based on continuous wavelet transform and first derivative transform methods," (in English), *Optik*, vol. 230, p. 166323, 2021.
- [11] M. Otto, "What is Chemometrics?," in *Chemometrics: statistics and computer application in analytical chemistry*: John Wiley & Sons, 2023, ch. 1, p. 1.
- [12] L. Wulandari, R. Idroes, T. R. Noviany, and G. Indrayanto, "Application of chemometrics using direct spectroscopic methods as a QC tool in pharmaceutical industry and their validation," (in English), *Profiles of Drug Substances, Excipients and Related Methodology*, vol. 47, pp. 327-379, 2022.
- [13] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," (in English), *Chemometrics and intelligent laboratory systems*, vol. 58, no. 2, pp. 109-130, 2001.
- [14] R. C. Rial, "AI in analytical chemistry: Advancements, challenges, and future directions," (in English), *Talanta*, p. 125949, 2024.

- [15] M. S. Elmasry, W. S. Hassan, M. Y. El-Mammli, and M. Badrawy, "Earth friendly spectrophotometric methods based on different manipulation approaches for simultaneous determination of aspirin and omeprazole in binary mixture and pharmaceutical dosage form: comparative statistical study," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 266, p. 120436, 2022.
- [16] L. H. Nurani *et al.*, "Chemometrics-assisted UV-vis spectrophotometry for quality control of pharmaceuticals: A review," (in English), *Indonesian Journal of Chemistry*, vol. 23, no. 2, pp. 542-567, 2023.
- [17] P. Gemperline, *Practical guide to chemometrics*. CRC press (in English), 2006.
- [18] L. Eriksson and E. Johansson, "Multivariate design and modeling in QSAR," (in English), *Chemometrics and intelligent laboratory systems*, vol. 34, no. 1, pp. 1-19, 1996.
- [19] G. Quintás, N. Portillo, J. C. García-Cañaveras, J. V. Castell, A. Ferrer, and A. Lahoz, "Chemometric approaches to improve PLSDA model outcome for predicting human non-alcoholic fatty liver disease using UPLC-MS as a metabolic profiling tool," (in English), *Metabolomics*, vol. 8, pp. 86-98, 2012.
- [20] C. Ratnasekhar, P. Rathor, and N. Birse, "The chemometric models in metabolomics," in *Comprehensive Analytical Chemistry*, vol. 104: Elsevier, 2024, pp. 21-42.
- [21] D. MacKay, "Chemometrics, econometrics, psychometrics—How best to handle hedonics?," (in English), *Food Quality and Preference*, vol. 17, no. 7-8, pp. 529-535, 2006.
- [22] R. G. Brereton, "Introduction to multivariate calibration in analytical chemistry Electronic Supplementary Information available. See <http://www.rsc.org/suppdata/an/b0/b003805i>," (in English), *Analyst*, vol. 125, no. 11, pp. 2125-2154, 2000.
- [23] E. M., C., Ariño, and JM., Díaz-Cruz,, "Chemometrics in Electrochemistry," in *Comprehensive chemometrics: chemical and biochemical data analysis*, vol. 4: Elsevier, 2019, ch. 1, pp. 425–458.
- [24] H.-P. Wang *et al.*, "Recent advances of chemometric calibration methods in modern spectroscopy: Algorithms, strategy, and related issues," (in English), *TrAC Trends in Analytical Chemistry*, vol. 153, p. 116648, 2022.
- [25] C. Y. Choo, N. A. Mohd Nor, N. Sahidan, and Y. Hitotsuyanagi, "Biomarkers identification of Lycopodiaceae and Huperziaceae species from peninsular Malaysia with HPLC chromatographic profiling and partial least square analysis," (in English), *Separation Science Plus*, vol. 1, no. 10, pp. 660-668, 2018.
- [26] K. Kumar, "Partial least square (PLS) analysis: Most favorite tool in chemometrics to build a calibration model," (in English), *Resonance*, vol. 26, pp. 429-442, 2021.
- [27] R. G. Brereton, *Chemometrics: data driven extraction for science*. John Wiley & Sons (in English), 2018.
- [28] Z. Hou, W. Song, and Z. Wang, "Partial Least Squares," (in English), *Chemometrics and Numerical Methods in LIBS*, pp. 277-302, 2022.
- [29] C. T. Lin, "Applications of partial least squares to spectral data for quantitative determinations of chemicals in a mixture," (in English), *Drug Information Journal*, vol. 24, no. 4, pp. 671-689, 1990.

- [30] A. Baum and L. Vermue, "Multiblock PLS: Block dependent prediction modeling for Python," (in English), *Journal of Open Source Software*, vol. 4, no. 34, p. 1190, 2019.
- [31] H. Abdi, "Partial least square regression (PLS regression)," (in English), *Encyclopedia for research methods for the social sciences*, vol. 6, no. 4, pp. 792-795, 2003.
- [32] Z. Chen, "Application of UV-vis spectroscopy in the detection and analysis of substances," in *3rd International Conference on Biotechnology, Life Science and Medical Engineering*, 2024.
- [33] G. Kaur, H. Singh, and J. Singh, "UV-vis spectrophotometry for environmental and industrial analysis," in *Green Sustainable Process for Chemical and Environmental Engineering and Science*: Elsevier, 2021, pp. 49-68.
- [34] A. Singhal, U. Saini, B. Chopra, A. K. Dhingra, A. Jain, and J. Chaudhary, "Uv-visible spectroscopy: A review on its pharmaceutical and bio-allied sciences applications," (in English), *Current Pharmaceutical Analysis*, vol. 20, no. 3, pp. 161-177, 2024, doi: <http://dx.doi.org/10.2174/0115734129300562240408042614>.
- [35] M. S. H. Akash, K. Rehman, M. S. H. Akash, and K. Rehman, "Ultraviolet-visible (UV-VIS) spectroscopy," in *Essentials of pharmaceutical analysis*: Springer, Singapore, 2020, pp. 29-56.
- [36] F. S. Rocha, A. J. Gomes, C. N. Lunardi, S. Kaliaguine, and G. S. Patience, "Experimental methods in chemical engineering: Ultraviolet visible spectroscopy—UV-Vis," (in English), *The Canadian Journal of Chemical Engineering*, vol. 96, no. 12, pp. 2512-2517, 2018, doi: <https://doi.org/10.1002/cjce.23344>.
- [37] R. Bro, "Multivariate calibration: What is in chemometrics for the analytical chemist?," (in English), *Analytica Chimica Acta*, vol. 500, no. 1-2, pp. 185-194, 2003, doi: [https://doi.org/10.1016/S0003-2670\(03\)00681-0](https://doi.org/10.1016/S0003-2670(03)00681-0).
- [38] P. Oliveri, C. Malegori, and M. Casale, "Chemometrics: Multivariate analysis of chemical data," in *Chemical analysis of food*: Elsevier, 2020, pp. 33-76.
- [39] M. Otto, *Chemometrics: statistics and computer application in analytical chemistry*. John Wiley & Sons (in English), 2023.
- [40] M. Bystrzanowska and M. Tobiszewski, "Chemometrics for selection, prediction, and classification of sustainable solutions for green chemistry—A review," (in English), *Symmetry*, vol. 12, no. 12, p. 2055, 2020, doi: <https://doi.org/10.3390/sym12122055>.
- [41] R. K. Ahmed, E. M. Saad, H. M. Fahmy, and R. M. El Nashar, "Multivariate experimental design: towards more reliable electrochemical detection," (in English), *Current Opinion in Electrochemistry*, vol. 31, p. 100880, 2022, doi: <https://doi.org/10.1016/j.coelec.2021.100880>.
- [42] F. A. N. Fernandes, S. Rodrigues, and E. G. Alves Filho, *Chemometrics: Data Treatment and Applications*. Elsevier (in English), 2024.
- [43] B. Lavine and J. Workman, "Chemometrics," (in English), *Analytical chemistry*, vol. 80, no. 12, pp. 4519-4531, 2008, doi: <https://doi.org/10.1021/ac800728t>.
- [44] O. Monago-Marana, I. Durán-Merás, A. M. de la Peña, and T. Galeano-Díaz, "Analytical techniques and chemometrics approaches in authenticating and identifying adulteration of paprika powder using fingerprints: A review," (in

- English), *Microchemical Journal*, vol. 178, p. 107382, 2022, doi: <https://doi.org/10.1016/j.microc.2022.107382>.
- [45] M. D. Peris-Díaz and A. Krężel, "A guide to good practice in chemometric methods for vibrational spectroscopy, electrochemistry, and hyphenated mass spectrometry," (in English), *TrAC Trends in Analytical Chemistry*, vol. 135, p. 116157, 2021, doi: <https://doi.org/10.1016/j.trac.2020.116157>.
- [46] M. C. Santos, P. A. M. Nascimento, W. N. Guedes, E. R. Pereira Filho, É. R. Filletti, and F. M. V. Pereira, "Chemometrics in analytical chemistry—an overview of applications from 2014 to 2018," (in English), *Eclética Química*, vol. 44, no. 2, pp. 11-25, 2019, doi: <https://doi.org/10.26850/1678-4618eqj.v44.2.11-25>.
- [47] K. Kalinowska, M. Bystrzanowska, and M. Tobiszewski, "Chemometrics approaches to green analytical chemistry procedure development," (in English), *Current Opinion in Green and Sustainable Chemistry*, vol. 30, p. 100498, 2021, doi: <https://doi.org/10.1016/j.cogsc.2021.100498>.
- [48] H. Martens and T. Næs, "Multivariate calibration. I. Concepts and distinctions," (in English), *TrAC Trends in Analytical Chemistry*, vol. 3, no. 8, pp. 204-210, 1984, doi: [https://doi.org/10.1016/0165-9936\(84\)85008-6](https://doi.org/10.1016/0165-9936(84)85008-6).
- [49] A. C. Olivieri, N. M. Faber, J. Ferré, R. Boqué, J. H. Kalivas, and H. Mark, "Uncertainty estimation and figures of merit for multivariate calibration (IUPAC Technical Report)," (in English), *Pure and Applied Chemistry*, vol. 78, no. 3, pp. 633-661, 2006, doi: <https://doi.org/10.1351/pac200678030633>.
- [50] A. C. Olivieri, *Introduction to multivariate calibration: A practical approach*. Springer Nature (in English), 2024.
- [51] G. Hanrahan, F. Udeh, and D. Patil, "Chemometrics and statistics| multivariate calibration techniques," (in English), pp. 27-32, 2005, doi: <https://doi.org/10.1016/B0-12-369397-7/00077-7>.
- [52] A. Diwan and M. R. Linford, "An introduction to classical least squares (CLS) and multivariate curve resolution (MCR) as applied to UV-VIS, FTIR, and ToF-SIMS," (in English), *Vac Technol Coat*, 2015.
- [53] R. Kramer, *Chemometric techniques for quantitative analysis*. CRC press (in English), 1998.
- [54] D. M. Haaland and E. V. Thomas, "Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information," (in English), *Analytical chemistry*, vol. 60, no. 11, pp. 1193-1202, 1988, doi: <https://doi.org/10.1021/ac00162a020>.
- [55] A. El-Gindy and G. M. Hadad, "Chemometrics in pharmaceutical analysis: an introduction, review, and future perspectives," (in English), *Journal of AOAC International*, vol. 95, no. 3, pp. 609-623, 2012, doi: https://doi.org/10.5740/jaoacint.SGE_El-Gindy.
- [56] M. Blanco, J. Coello, H. Iturriaga, S. Maspocho, and S. Alaoui-Ismaïli, "Use of inverse multiple linear regression (ILS) for the analytical control of pharmaceutical preparations. UV-visible spectrophotometric quantitation of an active principal in the presence of absorbing excipients," (in English), vol. 32, no. 6, 1999, doi: <https://doi.org/10.1080/00032719908542886>.
- [57] J. L. Godoy, J. R. Vega, and J. L. Marchetti, "Relationships between PCA and PLS-regression," (in English), *Chemometrics and Intelligent Laboratory*

- Systems, vol. 130, pp. 182-191, 2014, doi: <https://doi.org/10.1016/j.chemolab.2013.11.008>.
- [58] J. Li, Y. Luo, and H. Dai, "Application of principal component regression and partial least squares regression in ultraviolet spectrum water quality detection," in *2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology*, 2018, vol. 10620: SPIE, pp. 235-244, doi: <https://doi.org/10.1117/12.2285011>.
 - [59] H. Artigue and G. Smith, "The principal problem with principal components regression," (in English), *Cogent Mathematics & Statistics*, vol. 6, no. 1, p. 1622190, 2019, doi: <https://doi.org/10.1080/25742558.2019.1622190>.
 - [60] K. Varmuza and P. Filzmoser, *Introduction to multivariate statistical analysis in chemometrics*. CRC press (in English), 2016.
 - [61] R. W. Pinheiro, "Chemometrics with R: multivariate data analysis in the natural sciences and life sciences," ed: Springer, Heidelberg, 2012.
 - [62] F. Cadet, D. Pérez-Guaita, S. Garrigues, and M. de la Guardia, "Quantitative analysis, infrared," (in English), *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, pp. 1-49, 2006.
 - [63] D. M. Pirouz, "An overview of partial least squares," (in English), *Available at SSRN 1631359*, 2006, doi: <https://dx.doi.org/10.2139/ssrn.1631359>.
 - [64] A. K. Sankar, T. Vetrichevan, and D. Venkappaya, "Simultaneous estimation of ramipril, acetylsalicylic acid and atorvastatin calcium by chemometrics assisted UV-spectrophotometric determinations in capsules," (in English), *Acta Pharmaceutica*, vol. 61, no. 3, pp. 283-296, 2011, doi: <https://doi.org/10.2478/v10007-011-0027-1>.
 - [65] M. Stocchero, "PLS for designed experiments," (in English), *Chemometrics and Intelligent Laboratory Systems*, vol. 240, p. 104928, 2023.
 - [66] M. Stocchero, M. De Nardi, and B. Scarpa, "PLS for classification," (in English), *Chemometrics and Intelligent Laboratory Systems*, vol. 216, p. 104374, 2021.
 - [67] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *International Statistical and Optimization Perspectives Workshop" Subspace, Latent Structure and Feature Selection"*, 2005: Springer, pp. 34-51.
 - [68] J. Beltran, R. Ferrer, and J. Guiteras, "Multivariate calibration of polycyclic aromatic hydrocarbon mixtures from excitation–emission fluorescence spectra," (in English), *Analytica chimica acta*, vol. 373, no. 2-3, pp. 311-319, 1998.
 - [69] K. A. Einarson *et al.*, "Predicting pectin performance strength using near-infrared spectroscopic data: A comparative evaluation of 1-D convolutional neural network, partial least squares, and ridge regression modeling," (in English), *Journal of Chemometrics*, vol. 36, no. 2, p. e3348, 2022.
 - [70] T. Mehmood, S. Sæbø, and K. H. Liland, "Comparison of variable selection methods in partial least squares regression," (in English), *Journal of Chemometrics*, vol. 34, no. 6, p. e3226, 2020.
 - [71] Y. Huang, "Linear Calibration Methods," in *Chemometric Methods in Analytical Spectroscopy Technology*: Springer, 2022, ch. 7, pp. 237-253.
 - [72] O. Silakari and P. K. Singh, "Statistical methods and parameters: Tools to generate and evaluate theoretical in silico models," in *Concepts and experimental protocols of modelling and informatics in drug design*: Academic Press, 2020, ch. 16, pp. 333-340.

- [73] S. Sato and Y. Numata, "Simultaneous quantitative analysis of quercetin and rutin in Tartary buckwheat flour by Raman spectroscopy and partial least square regression," (in English), *Journal of Food Composition and Analysis*, vol. 128, p. 105991, 2024, doi: <https://doi.org/10.1016/j.jfca.2024.105991>.
- [74] Z. Ali *et al.*, "Determination of florfenicol by Raman spectroscopy with principal component analysis (PCA) and partial least squares regression (PLSR)," (in English), *Analytical Letters*, vol. 57, no. 1, pp. 30-40, 2024, doi: <https://doi.org/10.1080/00032719.2023.2192942>.
- [75] K. Yaginuma, K. Matsunami, L. Descamps, A. Ryckaert, and T. De Beer, "Hybrid modeling of T-shaped partial least squares regression and transfer learning for formulation and manufacturing process development of new drug products," (in English), *International Journal of Pharmaceutics*, vol. 662, p. 124463, 2024, doi: <https://doi.org/10.1016/j.ijpharm.2024.124463>.
- [76] R. M. Mahmood, S. A. Darweesh, N. A. Alassaf, and R. S. Al-Khalisy, "Simultaneous Spectrophotometric Determination of Piroxicam, Naproxen, Diclofenac Sodium and Mefenamic Acid in Pharmaceutical Formulations by Partial Least Squares Method," (in English), *Methods & Objects of Chemical Analysis/Metody & Obekty Himičeskogo Analiza*, vol. 19, no. 2, 2024, doi: <https://doi.org/10.17721/moca.2024.101-110>.
- [77] K. F. Amin, R. H. Obaydo, and A. M. Abdullah, "Eco-friendly chemometric analysis: Sustainable quantification of five pharmaceutical compounds in bulk, tablets, and spiked human plasma," (in English), *Results in Chemistry*, vol. 11, p. 101761, 2024, doi: <https://doi.org/10.1016/j.rechem.2024.101761>.
- [78] Y.-Q. Zhong, J.-Q. Li, X.-L. Li, S.-Y. Dai, and F. Sun, "Near infrared spectroscopy for simultaneous quantification of five chemical components in arnebiae radix (AR) with partial least squares and support vector machine algorithms," (in English), *Vibrational Spectroscopy*, vol. 127, p. 103556, 2023, doi: <https://doi.org/10.1016/j.vibspec.2023.103556>.
- [79] E. Guzelmeric, D. Özdemir, N. B. Sen, C. Celik, and E. Yesilada, "Quantitative determination of phenolic compounds in propolis samples from the Black Sea Region (Türkiye) based on HPTLC images using partial least squares and genetic inverse least squares methods," (in English), *Journal of Pharmaceutical and Biomedical Analysis*, vol. 229, p. 115338, 2023, doi: <https://doi.org/10.1016/j.jpba.2023.115338>.
- [80] N. W. Xuejia Zhao, Yitong Liu, Yifei Jiang, Shengnan Sun, Xiaohui Wang, Jing Yao, Guangzhi Shan, "Fast quantification of diclofenac sodium content in commercially available tablets by transmission Raman spectroscopy with PLS," (in English), *Journal of Chemometrics*, vol. 37, no. 2, 2022, doi: <https://doi.org/10.1002/cem.3462>.
- [81] Z. Z. Yuwen Zhao, Ying Ning, Peiqi Miao, Zheng Li, Haixia Wang, "Simultaneous quantitative analysis of Escherichia coli, Staphylococcus aureus and Salmonella typhimurium using surface-enhanced Raman spectroscopy coupled with partial least squares regression and artificial neural networks," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 295, p. 122510, 2023, doi: <https://doi.org/10.1016/j.saa.2023.122510>.
- [82] G. A. C. Purba, Muchlisyam, M., & Putra, E. D. L., "Determination of The Level of Paracetamol and Tramadol Hcl Mixture in Tablets by Chemometric Method

- by Fourier Transform Infrared Spectrophotometry," (in English), *Eduvest - Journal of Universal Studies*, vol. 5, no. 1, pp. 297-304, 2025, doi: <https://doi.org/10.59188/eduvest.v5i1.50283>.
- [83] T. H. Meng Wang, Yuhang Li, Rui Wang, Yudie Xu, Yabo Shi, Huangjin Tong, Mengting Yu, Yuwen Qin, and L. S. Xi Mei, Chunqin Mao, Tulin Lu, Lin Li, De Ji, Chengxi Jiang,, "An integrated and rapid evaluation of Curcumae Radix from different botanical origins based on chemical components, antiplatelet aggregation effect and Fourier transform near-infrared spectroscopy," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 324, p. 124992, 2025, doi: <https://doi.org/10.1016/j.saa.2024.124992>.
- [84] N. S. Katamesh, A. E. F. Abbas, and S. A. Mahmoud, "Four chemometric models enhanced by Latin hypercube sampling design for quantification of anti-COVID drugs: sustainability profiling through multiple greenness, carbon footprint, blueness, and whiteness metrics," (in English), *BMC chemistry*, vol. 18, no. 1, p. 54, 2024, doi: <https://doi.org/10.1186/s13065-024-01158-7>.
- [85] M. Liu *et al.*, "Quantitative Analysis of Amorphous Form in Indomethacin by Near Infrared Spectroscopy Combined with Partial Least Squares Regression Analysis," (in English), *Molecules*, vol. 29, no. 22, p. 5290, 2024, doi: <https://doi.org/10.3390/molecules29225290>.
- [86] A. Alqahtani, T. Alqahtani, A. Al Fatease, and E. H. Tolba, "Rapid UV-Vis spectrophotometric method aided by firefly-PLS models for the simultaneous quantification of ciprofloxacin, lomefloxacin, and enrofloxacin in their laboratory mixture, dosage forms and water samples: greenness and blueness assessment," (in English), *BMC chemistry*, vol. 18, no. 1, p. 172, 2024, doi: <https://doi.org/10.1186/s13065-024-01286-0>.
- [87] M. Kachalkin, T. Ryazanova, and I. Sokolova, "Quantitative determination of ademetionine in tablets utilizing ATR-FTIR and partial least squares methods approaches," (in English), *Journal of Pharmaceutical and Biomedical Analysis*, vol. 241, p. 115991, 2024, doi: <https://doi.org/10.1016/j.jpba.2024.115991>.
- [88] H. Godzo *et al.*, "PLS based quantitative determination of insulin aspart in solution using Raman spectroscopy," (in English), *Archives of Pharmacy*, vol. 74, no. Notebook 2, pp. 206-215, 2024, doi: <https://doi.org/10.5937/arhfarm74-49901>.
- [89] F. Nouri and N. Alassaf, "Partial least squares method for the multicomponent analysis of antibacterial mixture," (in English), *Methods Objects Chem Anal*, vol. 18, no. 2, pp. 92-100, 2023, doi: <https://doi.org/10.17721/moca.2023.92-100>.
- [90] M. A. Mansouri, M. Kharbach, M. El Maouardi, I. Barra, and A. Bouklouze, "Quantification of ciprofloxacin in pharmaceutical products from various brands using FT-NIR: A comparative investigation of PLS and MCR-ALS," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 303, p. 123268, 2023, doi: <https://doi.org/10.1016/j.saa.2023.123268>.
- [91] Y. Chiba, K. Okada, Y. Hayashi, S. Kumada, and Y. Onuki, "Usefulness of applying partial least squares regression to T2 relaxation curves for predicting the solid form content in binary physical mixtures," (in English), *Journal of Pharmaceutical Sciences*, vol. 112, no. 4, pp. 1041-1051, 2023, doi: <https://doi.org/10.1016/j.xphs.2022.11.028>.

- [92] A. Biancolillo, C. Scappaticci, M. Foschi, C. Rossini, and F. Marini, "Coupling of NIR spectroscopy and chemometrics for the quantification of dexamethasone in pharmaceutical formulations," (in English), *Pharmaceuticals*, vol. 16, no. 2, p. 309, 2023, doi: <https://doi.org/10.3390/ph16020309>.
- [93] R. C. Castro, D. S. Ribeiro, J. L. Santos, C. Nunes, S. Reis, and R. N. Páscoa, "Chemometric-assisted surface-enhanced Raman spectroscopy for metformin determination using gold nanoparticles as substrate," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 287, p. 122118, 2023, doi: <https://doi.org/10.1016/j.saa.2022.122118>.
- [94] B. Sabbagh, P. Kumar, Y. Chew, J. Chin, and G. Akowuah, "Determination of metformin in fixed-dose combination tablets by ATR-FTIR spectroscopy," (in English), *Chemical Data Collections*, vol. 39, p. 100868, 2022, doi: <https://doi.org/10.1016/j.cdc.2022.100868>.
- [95] L. Gozdziński, M. Ramsay, A. Larnder, B. Wallace, and D. K. Hore, "Fentanyl detection and quantification using portable Raman spectroscopy in community drug checking," (in English), *Journal of Raman Spectroscopy*, vol. 52, no. 7, pp. 1308-1316, 2021, doi: <https://doi.org/10.1002/jrs.6133>.
- [96] L. Makraduli, P. Makreski, K. Goracinova, S. Stefov, M. Anevska, and N. Geskovski, "A comparative approach to screen the capability of Raman and infrared (mid-and near-) spectroscopy for quantification of low-active pharmaceutical ingredient content solid dosage forms: the case of alprazolam," (in English), *Applied Spectroscopy*, vol. 74, no. 6, pp. 661-673, 2020, doi: <https://doi.org/10.1177/0003702820905367>.
- [97] P. F. A. Kambira *et al.*, "Combination UV-Vis spectroscopy and partial least square for detecting adulteration paracetamol and piroxicam in traditional medicines," (in English), *J Pharm Sci Community*, vol. 17, no. 1, pp. 41-50, 2020, doi: <http://dx.doi.org/10.24071/jpsc.002354>.
- [98] Y. Chen, H. Ma, Q. Zhang, S. Zhang, M. Chen, and Y. Wu, "Comparison of several variable selection methods for quantitative analysis and monitoring of the Yangxinshi tablet process using near-infrared spectroscopy," (in English), *Infrared Physics & Technology*, vol. 105, p. 103188, 2020, doi: <https://doi.org/10.1016/j.infrared.2020.103188>.
- [99] A. A. Makki *et al.*, "Qualitative and quantitative analysis of therapeutic solutions using Raman and infrared spectroscopy," (in English), *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 218, pp. 97-108, 2019, doi: <https://doi.org/10.1016/j.saa.2019.03.056>.
- [100] M. Edinger *et al.*, "Quantification of inkjet-printed pharmaceuticals on porous substrates using Raman spectroscopy and near-infrared spectroscopy," (in English), *AAPS PharmSciTech*, vol. 20, pp. 1-10, 2019, doi: <https://doi.org/10.1208/s12249-019-1423-y>.
- [101] M. Ostra, C. Ubide, M. Vidal, and J. Zuriarrain, "Detection limit estimator for multivariate calibration by an extension of the IUPAC recommendations for univariate methods," (in English), *Analyst*, vol. 133, no. 4, pp. 532-539, 2008, doi: <https://doi.org/10.1039/B716965P>.
- [102] R. Boqué and Y. Vander Heyden, "The limit of detection," (in English), 2009.
- [103] J. Miller and J. C. Miller, *Statistics and chemometrics for analytical chemistry*. Pearson education (in English), 2018.

- [104] R. Boqué, M. Larrechi, and F. Rius, "Multivariate detection limits with fixed probabilities of error," (in English), *Chemometrics and Intelligent Laboratory Systems*, vol. 45, no. 1-2, pp. 397-408, 1999, doi: [https://doi.org/10.1016/S0169-7439\(98\)00195-6](https://doi.org/10.1016/S0169-7439(98)00195-6).
- [105] A. Lorber, "Error propagation and figures of merit for quantification by solving matrix equations," (in English), *Analytical Chemistry*, vol. 58, no. 6, pp. 1167-1172, 1986, doi: <https://doi.org/10.1021/ac00297a042>.
- [106] A. Lorber, K. Faber, and B. R. Kowalski, "Net analyte signal calculation in multivariate calibration," (in English), *Analytical Chemistry*, vol. 69, no. 8, pp. 1620-1626, 1997, doi: <https://doi.org/10.1021/ac960862b>.
- [107] J. Ferré and N. K. M. Faber, "Net analyte signal calculation for multivariate calibration," (in English), *Chemometrics and Intelligent Laboratory Systems*, vol. 69, no. 1-2, pp. 123-136, 2003, doi: [https://doi.org/10.1016/S0169-7439\(03\)00118-7](https://doi.org/10.1016/S0169-7439(03)00118-7).
- [108] M. C. Sarraguça and J. A. Lopes, "The use of net analyte signal (NAS) in near infrared spectroscopy pharmaceutical applications: Interpretability and figures of merit," (in English), *Analytica chimica acta*, vol. 642, no. 1-2, pp. 179-185, 2009, doi: <https://doi.org/10.1016/j.aca.2008.10.006>.
- [109] F. Allegrini and A. C. Olivieri, "IUPAC-consistent approach to the limit of detection in partial least-squares calibration," (in English), *Analytical chemistry*, vol. 86, no. 15, pp. 7858-7866, 2014, doi: <https://doi.org/10.1021/ac501786u>.
- [110] N. K. M. Faber and R. Bro, "Standard error of prediction for multiway PLS: 1. Background and a simulation study," (in English), *Chemometrics and Intelligent Laboratory Systems*, vol. 61, no. 1-2, pp. 133-149, 2002, doi: [https://doi.org/10.1016/S0169-7439\(01\)00204-0](https://doi.org/10.1016/S0169-7439(01)00204-0).
- [111] M. Blanco, M. Castillo, A. Peinado, and R. Beneyto, "Determination of low analyte concentrations by near-infrared spectroscopy: effect of spectral pretreatments and estimation of multivariate detection limits," (in English), *Analytica chimica acta*, vol. 581, no. 2, pp. 318-323, 2007, doi: <https://doi.org/10.1016/j.aca.2006.08.018>.
- [112] Z. Wu *et al.*, "Multivariate detection limits of on-line NIR model for extraction process of chlorogenic acid from *Lonicera japonica*," (in English), *Journal of Pharmaceutical and Biomedical Analysis*, vol. 77, pp. 16-20, 2013, doi: <https://doi.org/10.1016/j.jpba.2012.12.026>.
- [113] M. Ortiz, L. Sarabia, and M. Sánchez, "Tutorial on evaluation of type I and type II errors in chemical analyses: From the analytical detection to authentication of products and process control," (in English), *Analytica Chimica Acta*, vol. 674, no. 2, pp. 123-142, 2010, doi: <https://doi.org/10.1016/j.aca.2010.06.026>.
- [114] M. Ortiz *et al.*, "Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares," (in English), *Chemometrics and intelligent laboratory systems*, vol. 69, no. 1-2, pp. 21-33, 2003, doi: [https://doi.org/10.1016/S0169-7439\(03\)00110-2](https://doi.org/10.1016/S0169-7439(03)00110-2).
- [115] K. Faber and B. R. Kowalski, "Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares," (in English), *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 11, no. 3, pp. 181-238, 1997, doi: [https://doi.org/10.1002/\(SICI\)1099-128X\(199705\)11:3%3C181::AID-CEM459%3E3.0.CO;2-7](https://doi.org/10.1002/(SICI)1099-128X(199705)11:3%3C181::AID-CEM459%3E3.0.CO;2-7).

- [116] J. A. F. Pierna, O. Abbas, V. Baeten, and P. Dardenne, "A Backward Variable Selection method for PLS regression (BVSPLS)," *Analytica chimica acta*, vol. 642, no. 1-2, pp. 89-93, 2009, doi: <https://doi.org/10.1016/j.aca.2008.12.002>.
- [117] Y.-H. Yun, H.-D. Li, B.-C. Deng, and D.-S. Cao, "An overview of variable selection methods in multivariate analysis of near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 113, pp. 102-115, 2019, doi: <https://doi.org/10.1016/j.trac.2019.01.018>.
- [118] S. Abe, "Modified backward feature selection by cross validation," in *ESANN*, 2005, pp. 163-168.
- [119] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157-1182, 2003.
- [120] T. Mehmood, H. Martens, S. Sæbø, J. Warringer, and L. Snipen, "A Partial Least Squares based algorithm for parsimonious variable selection," *Algorithms for Molecular Biology*, vol. 6, pp. 1-12, 2011.
- [121] T. Mehmood, K. H. Liland, L. Snipen, and S. Sæbø, "A review of variable selection methods in partial least squares regression," *Chemometrics and intelligent laboratory systems*, vol. 118, pp. 62-69, 2012, doi: <http://dx.doi.org/10.1016/j.chemolab.2012.07.010>.
- [122] M. Rogers, J. Blanc-Talon, M. Urschler, and P. Delmas, "Wavelength and texture feature selection for hyperspectral imaging: a systematic literature review," *Journal of Food Measurement and Characterization*, vol. 17, no. 6, pp. 6039-6064, 2023, doi: <https://doi.org/10.1007/s11694-023-02044-x>.
- [123] I. Ilhan and G. Tezel, "A genetic algorithm–support vector machine method with parameter optimization for selecting the tag SNPs," *Journal of biomedical informatics*, vol. 46, no. 2, pp. 328-340, 2013, doi: <https://doi.org/10.1016/j.jbi.2012.12.002>.
- [124] H. Chen, X. Liu, Z. Jia, Z. Liu, K. Shi, and K. Cai, "A combination strategy of random forest and back propagation network for variable selection in spectral calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 182, pp. 101-108, 2018, doi: <https://doi.org/10.1016/j.chemolab.2018.09.002>.
- [125] H.-D. Li, Q.-S. Xu, and Y.-Z. Liang, "Random frog: An efficient reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification," *Analytica Chimica Acta*, vol. 740, pp. 20-26, 2012, doi: <https://doi.org/10.1016/j.aca.2012.06.031>.
- [126] A. Santos-Rufo, F.-J. Mesas-Carrascosa, A. García-Ferrer, and J. E. Meroño-Larriva, "Wavelength selection method based on partial least square from hyperspectral unmanned aerial vehicle orthomosaic of irrigated olive orchards," *Remote Sensing*, vol. 12, no. 20, p. 3426, 2020, doi: <https://doi.org/10.3390/rs12203426>.
- [127] T. Wang, Y. Zheng, L. Xu, and Y.-H. Yun, "Comprehensive comparison on different wavelength selection methods using several near-infrared spectral datasets with different dimensionalities," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, p. 125767, 2025, doi: <https://doi.org/10.1016/j.saa.2025.125767>.
- [128] W. Lin, H. Hang, Y. Zhuang, and S. Zhang, "Variable selection in partial least squares with the weighted variable contribution to the first singular value of the covariance matrix," *Chemometrics and Intelligent Laboratory Systems*, vol. 183, pp. 113-121, 2018, doi: <https://doi.org/10.1016/j.chemolab.2018.11.003>.

- [129] H. Xu, Z. Liu, W. Cai, and X. Shao, "A wavelength selection method based on randomization test for near-infrared spectral analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 97, no. 2, pp. 189-193, 2009, doi: <https://doi.org/10.1016/j.chemolab.2009.04.006>.
- [130] U. Talukdar, S. M. Hazarika, and J. Q. Gan, "A kernel partial least square based feature selection method," *Pattern Recognition*, vol. 83, pp. 91-106, 2018, doi: <https://doi.org/10.1016/j.patcog.2018.05.012>.
- [131] D. Kepplinger, P. Filzmoser, and K. Varmuza, "Variable selection with genetic algorithms using repeated cross-validation of PLS regression models as fitness measure," 2017, doi: <https://doi.org/10.48550/arXiv.1711.06695>.
- [132] B. D. P. Vijay Kotu, "Feature Selection," in *Predictive Analytics and Data Mining*, 1st ed.: Elsevier, 2015, ch. 12, pp. 347-370.
- [133] S. Wang *et al.*, "An efficient method for tracing the geographic origin of Enshi Yulu fresh tea leaves based on near infrared spectroscopy combined with synergy interval PLS and genetic algorithm," *Lwt*, vol. 203, p. 116372, 2024, doi: <https://doi.org/10.1016/j.lwt.2024.116372>.
- [134] M. Salim, M. E. El Sharkasy, M. Walash, and F. Belal, "Genetic algorithm with model-updating-based PLS regression for the spectrophotometric determination of clopidogrel, atorvastatin, and aspirin in the presence of its degradation product," *Journal of Applied Spectroscopy*, vol. 87, pp. 568-578, 2020, doi: <https://doi.org/10.1007/s10812-020-01040-z>.
- [135] A. Elsonbaty, A. Serag, S. Abdulwahab, W. S. Hassan, and M. S. Eissa, "Analysis of quinary therapy targeting multiple cardiovascular diseases using UV spectrophotometry and chemometric tools," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 238, p. 118415, 2020, doi: <https://doi.org/10.1016/j.saa.2020.118415>.
- [136] H. W. Darwish, A. A. Al Majed, I. A. Al-Suwaidan, I. A. Darwish, A. H. Bakheit, and H. H. Al-Shehri, "Full spectrum and genetic algorithm-selected spectrum-based chemometric methods for simultaneous determination of azilsartan medoxomil, chlorthalidone, and azilsartan: development, validation, and application on commercial dosage form," *Open Chemistry*, vol. 19, no. 1, pp. 205-213, 2021, doi: <https://doi.org/10.1515/chem-2021-0022>.
- [137] F. Assadian and A. Niazi, "Genetic Algorithms-Based Approach for Wavelength Selection in Simultaneous Spectrofluorimetric Determination of Ofloxacin and Riboflavin in Biological Samples by Partial Least-Squares," *Journal of the Mexican Chemical Society*, vol. 63, no. 4, pp. 218-228, 2019, doi: <https://doi.org/10.29356/jmcs.v63i4.704>.
- [138] P. Kingwattanakul, T. Phatrapornnant, C. Theerawitaya, T. Kasetkasem, and M. Yamagushi, "Determination Lactone Composition in *Andrographis paniculata* (burm. f.) Wall, ex Nees using hyperspectral imaging," in *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, 2022: IEEE, pp. 1-4, doi: <https://doi.org/10.1109/ECTI-CON54298.2022.9795599>.
- [139] T. Merino, A. Criollo, G. La Rosa, J. Valdiviezo, and J. Soto, "Selection of spectral bands for estimating fishmeal quality parameters using ML models," in *2024 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI)*, 2024: IEEE, pp. 1-5, doi: <https://doi.org/10.1109/ColCACI63187.2024.10666581>.

- [140] A. M. Ibrahim, H. A. Hendawy, W. S. Hassan, A. Shalaby, and M. S. ElMasry, "Determination of terazosin in the presence of prazosin: Different state-of-the-art machine learning algorithms with UV spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 236, p. 118349, 2020, doi: <https://doi.org/10.1016/j.saa.2020.118349>.
- [141] R. Joshi *et al.*, "Evaluation of benzene residue in edible oils using Fourier transform infrared (FTIR) spectroscopy," *Korean Journal of Agricultural Science*, vol. 46, no. 2, pp. 257-271, 2019, doi: <https://doi.org/10.7744/kjoas.20190012>.
- [142] B. Krakowska, D. Custers, E. Deconinck, and M. Daszykowski, "The Monte Carlo validation framework for the discriminant partial least squares model extended with variable selection methods applied to authenticity studies of Viagra® based on chromatographic impurity profiles," *Analyst*, vol. 141, no. 3, pp. 1060-1070, 2016, doi: <https://doi.org/10.1039/C5AN01656H>.
- [143] Bộ Y tế, *Hóa dược tập 1*. Hà Nội: Nhà xuất bản Y học (in Vietnam), 2007.
- [144] Bộ Y tế, *Dược điển Việt Nam V*. Hà Nội: Nhà xuất bản Y học (in Vietnam), 2017.
- [145] Bộ Y tế, *Dược thư quốc gia Việt Nam*. Hà Nội: Nhà xuất bản Y học (in Vietnam), 2018.
- [146] R. N. Shah, D. B. Gandhi, and M. M. Patel, "Simultaneous determination of amlodipine besylate and indapamide in tablet dosage form by absorption correction method and first-order derivative UV spectrophotometry," (in English), *Int J PharmTech Res*, vol. 4, no. 3, pp. 1018-24, 2012.
- [147] A. K. Mohanty, R. Saraswat, and S. Puranik, "RP-HPLC Method for the Simultaneous Estimation of Calcium Channel Blocker with Diuretic and Antihypertensive in Bulk Drugs and Marketed Formulation," (in English), *Journal of Current Pharma Research*, vol. 16, no. 3, pp. 29-39, 2022.
- [148] A. Desai, R. Chauhan, S. Shah, and D. Shah, "HPTLC method for the simultaneous estimation of Amlodipine Besylate and Indapamide in tablet formulation," (in English), *Asian Journal of Research in Chemistry*, vol. 5, no. 4, pp. 510-514, 2012.
- [149] "Atorvastatin calcium." Drug bank. <https://go.drugbank.com/salts/DBSALT000011> (accessed).
- [150] N. Đ. Thiện, N. T. M. Linh, and Đ. D. Lê Thị Phương Anh, "Hai phương pháp định lượng Amlodipine besilate và Atorvastatin calcium trong chế phẩm bằng phép đo quang phổ bước sóng kép," (in Vietnam), *Bản B của Tạp chí Khoa học và Công nghệ Việt Nam*, vol. 64, no. 2, 2022.
- [151] H. Hafez, A. Elshanawany, L. Abdelaziz, and M. Mohram, "Development of a stability-indicating HPLC method for simultaneous determination of amlodipine besylate and atorvastatin calcium in tablets," (in English), *Austin Journal of Analytical and Pharmaceutical Chemistry*, vol. 1, no. 6, p. 1028, 2014.
- [152] A. Mohammadi, N. Rezanour, M. A. Dogahneh, F. G. Bidkorbeh, M. Hashem, and R. Walker, "A stability-indicating high performance liquid chromatographic (HPLC) assay for the simultaneous determination of atorvastatin and amlodipine in commercial tablets," (in English), *Journal of chromatography B*, vol. 846, no. 1-2, pp. 215-221, 2007, doi: <https://doi.org/10.1016/j.jchromb.2006.09.007>.
- [153] A. H. Sarrafi, E. Konozy, and M. Ghiyasvand, "Simultaneous determination of atorvastatin calcium and amlodipine besylate by spectrophotometry and multivariate calibration methods in pharmaceutical formulations," (in English),

- Journal of Chemistry*, vol. 8, no. 4, pp. 1670-1679, 2011, doi: <https://doi.org/10.1155/2011/292346>.
- [154] A. K. Muthu, T. R. Gupta, S. Sharma, A. A. Smith, R. Manavalan, and N. Kannappan, "Simultaneous estimation of amlodipine and atorvastatin in tablets using orthogonal function ratio spectrometry," (in English), *Int J Chem Sci*, vol. 6, no. 4, pp. 2233-2241, 2008.
- [155] H. Danafar and M. Hamidi, "Method validation of amlodipine and atorvastatin by liquid chromatography–mass spectrometry (LC–MS) method in human plasma," (in English), *Cogent Medicine*, vol. 3, no. 1, p. 1129790, 2016, doi: <https://doi.org/10.1080/2331205X.2015.1129790>.
- [156] N. C. f. B. Information. "PubChem Compound Summary for CID 5833, Spironolactone." (accessed).
- [157] N. C. f. B. Information. "PubChem Compound Summary for CID 3440, Furosemide." <https://pubchem.ncbi.nlm.nih.gov/compound/furosemide> (accessed).
- [158] R. R. Chavan, S. D. Bhinge, M. A. Bhutkar, and D. S. Randive, "Development and validation of spectrophotometric methods for simultaneous estimation of furosemide and spironolactone by Vierordt's method in bulk and combined tablet dosage form," (in English), 2018.
- [159] V. R. Ram, P. N. Dave, and H. S. Joshi, "Development and validation of a stability-indicating HPLC assay method for simultaneous determination of spironolactone and furosemide in tablet formulation," (in English), *Journal of chromatographic science*, vol. 50, no. 8, pp. 721-726, 2012, doi: <https://doi.org/10.1093/chromsci/bms062>.
- [160] H. M. Lotfy, S. El-Hanboushy, Y. M. Fayz, and M. Abdelkawy, "Smart spectrophotometric methods for concurrent determination of furosemide and spironolactone mixture in their pharmaceutical dosage forms," (in English), *Brazilian Journal of Pharmaceutical Sciences*, vol. 58, p. e19487, 2022, doi: <https://doi.org/10.1590/s2175-97902022e19487>.
- [161] S. Israt, M. Uddin, R. Jahan, and M. Karim, "Simultaneous determination of furosemide and spironolactone in pharmaceutical formulations by spectrophotometric method using principal component regression," (in English), *Bangladesh Journal of Scientific and Industrial Research*, vol. 51, no. 4, pp. 297-306, 2016, doi: <https://doi.org/10.3329/bjsir.v51i4.30450>.
- [162] N. C. f. B. Information. "PubChem Compound Summary for CID 4173, Metronidazole." <https://pubchem.ncbi.nlm.nih.gov/compound/Metronidazole> (accessed).
- [163] N. C. f. B. Information. "PubChem Compound Summary for CID 5289394, Spiramycin." <https://pubchem.ncbi.nlm.nih.gov/compound/Spiramycin> (accessed).
- [164] F. I. Khattab, N. K. Ramadan, M. A. Hegazy, and N. S. Ghoniem, "Simultaneous determination of metronidazole and spiramycin in bulk powder and in tablets using different spectrophotometric techniques," (in English), *Drug testing and analysis*, vol. 2, no. 1, pp. 37-44, 2010, doi: <https://doi.org/10.1002/dta.83>.
- [165] H. Ebrahimi-Najafabadi, R. Leardi, and M. Jalali-Heravi, "Experimental design in analytical chemistry—Part II: Applications," (in English), *Journal of AOAC International*, vol. 97, no. 1, pp. 12-18, 2014, doi: <https://doi.org/10.5740/jaoacint.SGEEbrahimi2>.

- [166] D. C. Montgomery, "Response Surface Methods and Designs," in *Design and analysis of experiments*: John Wiley & Sons, 2017, ch. 11, p. 511.
- [167] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, "DESIGN OF EXPERIMENTS FOR FITTING RESPONSE SURFACES," in *Response surface methodology: process and product optimization using designed experiments*, Fourth ed.: John Wiley & Sons, 2016, ch. 8, p. 369.
- [168] H. Ebrahimi-Najafabadi, R. Leardi, and M. Jalali-Heravi, "Experimental design in analytical chemistry—part I: theory," (in English), *Journal of AOAC International*, vol. 97, no. 1, pp. 3-11, 2014, doi: <https://doi.org/10.5740/jaoacint.SGEEbrahimi1>.
- [169] A. C. Olivieri, *Introduction to multivariate calibration: A practical approach*, 1st ed. Springer Nature (in English), 2018.
- [170] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," (in English), *Pattern recognition*, vol. 48, no. 9, pp. 2839-2846, 2015, doi: <https://doi.org/10.1016/j.patcog.2015.03.009>.
- [171] X. Zhang and C.-A. Liu, "Model averaging prediction by K-fold cross-validation," (in English), *Journal of Econometrics*, vol. 235, no. 1, pp. 280-301, 2023, doi: <https://doi.org/10.1016/j.jeconom.2022.04.007>.
- [172] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," (in English), *Computational Statistics*, vol. 36, no. 3, pp. 2009-2031, 2021, doi: <https://doi.org/10.1007/s00180-020-00999-9>.
- [173] R. Sjögren, "Synergies between Chemometrics and Machine Learning," Umeå Universitet, 2021.
- [174] M. C. Ferraro, P. M. Castellano, and T. S. Kaufman, "Chemometric determination of amiloride hydrochloride, atenolol, hydrochlorothiazide and timolol maleate in synthetic mixtures and pharmaceutical formulations," (in English), *Journal of Pharmaceutical and Biomedical Analysis*, vol. 34, no. 2, pp. 305-314, 2004, doi: [https://doi.org/10.1016/S0731-7085\(03\)00521-1](https://doi.org/10.1016/S0731-7085(03)00521-1).
- [175] A. C. Olivieri, S. Bortolato, and F. Allegrini, "Figures of merit in multiway calibration," in *Data Handling in Science and Technology*, vol. 29: Elsevier, 2015, pp. 541-575.
- [176] O. M. Kvalheim, "Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots," *Journal of Chemometrics*, vol. 24, no. 7-8, pp. 496-504, 2010, doi: <https://doi.org/10.1002/cem.1289>.
- [177] X. Huang, Y.-P. Luo, and L. Xia, "An efficient wavelength selection method based on the maximal information coefficient for multivariate spectral calibration," *Chemometrics and Intelligent Laboratory Systems*, vol. 194, p. 103872, 2019, doi: <https://doi.org/10.1016/j.chemolab.2019.103872>.
- [178] T. Rajalahti, R. Arneberg, A. C. Kroksveen, M. Berle, K.-M. Myhr, and O. M. Kvalheim, "Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles," *Analytical chemistry*, vol. 81, no. 7, pp. 2581-2590, 2009, doi: <https://doi.org/10.1021/ac802514y>.
- [179] G. M. Merola, "Least squares sparse principal component analysis: A backward elimination approach to attain large loadings," *Australian & New Zealand Journal of Statistics*, vol. 57, no. 3, pp. 391-429, 2015, doi: <https://doi.org/10.1111/anzs.12128>.

- [180] I. E. Frank, "Intermediate Least Squares Regression Method," *Chemometrics and Intelligent Laboratory Systems*, vol. 1, pp. 233-242, 1987.
- [181] S. Saha, A. S. Sairam, A. Yadav, and A. Ekbal, "Genetic algorithm combined with support vector machine for building an intrusion detection system," in *Proceedings of the International Conference on Advances in Computing, Communications and Informatics*, 2012, pp. 566-572, doi: <https://doi.org/10.1145/2345396.2345489>.
- [182] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decision Analytics Journal*, vol. 3, p. 100071, 2022, doi: <https://doi.org/10.1016/j.dajour.2022.100071>.
- [183] A. Niazi and R. Leardi, "Genetic algorithms in chemometrics," *Journal of Chemometrics*, vol. 26, no. 6, pp. 345-351, 2012, doi: <https://doi.org/10.1002/cem.2426>.
- [184] A. Lambora, K. Gupta, and K. Chopra, "Genetic algorithm-A literature review," in *2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon)*, 2019: IEEE, pp. 380-384, doi: <https://doi.org/10.1109/COMITCon.2019.8862255>.
- [185] P. X. Đ. Trần Cao Sơn, Lê Thị Hồng Hào, Nguyễn Thành Trung,, *Thẩm định phương pháp trong phân tích hóa học và vi sinh vật*. Nhà xuất bản Khoa học và Kỹ thuật (in Vietnam), 2010.
- [186] A. International, "Appendix F: Guidelines for Standard Method Performance Requirements," ed, 2017.
- [187] S. Audino *et al.*, "AOAC SMPR® 2017.001: Standard Method Performance Requirements (SMPRs) for Quantitation of Cannabinoids in Cannabis concentrates," (in English), *Journal of AOAC International*, vol. 100, no. 4, pp. 1200-1203, 2017, doi: https://doi.org/10.5740/jaoacint.SMPR2017_001.
- [188] J. C. M. James N. Miller, Robert D. Miller, *Statistics and Chemometrics for Analytical Chemistry*, Seventh ed. Pearson Education Limited (in English), 2018.
- [189] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," (in English), *Biometrika*, vol. 52, no. 3-4, pp. 591-611, 1965, doi: <https://doi.org/10.1093/biomet/52.3-4.591>.
- [190] N. M. Razali and Y. B. Wah, "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests," (in English), *Journal of statistical modeling and analytics*, vol. 2, no. 1, pp. 21-33, 2011.

PHỤ LỤC

Phụ lục 1. Chương trình tính toán $LOD_{min-max}$ trong R

```
model.LOD = function(A.sample, C.sample, ncomp){  
  X<-A.sample  
  X=t(t(X))  
  I<- dim(X)[2]  
  J<- dim(X)[1]  
  Y<-C.sample  
  Y = Y[1,]  
  Y= as.matrix(t(Y))  
  # Centers  
  meanX <- rowMeans(X, dims = 1)  
  Xm <- X - meanX %*% matrix(data = 1, nrow = 1, ncol = I)  
  meanY = mean(Y)  
  Ym <- Y - meanY  
  Xpls <- Xm  
  Ypls <- Ym  
  A = ncomp #chọn theo mô hình PLS đã xây dựng  
  Wn <- matrix(data = NA, nrow = J, ncol = A)  
  Ts <- matrix(data = NA, nrow = I, ncol = A)  
  P <- matrix(data = NA, nrow = J, ncol = A)  
  v <- matrix(data = NA, nrow = 1, ncol = A)  
  for (i in 1:A) {  
    W = (Xpls) %*% Ypls %*% solve(t(Ypls) %*% Ypls)  
    Wn[,i] = scale(W, center = TRUE, scale = T)  
    Ts[,i] = t(Xpls) %*% Wn[,i]  
    v[i] <- t(Ts[,i]) %*% Ypls %*% solve(t(Ts[,i]) %*% Ts[,i])  
    P[,i] <- (Xpls) %*% Ts[,i]/as.numeric((t(Ts[,i]) %*% Ts[,i]))  
    Xpls <- Xpls - P[,i]%*% t(Ts[,i])  
    Ypls <- Ypls- Ts[,i]*v[i]
```

```

}
lege <- matrix(data = NA, nrow = 1, ncol = A)
for (i in 1:A) {
  b = as.character(i)
  lege[i] <- b
}
# matplot(P,type='l',main="Loadings")
# legend(x = "topright",legend = lege,lty = c(seq(1,A)))
b1 <- Wn %*% solve(t(P) %*% Wn) %*% t(v)
plot(b1,type='l',main="Regression coefficients b1")
SEN <- 1/sqrt(sum(b1^2))
Xcalmod <- P %*% solve(t(Wn) %*% P) %*% (t(Wn) %*% (Xm))
Xcalres <- P %*% solve(t(Wn) %*% P) %*% (t(Wn) %*% (Xm)) - (Xm)
Sx <- sqrt(sum(Xcalres^2))/sqrt(I*J)
Ycalres <- Ts %*% t(v)-(Ym)
Sy <- sqrt(sum(Ycalres^2))/sqrt(I-1)
huni <- (mean(Y)^2)/sum(Ym^2)
h0min <- huni+1/I
LODmin <- 3.3*(Sx^2/(SEN^2)+ h0min*Sx^2/(SEN^2) + h0min*Sy^2)^0.5
h0cal=matrix(data = NA, nrow = I, ncol = 1)
for (i in 1:I) {
  mYmY<- 1-(Ym[i])^2/mean(Y)^2
  TtT<- Ts%*%solve(t(Ts)%*%Ts)%*% Ts[i,]
  hcal<-sum(TtT^2)
  h0cal[i] <- hcal+huni*mYmY
}
h0max=max(h0cal)+1/I
LODmax <- 3.3*(Sx^2/(SEN^2)+ h0max*Sx^2/(SEN^2) + h0max*Sy^2)^0.5
return(c(SEN, LODmin, LODmax))
}

```

Phụ lục 2.1. Hệ số a_i trong kiểm định Shapiro-Wilk

[illegible]

[illegible]

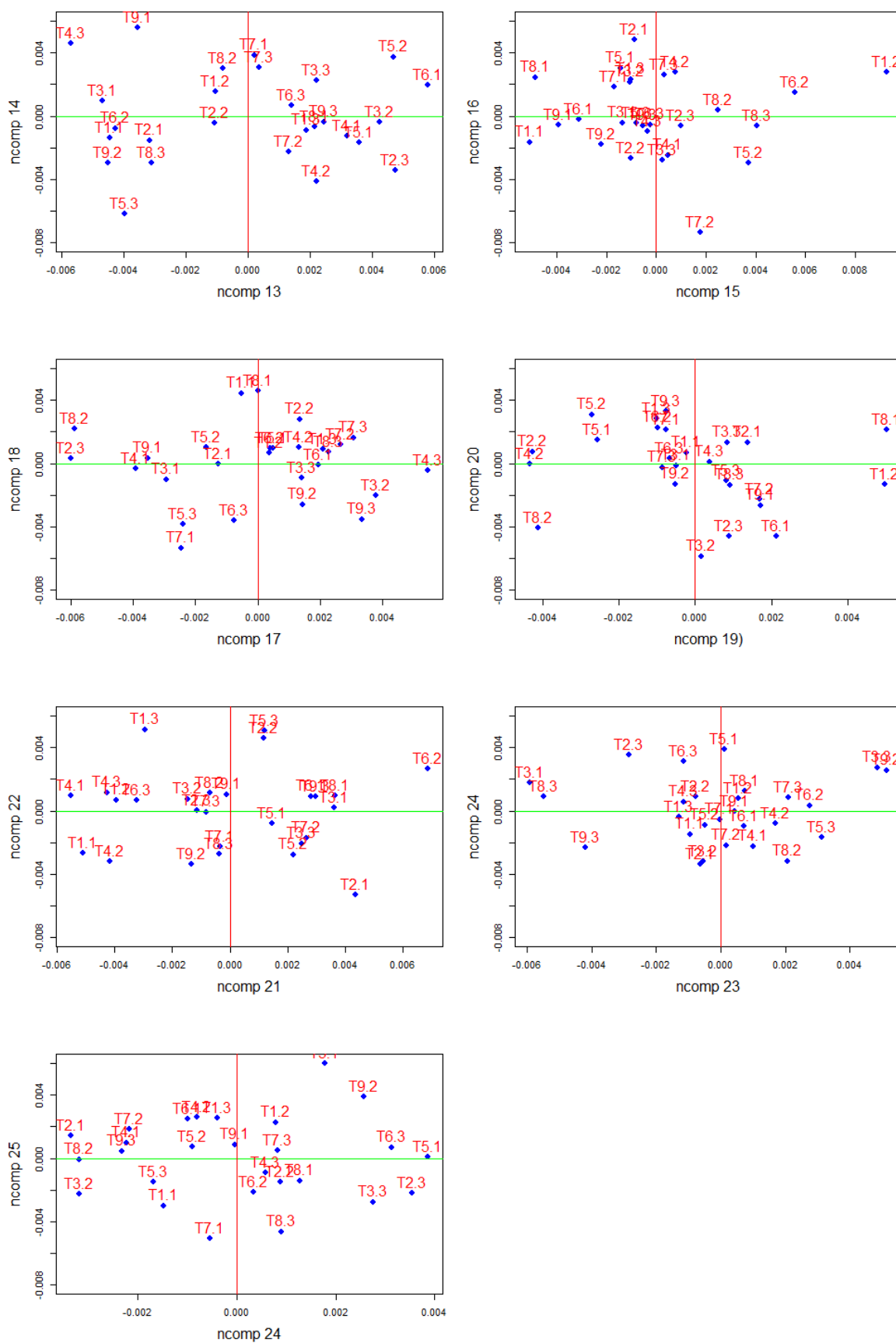
[illegible]

Phụ lục 2.2 Giá trị phân phối W trong kiểm định Shapiro-Wilk

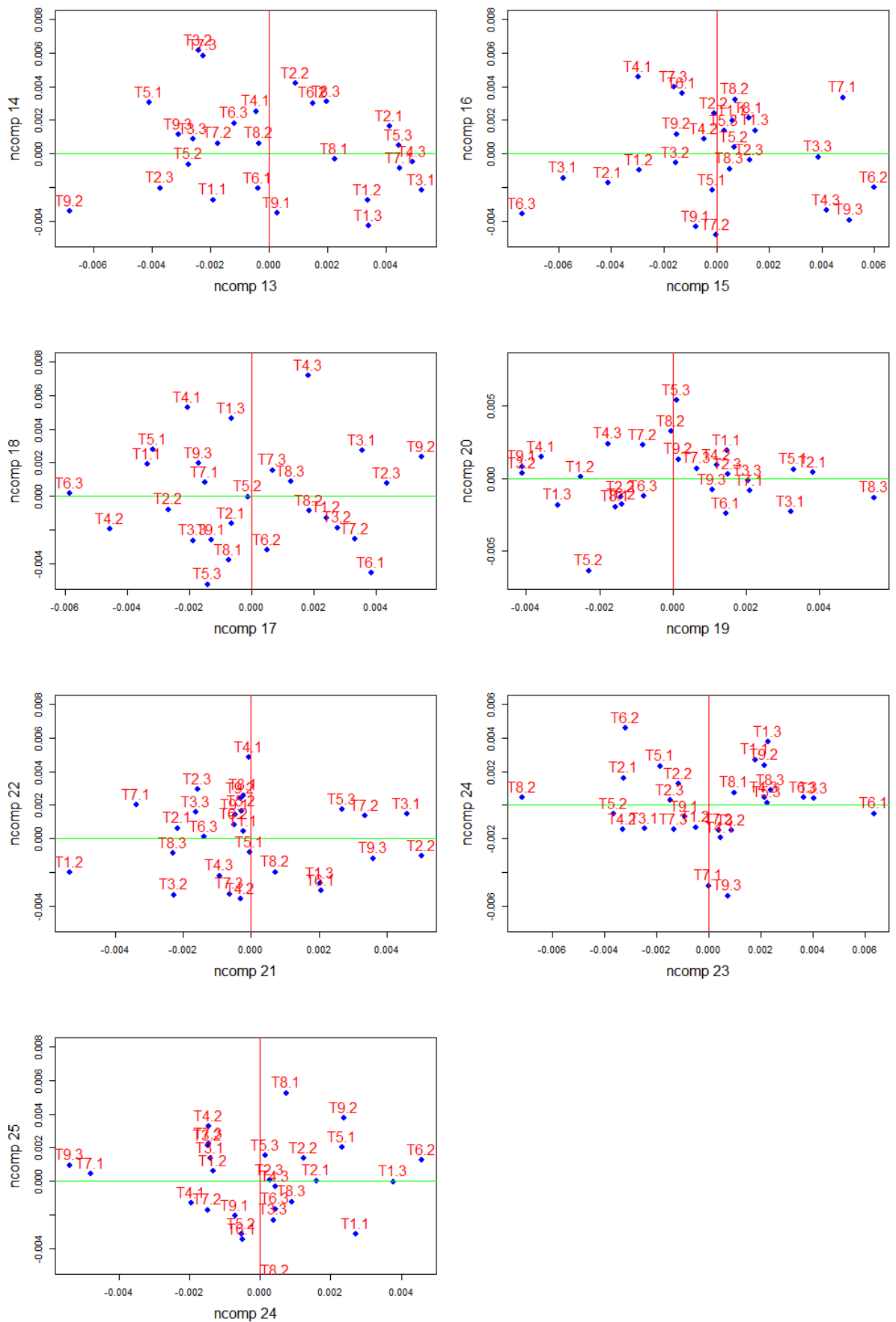
$n \backslash P$	0.01	0.02	0.05	0.1	0.5	0.9	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989
24	0.884	0.898	0.916	0.930	0.963	0.981	0.984	0.987	0.989
25	0.888	0.901	0.918	0.931	0.964	0.981	0.985	0.988	0.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989

27	0.894	0.906	0.923	0.935	0.965	0.982	0.985	0.988	0.990
28	0.896	0.908	0.924	0.936	0.966	0.982	0.985	0.988	0.990
29	0.898	0.910	0.926	0.937	0.966	0.982	0.985	0.988	0.990
30	0.900	0.912	0.927	0.939	0.967	0.983	0.985	0.988	0.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	0.904	0.915	0.930	0.941	0.968	0.983	0.986	0.988	0.990
33	0.906	0.917	0.931	0.942	0.968	0.983	0.986	0.989	0.990
34	0.908	0.919	0.933	0.943	0.969	0.983	0.986	0.989	0.990
35	0.910	0.920	0.934	0.944	0.969	0.984	0.986	0.989	0.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	0.914	0.924	0.936	0.946	0.970	0.984	0.987	0.989	0.990
38	0.916	0.925	0.938	0.947	0.971	0.984	0.987	0.989	0.990
39	0.917	0.927	0.939	0.948	0.971	0.984	0.987	0.989	0.991
40	0.919	0.928	0.940	0.949	0.972	0.985	0.987	0.989	0.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	0.922	0.930	0.942	0.951	0.972	0.985	0.987	0.989	0.991
43	0.923	0.932	0.943	0.951	0.973	0.985	0.987	0.990	0.991
44	0.924	0.933	0.944	0.952	0.973	0.985	0.987	0.990	0.991
45	0.926	0.934	0.945	0.953	0.973	0.985	0.988	0.990	0.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	0.928	0.936	0.946	0.954	0.974	0.985	0.988	0.990	0.991
48	0.929	0.937	0.947	0.954	0.974	0.985	0.988	0.990	0.991
49	0.929	0.939	0.947	0.955	0.974	0.985	0.988	0.990	0.991
50	0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991

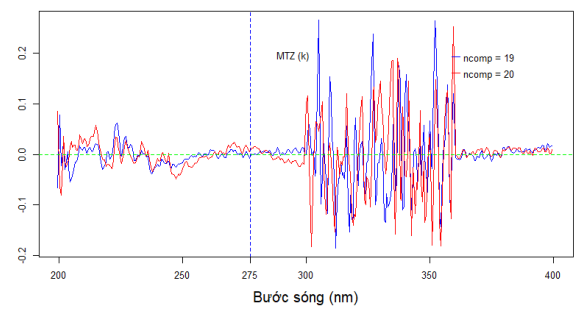
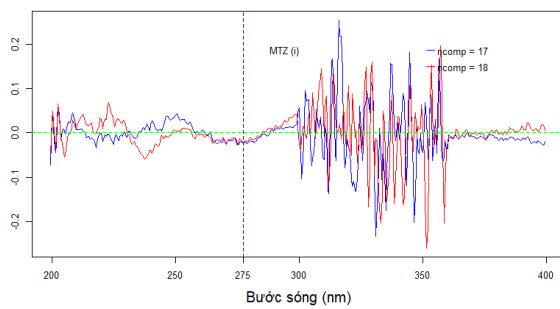
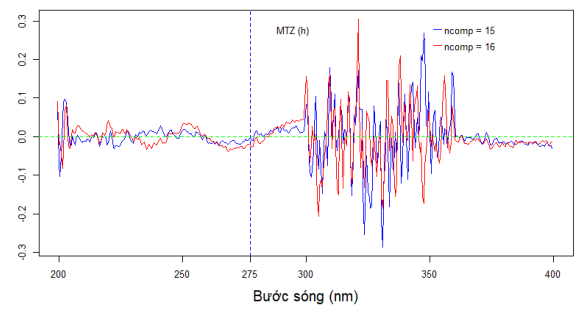
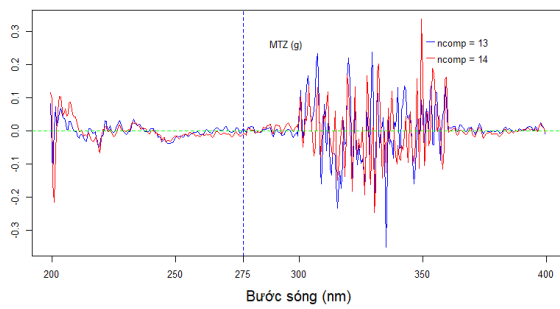
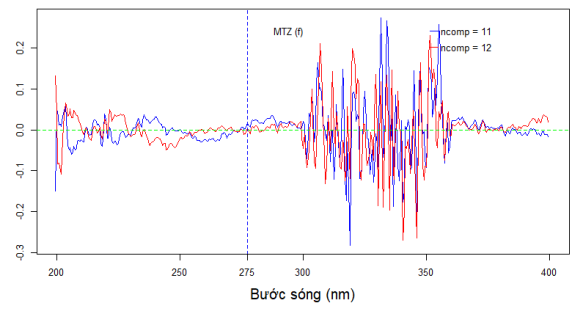
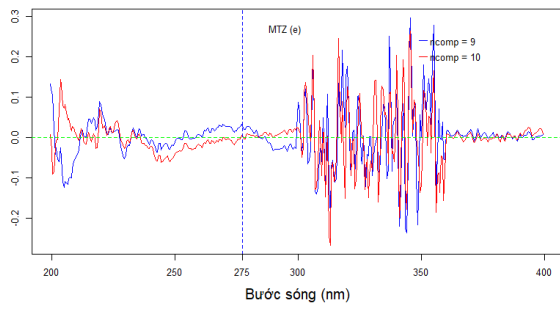
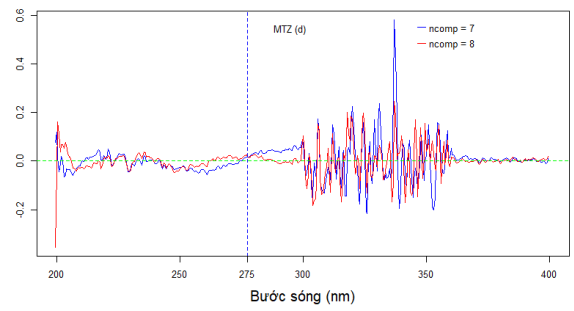
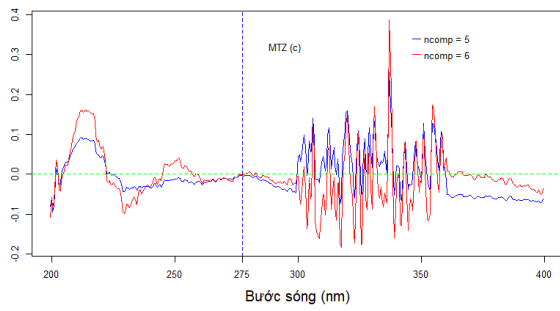
Phụ lục 3.1. Biểu đồ điểm của 13 biến tiềm ẩn của MTZ

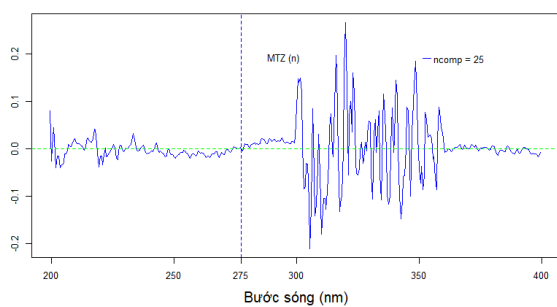
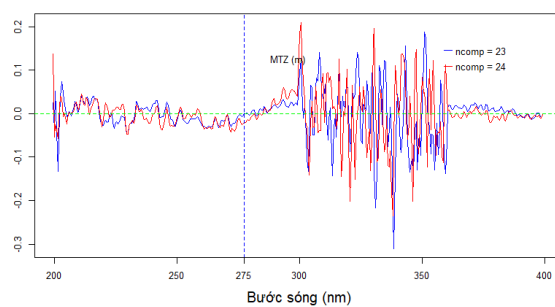
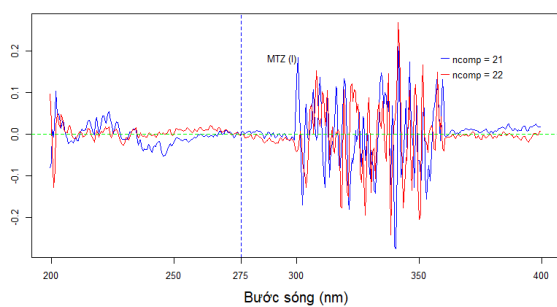


Phụ lục 3.2. Biểu đồ điểm của 13 biến tiềm ẩn của SPI

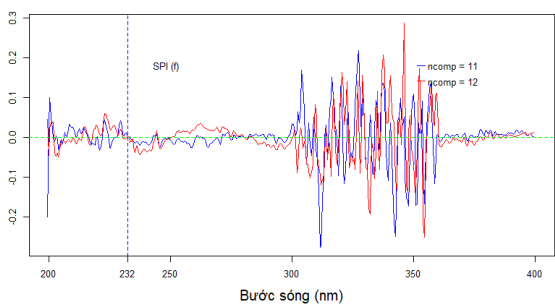
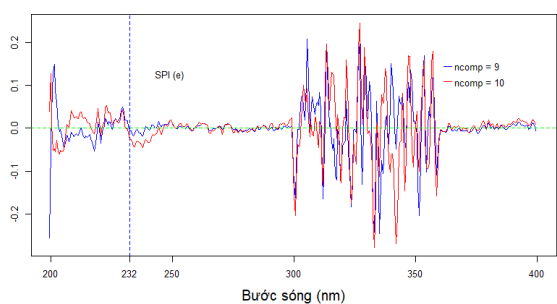
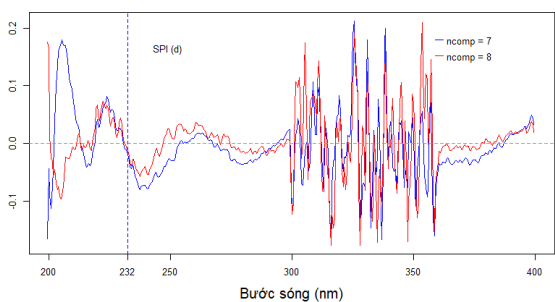
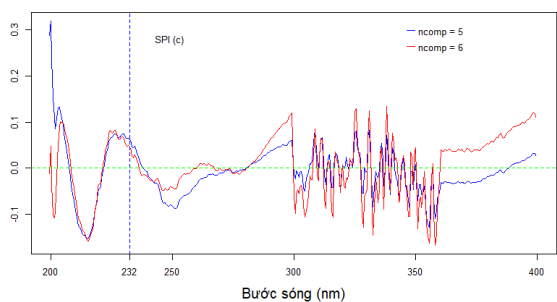


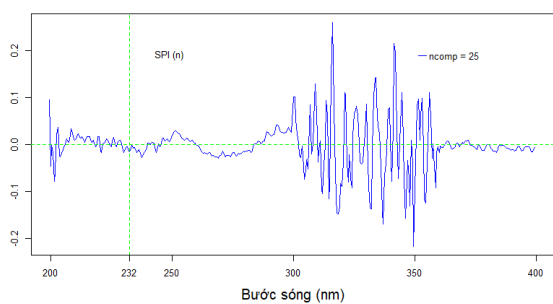
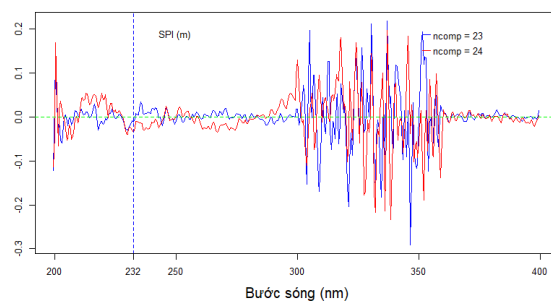
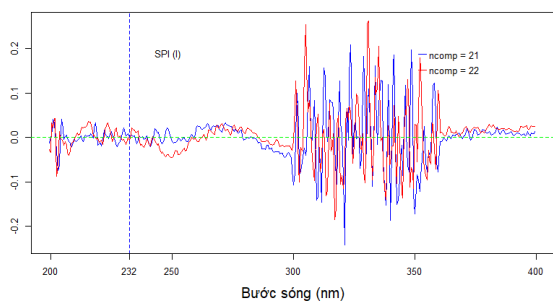
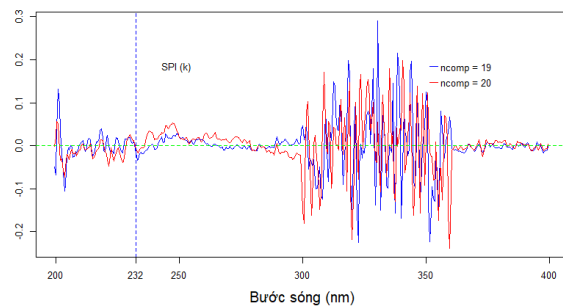
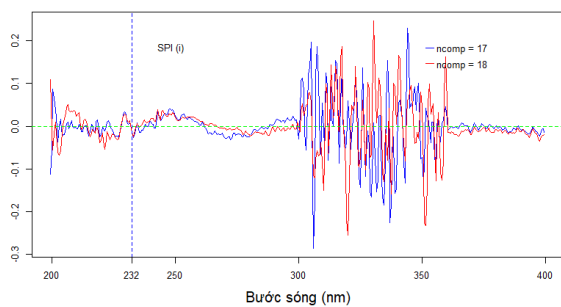
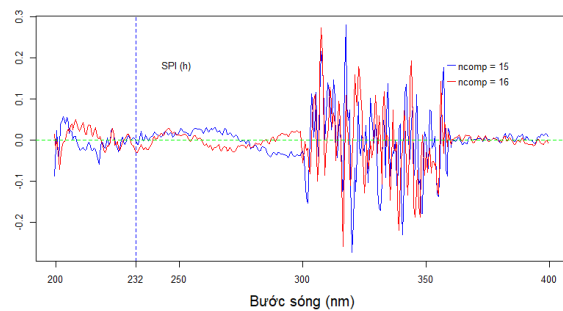
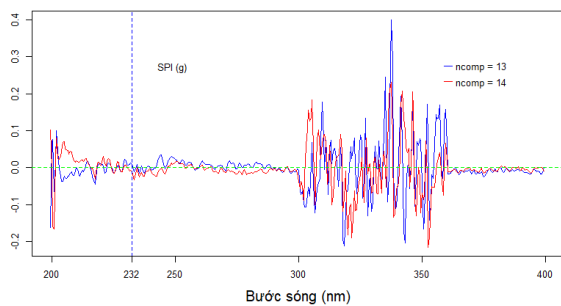
Phụ lục 4.1. Phổ của các vector chiếu dữ liệu quang phổ gốc lên không gian tiềm ẩn PLS của MTZ



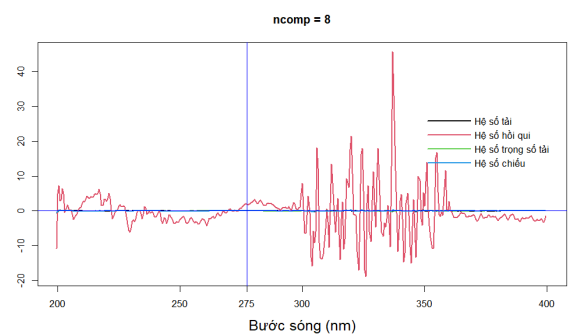
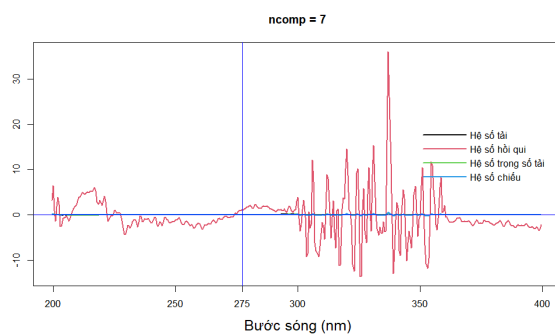
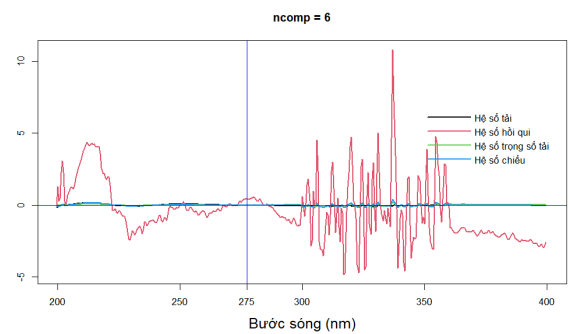
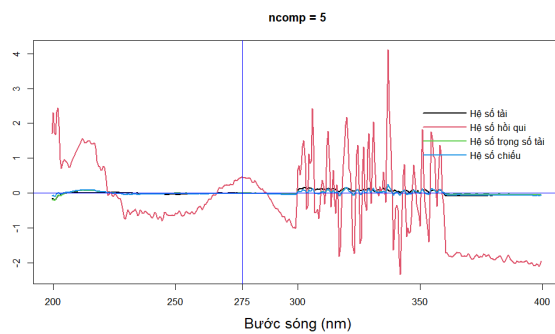
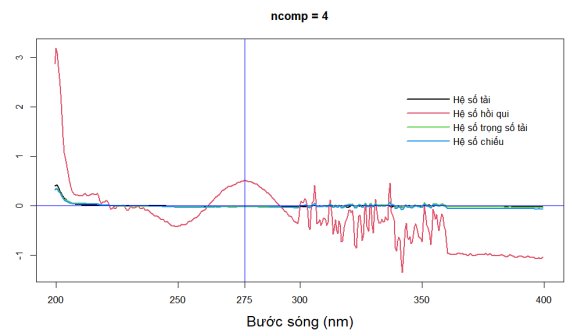
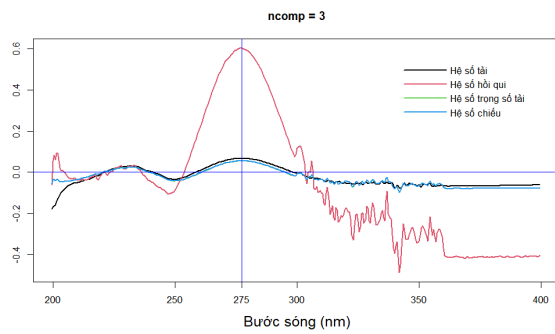
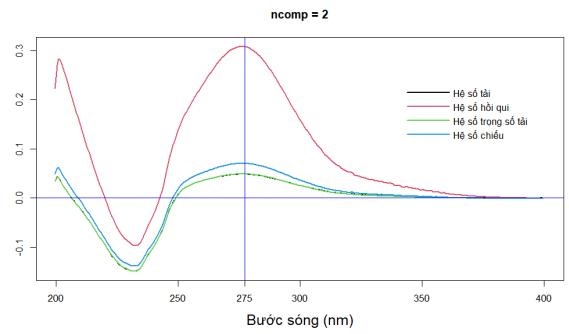
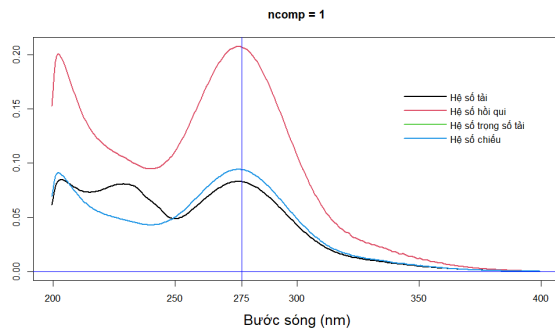


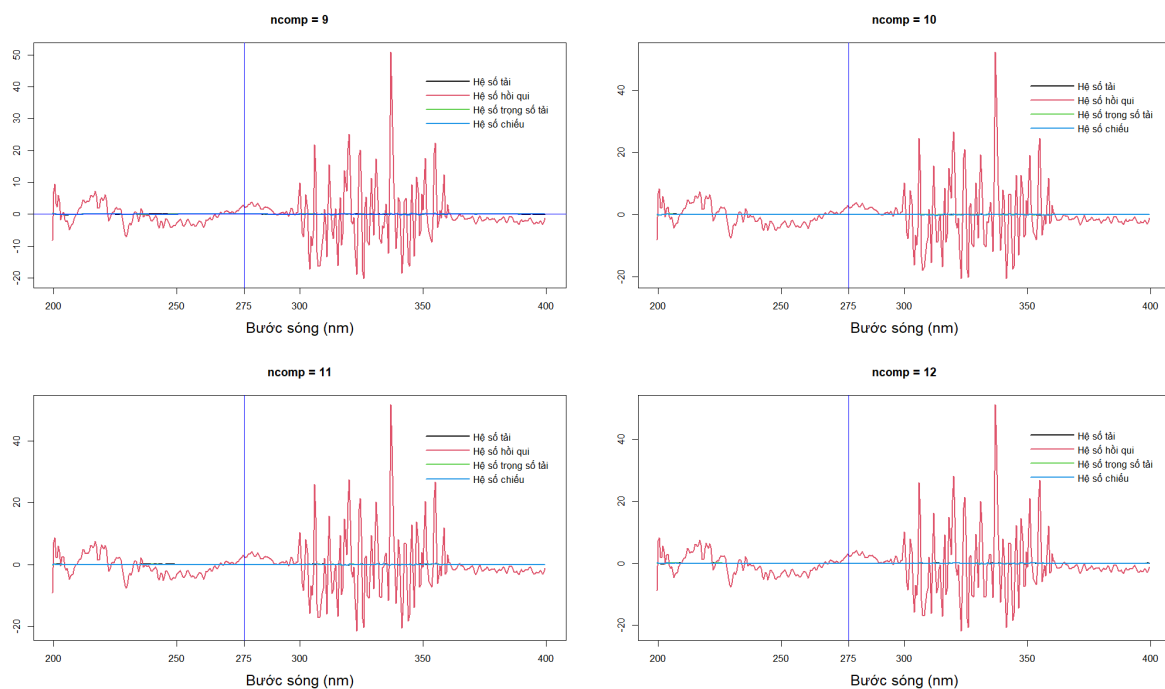
Phụ lục 4.2. Phổ của các vector chiếu dữ liệu quang phổ gốc lên không gian tiềm ẩn PLS của SPI



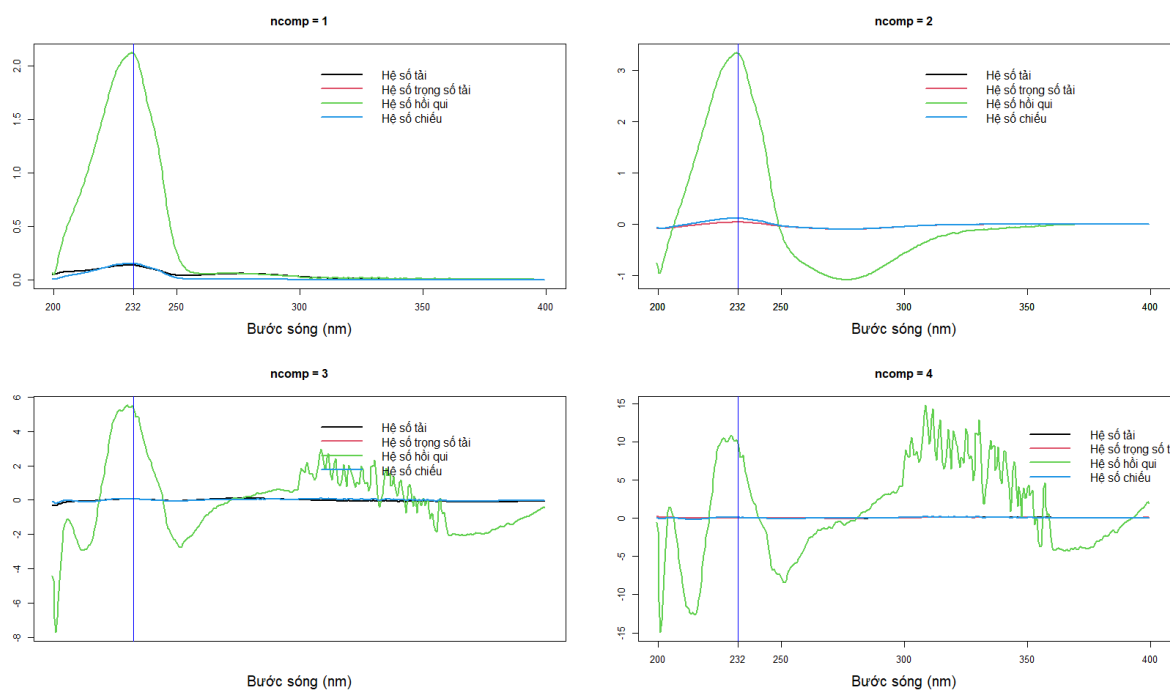


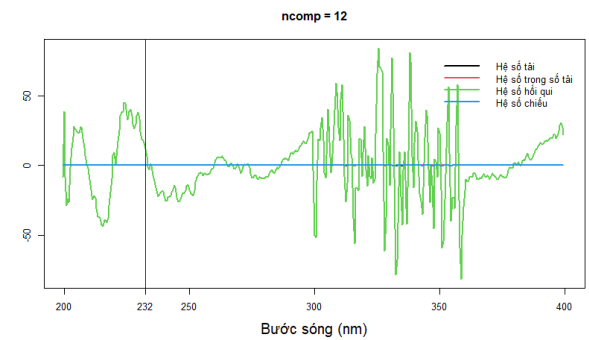
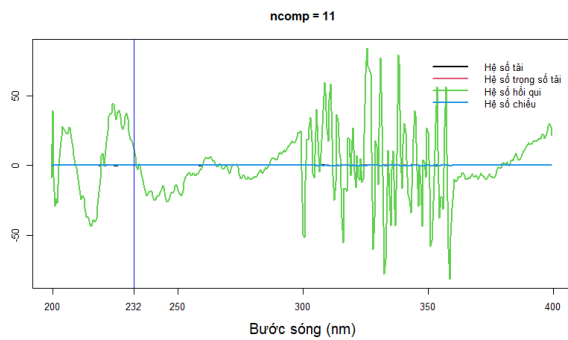
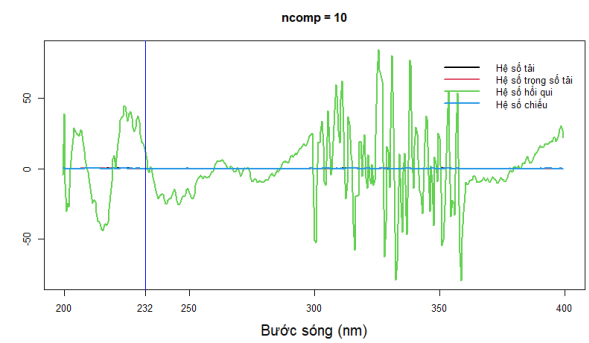
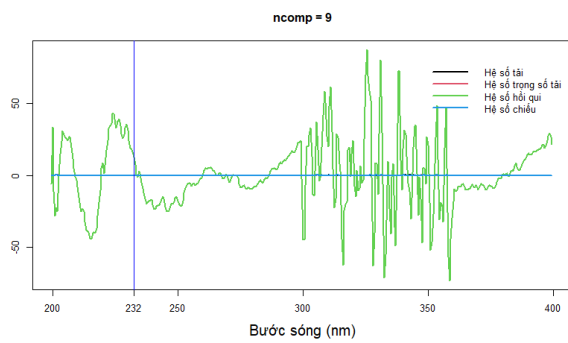
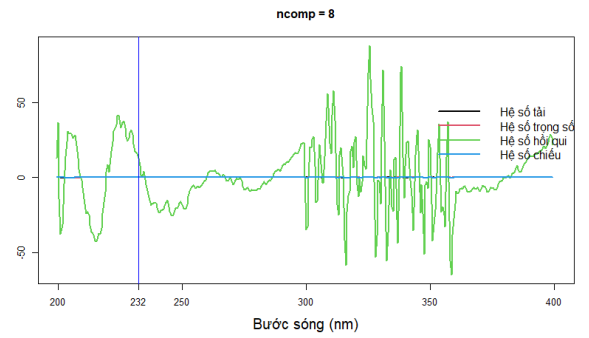
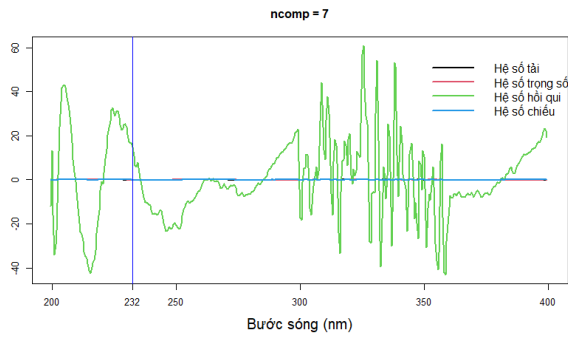
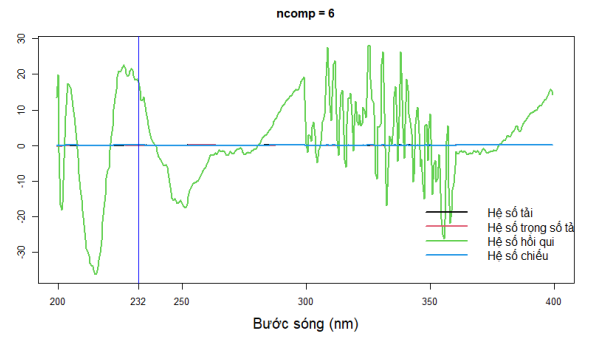
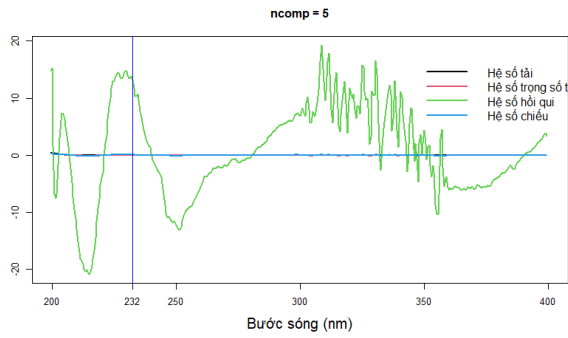
Phụ lục 5.1 Hệ số tải, hệ số hồi qui, hệ số trọng số tải, hệ số chiếu của 12 biến tiềm ẩn đầu tiên của mô hình PLS cho MTZ





Phụ lục 5.2 Hệ số tải, hệ số hồi qui, hệ số trọng số tải, hệ số chiếu của 12 biến tiềm ẩn đầu tiên của mô hình PLS cho SPI





Phụ lục 6. Chương trình lựa chọn số biến tiềm ẩn theo Haaland và Thomas.

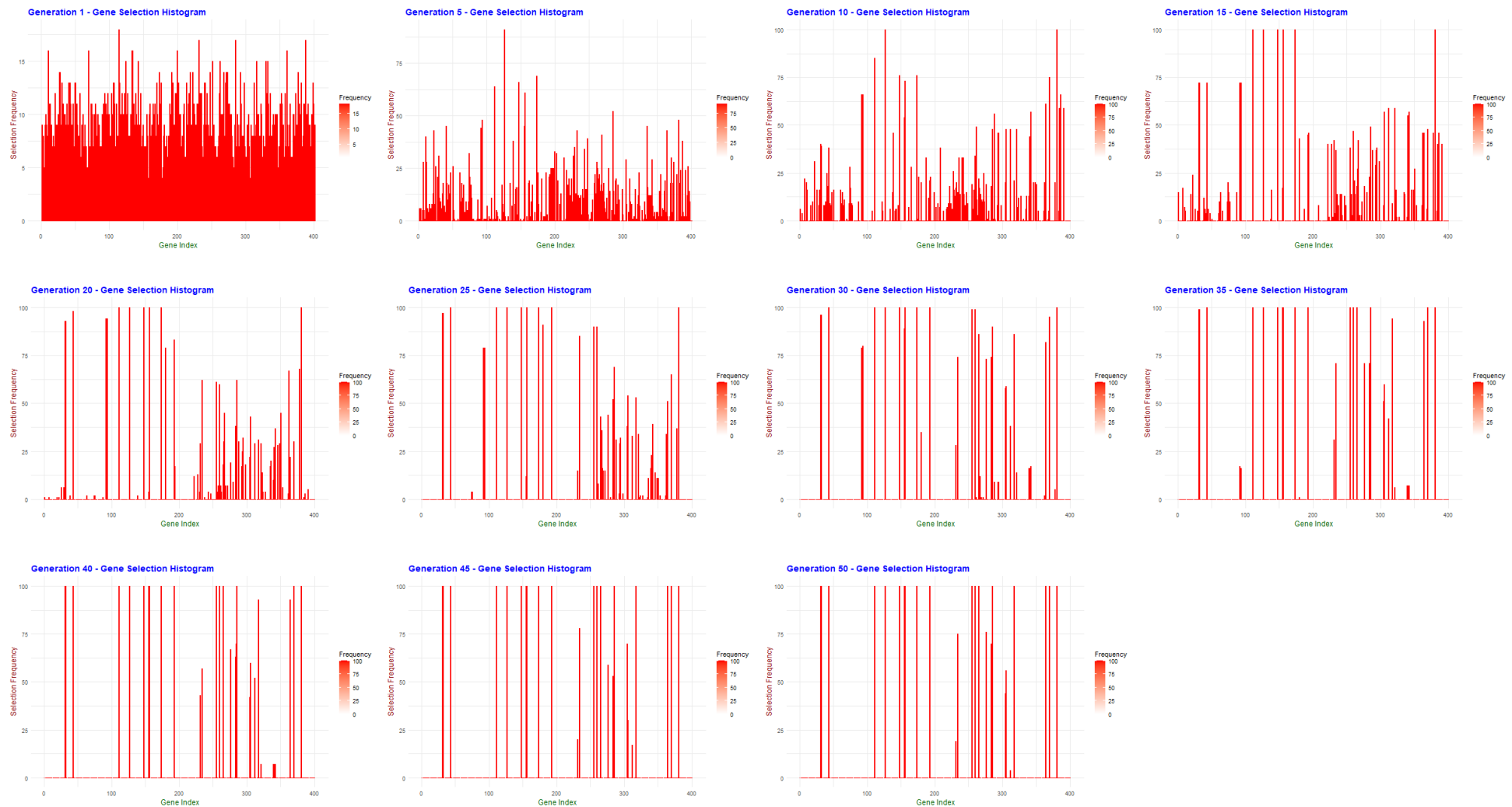
```
halaand.thomas.test <- function(pls.model) {  
  # Bước 1: Trích xuất giá trị PRESS từ mô hình pls.model  
  if (is.null(pls.model$validation) || is.null(pls.model$validation$PRESS)) {  
    stop("The PLS model does not contain validation PRESS values.")  
  }  
  PRESS <- pls.model$validation$PRESS[1, ] # Lấy giá trị PRESS  
  ncomp <- length(PRESS) # Số thành phần chính  
  # Tìm giá trị PRESS nhỏ nhất  
  PRESS_min <- min(PRESS)  
  min_comp <- which.min(PRESS)  
  # Bước 2: Tính giá trị thống kê F và p-value  
  Fcalc <- PRESS / PRESS_min  
  m <- nrow(pls.model$model) # Số mẫu hiệu chuẩn  
  # Tính giá trị p-value  
  p_values <- pf(Fcalc, m, m) # Sử dụng hàm phân phối F  
  # Bước 3: Chọn số thành phần chính nhỏ nhất sao cho p-value < 0.75  
  selected_comp <- which(p_values < 0.75)[1]  
  # Kiểm tra nếu không tìm thấy giá trị nào thỏa mãn điều kiện  
  if (is.na(selected_comp)) {  
    warning("No components satisfy the p-value < 0.75 criterion.")  
    selected_comp <- NA  
  }  
  # Kết quả trả về: Số thành phần chính được chọn và bảng dữ liệu  
  results <- data.frame(  
    Component = 1:ncomp,  
    PRESS = PRESS,  
    Fcalc = Fcalc,  
    p_value = p_values  
  )  
  
  list(  
    selected_comp = selected_comp,  
    results = results  
  )  
}
```

Phụ lục 7. Chương trình lựa chọn khoảng bước sóng trong phần mềm R bằng phương pháp PLS của thuốc Turbezid

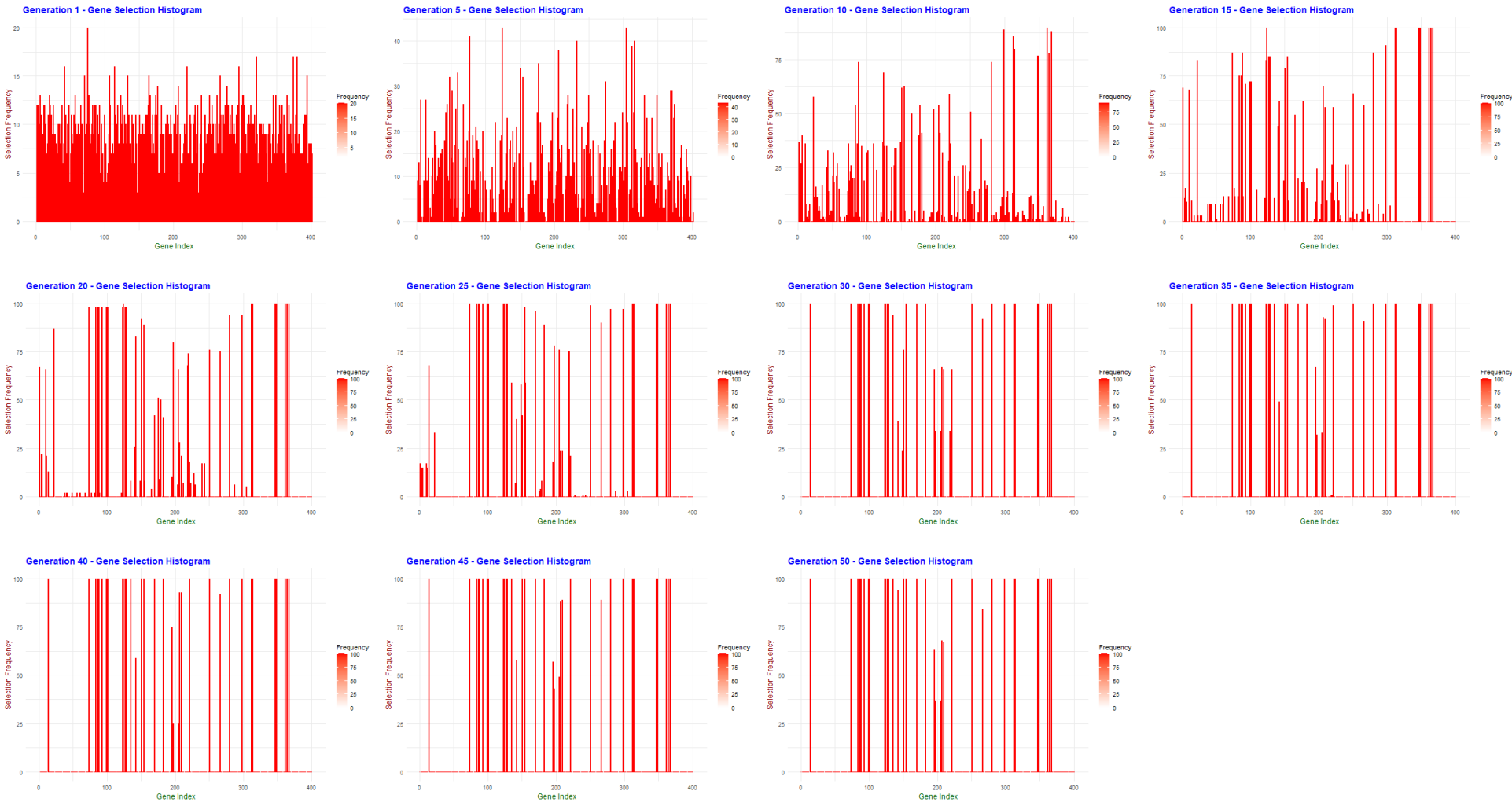
```
find_best_wavelength.turbezid <- function(A.train, C.train, A.val, C.val, a, b, c) {  
  # Khởi tạo data frame để lưu khoảng bước sóng và giá trị RMSEP  
  results <- data.frame(Wavelength_Range = character(), RMSEP_Value = numeric(),  
                        stringsAsFactors = FALSE)  
  
  # Vòng lặp qua các khoảng bước sóng từ a đến b với bước nhảy c  
  while (a + c <= b - 10) {  
    # Xác định khoảng bước sóng hiện tại  
    current_range <- a:(b)  
  
    # Bước 1: Chạy mô hình PLS trên tập huấn luyện cho khoảng bước sóng hiện tại  
    model.pls <- pls(C.train ~ A.train[,current_range],scale = TRUE,  
                    method = "kernelpls", validation = "CV")  
  
    # Bước 2: Chọn số thành phần tối ưu cho mô hình PLS  
    optimal_ncomp <- selectNcomp(model.pls, method = "randomization", plot =  
FALSE)  
    if (optimal_ncomp < 2){  
      optimal_ncomp = 2  
    }  
    else {optimal_ncomp = optimal_ncomp}  
  
    # Bước 3: Dự đoán trên tập kiểm tra với số thành phần tối ưu  
    predictions <- as.vector(round(predict(model.pls, newdata = A.val[,current_range],  
                                         ncomp = optimal_ncomp),2))  
  
    # Bước 4: Tính giá trị RMSEP cho khoảng bước sóng hiện tại  
    rmsep_value <- calculate_rmsep(C.val, predictions)  
  
    # Lưu kết quả vào data frame  
    results <- rbind(results, data.frame(Wavelength_Range = paste(a, b, sep = ":"),  
RMSEP_Value = rmsep_value))  
  
    # Tăng bước sóng bắt đầu `a` thêm `c`  
    a <- a + c  
  }
```

```
}  
  
# Tìm khoảng bước sóng có RMSEP nhỏ nhất  
best_row <- results[which.min(results$RMSEP_Value), ]  
  
# Kết quả trả về bao gồm tất cả các khoảng và khoảng bước sóng có RMSEP nhỏ  
nhất  
list(All_Results = results, Best_Range = best_row)  
}
```

Phụ lục 8.1 Tần suất các bước sóng xuất hiện qua 50 thế hệ của thuật toán di truyền của MTZ



Phụ lục 8.2 Tần suất các bước sóng xuất hiện qua 50 thế hệ của thuật toán di truyền của SPI



Phụ lục 9. Chương trình tìm SR.threshold tối ưu

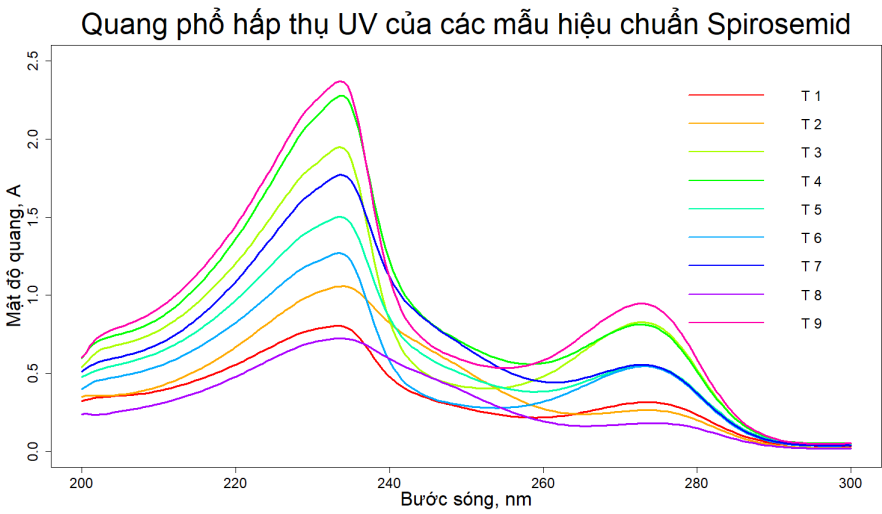
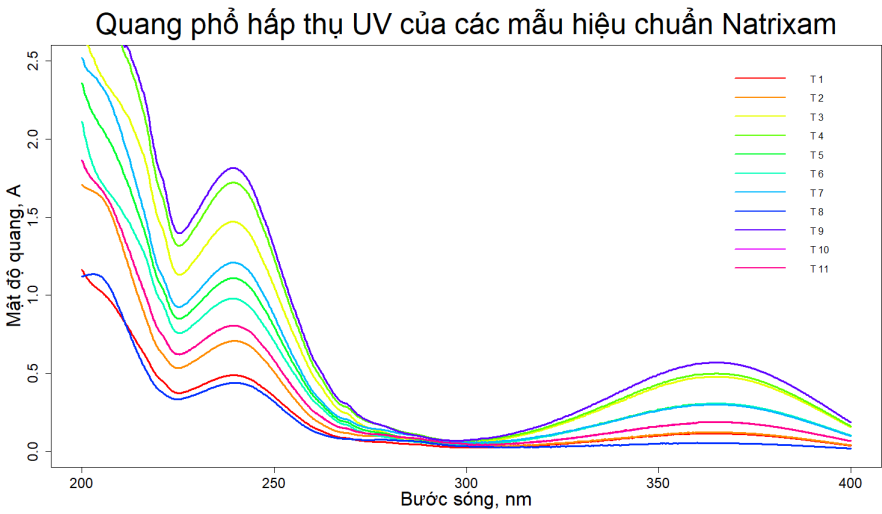
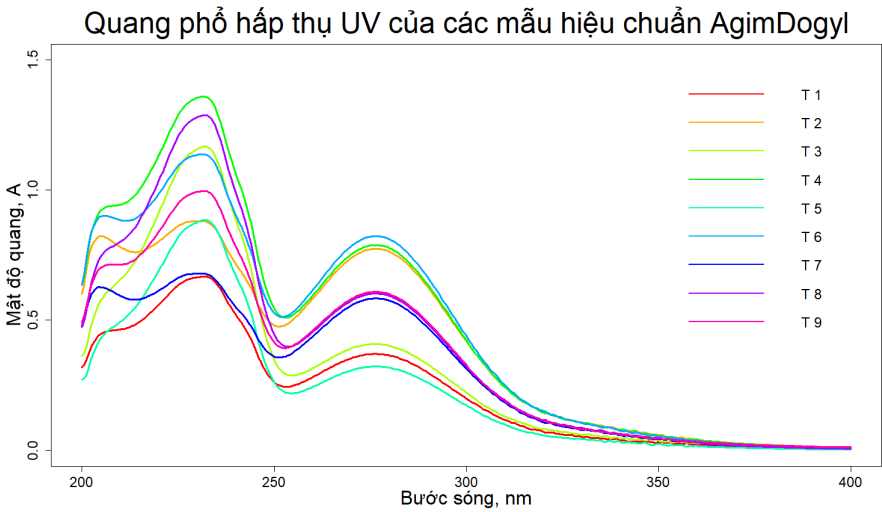
```
find_best_sr_threshold <- function(C.train, A.train, a, b, c) {  
  # Khởi tạo các biến để lưu trữ kết quả.  
  thresholds <- seq(a, b, by = c)  
  rmsep_values <- numeric(length(thresholds))  
  num_wavelengths <- integer(length(thresholds))  
  best_rmsep <- Inf  
  best_threshold <- NA  
  best_wavelengths <- NULL  
  iteration <- 1  
  # Khung dữ liệu để lưu trữ kết quả cho mỗi ngưỡng.  
  results_df <- data.frame(  
    Iteration = integer(),  
    SR.threshold = numeric(),  
    RMSEP = numeric(),  
    Num_Wavelengths_Selected = integer(),  
    stringsAsFactors = FALSE  
  )  
  # Lặp qua từng giá trị SR.threshold.  
  for (threshold in thresholds) {  
    # Chạy filterPLSR với SR.threshold hiện tại.  
    model.SR <- filterPLSR(C.train, A.train, ncomp = 10, criterion = "minimum",  
                          validation = "LOO", SR.threshold = threshold)  
    # Kiểm tra xem có bước sóng nào được chọn không.  
    selected_wavelengths <- model.SR$SR  
    if (length(selected_wavelengths) == 0) {  
      message(paste("No wavelengths selected for SR.threshold =", threshold))  
      next # Bỏ qua và chuyển sang ngưỡng tiếp theo nếu không có bước sóng nào được  
        chọn.  
    }  
  
    # Tạo lại A.new với các bước sóng đã được chọn.  
    A.new <- A.train[, selected_wavelengths, drop = FALSE]  
    # Khớp mô hình PLS với các bước sóng đã chọn.  
    pls.model <- plsr(C.train ~ A.new, method = "kernelpls", validation = "LOO")  
    # Xác định số lượng thành phần tối ưu dựa trên kiểm tra chéo.  
    ncomp_opt <- selectNcomp(pls.model, method = "randomization", plot = FALSE)  
    # Tính toán RMSEP tại số lượng thành phần tối ưu.  
    rmsep_value <- RMSEP(pls.model)$val["CV", , ncomp_opt+1]  
    # Thêm kết quả cho lần lặp này.
```

```

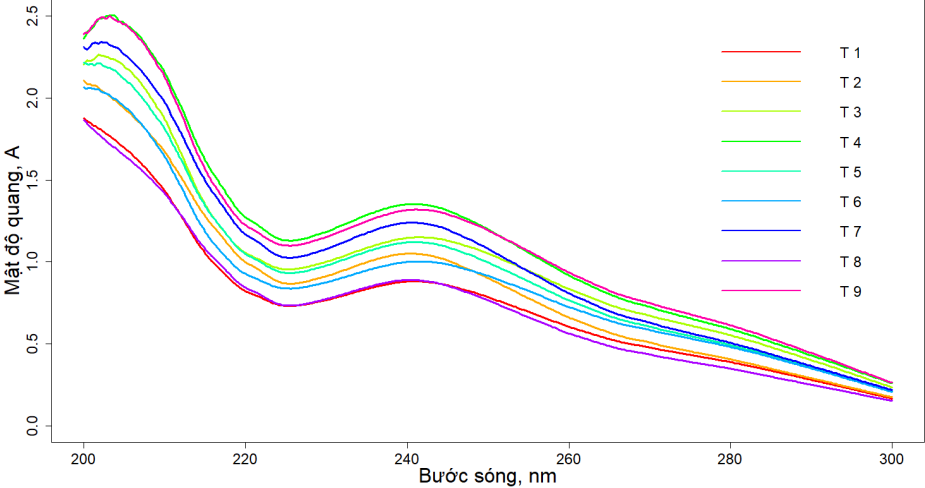
results_df <- rbind(results_df, data.frame(
  Iteration = iteration,
  SR.threshold = threshold,
  RMSEP = rmsep_value,
  Num_Wavelengths_Selected = length(selected_wavelengths)
))
# Cập nhật RMSEP tốt nhất và ngưỡng nếu đây là RMSEP thấp nhất.
if (rmsep_value < best_rmsep) {
  best_rmsep <- rmsep_value
  best_threshold <- threshold
  best_wavelengths <- selected_wavelengths
}
iteration <- iteration + 1
}
# TRích xuất dữ liệu cuối cùng
list(
  best_rmsep = best_rmsep,
  best_threshold = best_threshold,
  best_wavelengths = best_wavelengths,
  results_df = results_df
)
}

```

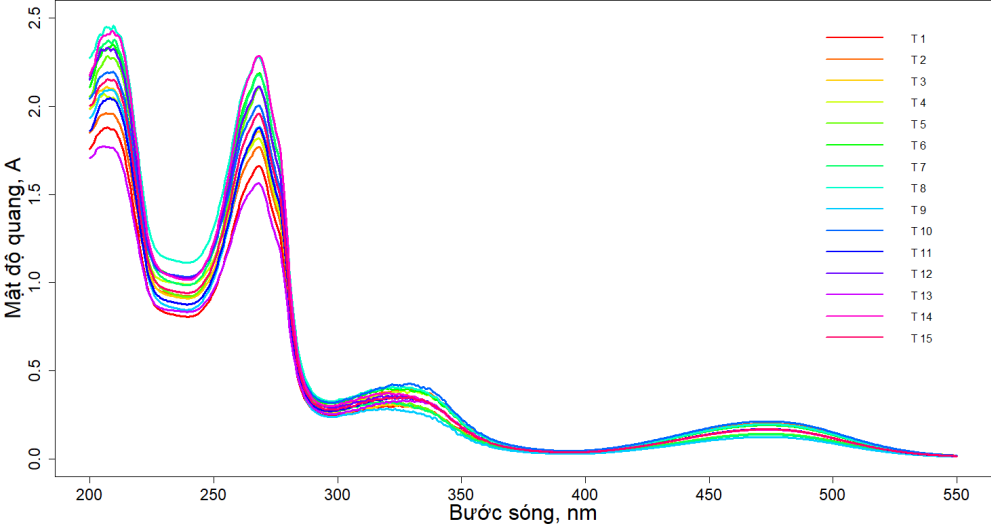
Phụ lục 10. Phổ hấp thụ của các mẫu trong tập hiệu chuẩn



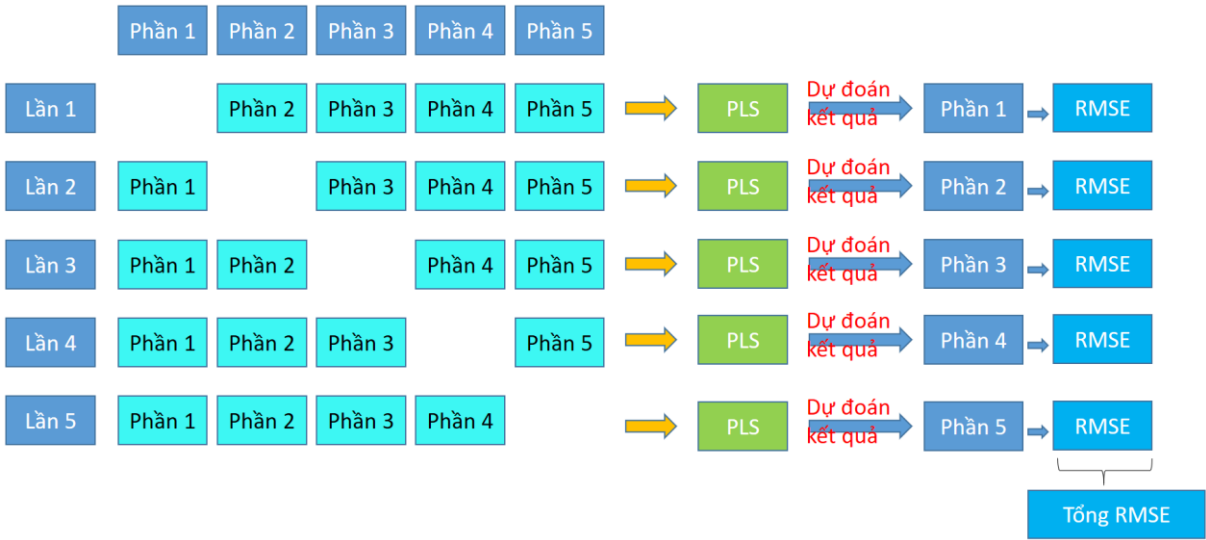
Quang phổ hấp thụ UV của các mẫu hiệu chuẩn Zoamco-A



Quang phổ hấp thụ UV của các mẫu hiệu chuẩn Turbidity



Phụ lục 11. Quy trình thực hiện phương pháp thẩm định chéo



Phụ lục 11. Phiếu kết quả kiểm nghiệm

SỞ Y TẾ TỈNH THỪA THIÊN HUẾ
TRUNG TÂM KIỂM NGHIỆM
THUỐC, MỸ PHẨM, THỰC PHẨM

17 Trương Định - TP Huế
ĐT/Fax: 0234.3936272

Số: 24DV744

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập - Tự do - Hạnh phúc



PHIẾU KẾT QUẢ KIỂM NGHIỆM

Tên mẫu: Viên nén bao phim AgimDogy1
Nơi sản xuất: Nguyễn Duy Lưu – Khoa Hóa, trường Đại học Khoa học Huế
Số lô: 020323 Ngày sản xuất: 280323 Hạn dùng: 280326
Mô tả mẫu: Mẫu được ép trong vỉ nhôm kín
Thời gian lưu mẫu: Theo thủ tục Xử lý đối tượng thử nghiệm KNH/TT/7.4-01
Nơi gửi mẫu: Nguyễn Duy Lưu – Khoa Hóa, trường Đại học Khoa học Huế
Ngày nhận mẫu: 06/05/2024
Thời gian kiểm nghiệm: từ 06/05/2024 đến 22/05/2024
Kết quả thử nghiệm:

TT	Chỉ tiêu thử nghiệm	Đơn vị	Phương pháp thử	Kết quả
1	Định lượng Metronidazol - Lần 1 - Lần 2 - Lần 3	mg/viên	HPLC	124,54 123,25 122,84
2	Định lượng Spiramycin - Lần 1 - Lần 2 - Lần 3	IU/viên	DDVN V	774.172,3 768.611,4 765.895,9

Ghi chú: -KPH. Không phát hiện (nghĩa là dưới ngưỡng định lượng/phát hiện của phương pháp thử)
Thừa Thiên Huế, ngày 24 tháng 05 năm 2024

CS. GIÁM ĐỐC
PHÓ GIÁM ĐỐC



Ngô Thị Tuyết Mai

(*) Các tiêu chuẩn thực hiện kiểm nghiệm theo quy định của pháp luật. (**) Phương pháp chuẩn được công nhận theo ISO 17025

Thương mại và xuất khẩu được thực hiện theo quy định của pháp luật

Được giao cho người sử dụng để kiểm tra chất lượng sản phẩm

Trong quá trình kiểm nghiệm, nếu có sự thay đổi về quy định hoặc tiêu chuẩn, người sử dụng phải cập nhật và áp dụng theo quy định mới

Đã được kiểm tra và đạt kết quả theo quy định của pháp luật

Đã được kiểm tra và đạt kết quả theo quy định của pháp luật

Trang 5/5

Số: 24DV743



PHIẾU KẾT QUẢ KIỂM NGHIỆM

Tên mẫu: Viên nén bao phim TURBEZID

Nơi sản xuất: Nguyễn Duy Lưu – Khoa Hóa, trường Đại học Khoa học Huế

Số lô: 23121

Ngày sản xuất: 20/08/23

Hạn dùng: 20/08/27

Mô tả mẫu: Mẫu được ép trong vỉ nhôm kín, đựng tiếp trong hộp giấy, nhãn rõ ràng

Thời gian lưu mẫu: Theo thủ tục Xử lý đối tượng thử nghiệm KNH/TT/7.4-01

Nơi gửi mẫu: Nguyễn Duy Lưu – Khoa Hóa, trường Đại học Khoa học Huế

Ngày nhận mẫu: 06/05/2024

Thời gian kiểm nghiệm: từ 06/05/2024 đến 14/05/2024

Kết quả thử nghiệm:

TT	Chỉ tiêu thử nghiệm	Đơn vị	Phương pháp thử	Kết quả
1	Định lượng Rifampicin - Lần 1 - Lần 2 - Lần 3	mg/viên	HPLC	156,48 154,95 154,83
2	Định lượng Isoniazid - Lần 1 - Lần 2 - Lần 3	mg/viên	HPLC	76,86 76,62 76,39
3	Định lượng Pyrazinamid - Lần 1 - Lần 2 - Lần 3	mg/viên	HPLC	402,86 403,95 402,69

Ghi chú: -KPH: Không phát hiện (nghĩa là dưới ngưỡng định lượng/phát hiện của phương pháp thử)

Thị trấn Huế, ngày 24 tháng 05 năm 2024

Đ. GIÁM ĐỐC

PHÓ GIÁM ĐỐC



Ngô Thị Tuyết Mai

(*) : Theo được chấp hành theo như mẫu thử (**): Phương pháp được công nhận (ISO 17025)

Thường có và mẫu do: Khoa Hóa, trường Đại học Khoa học Huế

Kiểm nghiệm theo yêu cầu: ... gửi từ địa chỉ mẫu thử

Trường hợp không nhận được mẫu trong trường hợp: ... không ... mẫu thử hoặc hết hạn bảo quản mẫu thử gửi đến

... do: hạn ... gửi từ địa chỉ không ... gửi từ địa chỉ không ... do: không ... Trường hợp không nhận được mẫu thử, kết quả thử nghiệm không được chấp nhận theo quy định

BM HD/TT/7.4-01 (02/05/2024)

Trang 1/1

Số: 23DV129



PHIẾU KẾT QUẢ THỬ NGHIỆM

Tên mẫu: Viên nén bao phim Spiromide-40

Nơi sản xuất: The Searle Company Limited

Số lô: C0033

Ngày sản xuất: 06-12-2021

Hạn dùng: 05-12-2023

Mô tả mẫu: Mẫu được ép trong vỉ kín, đựng tiếp trong hộp giấy, phân rõ ràng.

Tiêu chuẩn áp dụng: TCCS số 1182- DS/A-2

Thời gian lưu mẫu: Theo thủ tục Xử lý đối tượng thử nghiệm KNH/TT/7.4-01

Nơi gửi mẫu: Trường Đại học khoa học

77 Nguyễn Huệ, phường Phú Nhuận, thành phố Huế.

Ngày nhận mẫu: 30/01/2023

Kết quả thử nghiệm:

TT	Chỉ tiêu thử nghiệm	Đơn vị	Kết quả
1	Định lượng Spironolacton:		
	- Lần 1	mg/viên	49,75
	- Lần 2		49,15
	- Lần 3		50,02
2	Định lượng Furosemide:		
	- Lần 1	mg/viên	39,65
	- Lần 2		39,97
	- Lần 3		40,25

Thừa Thiên Huế, ngày 10 tháng 02 năm 2023



- (*) : Chỉ tiêu được thực hiện bởi nhà thầu phụ. (**): Phép thử chưa được công nhận theo ISO 17025
- Thông tin về mẫu và nơi gửi mẫu do khách hàng cung cấp
 - Kết quả thử nghiệm chỉ có giá trị đối với mẫu đem thử
 - Trung tâm không nhận khiếu nại trong trường hợp không có mẫu lưu hoặc hết thời hạn lưu mẫu theo quy định
 - Các bản trích sao kết quả này sẽ không có giá trị nếu không có sự đồng ý của Trung tâm Kiểm nghiệm thuốc, mỹ phẩm, thực phẩm Thừa Thiên Huế

BM TT/7.8-01.04 (01/05/2019)

Trang 1/1



Số: 23DV130

PHIẾU KẾT QUẢ THỬ NGHIỆM

Tên mẫu: Viên nén bao phim Zoamco-A

Nơi sản xuất: Công ty cổ phần Pymepharco

Số lô: 201220

Ngày sản xuất: 16/12/20

Hạn dùng: 161223

Mô tả mẫu: Mẫu được ép trong vỉ kín, đựng tiếp trong hộp giấy, phân rõ ràng.

Tiêu chuẩn áp dụng: TCCS số 057-A-072-10

Thời gian lưu mẫu: Theo thủ tục Xử lý đối tượng thử nghiệm KNH/TT/7.4-01

Nơi gửi mẫu: Trường Đại học khoa học

77 Nguyễn Huệ, phường Phú Nhuận, thành phố Huế.

Ngày nhận mẫu: 30/01/2023

Kết quả thử nghiệm:

TT	Chỉ tiêu thử nghiệm	Đơn vị	Kết quả
1	Định lượng Amlodipin:		
	- Lần 1	mg/viên	5,20
	- Lần 2		5,24
	- Lần 3		5,17
2	Định lượng Atorvastatin:		
	- Lần 1	mg/viên	10,49
	- Lần 2		10,54
	- Lần 3		10,46

Thừa Thiên Huế, ngày 10 tháng 02 năm 2023

GIÁM ĐỐC



Ngô Thị Thanh Xuân

- (*): Chỉ tiêu được thực hiện bên nhà thầu phụ; (**): Phép thử chưa được công nhận theo ENO HA 170,5

- Thông tin về mẫu và nơi gửi mẫu do khách hàng cung cấp

- Kết quả thử nghiệm chỉ có giá trị đối với mẫu đem thử

- Trung tâm không nhận khiếu nại trong trường hợp không có mẫu lưu hoặc hết thời hạn lưu mẫu theo quy định

- Các bản trích sao kết quả này sẽ không có giá trị nếu không có sự đồng ý của Trung tâm Kiểm nghiệm thuốc, mỹ phẩm, thực phẩm Thừa Thiên Huế