



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΡΟΦΙΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΩΝ ΟΙΝΟΥ, ΑΜΠΕΛΟΥ ΚΑΙ ΠΟΤΩΝ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

**Χρήση υπερφασματικών δεδομένων του ελάσματος
για την διάκριση ποικιλιών της αμπέλου**

Μητρόπουλος Γεώργιος

ΑΜ: 171055

Επιβλέπων

Όνοματεπώνυμο:

Καθ. Κόρκας Ηλίας

ΑΘΗΝΑ, Φεβρουάριος 2023



**UNIVERSITY OF WEST ATTICA
SCHOOL OF FOOD SCIENCE
DEPARTMENT OF WINE, VINE AND BEVERAGE SCIENCES**

BACHELOR THESIS

**Using leaf hyperspectral data for grapevine
cultivar discrimination**

Georgios Mitropoulos

Registration Number: 171055

**Supervisor
name and surname: Prof. Elias Korkas**

ATHENS, February 2023



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΤΡΟΦΙΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΩΝ ΟΙΝΟΥ, ΑΜΠΕΛΟΥ ΚΑΙ ΠΟΤΩΝ**

ΔΗΛΩΣΗ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ

Οι υπογράφοντες δηλώνουμε ότι έχουμε εξετάσει τη διπλωματική εργασία με τίτλο:

«Χρήση υπερφασματικών δεδομένων του ελάσματος

για την διάκριση ποικιλιών της αμπέλου»

και βεβαιώνουμε ότι γίνεται δεκτή.

Ψηφιακή Υπογραφή Επιβλέποντα Καθηγητή (1^ο Μέλους Επιτροπής)	Καθ. Κόρκας Ηλίας
Ψηφιακή Υπογραφή Καθηγητή (2^ο Μέλους Επιτροπής)	Δρ. Τάσκος Δημήτριος
Ψηφιακή Υπογραφή Καθηγητή (3^ο Μέλους Επιτροπής)	Αναπλ. Καθ. Μπανίλας Γεώργιος

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογράφων Μητρόπουλος Γεώργιος του Αθανασίου, με αριθμό μητρώου 171055 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Επιστημών Τροφίμων του Τμήματος Επιστημών Οίνου, Αμπέλου και Ποτών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



Μητρόπουλος Γεώργιος

ΠΕΡΙΛΗΨΗ

Ο προσδιορισμός της ποικιλίας της αμπέλου είναι ένα θέμα μεγάλου ενδιαφέροντος στην αμπελουργία. Οι μέθοδοι που υπάρχουν σήμερα για τον προσδιορισμό της ποικιλίας της αμπέλου είναι η αμπελογραφία, η υγρή χημεία μέσω μεθόδων που βασίζονται σε ισοένζυμα και ή DNA ανάλυση. Σε αυτή την εργασία θα γίνει μία απόπειρα ταυτοποίησης των ποικιλιών της αμπέλου με την χρήση φασματοσκοπίας στα φύλλα που στοχεύει σε μία γρήγορη ,αυτόματη και φιλική προς το περιβάλλον μεθόδου διάκρισης των ποικιλιών της αμπέλου. Η διάκριση των ποικιλιών της αμπέλου έγινε με τον NIPALS-PLS που είναι ένας αλγόριθμος μερικών ελαχίστων τετραγώνων του SAS Studio χρησιμοποιώντας τα υπερφασματικά δεδομένα τις ανάκλασης των φύλλων. Η υπερφασματική απεικόνιση πραγματοποιήθηκε με το φασματοραδιόμετρο PSR+ 3500 Spectroradiometer της Spectral Evolution που έχει φασματικό εύρος καταγραφής τα 350 nm έως 2500 nm. Ο ταξινομητής δημιουργήθηκε χρησιμοποιώντας 180 φύλλα, 10 από κάθε μία από τις ποικιλίες Ασπρούδι, Ασπρούδα, Γκουμάσι Λιανόρραγο, Λευκό Γιάννενα, Μαυρούδι, Μελίσσα, Μοσχούδι, Πατίνες, Σάμια, Βόλιτσα Λευκή έγιναν 6 διαφορετικές παραλλαγές αρχικοποίησης της ταξινόμησης με καλύτερη απόδοση ταξινόμησης 80,8%. Η προτεινόμενη μέθοδος φαίνεται πολλά υποσχόμενη για την διάκριση των ποικιλιών της αμπέλου , αλλά σίγουρα ένα ευρύτερο φάσμα ποικιλιών αμπέλου θα πρέπει να δοκιμαστεί πριν εφαρμοστεί για την ανίχνευση τους.

Λέξεις κλειδιά: Μορφοχρωματομετρία φύλλων, Υπερφασματική εικόνα, *Vitis vinifera* L., Ταξινομητής Ποικιλίας αμπέλου

ABSTRACT

Vine variety identification is a topic of great interest in viticulture. The methods that exist today to determine the variety of the vine are the viticulture, liquid chemistry through methods based on isozymes and or DNA analysis. In this work an attempt will be made to identify the vine varieties using spectroscopy in the leaves which aims at a fast, automatic and environmentally friendly method of distinguishing grape varieties. The distinction of grape varieties was done by NIPALS-PLS which is a partial least algorithm quadrats of SAS Studio using leaf reflectance hyperspectral data. Hyperspectral imaging was performed with Spectral Evolution's PSR+ 3500 Spectroradiometer which has a spectral recording range of 350 nm to 2500 nm. The classifier was created using 180 leaves, 10 of each of the varieties Asprudi, Aspruda, Goumasi Lianorrigo, Lefko Yannena, Mavroudi, Melissa, Moschoudi, Patinas, Samia ,Volitsa Lefki 6 different initialization variants of the classification were made with a better classification performance of 80.8%. The proposed method looks promising for distinguishing grape varieties, but certainly a wider range of grape varieties should be tested before being applied to detect them.

Keywords: Leaf morpho-colorimetry ,Hyperspectral image, *Vitis vinifera* L. , Classifier Grapevine cultivar

Ευχαριστίες

Πριν να προχωρήσω στην παρουσίαση της πτυχιακής μου εργασίας, θα ήθελα να εκφράσω τις ευχαριστίες μου στους ανθρώπους που συνέβαλαν σε αυτή τη προσπάθεια. Καταρχήν ευχαριστώ τον κ. Κόρκα Ηλία καθηγητή και προέδρου του τμήματος Οίνου, Αμπέλου και Ποτών του Πανεπιστημίου Δυτικής Αττικής επιβλέπων στην πτυχιακή μου εργασία. Επιπλέον τις ευχαριστίες μου θα ήθελα να εκφράσω στον κ. Τάσκο Δημήτριο ερευνητή του Ινστιτούτου Αμπέλου με έδρα την Λυκόβρυση, για την υποστήριξη και βοήθεια που μου προσέφερε καθ' όλη την διάρκεια, καθώς και για την επιστημονική υποστήριξη και καθοδήγησή του. Επίσης θα ήθελα να ευχαριστώ τον κ. Κανδυλάκη Ζαχαρία από το ΕΜΠ για την παροχή εξοπλισμού και τεχνογνωσίας. Τέλος θα ήθελα να ευχαριστήσω όλους του προπτυχιακούς καθηγητές του τμήματος μου για τις γνώσεις που μου προσέφεραν στα χρόνια της φοίτησής μου.

Πίνακας περιεχομένων

Κατάλογος Πινάκων	vi
Κατάλογος Σχημάτων.....	vii
1 Εισαγωγή.....	1
2 Βιβλιογραφική Ανασκόπηση.....	3
2.1 Αμπελογραφία	3
2.1.1 Κλασικές μέθοδοι Αμπελογραφίας	3
2.1.2 Σύγχρονες μέθοδοι Αμπελογραφίας.....	4
2.2 Ανάκλαση φωτός.....	5
2.2.1 Ανάκλαση στα φύλλα.....	6
2.3 Μοντέλα Πρόβλεψης.....	9
2.3.1 Στατιστικό μοντέλο nirs	10
2.4 Χρήση Φασματικών μεθόδων στην Άμπελο.....	10
2.4.1 Φασματοσκοπία στο εγγύς υπέρυθρο φάσμα(NIRS) για τον προσδιορισμό ποικιλιών της αμπέλου	11
2.4.2 OntheGO Φασματοσκοπία για τον προσδιορισμό ποικιλιών της αμπέλου	11
2.4.3 Φασματοσκοπία για τον προσδιορισμό του επίπεδου νερού στην άμπελο.....	11
2.4.4 Φασματοσκοπία μέσω δορυφόρου για τον προσδιορισμό ποικιλιών της αμπέλου	12
3 Υλικά και Μέθοδοι	13
3.1 Δειγματοληψία	13
3.2 Υλικά.....	13
3.3 Πρώτη επεξεργασία δεδομένων	14
3.4 Ανάλυση δεδομένων.....	16
4 Αποτελέσματα & Συζήτηση.....	18
4.1 Χρήση του Rank350_1365 σε όλο το σύνολο δεδομένων (παραλλαγή 1α)	18
4.2 Χρήση του Rank350_1365μόνο στα φύλλα που βρίσκονται στην κορυφή του φυτού (παραλλαγή 1β)	21
4.3 Χρήση του Rank350_1365μόνο στα φύλλα των φύλλων που βρίσκονται στην μέση του φυτού(παραλλαγή 1γ)	24
4.4 Χρήση της δυαδικής κωδικοποίησης σε όλο το σύνολο δεδομένων (παραλλαγή 2α).....	27

4.5	Χρήση της δυαδικής κωδικοποίησης μόνο στα φύλλα που βρίσκονται στην κορυφή του φυτού (παραλλαγή 2β).....	30
4.6	Χρήση της δυαδικής κωδικοποίησης μόνο στα φύλλα που βρίσκονται στην μέση του φυτού (παραλλαγή 2γ)	33
4.7	Σύγκριση παραλλαγών	37
4.8	Αξιολόγηση του αλγορίθμου NIPALS για τις παραλλαγές 1β και 1γ (χρήση του rank 350-1365 για τα φύλλα της κορυφής και της μέσης αντίστοιχα).....	38
4.8.1	Αξιολόγηση αλγορίθμου NIPALS για την παραλλαγή 1β.....	38
4.8.2	Αξιολόγηση αλγορίθμου NIPALS για την παραλλαγή 1γ.....	40
5	Συμπεράσματα	42
6	Βιβλιογραφία.....	44
7	Παράρτημα.....	47

Κατάλογος Πινάκων

Πίνακας 1	Ποικιλίες Δειγματοληψίας	13
Πίνακας 2	Χαρακτηριστικά ομαδοποιημένων ποσοστών ανακλάσεων	15
Πίνακας 3	Επεξηγηματικότητα 1α.....	18
Πίνακας 4	Root Mean PRESS 1α.....	19
Πίνακας 5	Επεξηγηματικότητα 1β.....	22
Πίνακας 6	Root Mean PRESS 1β.....	22
Πίνακας 7	Επεξηγηματικότητα 1γ.....	25
Πίνακας 8	Root Mean PRESS 1γ.....	25
Πίνακας 9	Επεξηγηματικότητα 2α.....	28
Πίνακας 10	Root Mean PRESS 2α.....	28
Πίνακας 11	Επεξηγηματικότητα 2β.....	31
Πίνακας 12	Root Mean PRESS 2β.....	31
Πίνακας 13	Επεξηγηματικότητα 2γ.....	34
Πίνακας 14	Root Mean PRESS 2γ.....	34
Πίνακας 15	Επεξηγηματικότητα Παραλλαγών	37
Πίνακας 16	Αξιολόγηση 1β πριν την επικύρωση.....	38
Πίνακας 17	Αξιολόγηση 1β μετά την επικύρωση.....	39
Πίνακας 18	Αξιολόγηση 1γ πριν την επικύρωση.....	40
Πίνακας 19	Αξιολόγηση 1γ μετά την επικύρωση.....	41
Πίνακας 20	Δείγμα Αρχικών Δεδομένων	47

Κατάλογος Σχημάτων

Εικόνα 1	Καμπύλη μέσων όρων ποσοστών ανάκλασης ανά ποικιλία	16
Εικόνα 2	R-Square Analysis 1a	20
Εικόνα 3	Variable Importance Plot 1a.....	21
Εικόνα 4	R-Square Analysis 1b.....	23
Εικόνα 5	Variable Importance Plot 1b	24
Εικόνα 6	R-Square Analysis 1c	26
Εικόνα 7	Variable Importance Plot 1c.....	27
Εικόνα 8	R-Square Analysis 2a	29
Εικόνα 9	Variable Importance Plot 2a.....	30
Εικόνα 10	R-Square Analysis 2b.....	32
Εικόνα 11	Variable Importance Plot 2b	33
Εικόνα 12	R-Square Analysis 2c	35
Εικόνα 13	Variable Importance Plot 2c.....	36

Συντμήσεις, ακρωνύμια, σύμβολα και ορισμοί

PLS Partial Least Squares

NIPALS Nonlinear Iterative Partial Least Squares

NIRS Near-infrared spectroscopy

1 Εισαγωγή

Η μεγάλη οικονομική και κοινωνική σημασία της αμπελουργίας και η βιομηχανία του κρασιού παγκοσμίως ενθαρρύνει την ανάπτυξη και την εφαρμογή καινοτόμων τεχνολογιών με στόχο την αντικειμενική παρακολούθηση του αμπελώνα για τη βελτίωση της ποιότητας των σταφυλιών και του κρασιού. Υπάρχουν πολλές χιλιάδες ποικιλίες αμπέλου *Vitis vinifera* L. παγκοσμίως και η καλλιέργειά του, η ποιότητα του κρασιού και η τιμή που καταβάλλεται για τα σταφύλια τους εξαρτάται από την ποικιλία. Επομένως, υπάρχει αυξανόμενο ενδιαφέρον για την ακριβή αναγνώριση των ποικιλιών αμπέλου: προκειμένου να εξασφαλίζεται η αληθινή ταυτότητα ως προς τον τύπο των μοσχευμάτων ενός φυτωρίου, να αναγνωρίζονται τα αμπέλια απαγορευμένων ποικιλιών σε περιοχές ονομασίας Προέλευσης παγκόσμιος, ή για να εκτιμηθούν τα στοιχεία προσφοράς και ζήτησης και οι τιμές σε μια δεδομένη αναπτυσσόμενη αμπελουργική περιοχή.

Στις μέρες μας, οι τρέχουσες τεχνικές για να γίνει η διάκριση μεταξύ των ποικιλιών της αμπέλου περιλαμβάνουν την αμπελομετρία, την υγρή χημεία μέσω μεθόδων που βασίζονται σε ισοένζυμο και ή DNA ανάλυση. Η αμπελομετρία χρησιμοποιεί μορφολογικές διαφορές μεταξύ των ποικιλιών για την αναγνώρισή τους. Ένα από τα μειονεκτήματά της είναι ότι αυτή η ταύτιση είναι οπτική και πρέπει να γίνει από ειδικό με εκτενή εκπαίδευση. Από την άλλη, οι μέθοδοι που βασίζονται στο ισοένζυμο ή στο DNA, αν και ακριβείς, μοιράζονται το ίδιο μειονέκτημα με την αμπελομετρία, καθώς μπορούν να εφαρμοστούν μόνο από ειδικά εκπαιδευμένα άτομα. Επιπλέον, αυτές οι τεχνικές υγρής χημείας είναι χρονοβόρες και απαιτητικές και δεν μπορούν παρέχουν μια αληθινή αναγνώριση χιλιάδων αμπέλων με γρήγορο και ρεαλιστικό τρόπο. Ως εκ τούτου, η ανάπτυξη μίας φιλικής προς το περιβάλλον (χωρίς διαλύτες), αυτόματης και γρήγορης μεθόδου για την αναγνώριση της ποικιλίας της αμπέλου θα μπορούσε να είναι πολύ πολύτιμη για τη βιομηχανία σταφυλιών και κρασιού, συμπεριλαμβανομένων φυτωρίων, οινοποιείων και αμπελουργών και προμηθευτών.

Τα αμπελόφυλλα διαφορετικών ποικιλιών ποικίλλουν ως προς τη χημική σύσταση και τη μορφολογία, όπως σχήμα, διάσταση, χρώμα και σχήμα άκρης. Αυτές οι διαφορές στα μορφομετρικά χαρακτηριστικά έχουν αποκτηθεί ως εξελικτικά χαρακτηριστικά που αντιστοιχούν σε συγκεκριμένες γονιδιακές εκφράσεις και στην αλληλεπίδρασή τους με το περιβάλλον στο οποίο έχει προσαρμοστεί κάθε ποικιλία. Επομένως, κάθε μορφολογία και χημική παράμετρος των φύλλων είναι μοναδική για όλες τις ποικιλίες.

Στο πλαίσιο της παρούσας εργασίας θα αξιοποιήσουμε την τεχνική της φασματοσκοπίας σε συνδυασμό με το στατιστικό μοντέλο np για να αποδείξουμε ότι αυτή η μοναδικότητα κληρονομείται και στην ανάκλαση του φωτός. Πιο συγκεκριμένα θα μετρήσουμε την ανάκλαση των φύλλων στα μήκη κύματος 350nm-1365nm, καθώς θεωρούμε ότι συγκαταλέγεται στην οικογένεια των εξελικτικών χαρακτηριστικών. Το αντικείμενο της φασματοσκοπίας είναι η ανάλυση της αλληλεπίδρασης της ηλεκτρομαγνητικής ακτινοβολίας με την ύλη σε διαφορετικά μήκη κύματος. Βρίσκει εφαρμογή στον προσδιορισμό της ποικιλίας διαφόρων φυτών σε καλλιέργειες, όπως στη ντομάτα και στη δάφνη. Ωστόσο, στα αμπέλια, η τοπική φασματοσκοπία έχει εφαρμοστεί σπάνια για τη διαφοροποίηση των ποικιλιών.

2 Βιβλιογραφική Ανασκόπηση

Στη παρούσα ενότητα γίνεται περιγραφή του θεωρητικού υπόβαθρου που απαιτείται για την διεξαγωγή των πειραμάτων. Θα αναφερθούμε στην έννοια της αμπελογραφίας, στην συνέχεια θα ακολουθήσει περιγραφή των κλασικών και των σύγχρονων μεθόδων αμπελογραφίας. Ύστερα θα γίνει εκτενείς ανάλυση, καθώς είναι το κύριο αντικείμενο μελέτης αυτής της μελέτης, στην ανάκλαση του φωτός, πως αυτή διαμορφώνεται στο εύρος του φάσματος και από τι παράγοντες επηρεάζεται. Τέλος θα γίνει μια αναφορά στο στατιστικό μοντέλο PLS-nipals που χρησιμοποιήσαμε για να πάρουμε τα αποτελέσματα μας.

2.1 Αμπελογραφία

Ο όρος Αμπελογραφία έχει ελληνική ρίζα από το άμπελος και γραφή και σημαίνει περιγραφή της αμπέλου. Πιο συγκεκριμένα η Αμπελογραφία είναι κλάδος της αμπελουργίας και το αντικείμενο της είναι η μελέτη και η περιγραφή των χαρακτηριστικών και των ιδιοτήτων των διάφορων ειδών, ποικιλιών και κλώνων της αμπέλου. Πιο συγκεκριμένα ασχολείται με την περιγραφή και βαθμολόγηση σε ένα διεθνές αποδεκτό σύστημα, των κυριότερων εξωτερικών χαρακτηριστικών ορισμένων οργάνων της αμπέλου, όπως φύλλων, βλαστών, σταφυλών, ραγών κ.α. Η Αμπελογραφία διαιρείται σε γενική και ειδική. Η γενική Αμπελογραφία έχει ως αντικείμενο μελέτης την μελέτη της οικογένειας των αμπελιδών (Vitaceae) με τα γένη και τα είδη που την απαρτίζουν με κύρια έμφαση στο γένος *Vitis* και τα είδη του. Η ειδική Αμπελογραφία που είναι και ο κύριος σκοπός της Αμπελογραφίας αφορά την διάκριση μεταξύ των καλλιεργούμενων ποικιλιών της αμπέλου. Σε αυτή την εργασία θα μας απασχολήσει η ειδική Αμπελογραφία.

2.1.1 Κλασικές μέθοδοι Αμπελογραφίας

Υπάρχουν χιλιάδες ποικιλίες αμπέλου *Vitis vinifera* L. παγκοσμίως, συνεπώς καθίσταται απαραίτητη η ανάγκη για την ύπαρξη ενός σαφώς καθορισμένου συστήματος περιγραφής. Μια ολοκληρωμένη αμπελογραφική μελέτη περιλαμβάνει επιγραμματικά τα παρακάτω:

- Το όνομα της ποικιλίας
- Τα συνώνυμα
- Την προέλευση και την διάδοση
- Την αμπελογραφική περιγραφή η οποία απαρτίζεται από 13 χαρακτήρες
 1. Εκβλάστημα (κορυφή νεαρού βλαστού μήκους 10-20cm)
 2. Ποώδης Βλαστός

3. Νέα φύλλα της κορυφής
 4. Νέα φύλλα της βάσης
 5. Τυπικό ανεπτυγμένο φύλλο
 6. Άνθος
 7. Έλικες
 8. Σταφυλή
 9. Ράγα
 10. Γίγαρτο
 11. Κληματίδα
 12. Κορμός
 13. Ρίζες ενός έτους
- Τις φαινολογικές παρατηρήσεις
 1. Εκβλάστηση (ημερομηνία εκβλάστησης του 50% των οφθαλμών)
 2. Άνθηση (αρχή- πλήρης- τέλος)
 3. Έναρξη ωρίμανσης (περκασμός)
 4. Ωρίμανση (ημερομηνία τεχνολογικής ωριμότητας)
 5. Φυλλόπτωση
 - Τις καλλιεργητικές ιδιότητες
 1. Ζωηρότητα
 2. Παραγωγικότητα
 3. Διαμόρφωση
 4. Κλάδεμα καρποφορίας
 5. Εδάφη στα οποία προσαρμόζεται η ποικιλία
 6. Γονιμότητα των ταχυφυών (μεσοκάρδιων)
 7. Συγγένεια (εμβολιασμός με τα υποκείμενα της αμπέλου)
 - Την αντοχή στους εχθρούς ,τις αντίξοες συνθήκες και τις ασθένειες
 - Τον προορισμό χρήσης
 - Την οικονομική σημασία της ποικιλίας

Σήμερα, από όλους τους αμπελογράφους χρησιμοποιείται ο τρόπος περιγραφής της I.B.P.G.R. (International Board for Plant Genetic Resources) που αποτελείται από την περιγραφή και βαθμολόγηση 125 βασικών χαρακτηριστικών της αμπέλου που μπορούν στην συνέχεια να καταχωρηθούν σε ηλεκτρονικό υπολογιστή και να μας δώσουν αποτέλεσμα σχετικά με την ποικιλία. Ένα από τα μειονεκτήματά της κλασικής Αμπελογραφίας είναι ότι τόσο η οπτική ταύτιση καθώς και η χρήση κατάλληλου λογισμικού μπορεί να γίνει μόνο από ειδικό με εκτενή εκπαίδευση.

2.1.2 Σύγχρονες μέθοδοι Αμπελογραφίας

Προχωρώντας από τις κλασσικές στις σύγχρονες μεθόδους, έχουν σημειωθεί αρκετές βελτιώσεις στην ταξινόμηση των ποικιλιών αμπέλου μέσω της χρήσης αντικειμενικών

μεθόδων. Αυτές επικεντρώνονται κυρίως σε βιοχημικές μεθόδους όπως η υγρή χημεία και μεθόδους που βασίζονται σε ισοένζυμα και φασματοραδιομετρικές και γενετικές τεχνικές αποτύπωσης δακτυλικών αποτυπωμάτων.

Οι βιοχημικές μέθοδοι βασίζονται στο ότι οι πρωτεΐνες ως το τελικό προϊόν έκφρασης των γονιδίων αντανακλούν την γενετική σύσταση του οργανισμού. Έτσι, εάν είμαστε σε θέση να προσδιορίσουμε τις πρωτεΐνες μπορούμε να τις αντιστοιχήσουμε στο αρχικό γονιδίωμα και να επιτύχουμε την διάκριση, την ταυτοποίηση και την ανίχνευση γενετικής συγγένειας μεταξύ ποικιλιών. Η πρωτεϊνική σύσταση ενός οργανισμού μπορεί να αποκαλυφθεί με την βοήθεια κατάλληλων τεχνικών που επιτυγχάνουν τον διαχωρισμό των πρωτεϊνών. Η μειωμένη ακρίβεια των βιοχημικών δεικτών, καθώς αυτοί επηρεάζονται από τους παράγοντες όπως η ηλικία, η κατάσταση του πρέμνου, το περιβάλλον ανάπτυξης κ.ά. οδήγησαν στην ανάγκη για την χρήση των μοριακών μεθόδων. Οι Μοριακές μέθοδοι με την χρήση δεικτών βασισμένων στο DNA, έχουν την δυνατότητα να αποκαλύψουν τον πολυμορφισμό του γενετικού υλικού και να επιτύχουν μια πιο ακριβή ταξινόμηση. Ωστόσο, αν και ακριβείς, αυτές οι τεχνικές απαιτούν υψηλό επίπεδο εξειδικευμένων δεξιοτήτων, ειδικά και δαπανηρά όργανα και κουραστική εργαστηριακή εργασία, γεγονός που καθιστά τη γενικευμένη πρακτική χρήση δύσκολη και απρόσιτη για τη συνήθη ταξινόμηση της ποικιλίας της αμπέλου. Σαν αποτέλεσμα, δεν δύναται να παρέχουν αληθινή αναγνώριση χιλιάδων αμπέλων με γρήγορο και ρεαλιστικό τρόπο.

2.2 Ανάκλαση φωτός

Η ανάκλαση του φωτός είναι το φαινόμενο εκείνο κατά το οποίο το φως (ή κάποιο άλλο είδος ηλεκτρομαγνητικής ακτινοβολίας) προσπίπτει σε μια επιφάνεια, η οποία δεν μπορεί να το απορροφήσει και "το επιστρέφει πίσω". Όλα τα σώματα ανακλούν φως είτε αυτό είναι το φυσικό φως του ήλιου είτε είναι η λάμπα του σπιτιού μας. Έτσι, στην πραγματικότητα έχουμε δύο ειδών ανακλάσεις:

- ολική ανάκλαση, όπως στην περίπτωση του καθρέφτη
- διάθλαση, όπως στις υπόλοιπες και πιο κοινές περιπτώσεις

Στην περίπτωση της διάθλασης όταν μια επιφάνεια ενός αντικειμένου χτυπάται με μια ακτινοβολία απορροφάει τα περισσότερα μήκη κύματος, που αντιστοιχούν σε κάποιο εύρος χρωμάτων ενώ ανακλά ένα μικρό εύρος του μήκους κύματος που αντιστοιχεί στο χρώμα που βλέπουμε εμείς να έχει το αντικείμενο.

2.2.1 Ανάκλαση στα φύλλα

2.2.1.1 Φωτοσυνθετικές Χρωστικές στα φύλλα.

Όλες οι φωτοσυνθετικές χρωστικές εντοπίζονται στους χλωροπλάστες, οι οποίοι είναι συνδεδεμένα με τη μεμβράνη με οργανίδια που κατανέμονται στο κυτταρόπλασμα των κυττάρων του μεσόφυλλου των φύλλων. Το μεσόφυλλο είναι το στρώμα κάτω από την επιδερμίδα που αποτελείται από αδιαφοροποίητα λεπτά τοιχώματα τα παρεγχυματικά κύτταρα που περιέχουν χλωροπλάστες που σχηματίζουν έναν συμπαγή ιστό (με λίγους χώρους αέρα). Αυτό το στρώμα έχει συνήθως πάχος ενός έως λίγα κύτταρα, εκτός από τα σημεία που συνδέονται τα στόμια στο αγγειακό σύστημα. Σε αυτόν τον ιστό βρίσκονται επίσης κανάλια ρητίνης, που επιτρέπουν την εξώθηση της ρητίνης από τις επιφάνειες των φύλλων ή των μίσχων, με αποτέλεσμα μερικές φορές επικαλύπτοντάς τις να αλλάζοντας την ανάκλαση της επιφάνειας. Η απορρόφηση φωτός στο ορατό φάσμα από φυτικές φωτοσυνθετικές χρωστικές παράγει μια μοναδική υπογραφή φασματικής ανάκλασης, που είναι υπεύθυνη για το χρώμα των φυτών. Το φως συλλαμβάνεται στη διαδικασία της φωτοσύνθεσης που παρέχει την απαιτούμενη ενέργεια για τα φυτά να συνθέσουν υδατάνθρακες (οργανική ύλη) από το CO₂, μέσω μιας σειράς μεταφοράς ηλεκτρονίων που συμβαίνουν στις στοίβες των θυλακοειδών μεμβρανών στους χλωροπλάστες. Η αντίδραση φωτοσύνθεσης είναι:



Στην μεμβράνη του χλωροπλάστη (η θυλακοειδής), τα σύμπλοκα χρωστικής-πρωτεΐνης οργανώνονται σε δύο φωτοσυστήματα: το φωτοσύστημα I (PSI) και φωτοσύστημα II (PSII), τα οποία ονομάζονται επίσης P700 και P680 αντίστοιχα, όπου τα μήκη κύματος που καθορίζουν την μέγιστη απορρόφησή τους in vivo.

2.2.1.2 Φασματικό πεδίο ηλιακού φωτός

Το φασματικό πεδίο που ξεκινά στα 400 nm και τελειώνει στα 3000 nm, αλλά πιο συχνά στα 2500 nm, αντιστοιχεί στο 90% του ηλιακού φωτός (συνολική ενέργεια). Αυτή η περιοχή χωρίζεται κλασικά σε τρεις υποτομείς :

- το ορατό (400–800 nm) που χαρακτηρίζεται από ισχυρή απορρόφηση φωτός από τις φωτοσυνθετικές χρωστικές σε ένα πράσινο φύλλο.
- το εγγύς υπέρυθρο (800–1100 nm) όπου η απορρόφηση περιορίζεται στα βιοχημικά που συνθέτουν την ξηρή φυτική ύλη των φύλλων όπου η πολλαπλή διασπορά εντός του φύλλου, που σχετίζεται με το κλάσμα των χώρων αέρα, δηλαδή με την εσωτερική δομή, οδηγεί στην ανάκλαση και τα επίπεδα μετάδοσης.
- το υπέρυθρο βραχέων κυμάτων (1100–2500 nm) που είναι επίσης μια ζώνη ισχυρής απορρόφησης, κυρίως από το νερό στα φρέσκα φύλλα και δευτερευόντως από την ξηρή ουσία όταν το φύλλο μαραίνεται.

Γενικά, οι καμπύλες για τα πράσινα φύλλα έχουν όλες παρόμοια σχήματα, που ποικίλλουν κυρίως σε μέγεθος. Οι μηχανισμοί εσωτερικής απορρόφησης και σκέδασης ευθύνονται για την ομοιότητα στο σχήμα της ανάκλασης και τα φάσματα διαπερατότητας. Το ορατό (φωτοσυνθετικά ενεργή ακτινοβολία), χαρακτηρίζεται από μια ισχυρή απορρόφηση φωτός από τις φωτοσυνθετικές χρωστικές σε ένα πράσινο φύλλο τις χλωροφύλλες, τα καροτενοειδή και των μη φωτοσυνθετικών χρωστικών όπως η ανθοκυανίνη, με μέγιστη απορρόφηση στην περιοχή του μπλέ και του κόκκινου φάσματος και παρουσιάζει ελάχιστο στο πράσινο ενδιάμεσα από αυτά. Το φάσμα ανάκλασης είναι συνήθως κάτω από 5% στα 400 nm (μπλε) και αυξάνονται σταδιακά σε μια κορυφή περίπου 12 έως 15% στα 550 nm (πράσινο) αυτή η κορυφή ευθύνεται για το πράσινο χρώμα των φυτών που γίνεται αντιληπτό από το ανθρώπινο μάτι και στη συνέχεια μειώνεται σε περίπου 5% κοντά στα 675 nm (κόκκινο). Από αυτή την μείωση παρουσιάζεται μια απότομη άνοδος κατά μήκος του ορίου μεταξύ του κόκκινου και του υπέρυθρου (κόκκινη άκρη) και φτάνουμε σε ένα ασυμπτωτικό πλάτωμα περίπου 40 έως 60% στα 780 nm (κοντά στο υπέρυθρο). Αυτή η μεγάλη άνοδος τις ανάκλασης στα 750nm σηματοδοτεί το τέλος της απορρόφησης του ορατού φωτός από τις χλωροφύλλες. Σε μήκη κύματος μεγαλύτερα από 1100 nm, η ανάκλαση και η διαπερατότητα αρχίζουν να πέφτουν. Η ακτινοβολία απορροφάται κυρίως από το νερό σε περίπου 1200 nm, 1400 nm, 1900 nm και 2500 nm σε φρέσκα φύλλα, και δευτερευόντως από ξηρή ουσία σε ξερά φύλλα. Η ανάκλαση ενός αποξηραμένου φύλλου είναι συνήθως μεγαλύτερη από αυτή του νωπού φύλλου σε όλα τα μήκη κύματος. Δεδομένου ότι περιέχουν τα ίδια συστατικά, τα πράσινα φύλλα εμφανίζουν σχετική ποιοτική σταθερότητα στις οπτικές τους ιδιότητες. Η ποσοτική διακύμανση αυτών των ιδιοτήτων σχετίζεται με τις διακυμάνσεις τους σε χρωστική ουσία, σε νερό και περιεκτικότητα σε ξηρά ουσία, και με την ανατομική δομή των φύλλων.

2.2.1.3. Παράγοντες που επηρεάζουν την ανάκλαση-απορρόφηση στα φύλλα

Οι φασματικές ιδιότητες που σχετίζονται με τα φύλλα εξαρτώνται από πολλούς παράγοντες όπως το είδος του φυτού, η φαινολογική ηλικία των φύλλων, η αναπτυξιακή ηλικία, το μικροκλίμα, η θέση του φύλλου στο φυτό κλπ. Ειδικότερα οι επιφάνειες των περισσότερων ραχιαίων φύλλων αντανακλούν συνήθως περισσότερο φως από την κορυφή. Κατά μέσο όρο, η διαφορά ανάκλασης είναι περίπου 6,6%, αλλά μπορεί να υπάρχει αύξηση έως και 20% για ορισμένα φύλλα, ανάλογα με το είδος ή την ποικιλία. Η παρουσία ενός στρώματος κεριού στην επιφάνεια του φύλλου μπορεί να αλλάξει τον τύπο της ανάκλασης του φωτός από κατοπτρική (γυαλιστερή), βρίσκεται σε φύλλα με λείες επιφάνειες, έως διάχυτα (θαμπό), που σχετίζονται με μικροσκοπικά τραχιά επιφάνειες. Το πάχος του φύλλου σχετίζεται με το μήκος κύματος της μέγιστης ανάκλασης παρόλα αυτά μία σειρά ερευνών καθιέρωσε ορισμένες σχέσεις μεταξύ του πάχους του φύλλου και της αντανάκλασης της κόκκινης ακμής/σχεδόν υπέρυθρης ακτινοβολίας αρκετών οινοποιήσιμων ποικιλιών αμπέλου, αλλά έδειξε εξάρτηση από την ποικιλία. Ο μεγαλύτερος αριθμός διεπαφών μεταξύ των κυτταρικών τοιχωμάτων και ο μεσοκυττάριος αέρας στα παχύτερα φύλλα έχει ως αποτέλεσμα υψηλότερη ανάκλαση και χαμηλότερη διαπερατότητα.

Η σκέδαση του φωτός σχετίζεται πιο στενά στην εσωτερική δομή (αναλογίες του παρεγχύματος) παρά στο πάχος κάθε αυτό, οπότε το πάχος του φύλλου είναι αλληλένδετο με την δομή του. Στο επίπεδο των φύλλων, ο μαρασμός που προκαλείται από την απώλεια της περιεκτικότητας σε νερό αλλάζει τόσο τις οπτικές ιδιότητες που σχετίζονται με την απορρόφηση νερού και τη διάθλαση από το εσωτερικό φως των φύλλων (σκέδαση). Τα φυτά που αναπτύσσονται σε θερμά ή ξηρά κλίματα στην εφηβεία τους μειώνουν σημαντικά το θερμικό φορτίο των φύλλων αυξάνοντας την ανάκλαση από την επιφάνειά τους, η οποία κατά συνέπεια μειώνει την ποσότητα της απορροφούμενης ακτινοβολίας. Το φαινόμενο αυτό παρουσιάζεται ως προσαρμοστικό χαρακτηριστικό που συμβάλλει στον έλεγχο της θερμοκρασίας της επιφάνειας των φύλλων. Έρευνες έχουν δείξει ότι τα πυκνά τριχώματα αυξάνουν την ανάκλαση σε όλα μήκη κύματος μεταξύ 400 nm και 3000 nm, μειώνοντας την απορρόφηση κατά τρεις φορές σε σύγκριση με γλαυκά φύλλα, μειώνοντας τη θερμοκρασία των φύλλων και τους ρυθμούς φωτοσυνθετικής και διαπνοής. Το μέγεθος της UV ανακλαστικότητας είναι σε άμεση συνάρτηση με το σχήμα, τη διαμέτρο και της κατανομής των επιδερμικών επιφανειακών δομών του κεριού των φύλλων. Τέλος Η υπερβολική ηλιακή υπεριώδη ακτινοβολία μπορεί να οδηγήσει στην ουσιαστική τροποποίηση της φασματικής ανάκλασης της επιδερμίδας, τόσο στο ορατό και σε μικρότερο βαθμό στο κοντινό υπέρυθρο.

2.3 Μοντέλα Πρόβλεψης

Τα μοντέλα πρόβλεψης χρησιμοποιούνται για την πρόβλεψη μελλοντικών γεγονότων ή αποτελεσμάτων μέσω της ανάλυσης μοτίβων σε ένα σύνολο δεδομένων. Είναι ένα εργαλείο της επαγωγικής στατιστικής που χρησιμοποιεί ιστορικά δεδομένα για να προβλέψει τη μελλοντική δραστηριότητα, συμπεριφορά και το αναμενόμενο αποτέλεσμα που προκύπτει από μια αλληλουχία δεδομένων. Τα μοντέλα πρόβλεψης δεν εστιάζουν μόνο στην πρόβλεψη ενός μελλοντικού γεγονότος, αλλά και στην απόκτηση άγνωστης μέχρι στιγμής πληροφορίας που μπορεί να συμπεραθεί από τα διαθέσιμα δεδομένα (π.χ. η πιθανότητα μια τραπεζικά συναλλαγή να είναι κομμάτι απάτης). Στην περίπτωση αυτή, το συμβάν έχει ήδη και ο στόχος είναι να προβλέψουμε εάν η μελλοντική ανάλυση θα διαπιστώσει ότι η συναλλαγή είναι δόλια.

Υπάρχουν πολλοί τρόποι κατηγοριοποίησης των μοντέλων πρόβλεψης. Πρακτικά, μπορούν να συνδυαστούν πολλαπλοί τύποι μοντέλων κατά τη διαδικασία βελτιστοποίησης των αποτελεσμάτων. Η πιο σημαντική διάκριση είναι μεταξύ μοντέλων επιβλεπόμενης μάθησης και μη επιβλεπόμενης.

Τα μοντέλα επιβλεπόμενης μάθησης είναι μια προσέγγιση που ορίζεται από τη χρήση επισημασμένων συνόλων δεδομένων. Αυτά τα σύνολα δεδομένων έχουν σχεδιαστεί για να εκπαιδεύουν αλγόριθμους για την ταξινόμηση δεδομένων ή την ακριβή πρόβλεψη των αποτελεσμάτων. Η μάθηση με επίβλεψη μπορεί να χωριστεί σε δύο τύπους προβλημάτων κατά την εξόρυξη δεδομένων: ταξινόμηση και παλινδρόμηση:

- Τα προβλήματα ταξινόμησης χρησιμοποιούν έναν αλγόριθμο για την ακριβή αντιστοίχιση των δεδομένων σε συγκεκριμένες κατηγορίες, όπως ο διαχωρισμός των μήλων από τα πορτοκάλια.
- Τα προβλήματα παλινδρόμησης χρησιμοποιούν έναν αλγόριθμο για την κατανόηση της σχέσης μεταξύ εξαρτημένων και ανεξάρτητων μεταβλητών. Είναι χρήσιμα για την πρόβλεψη αριθμητικών τιμών με βάση διαφορετικούς συνδυασμούς δεδομένων, όπως οι προβλέψεις εσόδων από πωλήσεις για μια δεδομένη επιχείρηση.

Η μεγαλύτερη διαφορά μεταξύ αυτών των προσεγγίσεων είναι ότι τα μοντέλα επιβλεπόμενης μάθησης απαιτούν περισσότερη προσοχή στη σωστή επισήμανση των συνόλων δεδομένων εκ των προτέρων.

2.3.1 Στατιστικό μοντέλο nipals

Το στατιστικό μοντέλο nipals είναι μια μέθοδος υπολογισμού κύριων συνιστωσών των δεδομένων που αξιοποιεί τα μερικά ελάχιστα τετράγωνα (PLS). Συγκεκριμένα, είναι μια επαναληπτική διαδικασία που χρησιμοποιείται για την κατασκευή μοντέλων PLS για την αναπαράσταση πολυμεταβλητών συνόλων δεδομένων χρησιμοποιώντας έναν αριθμό κύριων συνιστωσών που έχουν μικρότερη διάσταση από τον αριθμό των αρχικών μεταβλητών. Είναι μια ευρέως χρησιμοποιούμενη μέθοδος καθώς παρέχει πιο ακριβή αποτελέσματα σε σύγκριση με άλλες μεθόδους.

Ο αλγόριθμος PLS μπορεί να θεωρηθεί ως παραλλαγή της πολλαπλής παλινδρόμησης, σε περιπτώσεις που ο αριθμός των μεταβλητών είναι μεγάλος. Σε τέτοιες περιπτώσεις, σπάνια υπάρχουν αρκετά δεδομένα για την κατασκευή ενός αξιόπιστου μοντέλου που μπορεί να χρησιμοποιηθεί για την πρόβλεψη της εξαρτημένης μεταβλητής Y από τις ανεξάρτητες μεταβλητές X , ένα ακόμη πρόβλημα είναι ότι συνήθως οι περισσότερες μεταβλητές του συνόλου δεδομένων προσθέτουν θόρυβο στο μοντέλο μειώνοντας την απόδοσή του, αντί να το βοηθούν. Το PLS καταπολεμά αυτό το πρόβλημα επιλέγοντας τα πιο σημαντικά στοιχεία και δημιουργώντας τις λανθάνουσες συνιστώσες. Σκοπός είναι η κατασκευή ενός νέου συνόλου δεδομένων (το οποίο βασίζεται στο αρχικό) που αντιπροσωπεύει όσο το δυνατόν περισσότερο την διακύμανση του αρχικού, ενώ παράλληλα προβλέπει καλά τη μεταβλητή Y . Έτσι, χτίζεται ένα νέο μοντέλο παλινδρόμησης, το οποίο αξιοποιεί το νέο σύνολο.

2.4 Χρήση Φασματικών μεθόδων στην Άμπελο.

Οι περισσότερες έρευνες επικεντρώνονται στις δια-ειδικές παραλλαγές, δηλαδή στη διαφορά μεταξύ των ειδών. Αν και η ενδοειδική μεταβλητότητα μπορεί να είναι υψηλότερη από τη διαειδική μεταβλητότητα, τα φάσματα ανάκλασης και διαπερατότητας, τα οποία περιέχουν πληροφορίες σχετικά με τις διαφορές στα χαρακτηριστικά των φύλλων, έχουν χρησιμοποιηθεί στη διάκριση ειδών για μεγάλο χρονικό διάστημα. Μετρήσεις φασματικής ανάκλασης στον οπτικό τομέα έχουν δείξει ότι αν και σχεδόν όλα τα υγιή φύλλα έχουν γενικά παρόμοια φασματική απόκριση, διαφορές σε χρωστικές όπως η χλωροφύλλη, καθώς και διαφορές στη δομή των φύλλων, την περιεκτικότητα σε νερό και η βιοχημική σύνθεση μπορούν να αποδώσουν φασματικές υπογραφές που είναι αποτελεσματικές για διάκριση μεταξύ των ειδών. Παρόλα αυτά τα τελευταία χρόνια η φασματοσκοπία έχει αναδειχθεί ως μια πολλά υποσχόμενη μέθοδος αναγνώρισης της ποικιλίας αμπέλου ή αναγνώριση κλώνου

εντός ενός δεδομένου γονότυπου. Παρακάτω θα αναφερθούν μέθοδοι που έχουν δοκιμαστεί για τον προσδιορισμό της ποικιλίας της αμπέλου:

2.4.1 Φασματοσκοπία στο εγγύς υπέρυθρο φάσμα(NIRS) για τον προσδιορισμό ποικιλιών της αμπέλου

Ο NIRS ,με εύρος μήκους κύματος από 780 nm έως 2500 nm αν και στην πραγματικότητα αντιστοιχεί στο εγγύς υπέρυθρο (760–1400 nm) και στο μεσαίο υπέρυθρο (1400–3000 nm) είναι ίσως η πιο δοκιμασμένη μέθοδος φασματογραφίας στην άμπελο. Οι μετρήσεις γίνονται με φασματογράφο στα φύλλα σε περιβάλλον εργαστηρίου όπου παίρνουμε σημειακές μετρήσεις από συγκεκριμένα φύλλα και στην συνέχεια με ένα στατιστικό πακέτο φτιάχνεται ένα μοντέλο πρόβλεψης της ποικιλίας. Αυτά τα μοντέλα έχουν κατά μέσο όρο 90% ποσοστό πρόβλεψης της ποικιλίας. Ο NIRS είναι σίγουρα μία πολύ υποσχόμενη μέθοδος για την πρόβλεψη των ποικιλιών της αμπέλου , το μεγαλύτερο πρόβλημα στις έρευνες που έχουν γίνει είναι ο πολύ μικρός αριθμός διαφορετικών ποικιλιών που έχουν αναλυθεί για την εξαγωγή αποτελέσματος (2-15 ποικιλίες) ενώ έχουμε χιλιάδες ποικιλίες και το δεύτερο πρόβλημα είναι ότι δεν έχει ληφθεί υπόψη πως το κλίμα από το οποίο έχει ληφθεί το δείγμα έχει επηρεάσει την κάθε ποικιλία.

2.4.2 OntheGO Φασματοσκοπία για τον προσδιορισμό ποικιλιών της αμπέλου

Αυτή η μέθοδος παίρνει δεδομένα από το φάσμα των 350nm-2500nm αυτό που την διαφοροποιεί όμως είναι ο τρόπος δειγματοληψίας που γίνεται από ειδική κάμερα ενσωματωμένη σε όχημα μέσα στον αμπελώνα και βγάζει τα δεδομένα από την ολική ανάκλαση του φυτού σε αντίθεση με της σημειακές μετρήσεις σε συγκεκριμένα φύλλα και στην συνέχεια με ένα στατιστικό πακέτο φτιάχνεται ένα μοντέλο πρόβλεψης της ποικιλίας. Τα αποτελέσματα αυτής της μεθόδου έχουν ποσοστά αναγνώρισης της κάθε ποικιλίας από 80% έως 90%. Σε αυτές τις έρευνες έχει μελετηθεί μεγαλύτερος αριθμός ποικιλιών (10-30) αλλά αυτός ο αριθμός παραμένει συγκριτικά μικρός και συνήθως τα δείγματα είναι όλα από τον ίδιο αμπελώνα και μικρόκλιμα και δεν έχουμε εικόνα πως αυτό επηρεάζει. Παρόλα αυτά τα αποτελέσματα είναι ενθαρρυντικά.

2.4.3 Φασματοσκοπία για τον προσδιορισμό του επίπεδου νερού στην άμπελο

Σκοπός αυτής της μεθόδου είναι με την χρήση της φασματοσκοπίας να εντοπίσει αρχικά σε πια μήκη κύματος (στο βραχύ υπέρυθρο 1100nm-2500nm) απεικονίζεται καλύτερα το επίπεδο του νερό στο αμπέλι αλλά και να προσδιοριστεί για μια πληθώρα

ποικιλιών . Το γενικευμένο συμπέρασμα των ερευνών που έχουν γίνει είναι ότι έχουμε καλύτερα αποτελέσματα για κάποιες ποικιλίες και ότι σε κάθε ποικιλία δεν αντανάκλατε απαραίτητα το επίπεδο νερού στο ίδιο μήκος κύματος για κάθε ποικιλία.

2.4.4 Φασματοσκόπια μέσω δορυφόρου για τον προσδιορισμό ποικιλιών της αμπέλου

Έχει γίνει μία μεγάλη προσπάθεια με την χρήση ειδικά εξοπλισμένων Drones με φασματοραδιόμετρα και άλλα όργανα για να πάρουν αεροφωτογραφίες με φασματικά δεδομένα με στόχο των διαχωρισμό ποικιλιών μέσα σε αμπελώνες. Στην συνέχεια τα δεδομένα επεξεργάστηκαν με στατιστικά μοντέλα και βρέθηκε μια συσχέτιση ποικιλιών 83% .Το μεγάλο πλεονέκτημα αυτής της έρευνας είναι η λήψη μεγάλου αριθμού δεδομένων και μάλιστα από διαφορετικούς αμπελώνες σε διαφορετικές τοποθεσίες που σημαίνει και διαφορετικές συνθήκες ανάπτυξης (κλίμα).Το μοντέλο ήταν σε θέση να συσχετίσει ίδιες ποικιλίες μεταξύ τους σε διαφορετικές περιοχές άλλα και σε κάποιες περιπτώσεις να διακρίνει και διαφορετικούς κλώνους μέσα σε μία ποικιλία. Τα αποτελέσματα σε αυτή την μέθοδο είναι πολύ ενθαρρυντικά και υποδεικνύουν ότι η φασματοσκοπία σε συνδυασμό με την τεχνολογία θα είναι σε θέση να διακρίνει διαφορετικές ποικιλίες εύκολα και γρήγορα.

3 Υλικά και Μέθοδοι

Στη παρούσα ενότητα γίνεται περιγραφή του συνόλου δεδομένων και των επιμέρους χαρακτηριστικών του, ενώ ταυτόχρονα δίνεται έμφαση στην περιγραφή της μεθοδολογίας ανάλυσης. Συγκεκριμένα περιγράφονται το κομμάτι της δειγματοληψίας, της προεπεξεργασίας, του μετασχηματισμού, της ανάλυσης των δεδομένων με χρήση αλγορίθμων.

3.1 Δειγματοληψία

Στα πλαίσια της παρούσας εργασίας μελετήθηκαν συνολικά 10 ποικιλίες (Πίνακας 1) από το μετεωρολογικό σταθμό του Ινστιτούτου Αμπέλου στην Λυκόβρυση όπου συλλέχτηκαν δείγματα τον Ιούνιο του 2021 .

Πίνακας 1 Ποικιλίες Δειγματοληψίας

Ποικιλία	Συντομογραφία
Ασπρούδι	ASI
Ασπρούδα	ASP
ΓκουμάσιΛιανόρραγο	GOU
ΛευκόΓιάννενα	LGI
Μαυρούδι	MAV
Μελίσσα	MEL
Μοσχούδι	MOS
Πατίνες	PAT
Σάμια	SAM
Βόλιτσα Λευκή	VOL

Για κάθε μία από τις προαναφερθείσες ποικιλίες επιλέχθηκαν συνολικά τρία φυτά και από κάθε φυτό ένας κύριος βλαστός. Στη συνέχεια από κάθε κύριο βλαστό επιλέχθηκαν έξι φύλλα, τρία από την μέση και τρία από την κορυφή. Τέλος, από κάθε φύλλο λήφθηκαν μετρήσεις από τρία διαφορετικά σημεία του φύλλου.

3.2 Υλικά

Για την συλλογή των δεδομένων χρησιμοποιήθηκε το φασματοραδιόμετρο PSR+ 3500 Spectroradiometer της SpectralEvolution που έχει φασματικό εύρος καταγραφής τα 350 nm έως 2500 nm. Επίσης χρησιμοποιήθηκε ειδικό πιστόλι συνδεδεμένο με οπτική ίνα με το φασματοραδιόμετρο που δίνει την δυνατότητα για ευκολότερες και γρηγορότερες μετρήσεις καθώς δεν χρειάζεται να κρατάμε το ογκώδες και βαρύ φασματοραδιόμετρο πάνω από κάθε φύλλο. Τέλος έγινε χρήση και Λαμπερτιανής πλάκας. Κατά την διαδικασία των μετρήσεων λήφθηκαν υπόψη οι παρακάτω παράμετροι : Οι μετρήσεις να γίνονται υπό γωνία ώστε να μην σκιάζεται το φύλλο , και η αποφυγή των ινών. Για την αποφυγή σφαλμάτων σε κάθε

σημείο χρησιμοποιήθηκε η προεπιλεγμένη ρύθμιση καταγραφής μέσου όρου πολλαπλών μετρήσεων ο οποίος καταγράφεται στην μνήμη ως τελική τιμή. Στην προκειμένη περίπτωση, επιλέχθηκε μέσος όρος 3 μετρήσεων. Για να εξασφαλιστεί μεγαλύτερη ορθότητα αποτελεσμάτων. Η Λαμπερτιανή πλάκα χρησιμοποιήθηκε για την καταγραφή της ακτινοβολίας αναφοράς της περιοχής του αμπελώνα που αποτελεί απαιτούμενο δεδομένο για τον υπολογισμό της ανακλαστικότητας που υπολογίστηκε κατά την μετέπειτα επεξεργασία, επιτρέποντας την ανεξαρτησία των αποτελεσμάτων από τις τοπικές ατμοσφαιρικές συνθήκες (ηλιοφάνεια, θερμοκρασία, υγρασία).

3.3 Πρώτη επεξεργασία δεδομένων

Το σύνολο δεδομένων που δημιουργήθηκε περιέχει κατηγορικές, ακέραιες και πραγματικές μεταβλητές. Η εξαρτημένη τιμή του συνόλου (Variety) είναι η ποικιλία της αμπέλου και μπορεί να πάρει τις δέκα τιμές που περιγράφηκαν στον πίνακα της Ενότητας 3.1. Το πλήθος των παρατηρήσεων που έχουν καταχωρηθεί στο προς μελέτη σύνολο είναι 538, ενώ το πλήθος των χαρακτηριστικών που έχουν καταγραφεί για κάθε μια από αυτές είναι 1016. Επιπλέον, για κάθε παρατήρηση δημιουργήθηκαν άλλες 13 μεταβλητές οι οποίες βοήθησαν στην κατανόηση του συνόλου δεδομένου και στην ορθή επεξεργασία του.

Η γραμμογράφηση των αρχικών μεταβλητών, καθώς και των δημιουργημένων (1029 στο σύνολο) παρατίθεται αναλυτικά παρακάτω:

1. **MesID**: μοναδικός αριθμός γραμμής
2. **VarAB**: συντομογραφία της ποικιλίας (τιμές όπως περιγράφονται στον Πίνακα 1)
3. **RV**: κωδικός κάθε φυτού (30 διαφορετικές τιμές, κατηγορική)
4. **VineID**: αύξον αριθμός φυτού στο αμπέλι
5. **Vine**: αύξον αριθμός φυτού (δείγματος) ανά ποικιλία (στο τελικό σύνολο δεδομένων)
6. **Section**: κατηγορία της θέσης του φύλλου (M: middle, U: up)
7. **Node**: αύξον αριθμός του φύλλου στον κύριο βλαστό (κατά την δειγματοληψία)
8. **LeafCode**: κωδικός του αύξοντα αριθμού του φύλλου στον κύριο βλαστό
9. **Leaf**: αύξον αριθμός φύλλου ανά κύριο βλαστό (στο τελικό σύνολο δεδομένων)
10. **Rank550_690**: αύξουσα σειρά των μέσων όρων των ομαδοποιημένων ποσοστών ανάκλασης για μήκη κύματος 550 έως 690
11. **Rank760_1300**: αύξουσα σειρά των μέσων όρων των ομαδοποιημένων ποσοστών ανάκλασης για μήκη κύματος 760 έως 1300
12. **Rank350_1365**: αύξουσα σειρά των μέσων όρων των ομαδοποιημένων ποσοστών ανάκλασης για μήκη κύματος 350 έως 1365
13. **BIN**: δυαδική κωδικοποίηση ποικιλιών
14. **B350-B1365**: ποσοστό ανάκλασης σε κάθε μήκος κύματος

Ο αναγνώστης μπορεί να δει και αναλυτικότερα ένα δείγμα από το υπό μελέτη σύνολο και σε μορφή πίνακα στο Παράρτημα Πίνακας 20. Τα πρώτα εννιά στοιχεία της λίστας χρησιμεύουν στην κατανόηση του συνόλου δεδομένων και στην ορθή επιλογή των βημάτων προεπεξεργασίας. Τα επόμενα τέσσερα στοιχεία (Rank550_690, Rank760_1300, Rank350_1365, BIN) χρησιμεύουν στον ορθό μετασχηματισμό της εξαρτημένης μεταβλητής για οποία αντιστοιχεί στην ποικιλία της αμπέλου.

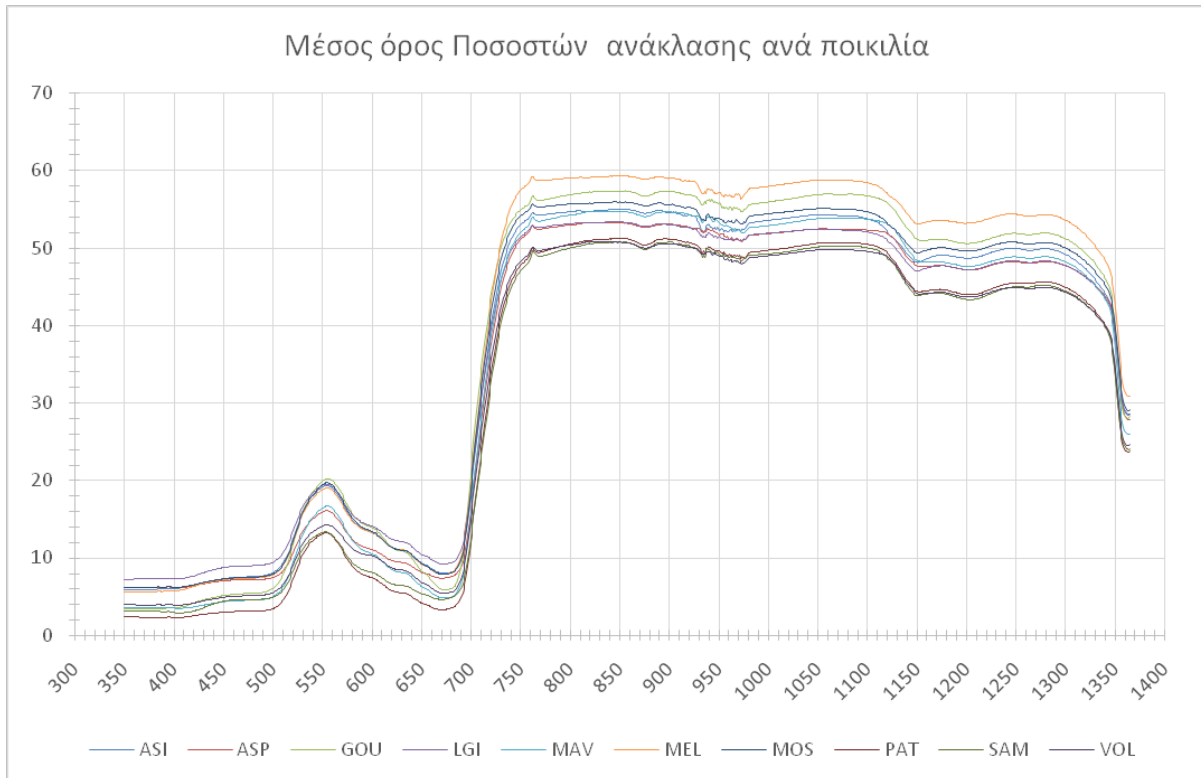
Στα πλαίσια της παρούσας εργασίας δοκιμάστηκαν δύο παραλλαγές όσον αφορά την κωδικοποίηση της κατηγορικής εξαρτημένης μεταβλητής. Η πρώτη παραλλαγή αξιοποιεί τον μέσο όρο του ποσοστού ανάκλασης ανά ποικιλία και ανά μήκος κύματος. Ύστερα, οι μέσοι όροι που προκύπτουν ανά ποικιλία, ομαδοποιούνται περαιτέρω (πάλι με χρήση μέσου όρου). Ο δεύτερος υπολογισμός έχει γίνει για τρία διαστήματα, για μήκη κύματος από 550 έως 690, από 760 έως 1300 και τέλος για όλα τα μήκη κύματος, 350 έως 1365 (ισάξιο του μέσου όρου ανά ποικιλία). Οι τρεις μεταβλητές που προέκυψαν εξετάστηκαν ως προς την διασπορά τους και τελικά επιλέχθηκε η Rank350_1365. Η δεύτερη παραλλαγή αντιστοιχεί σε κάθε κατηγορία έναν αριθμό χρησιμοποιώντας το δυαδικό σύστημα.

Ακολουθεί πίνακας με τους μέσους όρους και την αύξουσα σειρά των ομαδοποιημένων ποσοστών ανάκλασης για διαστήματα των μηκών κύματος 550-690, 760-1300, 350-1365.

Πίνακας 2 Χαρακτηριστικά ομαδοποιημένων ποσοστών ανακλάσεων

VarAb	Mean 550-690	Mean 760-1300	Mean 350-1365	Rank 550-690-1300	Rank 760-1300	Rank 350-1365
MOS	6.777	48.611	32.312	1	3	1
GOU	7.597	48.145	32.322	2	2	2
MEL	9.162	48.104	32.816	3	1	3
SAM	10.476	51.033	35.340	5	5	4
ASP	9.423	52.135	35.347	4	6	5
LGI	13.182	50.966	36.143	10	4	6
VOL	12.316	52.494	36.679	8	7	7
ASI	12.369	53.449	37.341	9	8	8
PAT	12.093	54.954	37.861	6	9	9
MAV	12.215	56.998	39.472	7	10	10

Τέλος, ακολουθεί το διάγραμμα με τον μέσο όρο των ποσοστών ανάκλασης ανά ποικιλία και ανά μήκος κύματος.



Εικόνα 1 Καμπύλη μέσων όρων ποσοστών ανάκλασης ανά ποικιλία

Στη συνέχεια έγινε κανονικοποίηση των ανεξάρτητων μεταβλητών B350-B1365 ώστε οι παρατηρήσεις να ανήκουν στο διάστημα $[0,1]$ και να βελτιωθεί η απόδοση του αλγορίθμου. Η διαδικασία κανονικοποίησης λαμβάνει σαν είσοδο ένα σύνολο τιμών, αντιστοιχεί στην μέγιστη την παρατήρηση του συνόλου την τιμή 1 και στην ελάχιστη την τιμή 0. Στη συνέχεια αντιστοιχεί όλες τις ενδιάμεσες τιμές με τιμές που ανήκουν στο διάστημα $[0,1]$ αναλογικά με τη θέση τους ως προς το μέγιστο και το ελάχιστο (Εξίσωση 1). Στα πλαίσια της παρούσας εργασίας, η διαδικασία που υλοποιήθηκε διαφοροποιείται ως προς την συνηθισμένη ως εξής: αντί να ληφθεί υπόψη κάθε φορά μια ολόκληρη στήλη (ανεξάρτητη μεταβλητή) ξεχωριστά, οι τιμές διαχωρίζονται και ανά ποικιλία και ύστερα εκτελείται η κανονικοποίηση σε κάθε υποσύνολο χωριστά.

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \#(1)$$

3.4 Ανάλυση δεδομένων

Σε αυτό το σημείο, γίνεται διαχωρισμός του συνόλου δεδομένων, σε δύο υποσύνολα. Το πρώτο θα χρησιμοποιηθεί για την εκπαίδευση του αλγορίθμου και το δεύτερο για τον έλεγχο ακρίβειας πρόβλεψης του εκπαιδευμένου μοντέλου. Συγκεκριμένα, η διαμέριση του αρχικού συνόλου δεδομένων έγινε με την αναλογία 66-33 (66% του συνόλου χρησιμοποιήθηκε για trainset και το 33% για testset). Η συγκεκριμένη αναλογία επιλέχθηκε

διότι διευκόλυνε τον διαχωρισμό του συνόλου. Έτσι, τα πρώτα δύο φυτά ανά ποικιλία εντάχθηκαν στο σύνολο εκπαίδευσης ενώ το τελευταίο στο σύνολο επικύρωσης.

Το υπό μελέτη πρόβλημα είναι ένα πρόβλημα παλινδρόμησης, αφού η κατηγορική εξαρτημένη μεταβλητή μετασχηματίστηκε σε αριθμητική με τη μέθοδο που αναλύθηκε στην προηγούμενη υποενότητα. Όπως αναφέρθηκε και στην προηγούμενη ενότητα, οι ανεξάρτητες μεταβλητές του συνόλου είναι 1016. Η μεγάλη διαστατικότητα του υπό μελέτη συνόλου καθιστά την επεξεργασία του από κάποιον αλγόριθμο παλινδρόμησης αδύνατη. Κατά συνέπεια πρέπει να γίνει μείωση της διάστασης των μεταβλητών και ύστερα να υλοποιηθεί ο αλγόριθμος παλινδρόμησης. Σε αυτό το πλαίσιο επιλέχθηκε ο αλγόριθμος NIPALS (Nonlinear Iterative Partial Least Squares) ο οποίος υλοποιεί επαναληπτικά μια σειρά από παραγοντοποιήσεις του αρχικού πίνακα ώστε να φτιάξει ένα νέο σύνολο από κύριες συνιστώσες (extracted factors) που μπορούν να αντικαταστήσουν τις αρχικές ανεξάρτητες μεταβλητές με τον βέλτιστο δυνατό τρόπο. Στόχος του αλγορίθμου είναι με τον ελάχιστο δυνατό αριθμό κύριων συνιστωσών να μεγιστοποιηθεί η επεξηγηματικότητα της εξαρτημένης μεταβλητής y .

Στα πλαίσια της παρούσας εργασίας ο αλγόριθμος NIPALS εφαρμόστηκε σε 6 παραλλαγές του αρχικού συνόλου δεδομένων οι οποίες περιγράφονται παρακάτω:

1. Αντικατάσταση της εξαρτημένης μεταβλητής με την κατάταξη που έχουν οι ποικιλίες με βάση τον υπολογισμό Mean350-1365 (χρήση του Rank350_1365)
 - a. Χρήση όλου του συνόλου δεδομένων
 - b. Χρήση μόνο των φύλλων που βρίσκονται στην κορυφή του φυτού
 - c. Χρήση μόνο των φύλλων που βρίσκονται στην μέση του φυτού
2. Αντικατάσταση της εξαρτημένης μεταβλητής με δυαδική κωδικοποίηση
 - a. Χρήση όλου του συνόλου δεδομένων
 - b. Χρήση μόνο των φύλλων που βρίσκονται στην κορυφή του φυτού
 - c. Χρήση μόνο των φύλλων που βρίσκονται στην μέση του φυτού

Για τις παραπάνω παραλλαγές αρχικά υπολογίστηκαν οι πρώτες 15 κύριες συνιστώσες και στη συνέχεια χρησιμοποιήθηκε το κριτήριο RootMeanPRESS ώστε να βρεθεί το βέλτιστο πλήθος κύριων συνιστωσών. Το συγκεκριμένο κριτήριο υπολογίζει τη ρίζα του μέσου σφάλματος. Ο τελικός αριθμός των extracted factors που επιλέγονται είναι αυτός για τον οποίο το p -value γίνεται μεγαλύτερο του 0.1. Στην επόμενη ενότητα ακολουθούν τα αποτελέσματα που πήραμε για κάθε μια από τις παραλλαγές που εξετάστηκαν.

4 Αποτελέσματα & Συζήτηση

4.1 Χρήση του Rank350_1365 σε όλο το σύνολο δεδομένων (παραλλαγή 1α)

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα του NIPALS αλγόριθμου για τους πρώτους 15 extracted factors. Στη στήλη Model Effects/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors (μέχρι την γραμμή που κοιτάμε) ως προς τις αρχικές ανεξάρτητες μεταβλητές. Στη στήλη Dependent Variables/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors (μέχρι την γραμμή που κοιτάμε) ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε όσο αυξάνουμε τον αριθμό των extracted factors, τόσο ανεβαίνουν οι επιμέρους επεξηγηματικότητες.

Πίνακας 3 Επεξηγηματικότητα 1α

Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	62.2546	62.2546	7.4058	7.4058
2	7.2333	69.4879	3.9371	11.3429
3	26.3018	95.7897	1.4218	12.7647
4	1.7225	97.5121	14.3672	27.1318
5	0.539	98.0511	9.5224	36.6542
6	0.641	98.6921	5.1267	41.7809
7	0.2603	98.9524	7.3263	49.1072
8	0.2248	99.1771	5.5917	54.6989
9	0.278	99.4551	2.6223	57.3212
10	0.1738	99.6289	3.1295	60.4507
11	0.0556	99.6844	7.0905	67.5412
12	0.0304	99.7148	7.8627	75.4039
13	0.0563	99.7711	2.8685	78.2725
14	0.06	99.8311	1.3872	79.6597
15	0.0166	99.8477	2.9775	82.6372

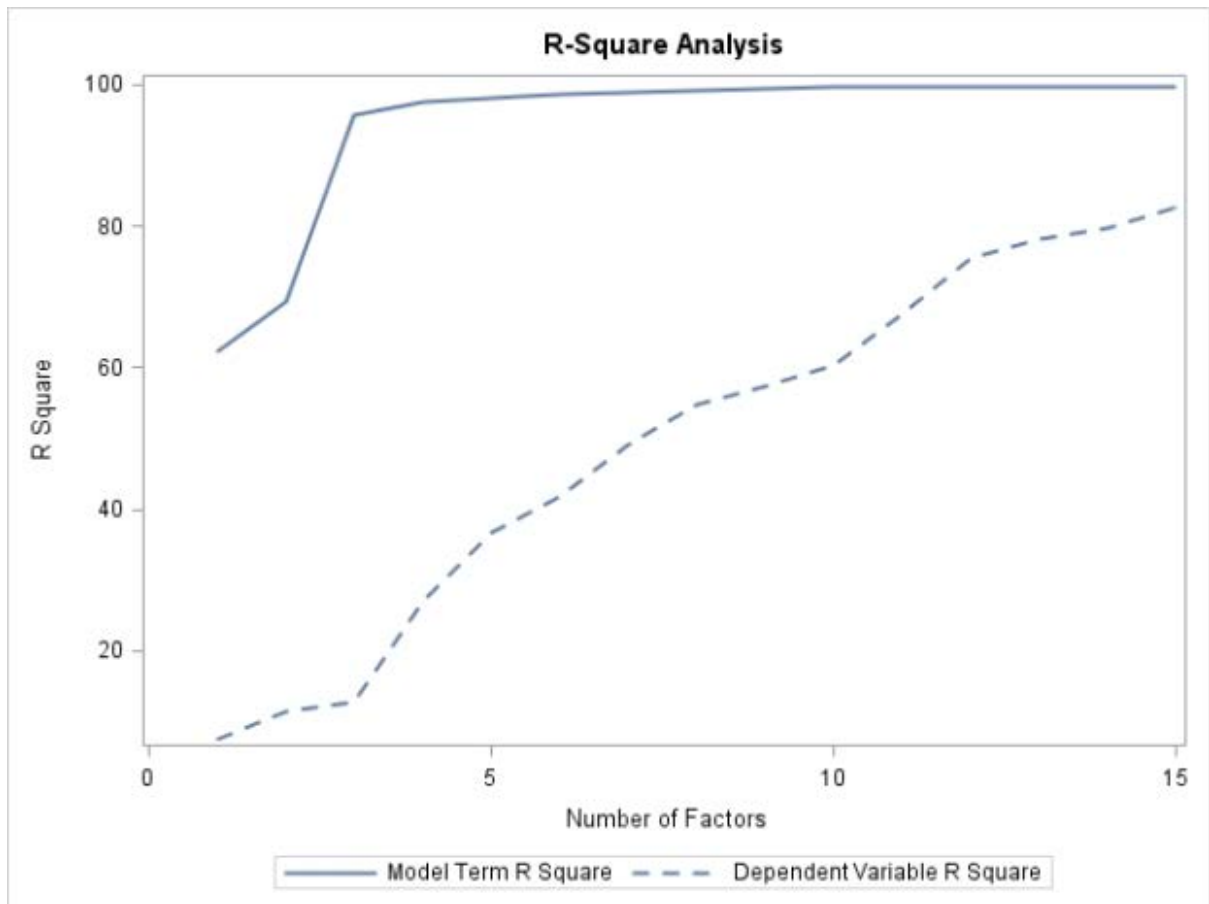
Στον πίνακα που ακολουθεί, στο 1^ο μέρος βλέπουμε το κριτήριο Root Mean PRESS για τους πρώτους 15 extracted factors, ενώ στο 2^ο βλέπουμε το βέλτιστο πλήθος factors που έχουν επιλεγεί βάσει του κριτηρίου.

Πίνακας 4 Root Mean PRESS 1α

Test Set Validation for the Number of Extracted Factors		
Number of Extracted Factors	Root Mean PRESS	Prob > PRESS
0	0.992779	<.0001
1	1.019903	<.0001
2	0.990799	<.0001
3	0.941851	<.0001
4	0.912229	0.005
5	0.841321	0.014
6	0.832405	0.015
7	0.860612	0.005
8	0.856615	0.006
9	0.808091	0.013
10	0.801726	0.008
11	0.784555	0.002
12	0.668873	1
13	0.69026	0.212
14	0.705549	0.172
15	0.781348	0.032
Minimum root mean PRESS		0.6689
Minimizing number of factors		12
Smallest number of factors with $p > 0.1$		12

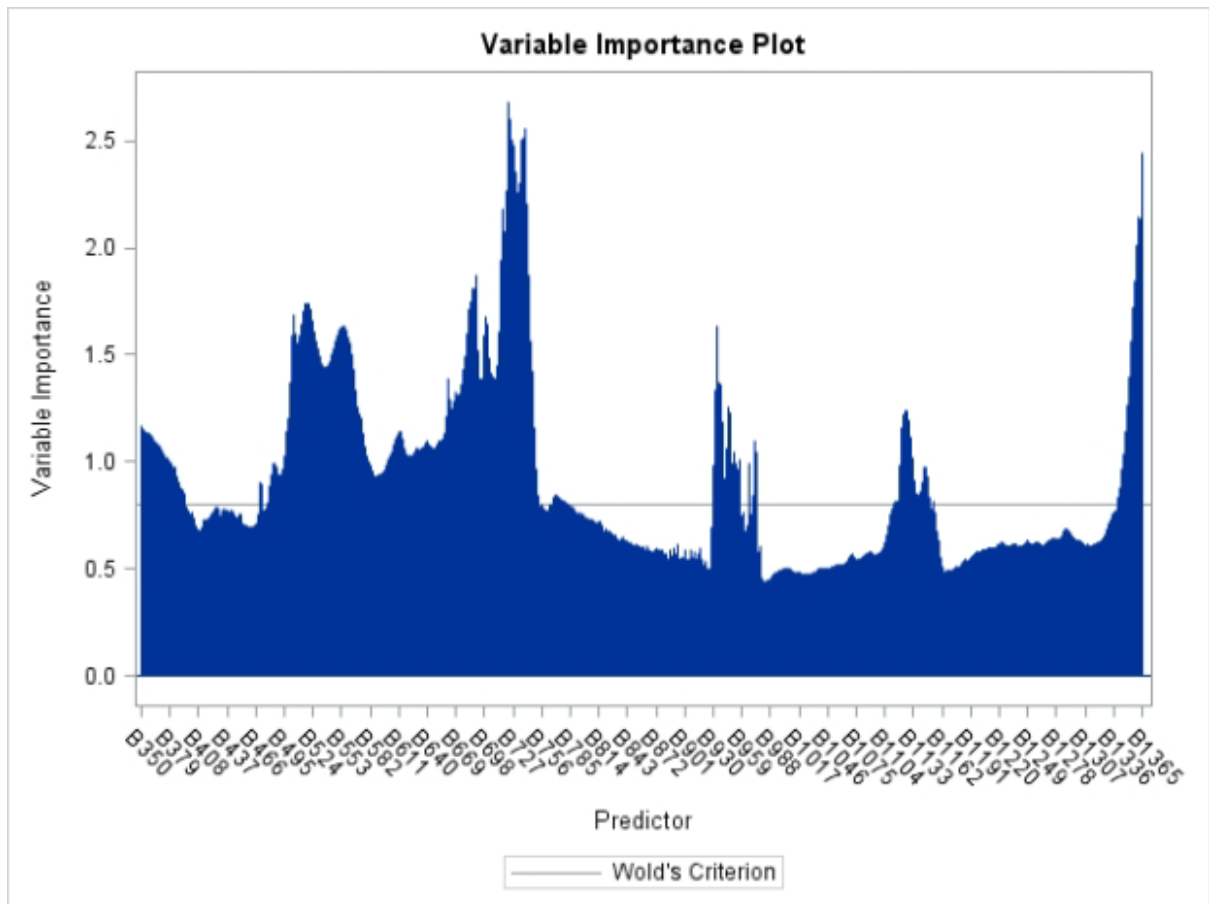
Τα αποτελέσματα του παραπάνω πίνακα επαληθεύουν τον προηγούμενο, αφού όπως βλέπουμε η επεξηγηματικότητα των πρώτων 12 factors είναι 75.4% ενώ των πρώτων 15 είναι 82.6%. Η διαφορά ανάμεσα σε αυτές τις δύο τιμές δεν είναι μεγάλη συγκριτικά με τη διαφορά που δημιουργείται αν από τους 12 πρώτους factors αφαιρέσουμε τον 12^ο.

Ακολουθεί το διάγραμμα του Πίνακας 3, όπου από την διακεκομμένη γραμμή που αντιστοιχεί στην επεξηγηματικότητα της εξαρτημένης μεταβλητής, μπορούμε να παρατηρήσουμε ότι μετά το σημείο 12, η κλίση μειώνεται.



Εικόνα 2 R-Square Analysis 1a

Τέλος, στο σχήμα που ακολουθεί βλέπουμε τη σημαντικότητα των αρχικών 1016 ανεξάρτητων μεταβλητών ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε τα ποσοστά ανάκλασης στο κάθε μήκος κύματος συσχετίζονται διαφορετικά στην άμπελο έχοντας κάποιες ζώνες με πολύ κοντινή και χαμηλή σημαντικότητα αλλά μεγαλύτερο ενδιαφέρον έχουν τα μήκη κύματος στα οποία έχουμε τις κορυφές με τις μεγαλύτερες σημαντικότητες που θα διαμορφώσουν τα factor μας.



Εικόνα 3 Variable Importance Plot 1a

4.2 Χρήση του Rank350_1365 μόνο στα φύλλα που βρίσκονται στην κορυφή του φυτού (παραλλαγή 1β)

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα του NIPALSαλγόριθμου για τους πρώτους 15 extracted factors. Στη στήλη Model Effects/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors(μέχρι την γραμμή που κοιτάμε) ως προς τις αρχικές ανεξάρτητες μεταβλητές. Στη στήλη Dependent Variables/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors(μέχρι την γραμμή που κοιτάμε) ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε όσο αυξάνουμε τον αριθμό των extracted factors, τόσο ανεβαίνουν οι επιμέρους επεξηγηματικότητες.

Πίνακας 5 Επεξηγηματικότητα 1β

Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	62.5611	62.5611	13.4700	13.4700
2	23.2846	85.8457	8.8922	22.3622
3	10.5168	96.3624	7.8660	30.2282
4	0.9163	97.2787	18.1090	48.3372
5	0.6769	97.9556	10.4608	58.7980
6	0.7514	98.7070	5.2812	64.0792
7	0.2972	99.0042	5.3863	69.4654
8	0.3356	99.3398	2.8951	72.3606
9	0.2539	99.5937	2.8957	75.2563
10	0.0903	99.6840	4.6081	79.8643
11	0.0622	99.7462	4.4803	84.3446
12	0.0337	99.7799	3.2700	87.6146
13	0.0504	99.8303	1.3328	88.9475
14	0.0336	99.8639	1.6908	90.6383
15	0.0420	99.9059	0.9580	91.5963

Στον πίνακα που ακολουθεί, στο 1^ο μέρος βλέπουμε το κριτήριο Root Mean PRESS για τους πρώτους 15 extracted factors, ενώ στο 2^ο βλέπουμε το βέλτιστο πλήθος factors που έχουν επιλεχθεί βάσει του κριτηρίου.

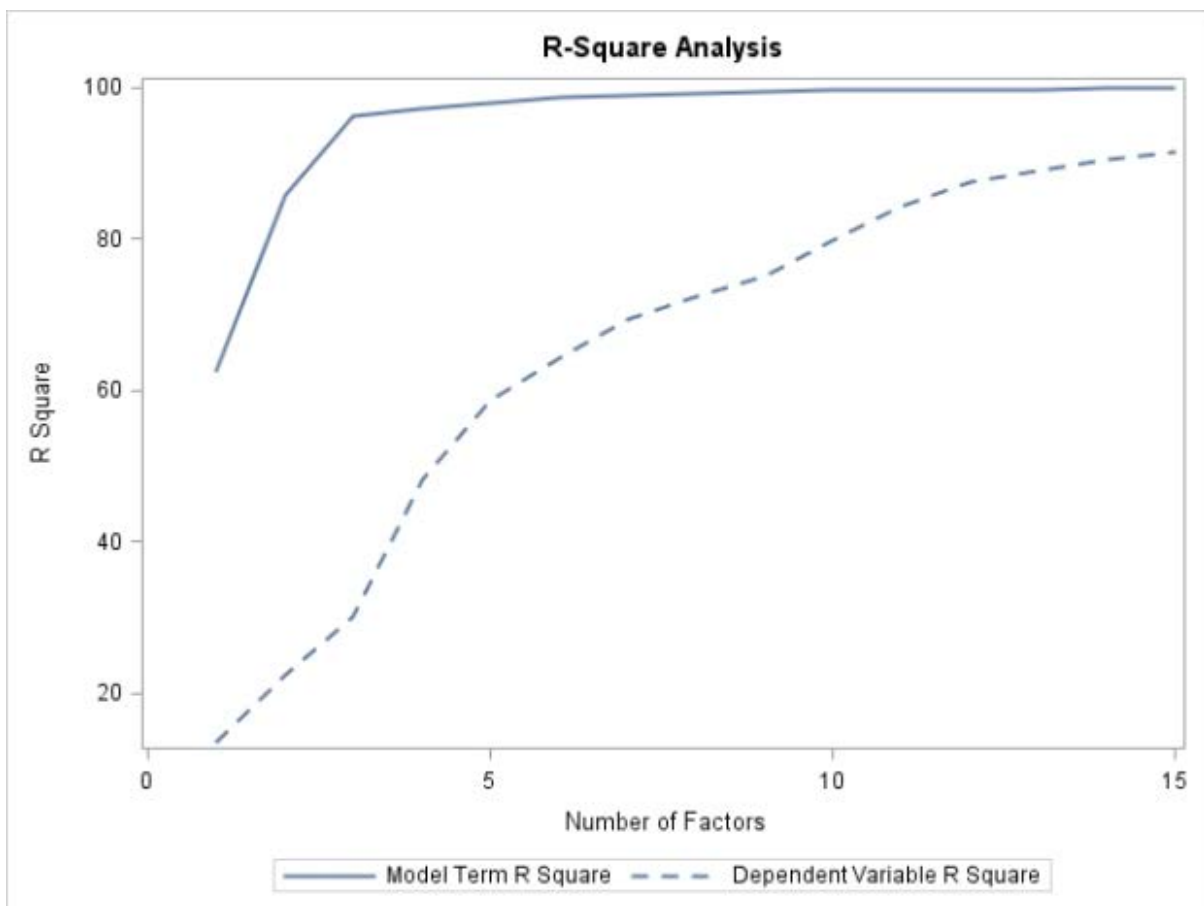
Πίνακας 6 Root Mean PRESS 1β

Test Set Validation for the Number of Extracted Factors		
Number of Extracted Factors	Root Mean PRESS	Prob > PRESS
0	0.991632	<.0001
1	0.991759	<.0001
2	0.897698	<.0001
3	0.862001	0.0040
4	0.824077	0.0100
5	0.768585	0.0250
6	0.707772	0.0330
7	0.699729	0.0170
8	0.708666	0.0170
9	0.675222	0.0270
10	0.564124	0.2200
11	0.537335	0.2480
12	0.551414	0.1690
13	0.515852	0.4030
14	0.526144	0.1890
15	0.507914	10.000
Minimum root mean PRESS		0.5079
Minimizing number of factors		15
Smallest number of factors with p > 0.1		10

Τα αποτελέσματα του παραπάνω πίνακα επαληθεύουν τον προηγούμενο, αφού όπως βλέπουμε η επεξηγηματικότητα των πρώτων 10 factors είναι 79.9% ενώ των πρώτων 15 είναι

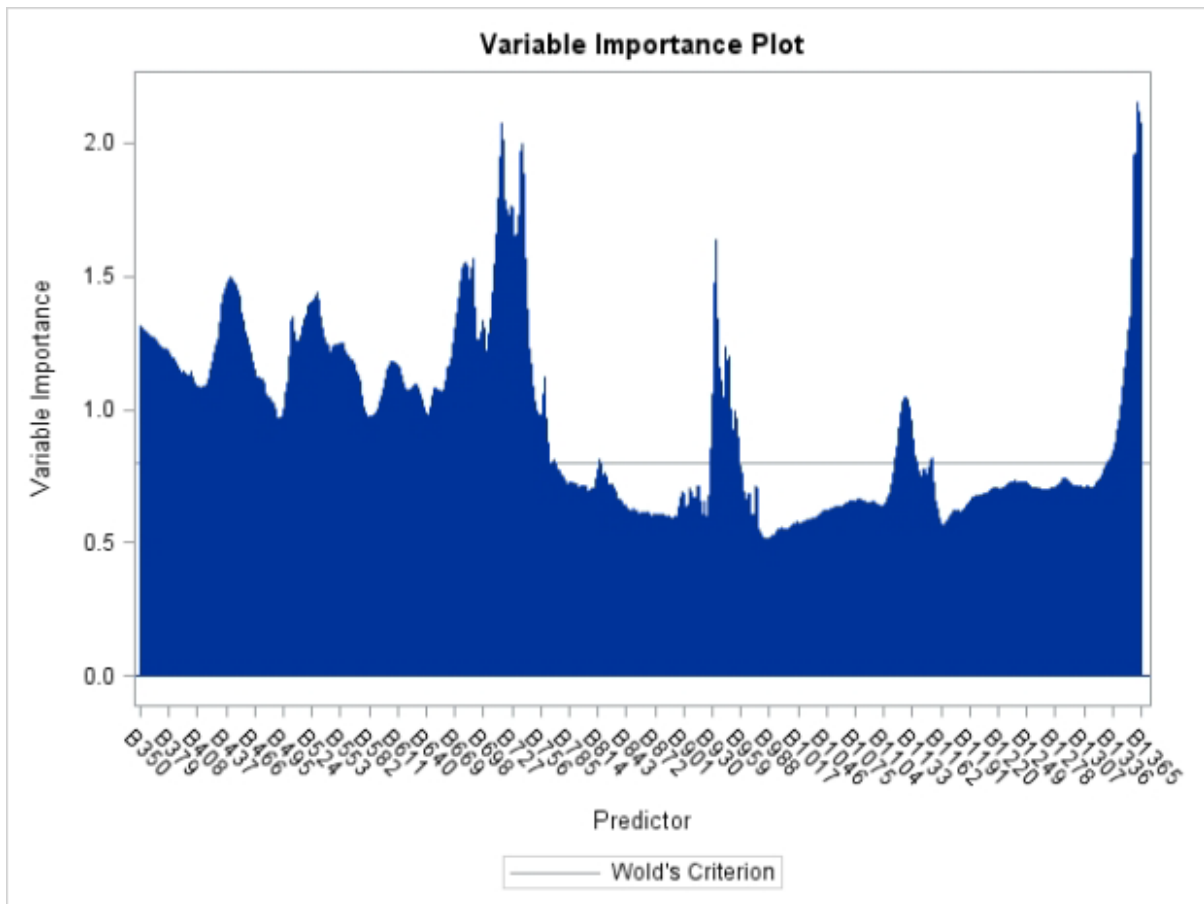
91.6%. Η διαφορά ανάμεσα σε αυτές τις δύο τιμές δεν είναι μεγάλη συγκριτικά με τη διαφορά που δημιουργείται αν από τους 10 πρώτους factors αφαιρέσουμε τον 10^ο.

Ακολουθεί το διάγραμμα του Πίνακας 5, όπου από την διακεκομμένη γραμμή που αντιστοιχεί στην επεξηγηματικότητα της εξαρτημένης μεταβλητής, μπορούμε να παρατηρήσουμε ότι μετά το σημείο 10, η κλίση μειώνεται.



Εικόνα 4 R-Square Analysis 1b

Τέλος, στο σχήμα που ακολουθεί βλέπουμε τη σημαντικότητα των αρχικών 1016 ανεξάρτητων μεταβλητών ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε τα ποσοστά ανάκλασης στο κάθε μήκος κύματος συσχετίζονται διαφορετικά στην άμπελο έχοντας κάποιες ζώνες με πολύ κοντινή και χαμηλή σημαντικότητα αλλά μεγαλύτερο ενδιαφέρον έχουν τα μήκη κύματος στα οποία έχουμε τις κορυφές με τις μεγαλύτερες σημαντικότητες που θα διαμορφώσουν τα factor μας .



Εικόνα 5 Variable Importance Plot 1b

4.3 Χρήση του Rank350_1365 μόνο στα φύλλα των φύλλων που βρίσκονται στην μέση του φυτού (παραλλαγή 1γ)

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα του NIPALS αλγόριθμου για τους πρώτους 15 extracted factors. Στη στήλη Model Effects/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors (μέχρι την γραμμή που κοιτάμε) ως προς τις αρχικές ανεξάρτητες μεταβλητές. Στη στήλη Dependent Variables/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors (μέχρι την γραμμή που κοιτάμε) ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε όσο αυξάνουμε τον αριθμό των extracted factors, τόσο ανεβαίνουν οι επιμέρους επεξηγηματικότητες.

Πίνακας 7 Επεξηγηματικότητα 1γ

Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	69.1501	69.1501	6.8721	6.8721
2	16.4240	85.5741	3.6939	10.5660
3	3.4507	89.0247	17.6299	28.1959
4	7.9888	97.0135	3.1847	31.3806
5	0.8221	97.8356	18.5375	49.9181
6	0.5605	98.3961	11.0180	60.9360
7	0.3540	98.7501	8.0932	69.0292
8	0.3136	99.0637	4.9990	74.0282
9	0.3185	99.3823	3.8286	77.8568
10	0.0622	99.4444	7.3191	85.1759
11	0.0579	99.5024	4.2335	89.4095
12	0.0756	99.5780	2.3528	91.7623
13	0.1004	99.6783	1.3455	93.1078
14	0.0847	99.7631	1.1584	94.2662
15	0.0686	99.8316	0.8650	95.1312

Στον πίνακα που ακολουθεί, στο 1^ο μέρος βλέπουμε το κριτήριο Root Mean PRESS για τους πρώτους 15 extracted factors, ενώ στο 2^ο βλέπουμε το βέλτιστο πλήθος factors που έχουν επιλεγεί βάσει του κριτηρίου.

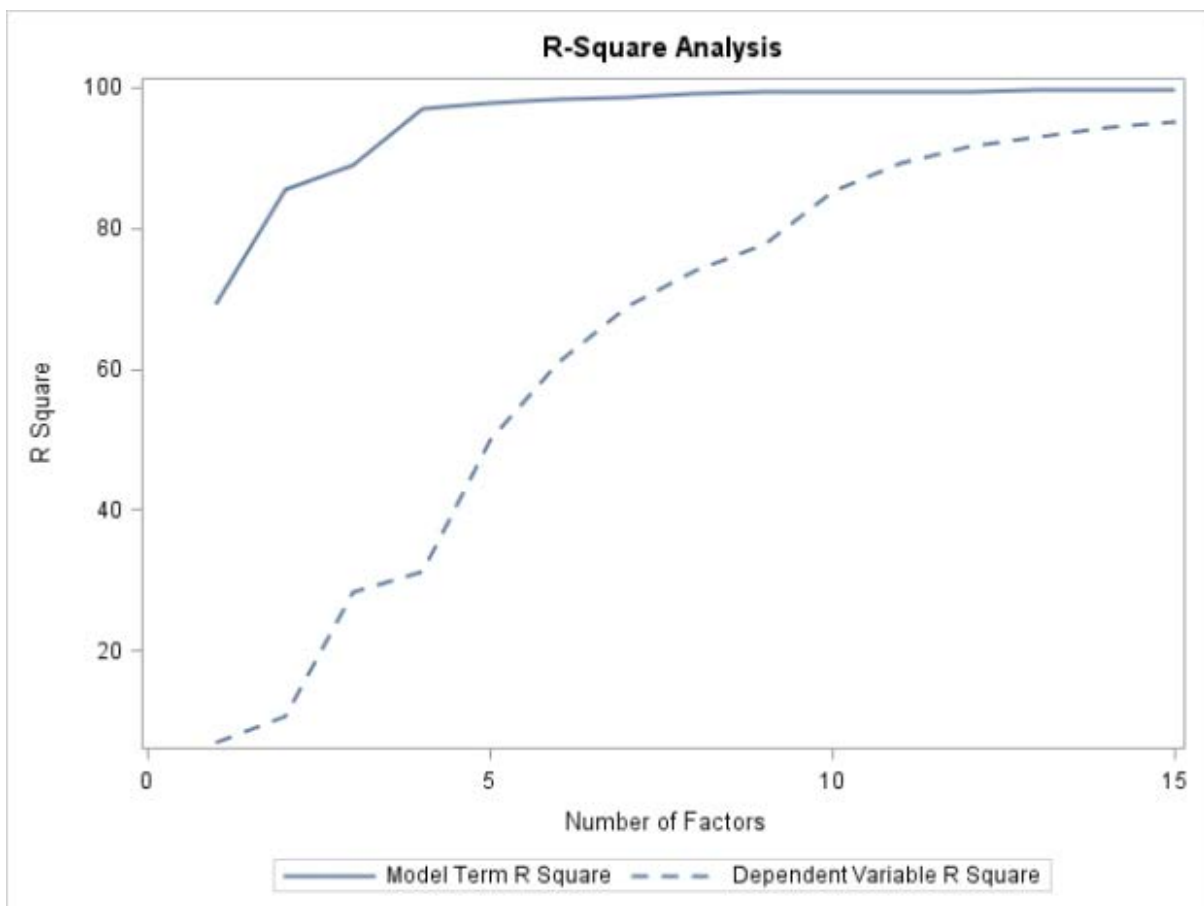
Πίνακας 8 Root Mean PRESS 1γ

Test Set Validation for the Number of Extracted Factors		
Number of Extracted Factors	Root Mean PRESS	Prob > PRESS
0	0.985511	<.0001
1	1.056.501	<.0001
2	0.903961	<.0001
3	0.860397	<.0001
4	0.774218	0.0020
5	0.700873	0.0030
6	0.860912	0.0060
7	0.750227	0.0070
8	0.559634	0.1520
9	0.514067	0.2380
10	0.561951	0.0530
11	0.559561	0.0140
12	0.481018	0.1480
13	0.451812	10.000
14	0.491349	0.1350
15	0.503492	0.2230
Minimum root mean PRESS		0.4518
Minimizing number of factors		13
Smallest number of factors with p > 0.1		8

Τα αποτελέσματα του παραπάνω πίνακα επαληθεύουν τον προηγούμενο, αφού όπως βλέπουμε η επεξηγηματικότητα των πρώτων 8 factors είναι 74% ενώ των πρώτων 15 είναι

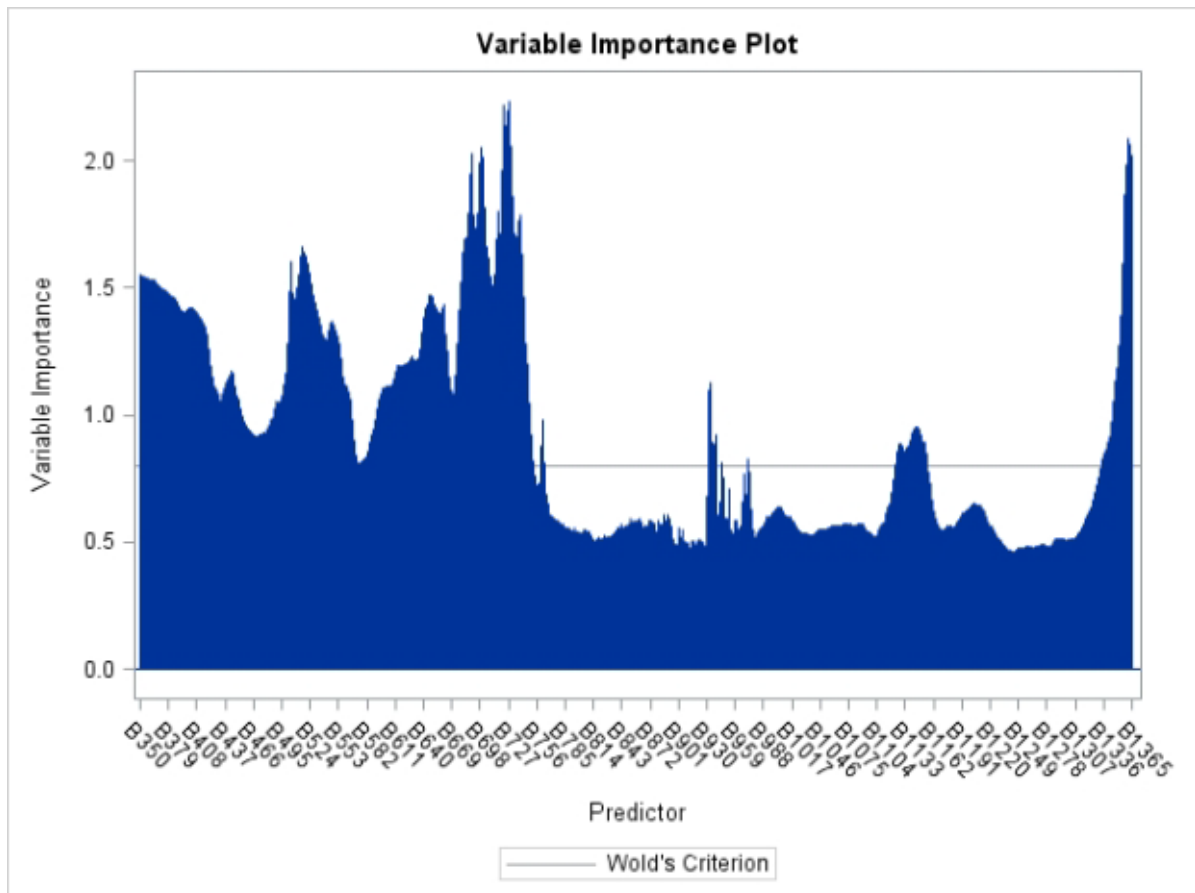
95,1%. Η διαφορά ανάμεσα σε αυτές τις δύο τιμές δεν είναι μεγάλη συγκριτικά με τη διαφορά που δημιουργείται αν από τους 8 πρώτους factors αφαιρέσουμε τον 8°.

Ακολουθεί το διάγραμμα του Πίνακας 7, όπου από την διακεκομμένη γραμμή που αντιστοιχεί στην επεξηγηματικότητα της εξαρτημένης μεταβλητής, μπορούμε να παρατηρήσουμε ότι μετά το σημείο 8, η κλίση μειώνεται.



Εικόνα 6 R-Square Analysis 1c

Τέλος, στο σχήμα που ακολουθεί βλέπουμε τη σημαντικότητα των αρχικών 1016 ανεξάρτητων μεταβλητών ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε τα ποσοστά ανάκλασης στο κάθε μήκος κύματος συσχετίζονται διαφορετικά στην άμπελο έχοντας κάποιες ζώνες με πολύ κοντινή και χαμηλή σημαντικότητα αλλά μεγαλύτερο ενδιαφέρον έχουν τα μήκη κύματος στα οποία έχουμε τις κορυφές με τις μεγαλύτερες σημαντικότητες που θα διαμορφώσουν τα factor μας .



Εικόνα 7 Variable Importance Plot 1c

4.4 Χρήση της δυαδικής κωδικοποίησης σε όλο το σύνολο δεδομένων (παραλλαγή 2α)

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα του NIPALSαλγόριθμου για τους πρώτους 15 extracted factors. Στη στήλη Model Effects/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors(μέχρι την γραμμή που κοιτάμε) ως προς τις αρχικές ανεξάρτητες μεταβλητές. Στη στήλη Dependent Variables/Total βλέπουμε την επεξηγηματικότητα των πρώτων x factors(μέχρι την γραμμή που κοιτάμε) ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε όσο αυξάνουμε τον αριθμό των extracted factors, τόσο ανεβαίνουν οι επιμέρους επεξηγηματικότητες.

Πίνακας 9 Επεξηγηματικότητα 2α

Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	45.8988	45.8988	0.8743	0.8743
2	40.6868	86.5856	1.8297	2.7040
3	6.6882	93.2738	9.5907	12.2947
4	4.5084	97.7822	3.5472	15.8419
5	0.5319	98.3141	13.4557	29.2976
6	0.2378	98.5519	23.5471	52.8447
7	0.3860	98.9378	10.2143	63.0590
8	0.2844	99.2222	6.1358	69.1948
9	0.2028	99.4251	2.8773	72.0721
10	0.1509	99.5759	3.1899	75.2620
11	0.0969	99.6728	2.8590	78.1210
12	0.0396	99.7124	4.6668	82.7878
13	0.0712	99.7836	1.7601	84.5480
14	0.0324	99.8160	3.4025	87.9505
15	0.0244	99.8404	1.6840	89.6345

Στον πίνακα που ακολουθεί, στο 1^ο μέρος βλέπουμε το κριτήριο Root Mean PRESS για τους πρώτους 15 extracted factors, ενώ στο 2^ο βλέπουμε το βέλτιστο πλήθος factors που έχουν επιλεγεί βάσει του κριτηρίου.

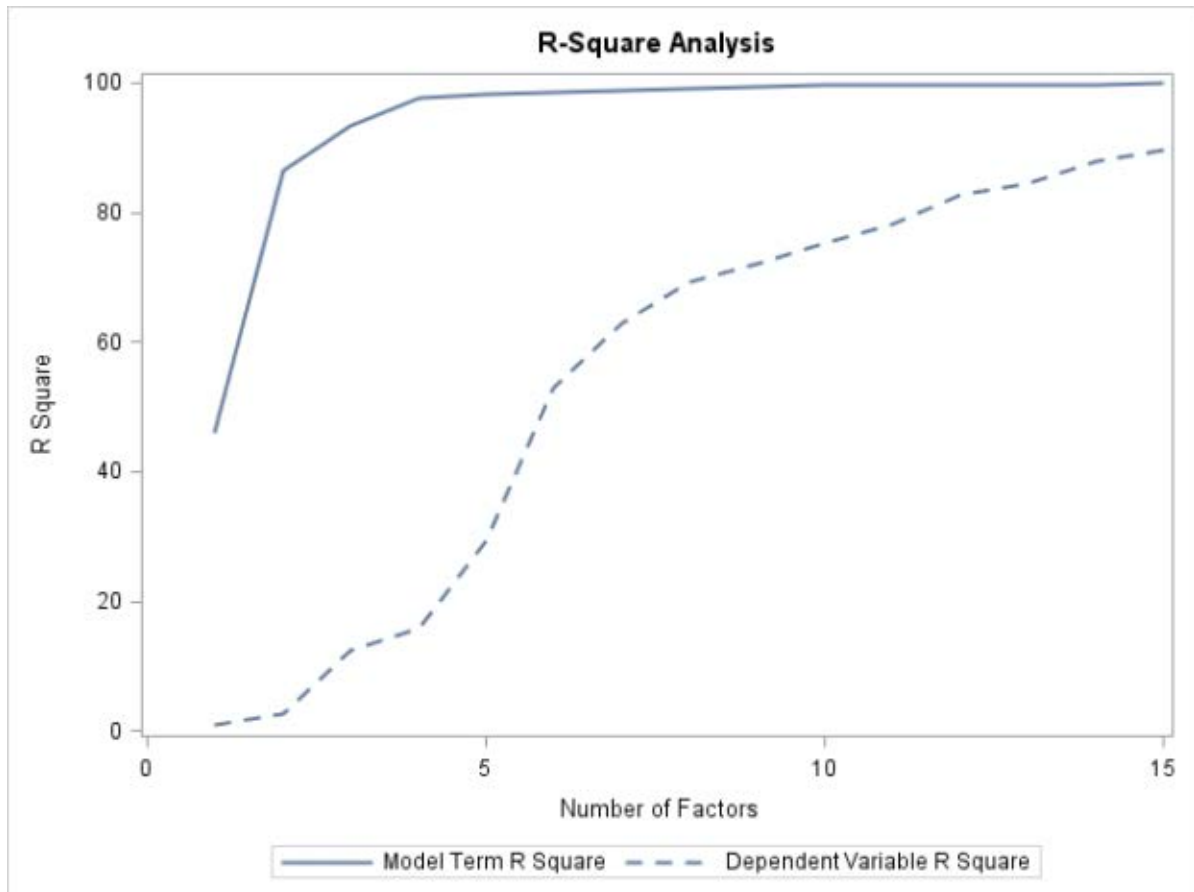
Πίνακας 10 Root Mean PRESS 2α

Test Set Validation for the Number of Extracted Factors		
Number of Extracted Factors	Root Mean PRESS	Prob > PRESS
0	1.002.106	0.0010
1	0.997397	0.0010
2	0.974261	0.0010
3	0.910431	0.0030
4	0.917998	0.0030
5	0.896924	0.0050
6	0.825487	0.0570
7	0.743279	0.1410
8	0.78535	0.0310
9	0.76175	0.0580
10	0.776804	0.0260
11	0.766054	0.0260
12	0.746022	0.0500
13	0.749679	0.0560
14	0.709922	0.0080
15	0.657882	1.0000
Minimum root mean PRESS		0.6579
Minimizing number of factors		15
Smallest number of factors with p > 0.1		7

Τα αποτελέσματα του παραπάνω πίνακα επαληθεύουν τον προηγούμενο, αφού όπως βλέπουμε η επεξηγηματικότητα των πρώτων 7 factors είναι 63% ενώ των πρώτων 15 είναι

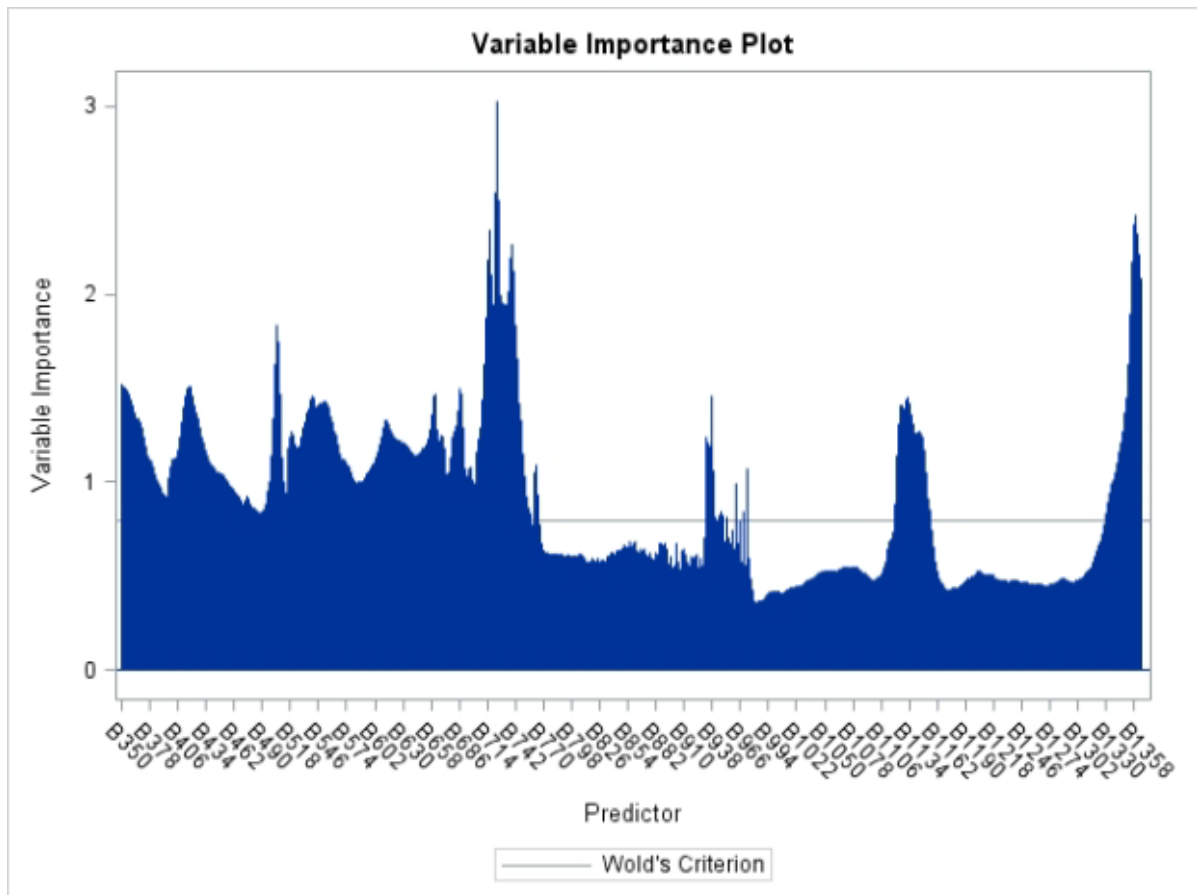
89.6%. Η διαφορά ανάμεσα σε αυτές τις δύο τιμές δεν είναι μεγάλη συγκριτικά με τη διαφορά που δημιουργείται αν από τους 7 πρώτους factors αφαιρέσουμε τον 7^ο.

Ακολουθεί το διάγραμμα του Πίνακας 9, όπου από την διακεκομμένη γραμμή που αντιστοιχεί στην επεξηγηματικότητα της εξαρτημένης μεταβλητής, μπορούμε να παρατηρήσουμε ότι μετά το σημείο 7, η κλίση μειώνεται.



Εικόνα 8 R-Square Analysis 2a

Τέλος, στο σχήμα που ακολουθεί βλέπουμε τη σημαντικότητα των αρχικών 1016 ανεξάρτητων μεταβλητών ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε τα ποσοστά ανάκλασης στο κάθε μήκος κύματος συσχετίζονται διαφορετικά στην άμπελο έχοντας κάποιες ζώνες με πολύ κοντινή και χαμηλή σημαντικότητα αλλά μεγαλύτερο ενδιαφέρον έχουν τα μήκη κύματος στα οποία έχουμε τις κορυφές με τις μεγαλύτερες σημαντικότητες που θα διαμορφώσουν τα factor μας .



Εικόνα 9 Variable Importance Plot 2a

4.5 Χρήση της δυαδικής κωδικοποίησης μόνο στα φύλλα που βρίσκονται στην κορυφή του φυτού (παραλλαγή 2β)

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα του NIPALSαλγόριθμου για τους πρώτους 15 extracted factors. Στη στήλη Model Effects/Totalβλέπουμε την επεξηγηματικότητα των πρώτων x factors(μέχρι την γραμμή που κοιτάμε) ως προς τις αρχικές ανεξάρτητες μεταβλητές. Στη στήλη Dependent Variables/Totalβλέπουμε την επεξηγηματικότητα των πρώτων x factors(μέχρι την γραμμή που κοιτάμε) ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε όσο αυξάνουμε τον αριθμό των extracted factors, τόσο ανεβαίνουν οι επιμέρους επεξηγηματικότητες.

Πίνακας 11 Επεξηγηματικότητα 2β

Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	42.8267	42.8267	4.4855	4.4855
2	45.4093	88.2360	3.2227	7.7082
3	5.4010	93.6370	17.6802	25.3884
4	3.9910	97.6280	9.0001	34.3885
5	0.4772	98.1052	15.8302	50.2187
6	0.4904	98.5956	7.9087	58.1273
7	0.5055	99.1011	5.4695	63.5968
8	0.2390	99.3401	10.1691	73.7660
9	0.2452	99.5853	4.9713	78.7372
10	0.0535	99.6388	5.0393	83.7765
11	0.0340	99.6728	6.0069	89.7835
12	0.0890	99.7618	1.5890	91.3725
13	0.0612	99.8230	1.2407	92.6132
14	0.0304	99.8534	2.2807	94.8939
15	0.0458	99.8993	0.8144	95.7083

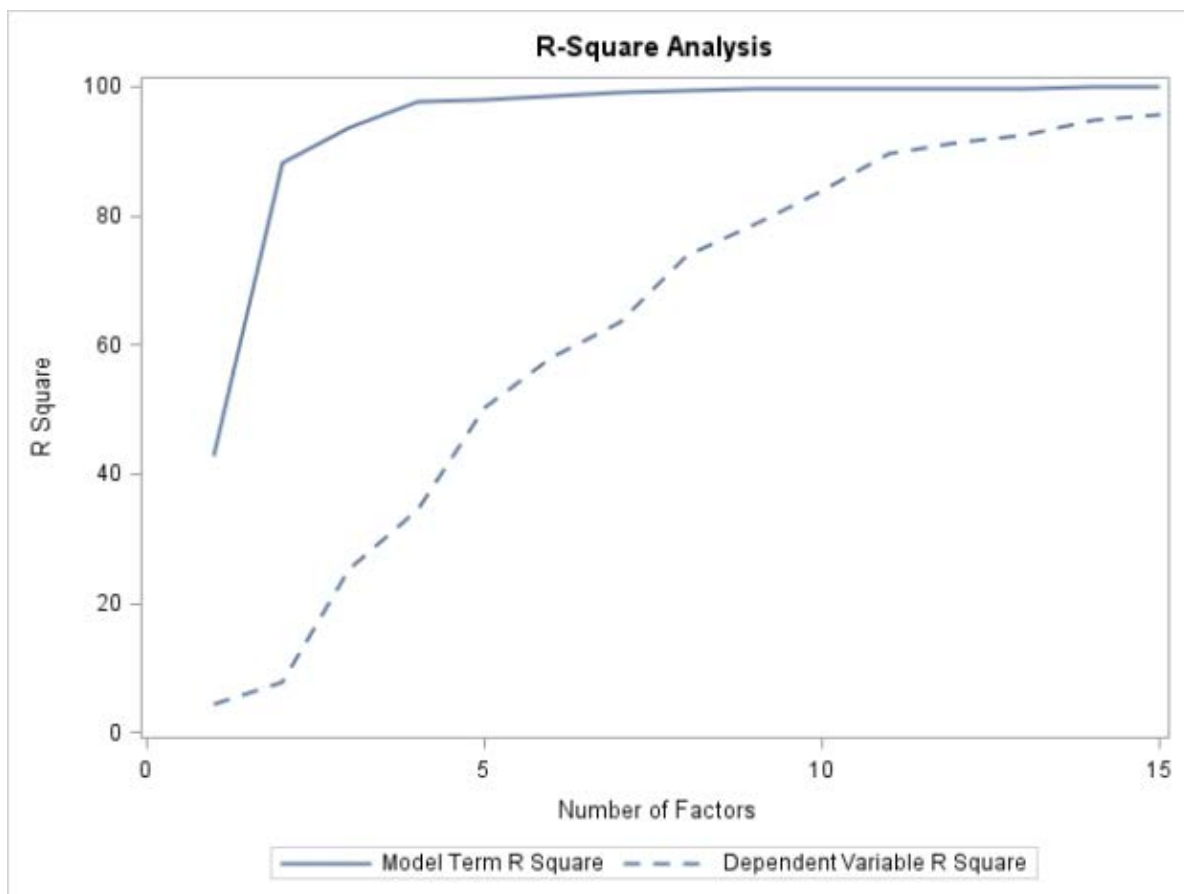
Στον πίνακα που ακολουθεί, στο 1^ο μέρος βλέπουμε το κριτήριο Root Mean PRESS για τους πρώτους 15 extracted factors, ενώ στο 2^ο βλέπουμε το βέλτιστο πλήθος factors που έχουν επιλεγεί βάσει του κριτηρίου.

Πίνακας 12 Root Mean PRESS 2β

Test Set Validation for the Number of Extracted Factors		
Number of Extracted Factors	RootMean PRESS	Prob> PRESS
0	0.991632	0.0010
1	0.985878	0.0030
2	0.970958	0.0040
3	0.886497	0.0640
4	0.90664	0.0360
5	0.815684	0.1060
6	0.733252	0.2360
7	0.74457	0.1880
8	0.761523	0.0410
9	0.766217	0.0160
10	0.73039	0.0460
11	0.747716	0.0140
12	0.743208	0.0070
13	0.716008	0.0080
14	0.646015	0.2310
15	0.629312	1.0000
Minimum root mean PRESS		0.6293
Minimizing number of factors		15
Smallest number of factors with p > 0.1		5

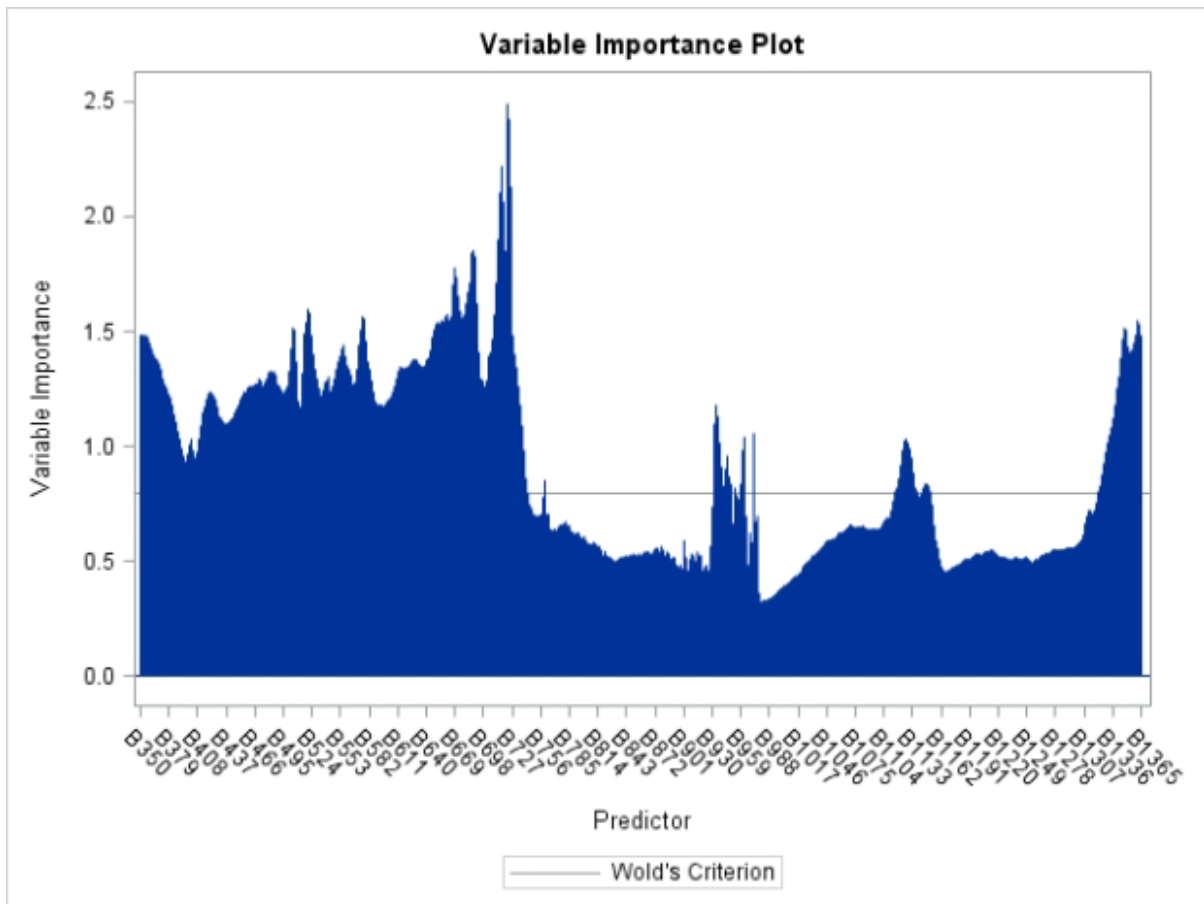
Τα αποτελέσματα του παραπάνω πίνακα επαληθεύουν τον προηγούμενο, αφού όπως βλέπουμε η επεξηγηματικότητα των πρώτων 5 factors είναι 50.2% ενώ των πρώτων 15 είναι 95.7%. Η διαφορά ανάμεσα σε αυτές τις δύο τιμές δεν είναι μεγάλη συγκριτικά με τη διαφορά που δημιουργείται αν από τους 5 πρώτους factors αφαιρέσουμε τον 5^ο.

Ακολουθεί το διάγραμμα του Πίνακας 11, όπου από την διακεκομμένη γραμμή που αντιστοιχεί στην επεξηγηματικότητα της εξαρτημένης μεταβλητής, μπορούμε να παρατηρήσουμε ότι μετά το σημείο 5, η κλίση μειώνεται.



Εικόνα 10 R-Square Analysis 2b

Τέλος, στο σχήμα που ακολουθεί βλέπουμε τη σημαντικότητα των αρχικών 1016 ανεξάρτητων μεταβλητών ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε τα ποσοστά ανάκλασης στο κάθε μήκος κύματος συσχετίζονται διαφορετικά στην άμπελο έχοντας κάποιες ζώνες με πολύ κοντινή και χαμηλή σημαντικότητα αλλά μεγαλύτερο ενδιαφέρον έχουν τα μήκη κύματος στα οποία έχουμε τις κορυφές με τις μεγαλύτερες σημαντικότητες που θα διαμορφώσουν τα factor μας .



Εικόνα 11 Variable Importance Plot 2b

4.6 Χρήση της δυαδικής κωδικοποίησης μόνο στα φύλλα που βρίσκονται στην μέση του φυτού (παραλλαγή 2γ)

Στον παρακάτω πίνακα βλέπουμε τα αποτελέσματα του NIPALSαλγόριθμου για τους πρώτους 15 extracted factors. Στη στήλη Model Effects/Totalβλέπουμε την επεξηγηματικότητα των πρώτων x factors (μέχρι την γραμμή που κοιτάμε) ως προς τις αρχικές ανεξάρτητες μεταβλητές. Στη στήλη Dependent Variables/Totalβλέπουμε την επεξηγηματικότητα των πρώτων x factors (μέχρι την γραμμή που κοιτάμε) ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε όσο αυξάνουμε τον αριθμό των extracted factors, τόσο ανεβαίνουν οι επιμέρους επεξηγηματικότητες.

Πίνακας 13 Επεξηγηματικότητα 2γ

Percent Variation Accounted for by Partial Least Squares Factors				
Number of Extracted Factors	Model Effects		Dependent Variables	
	Current	Total	Current	Total
1	25.6542	25.6542	19.1982	19.1982
2	61.9224	87.5766	2.0230	21.2212
3	4.8709	92.4475	10.1081	31.3292
4	4.9052	97.3526	7.7980	39.1273
5	0.6422	97.9948	10.0138	49.1411
6	0.3238	98.3186	12.4596	61.6007
7	0.5778	98.8965	4.9142	66.5149
8	0.1995	99.0959	11.1826	77.6975
9	0.2693	99.3653	3.1394	80.8369
10	0.0870	99.4523	2.3101	83.1470
11	0.0779	99.5302	2.7290	85.8760
12	0.0886	99.6188	2.8767	88.7527
13	0.0844	99.7032	2.0596	90.8123
14	0.0757	99.7788	1.0294	91.8416
15	0.0299	99.8088	1.9490	93.7906

Στον πίνακα που ακολουθεί, στο 1^ο μέρος βλέπουμε το κριτήριο Root Mean PRESS για τους πρώτους 15 extracted factors, ενώ στο 2^ο βλέπουμε το βέλτιστο πλήθος factors που έχουν επιλεγεί βάσει του κριτηρίου.

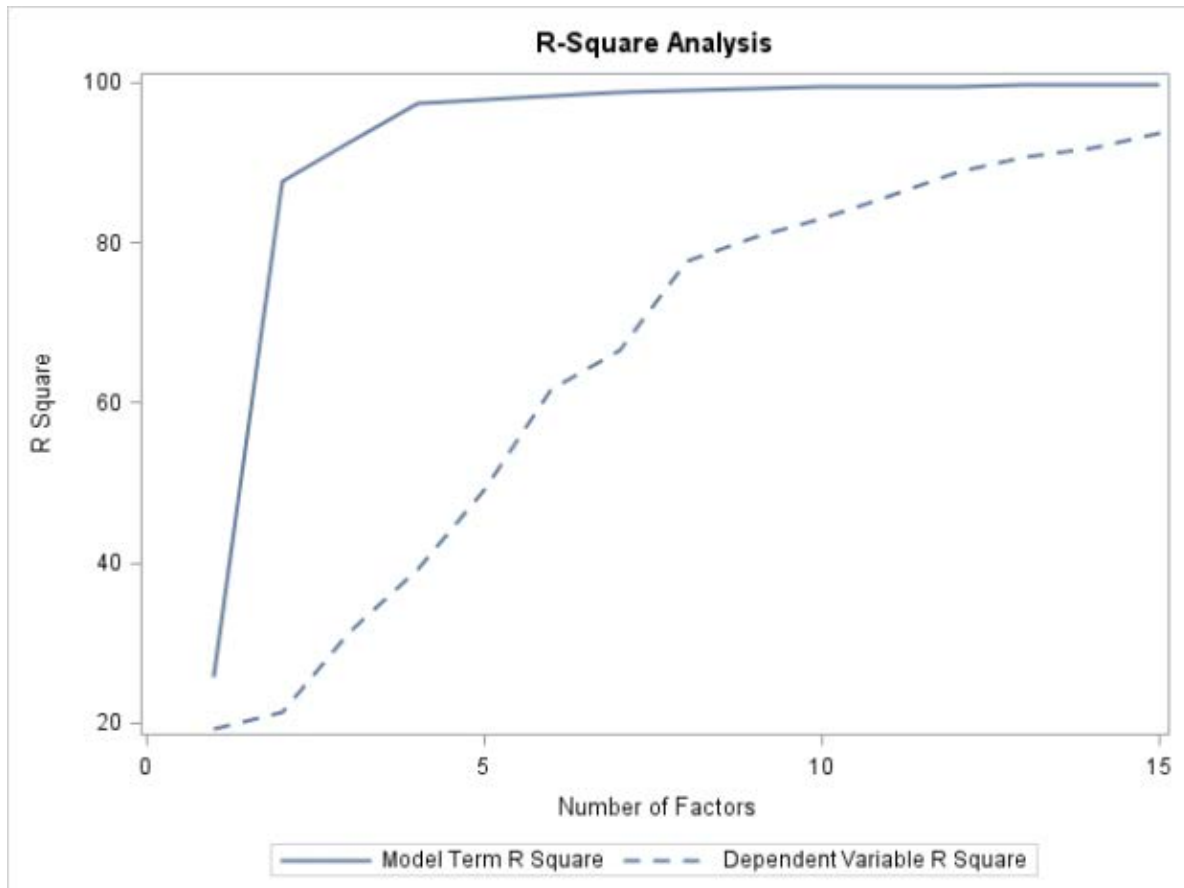
Πίνακας 14 Root Mean PRESS 2γ

Test Set Validation for the Number of Extracted Factors		
Number of Extracted Factors	RootMean PRESS	Prob> PRESS
0	1.004.659	<.0001
1	1.010.799	0.0010
2	1.030.118	<.0001
3	1.037.332	0.0010
4	0.979489	0.0030
5	1.016.239	0.0080
6	1.079.785	0.0080
7	0.881985	0.0250
8	0.762313	0.0610
9	0.633302	0.3960
10	0.613674	10.000
11	0.634134	0.2830
12	0.636883	0.3890
13	0.676286	0.2090
14	0.661803	0.2110
15	0.701966	0.1360
Minimum rootmean PRESS		0.6137
Minimizing number of factors		10
Smallest number of factors with p > 0.1		9

Τα αποτελέσματα του παραπάνω πίνακα επαληθεύουν τον προηγούμενο, αφού όπως βλέπουμε η επεξηγηματικότητα των πρώτων 9 factors είναι 80.8% ενώ των πρώτων 15 είναι

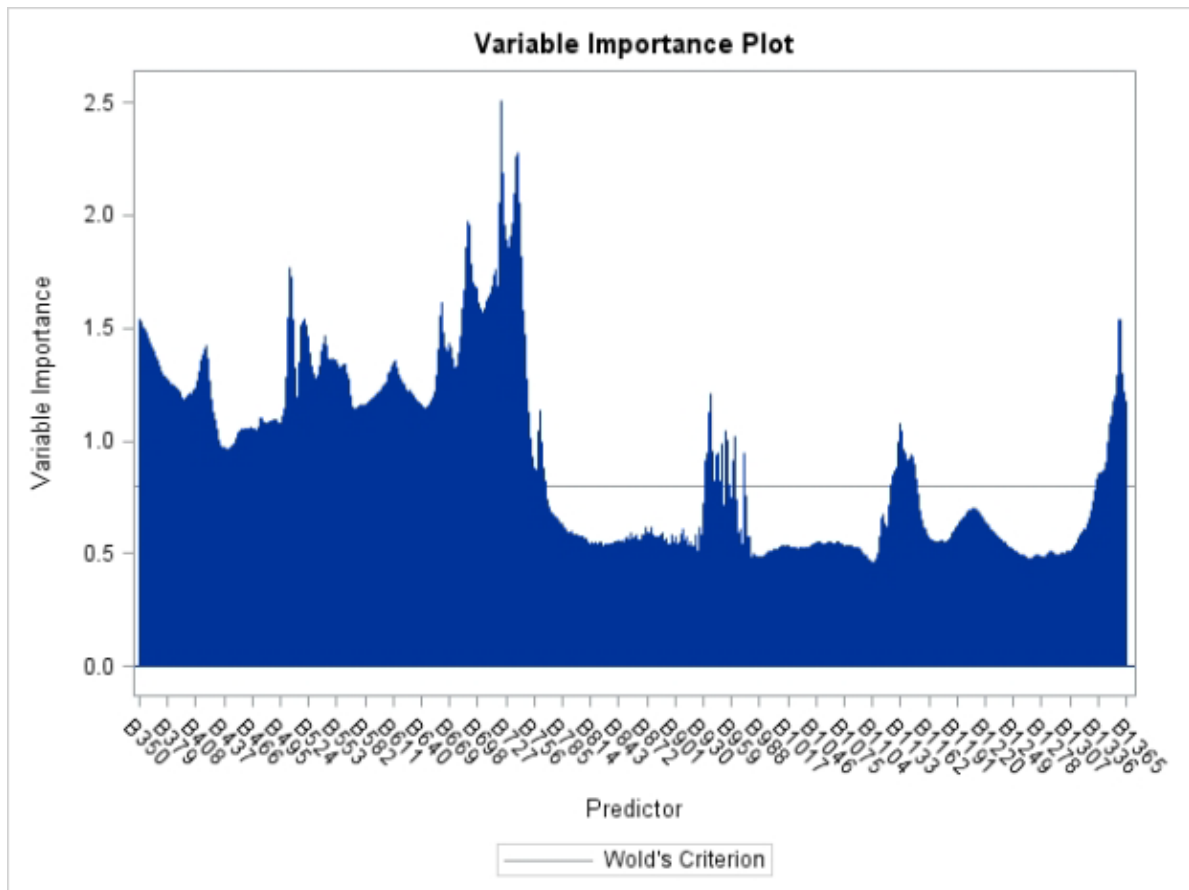
93.8%. Η διαφορά ανάμεσα σε αυτές τις δύο τιμές δεν είναι μεγάλη συγκριτικά με τη διαφορά που δημιουργείται αν από τους 9 πρώτους factors αφαιρέσουμε τον 9^ο.

Ακολουθεί το διάγραμμα του Πίνακα 13, όπου από την διακεκομμένη γραμμή που αντιστοιχεί στην επεξηγηματικότητα της εξαρτημένης μεταβλητής, μπορούμε να παρατηρήσουμε ότι μετά το σημείο 9, η κλίση μειώνεται.



Εικόνα 12 R-Square Analysis 2c

Τέλος, στο σχήμα που ακολουθεί βλέπουμε τη σημαντικότητα των αρχικών 1016 ανεξάρτητων μεταβλητών ως προς την εξαρτημένη μεταβλητή. Όπως μπορούμε να παρατηρήσουμε τα ποσοστά ανάκλασης στο κάθε μήκος κύματος συσχετίζονται διαφορετικά στην άμπελο έχοντας κάποιες ζώνες με πολύ κοντινή και χαμηλή σημαντικότητα αλλά μεγαλύτερο ενδιαφέρον έχουν τα μήκη κύματος στα οποία έχουμε τις κορυφές με τις μεγαλύτερες σημαντικότητες που θα διαμορφώσουν τα factor μας .



Εικόνα 13 Variable Importance Plot 2c

4.7 Σύγκριση παραλλαγών

Σε αυτή την ενότητα θα συγκρίνουμε τα αποτελέσματα των έξι παραλλαγών ως προς την επεξηγηματικότητα και τον αριθμό των factors, ακολουθεί Πίνακας με τα δεδομένα.

Πίνακας 15 Επεξηγηματικότητα Παραλλαγών

Επεξηγηματικότητα Παραλλαγών			
Παραλλαγή	minfactors	Επεξηγηματικότητα minfactors	Επεξηγηματικότητα 15 factors
1α	12	75.4	82.6
1β	10	79.9	91.6
1γ	8	74.0	95.1
2α	7	63.0	89.6
2β	5	50.2	95.7
2γ	9	80.8	93.8

Ιδανικά θα θέλαμε να έχουμε τον ελάχιστο αριθμό factors με την μεγίστη δυνατή επεξηγηματικότητα μεγαλύτερη από 95%.

Ως προς την επεξηγηματικότητα στα 15 factors παρατηρούμε ότι έχουμε σχετικά μεγάλες τιμές ειδικά όταν εξετάζουμε τα φύλλα κορυφής και μέσης ξεχωριστά παραλλαγές 1β ,1γ, 2β ,2γ . Επίσης παρατηρούμε στις δυο παραλλαγές με τους λιγότερους factors 2α , 2β με 7 factors και 5 factors αντίστοιχα η επεξηγηματικότητα είναι πολύ μικρή και όχι ικανοποιητική. Την καλύτερη επεξηγηματικότητα την έχουμε στην παραλλαγή 2γ με 80.8% και 9 factors και στην συνέχεια στην παραλλαγή 1β με 79.9% και 10 factors.

4.8 Αξιολόγηση του αλγορίθμου NIPALS για τις παραλλαγές 1β και 1γ (χρήση του rank 350-1365 για τα φύλλα της κορυφής και της μέσης αντίστοιχα)

Σκοπός αυτής της ενότητας είναι αφού κατασκευάσαμε με την χρήση του NIPALS τα μοντέλα ταξινόμησης των ποικιλιών μέσω των factors , να εισάγουμε ένα αριθμό δεδομένων από φύλλα που γνωρίζουμε την ποικιλία τους και να παρακολουθήσουμε πως αυτά θα ταξινομηθούν για την κάθε ποικιλία.

4.8.1 Αξιολόγηση αλγορίθμου NIPALS για την παραλλαγή 1β

Η αξιολόγηση για την παραλλαγή 1β έγινε σε 60 φύλλα 6 για την κάθε ποικιλία και μία πρώτη αξιολόγηση φαίνεται στον Πίνακα 16 .Στον Πίνακα αυτό βλέπουμε την ποσοτική και ποσοστιαία ταξινόμηση ανά ποικιλία των φύλλων.

Πίνακας 16 Αξιολόγηση 1β πριν την επικύρωση

Number of Observations and Percent Classified into VarAb											
Ποικιλία φύλλου / Ποικιλία που ταξινομήθηκε	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
ASI	5	0	0	1	0	0	0	0	0	0	6
	83.33%	0	0	16.67%	0	0	0	0	0	0	100%
ASP	0	6	0	0	0	0	0	0	0	0	6
	0	100%	0	0	0	0	0	0	0	0	100%
GOU	0	0	4	1	0	0	0	0	1	0	6
	0	0	66.67%	16.67%	0	0	0	0	16.67%	0	100%
LGI	0	0	0	6	0	0	0	0	0	0	6
	0	0	0	100%	0	0	0	0	0	0	100%
MAV	0	1	0	0	5	0	0	0	0	0	6
	0	16.67%	0	0	83.33%	0	0	0	0	0	100%
MEL	0	0	0	0	0	6	0	0	0	0	6
	0	0	0	0	0	100%	0	0	0	0	100%
MOS	0	0	0	0	0	0	6	0	0	0	6
	0	0	0	0	0	0	100%	0	0	0	100%
PAT	0	0	0	0	0	0	0	6	0	0	6
	0	0	0	0	0	0	0	100%	0	0	100%
SAM	0	0	1	0	0	0	0	0	5	0	6
	0	0	16.67%	0	0	0	0	0	83.33%	0	100%
VOL	0	0	0	0	1	0	0	0	0	5	6
	0	0	0	0	16.67%	0	0	0	0	83.33%	100%
Total	5	7	5	8	6	6	6	6	6	5	60
	8.33%	11.67%	8.33%	13.33%	10%	10%	10%	10%	10%	8.33%	100%
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	
Error Count Estimates for VarAb											
	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
Rate	0.1667	0	0.3333	0	0.1667	0	0	0	0.1667	0.1667	0.1
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

Στην συνέχεια με μία διαδικασία εσωτερικής ανάλυσης κατά την οποία κυκλικά αφαιρείται από το δείγμα ένα φύλλο και ταξινομούνται από την αρχή έως ότου αυτό γίνει για όλα τα φύλλα στην προκειμένη περίπτωση 60 επαναλήψεις παίρνουμε τους μέσους όρους αυτών των επαναλήψεων και έτσι επιτυγχάνουμε επικύρωση των αποτελεσμάτων που τα βλέπουμε στον Πίνακα 17.

Πίνακας 17 Αξιολόγηση 1β μετά την επικύρωση

Number of Observations and Percent Classified into VarAb											
Ποικιλία φύλλου / Ποικιλία που ταξινομήθηκε	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
ASI	3	0	0	1	1	0	0	1	0	0	6
	50%	0	0	16.67%	16.67%	0	0	16.67%	0	0	100%
ASP	0	3	1	1	0	0	0	0	0	1	6
	0	50%	16.67%	16.67%	0	0	0	0	0	16.67%	100%
GOU	0	1	2	1	0	0	0	0	2	0	6
	0	16.67%	33.33%	16.67%	0	0	0	0	33.33%	0	100%
LGI	2	0	0	4	0	0	0	0	0	0	6
	33.33%	0	0	66.67%	0	0	0	0	0	0	100%
MAV	0	1	0	0	5	0	0	0	0	0	6
	0	16.67%	0	0	83.33%	0	0	0	0	0	100%
MEL	0	0	0	0	0	5	0	0	1	0	6
	0	0	0	0	0	83.33%	0	0	16.67%	0	100%
MOS	0	0	0	0	0	0	6	0	0	0	6
	0	0	0	0	0	0	100%	0	0	0	100%
PAT	0	0	0	0	0	0	0	6	0	0	6
	0	0	0	0	0	0	0	100%	0	0	100%
SAM	0	0	1	0	0	0	0	0	5	0	6
	0	0	16.67%	0	0	0	0	0	83.33%	0	100%
VOL	0	1	0	0	2	0	0	0	0	3	6
	0	16.67%	0	0	33.33%	0	0	0	0	50%	100%
Total	5	6	4	7	8	5	6	7	8	4	60
	8.33%	10%	6.67%	11.67%	13.33%	8.33%	10%	11.67%	13.33%	6.67%	100%
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	
Error Count Estimates for VarAb											
	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
Rate	0.5	0.5	0.6667	0.3333	0.1667	0.1667	0	0	0.1667	0.5	0.3
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

Στον Πίνακα αυτό βλέπουμε την ποσοτική και ποσοστιαία ταξινόμηση ανά ποικιλία των φύλλων. Παρατηρούμε ότι ενώ αρχικά είχαμε μόνο 6 λάθος ταξινομήσεις (10%) με την διαδικασία επικύρωσης τα λάθη αυξήθηκαν σε 18 ταξινομήσεις (30%) αυτή η αύξηση ήταν αναμενόμενη καθώς για την παραλλαγή 1β έχουμε επεξηγηματικότητα 79,9% οπότε θεωρητικά θα περιμέναμε να ταξινομηθεί σωστά το 80%. Επίσης παρατηρούμε ότι για τις ποικιλίες Μοσχούδι (MOS) και Πατίνες (PAT) όλα τα φύλλα τους ταξινομήθηκαν σωστά ενώ για τις υπόλοιπες ποικιλίες το ποσοστό επιτυχίας ήταν από 33,33% έως 83,33%.

4.8.2 Αξιολόγηση αλγορίθμου NIPALS για την παραλλαγή 1γ

Η αξιολόγηση για την παραλλαγή 1γ έγινε σε 59 φύλλα 6 για την κάθε ποικιλία με εξαίρεση την Βόλιτσα Λεύκη (VOL) που έγινε σε 5 φύλλα και μία πρώτη αξιολόγηση φαίνεται στον Πίνακα 18 .Στον Πίνακα αυτό βλέπουμε την ποσοτική και ποσοστιαία ταξινόμηση ανά ποικιλία των φύλλων.

Πίνακας 18 Αξιολόγηση 1γ πριν την επικύρωση

Number of Observations and Percent Classified into VarAb											
Ποικιλία φύλλου / Ποικιλία που ταξινομήθηκε	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
ASI	6	0	0	0	0	0	0	0	0	0	6
	100%	0	0	0	0	0	0	0	0	0	100%
ASP	0	6	0	0	0	0	0	0	0	0	6
	0	100%	0	0	0	0	0	0	0	0	100%
GOU	0	0	6	0	0	0	0	0	0	0	6
	0	0	100%	0	0	0	0	0	0	0	100%
LGI	0	0	0	6	0	0	0	0	0	0	6
	0	0	0	100%	0	0	0	0	0	0	100%
MAV	2	0	0	0	4	0	0	0	0	0	6
	33.33%	0	0	0	66.67%	0	0	0	0	0	100%
MEL	0	1	0	0	0	5	0	0	0	0	6
	0	16.67%	0	0	0	83.33%	0	0	0	0	100%
MOS	0	0	0	0	0	0	6	0	0	0	6
	0	0	0	0	0	0	100%	0	0	0	100%
PAT	0	0	0	0	0	0	0	6	0	0	6
	0	0	0	0	0	0	0	100%	0	0	100%
SAM	0	0	0	0	0	0	0	0	6	0	6
	0	0	0	0	0	0	0	0	100%	0	100%
VOL	0	0	0	0	0	0	0	0	0	5	5
	0	0	0	0	0	0	0	0	0	100%	100%
Total	8	7	6	6	4	5	6	6	6	5	59
	13.56%	11.86%	10.17%	10.17%	6.78%	8.47%	10.17%	10.17%	10.17%	8.47%	100%
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	
Error Count Estimates for VarAb											
	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
Rate	0	0	0	0	0.3333	0.1667	0	0	0	0	0.05
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

Στην συνέχεια με μία διαδικασία εσωτερικής ανάλυσης κατά την οποία κυκλικά αφαιρείται από το δείγμα ένα φύλλο και ταξινομούνται από την αρχή έως ότου αυτό γίνει για όλα τα φύλλα στην προκειμένη περίπτωση 59 επαναλήψεις παίρνουμε τους μέσους όρους αυτών των επαναλήψεων και έτσι επιτυγχάνουμε επικύρωση των αποτελεσμάτων που τα βλέπουμε στον Πίνακα 19.

Πίνακας 19 Αξιολόγηση 1γ μετά την επικύρωση

Number of Observations and Percent Classified into VarAb											
Ποικιλία φύλλου / Ποικιλία που ταξινομήθηκε	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
ASI	4	0	0	0	2	0	0	0	0	0	6
	66.67%	0	0	0	33.33%	0	0	0	0	0	100%
ASP	0	6	0	0	0	0	0	0	0	0	6
	0	100%	0	0	0	0	0	0	0	0	100%
GOU	0	0	6	0	0	0	0	0	0	0	6
	0	0	100%	0	0	0	0	0	0	0	100%
LGI	0	0	0	6	0	0	0	0	0	0	6
	0	0	0	100%	0	0	0	0	0	0	100%
MAV	2	0	0	0	4	0	0	0	0	0	6
	33.33%	0	0	0	66.67%	0	0	0	0	0	100%
MEL	0	1	0	1	0	4	0	0	0	0	6
	0	16.67	0	16.67	0	66.67	0	0	0	0	100%
MOS	0	0	0	0	0	0	6	0	0	0	6
	0	0	0	0	0	0	100%	0	0	0	100%
PAT	0	0	0	0	0	1	0	5	0	0	6
	0	0	0	0	0	16.67%	0	83.33%	0	0	100%
SAM	0	0	1	0	0	0	0	0	5	0	6
	0	0	16.67%	0	0	0	0	0	83.33%	0	100%
VOL	0	0	0	0	1	0	0	0	0	4	5
	0	0	0	0	20%	0	0	0	0	80%	100%
Total	6	7	7	7	7	5	6	5	5	4	59
	10.17%	11.86%	11.86%	11.86%	11.86%	8.47%	10.17%	8.47%	8.47%	6.78%	100%
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	
Error Count Estimates for VarAb											
	ASI	ASP	GOU	LGI	MAV	MEL	MOS	PAT	SAM	VOL	Total
Rate	0.3333	0	0	0	0.3333	0.3333	0	0.1667	0.1667	0.2	0.1533
Priors	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	

Στον Πίνακα αυτό βλέπουμε την ποσοτική και ποσοστιαία ταξινόμηση ανά ποικιλία των φύλλων. Παρατηρούμε ότι ενώ αρχικά είχαμε μόνο 3 λάθος ταξινομήσεις (5%) με την διαδικασία επικύρωσης τα λάθη αυξήθηκαν σε 9 ταξινομήσεις (15%) αυτή η αύξηση ήταν αναμενόμενη καθώς για την παραλλαγή 1γ έχουμε επεξηγηματικότητα 74,0% οπότε θεωρητικά θα περιμέναμε να ταξινομηθεί σωστά το 74%. Επίσης παρατηρούμε ότι για τις ποικιλίες Ασπρούδα (ASP), Γκούμασι Λιανόρραγο (GOU) , Λευκό Γιάννενα (LGI) και Μοσχούδι (MOS) όλα τα φύλλα τους ταξινομήθηκαν σωστά ενώ για τις υπόλοιπες ποικιλίες το ποσοστό επιτυχίας ήταν από 66,67%έως 83,33%.

5 Συμπεράσματα

Σε αυτή την ενότητα θα σχολιάσουμε τα αποτελέσματα αυτής της εργασίας ,τις δυσκολίες που αντιμετωπίσαμε και της παρατηρήσεις που κάναμε πάνω στο αντικείμενο της φασματοσκοπίας ως προς την διάκριση των ποικιλιών της αμπέλου.

Είδη από την αρχή της έρευνας κατά τον ορισμό της μεταβλητής rank 350-1365 σε μία προσπάθεια να ιεραρχήσουμε τις δέκα ποικιλίες μας βάση την μέση ανάκλαση τους παρατηρήσαμε ότι κάποιες από τις ποικιλίες έχουν μεταξύ τους πολύ κοντινές τιμές μέσης ανάκλασης σχεδόν ταυτόσημες σε κάποιες περιπτώσεις (Πίνακας 2), για αυτό το λόγω έγινε και η προσπάθεια της δυαδικής κωδικοποίησης διότι στον άξονα των y που αντιστοιχεί η εξαρτημένη μεταβλητή της ποικιλίας που θέτουμε τις τιμές 1 έως 10 πού απέχουν μεταξύ τους κατά ένα αλλά αυτό δεν αντιστοιχεί στην πραγματικότητα αυτό θα δημιουργήσει προβλήματα κατά την ταξινόμηση.

Σε αυτή την μελέτη δοκιμάστηκαν έξι διαφορετικές παραλλαγές αρχικοποίησης του αλγορίθμου NIPALS. Τρεις από αυτές έχουν στον άξονα των y τη μεταβλητή rank 350-1365 ενώ οι άλλες τρεις είναι με δυαδική κωδικοποίηση για α όλα τα φύλλα β τα φύλλα κορυφής γ τα φύλλα μέσης αντίστοιχα. Τα συγκεντρωτικά αποτελέσματα αυτών των παραλλαγών βρίσκονται στον Πίνακα 15. Ιδανικά θα θέλαμε με 5 factors να έχουμε επεξηγηματικότητα μεγαλύτερη από 95% . Στην μελέτη αυτή τις καλύτερες επεξηγηματικότητες τις είχαμε στις παραλλαγές 2 γ με 80.8% και 9 factors και στην 1 β με 79.9% και 10 factors.

Σε μία αξιολόγηση του αλγορίθμου NIPALS για τις παραλλαγές 1 β και 1 γ βρήκαμε ότι για τον 1 β ταξινομούνται σωστά το 70% των φύλλων ενώ περιμέναμε το 80% και για τον 1 γ ταξινομήθηκε σωστά το 85% ενώ περιμέναμε το 74% .Αυτές οι αποκλίσεις φαίνονται σχετικά μεγάλες αλλά στην πραγματικότητα κυρίως οφείλονται στον μικρό αριθμό των φύλλων που χρησιμοποιήθηκαν για την αξιολόγηση των αλγορίθμων αλλά και στην ποσότητα των δεδομένων που χρησιμοποιήσαμε για να χτίσουμε τα μοντέλα μας.

Τα γενικά συμπεράσματα που καταλήξαμε σε αυτή την έρευνα είναι ότι ο αριθμός των ποικιλιών αλλά και το ποίες ποικιλίες επιλέξαμε για αυτή την έρευνα έπαιξε σημαντικό ρόλο ως προς τα αποτελέσματα που πήραμε ,εάν είχαμε κάνει την ίδια διαδικασία για διαφορετικό αριθμό ποικιλιών με μεγάλες διαφοροποιήσεις ως προς την ανακλαστικότητα θα είχαμε πολύ διαφορετικά αποτελέσματα . Παρόλα αυτά η φασματοσκοπία είναι μία πολύ υποσχόμενη μέθοδος αμπελογραφίας για την διάκριση των ποικιλιών της αμπέλου ,πρέπει όμως να γίνει μελέτη σε ένα πολύ μεγαλύτερο αριθμό ποικιλιών σε πολλά φυτά από την

καθεμία και από διαφορετικές τοποθεσίες με διαφορετικό κλίμα για να δούμε πως και αυτό επηρεάζει για να καταλήξουμε σε ασφαλή συμπεράσματα.

6 Βιβλιογραφία

- Abbasi, M.; Verrelst, J.; Mirzaei, M.; Marofi, S.; Bakhtiari, H.R.R. (2020): Optimal spectral wavelengths for discriminating orchard species using multivariate statistical techniques. In *Remote Sens.* 12 (1). DOI: 10.3390/RS12010063.
- Álvarez, Á.; Yáñez, J.; Neira, Y.; Castillo-Felices, R.; Hinrichsen, P. (2020): Simple distinction of grapevine (*Vitis vinifera* L.) genotypes by direct ATR-FTIR. In *Food Chem.* 328. DOI: 10.1016/j.foodchem.2020.127164.
- Barker, Matthew; Rayens, William (2003): Partial least squares for discrimination. In *J. Chemometrics* 17 (3), pp. 166–173. DOI: 10.1002/cem.785.
- Barnes, R. J.; Dhanoa, M. S.; Lister, Susan J. (1989): Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. In *Appl Spectrosc* 43 (5), pp. 772–777. DOI: 10.1366/0003702894202201.
- Cheng, Y. L.; Yang, S. Q.; Liu, X.; Zhang, E. Y.; Song, Z. S. (2019): Identification of wine grape varieties based on near-infrared hyperspectral imaging. In *Applied Engineering in Agriculture* 35 (6), pp. 959–967. DOI: 10.13031/aea.13452959.
- Cozzolino, Daniel; Smyth, Heather Eunice; Gishen, Mark (2003): Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. In *Journal of agricultural and food chemistry* 51 (26), pp. 7703–7708. DOI: 10.1021/jf034959s.
- Crusiol, L.G.T.; Nanni, M. R.; Furlanetto, R. H.; Sibaldelli, R.N.R.; Cezar, E.; Sun, L. et al. (2021): Classification of soybean genotypes assessed under different water availability and at different phenological stages using leaf-based hyperspectral reflectance. In *Remote Sens.* 13 (2), pp. 1–24. DOI: 10.3390/rs13020172.
- Diago, Maria P.; Fernandes, A. M.; Millan, B.; Tardaguila, J.; Melo-Pinto, P. (2013): Identification of grapevine varieties using leaf spectroscopy and partial least squares. In *Computers and Electronics in Agriculture* 99, pp. 7–13. DOI: 10.1016/j.compag.2013.08.021.
- Example 51.1: Examining Model Details (2000). Available online at <https://www.sfu.ca/sasdoc/sashtml/stat/chap51/sect19.htm>, updated on 1/6/2000, checked on 1/5/2023.
- Feng, Lei; Zhu, Susu; Zhang, Chu; Bao, Yidan; Gao, Pan; He, Yong (2018): Variety Identification of Raisins Using Near-Infrared Hyperspectral Imaging. In *Molecules* (Basel, Switzerland) 23 (11). DOI: 10.3390/molecules23112907.
- Fernandes, A.; Utkin, A.; Eiras-Dias, J.; Silvestre, J.; Cunha, J.; Melo-Pinto, P. (2018): Assessment of grapevine variety discrimination using stem hyperspectral data and AdaBoost of random weight neural networks. In *Appl. Soft Comput. J.* 72, pp. 140–155. DOI: 10.1016/j.asoc.2018.07.059.
- Fernandes, Armando M.; Utkin, Andrei B.; Eiras-Dias, José; Cunha, Jorge; Silvestre, José; Melo-Pinto, Pedro (2019): Grapevine variety identification using “Big Data” collected with miniaturized spectrometer combined with support vector machines and convolutional neural networks. In *Computers and Electronics in Agriculture* 163, p. 104855. DOI: 10.1016/j.compag.2019.104855.
- Fuentes, S.; Hernández-Montes, E.; Escalona, J. M.; Bota, J.; Gonzalez Viejo, C.; Poblete-Echeverría, C. et al. (2018): Automated grapevine cultivar classification based on

- machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters. In *Computers and Electronics in Agriculture* 151, pp. 311–318. DOI: 10.1016/j.compag.2018.06.035.
- Ghule, A. N.; Deshmukh, R. R. (2021): Wavelength selection and classification of hyperspectral non-imagery data to discriminate healthy and unhealthy vegetable leaves. In *Curr. Sci.* 120 (5), pp. 936–941. DOI: 10.18520/cs/v120/i5/936-941.
- González-Fernández, Ana Belén; Rodríguez-Pérez, José Ramón; Marcelo, Victoriano; Valenciano, José B. (2015): Using field spectrometry and a plant probe accessory to determine leaf water content in commercial vineyards. In *Agricultural Water Management* 156, pp. 43–50. DOI: 10.1016/j.agwat.2015.03.024.
- Gutiérrez, S.; Tardaguila, J.; Fernández-Novales, J.; Diago, M. P. (2015): Support vector machine and artificial neural network models for the classification of grapevine varieties using a portable NIR spectrophotometer. In *PloS one* 10 (11). DOI: 10.1371/journal.pone.0143197.
- Gutiérrez, S.; Tardaguila, J.; Fernández-Novales, J.; Diago, M. P. (2016): Data mining and NIR spectroscopy in viticulture: Applications for plant phenotyping under field conditions. In *Sensors* 16 (2). DOI: 10.3390/s16020236.
- Gutiérrez, Salvador; Fernández-Novales, Juan; Diago, Maria P.; Tardaguila, Javier (2018): On-The-Go Hyperspectral Imaging Under Field Conditions and Machine Learning for the Classification of Grapevine Varieties. In *Frontiers in Plant Science* 9, p. 1102. DOI: 10.3389/fpls.2018.01102.
- Jacquemoud, S.; Ustin, S. (2019): *Leaf optical properties*: Cambridge University Press (Leaf Optical Properties). Available online at <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086301414&doi=10.1017%2F9781108686457&partnerID=40&md5=8e16948bcfa5f86fb0cd574576e49b42>.
- Jacquemoud, Stéphane; Ustin, Susan (Eds.) (2019): *Leaf Optical Properties*: Cambridge University Press.
- Jacquemoud, Stéphane; Ustin, Susan (Eds.) (2019): *Leaf Optical Properties*: Cambridge University Press.
- Jacquemoud, Stéphane; Ustin, Susan (Eds.) (2019): *Leaf Optical Properties*: Cambridge University Press.
- Karabourniotis, G.; Bornman, J. F.; Liakoura, V. (1999): Different leaf surface characteristics of three grape cultivars affect leaf optical properties as measured with fibre optics: possible implication in stress tolerance. In *Functional Plant Biology* 26 (1), p. 47. DOI: 10.1071/PP98052.
- Karakizi, C.; Oikonomou, M.; Karantzalos, K. (2015): Spectral discrimination and reflectance properties of various vine varieties from satellite, UAV and proximate sensors. In Schreier G., Skrovseth P.E., Staudenrausch H. (Eds.): *International Society for Photogrammetry and Remote Sensing* (40), pp. 31–37. Available online at <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84930406894&doi=10.5194%2Fisprsarchives-XL-7-W3-31-2015&partnerID=40&md5=2ab161865cebaad975820db0b8956d61>.
- Leaf Biophysics (2019). In Stéphane Jacquemoud, Susan Ustin (Eds.): *Leaf Optical Properties*: Cambridge University Press, pp. 12–47.

- Leaf Optical Properties in Different Wavelength Domains (2019). In Stéphane Jacquemoud, Susan Ustin (Eds.): *Leaf Optical Properties*: Cambridge University Press, pp. 124–169.
- Lo, Yen-Li; Pan, Wen-Harn; Hsu, Wan-Lun; Chien, Yin-Chu; Chen, Jen-Yang; Hsu, Mow-Ming et al. (2016): Partial Least Square Discriminant Analysis Discovered a Dietary Pattern Inversely Associated with Nasopharyngeal Carcinoma Risk. In *PloS one* 11 (6), e0155892. DOI: 10.1371/journal.pone.0155892.
- Maimaitiyiming, M.; Miller, A. J.; Ghulam, A. (2016): Discriminating spectral signatures among and within two closely related grapevine species. In *Photogramm. Eng. Remote Sens.* 82 (1), pp. 51–62. DOI: 10.14358/PERS.82.2.51.
- Mirzaei, M.; Marofi, S.; Abbasi, M.; Solgi, E.; Karimi, R.; Verrelst, J. (2019): Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. In *Int. J. Appl. Earth Obs. Geoinformation* 80, pp. 26–37. DOI: 10.1016/j.jag.2019.04.002.
- Pérez-Enciso, Miguel; Tenenhaus, Michel (2003): Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. In *Human genetics* 112 (5-6), pp. 581–592. DOI: 10.1007/s00439-003-0921-9.
- Randolph, Timothy W. (2006): Scale-based normalization of spectral data. In *Cancer biomarkers : section A of Disease markers* 2 (3-4), pp. 135–144. DOI: 10.3233/cbm-2006-23-405.
- Rodríguez-Pérez, José R.; Riaño, David; Carlisle, Eli; Ustin, Susan; Smart, David R. (2007): Evaluation of Hyperspectral Reflectance Indexes to Detect Grapevine Water Status in Vineyards. In *Am J Enol Vitic.* 58 (3), pp. 302–317. DOI: 10.5344/ajev.2007.58.3.302.
- Ruiz-Perez, Daniel; Guan, Haibin; Madhivanan, Purnima; Mathee, Kalai; Narasimhan, Giri (2020): So you think you can PLS-DA? In *BMC bioinformatics* 21 (Suppl 1), p. 2. DOI: 10.1186/s12859-019-3310-7.
- Schreier G.; Skrovseth P.E.; Staudenrausch H. (Eds.) (2015): *International Society for Photogrammetry and Remote Sensing* (40).
- The PLS Procedure (SAS Help Center) (2019), updated on 12/12/2022, checked on 12/15/2022.
- van Wyngaard, E.; Blancquaert, E.; Nieuwoudt, H.; Aleixandre-Tudo, J. L. (2021): Infrared Spectroscopy and Chemometric Applications for the Qualitative and Quantitative Investigation of Grapevine Organs. In *Frontiers in Plant Science* 12. DOI: 10.3389/fpls.2021.723247.
- Variation Due to Leaf Structural, Chemical, and Physiological Traits (2019). In Stéphane Jacquemoud, Susan Ustin (Eds.): *Leaf Optical Properties*: Cambridge University Press, pp. 170–194.
- ΑΜΠΕΛΟΓΡΑΦΙΑ 2η έκδοση Δημήτριος Ευστ. Σταύρακας Ομότιμος Καθηγητής Α.Π.Θ.

7 Παράρτημα

Πίνακας 20 Δείγμα Αρχικών Δεδομένων

MesID	Variety	VarAb	RV	VineID	Vine	Section	Node	LeafCode	Leaf	Rank550_690	Rank750_1300
491	Ασπρούδι	ASI	2_6	4350	1	M	6	M1	1	9	8
492	Ασπρούδι	ASI	2_6	4350	1	M	6	M1	1	9	8
493	Ασπρούδι	ASI	2_6	4350	1	M	6	M1	1	9	8
494	Ασπρούδι	ASI	2_6	4350	1	M	7	M2	2	9	8
495	Ασπρούδι	ASI	2_6	4350	1	M	7	M2	2	9	8
496	Ασπρούδι	ASI	2_6	4350	1	M	7	M2	2	9	8
497	Ασπρούδι	ASI	2_6	4350	1	M	8	M3	3	9	8
498	Ασπρούδι	ASI	2_6	4350	1	M	8	M3	3	9	8
499	Ασπρούδι	ASI	2_6	4350	1	M	8	M3	3	9	8
506	Ασπρούδι	ASI	2_6	4350	1	U	16	K1	1	9	8
507	Ασπρούδι	ASI	2_6	4350	1	U	16	K1	1	9	8
508	Ασπρούδι	ASI	2_6	4350	1	U	16	K1	1	9	8
503	Ασπρούδι	ASI	2_6	4350	1	U	15	K2	2	9	8
504	Ασπρούδι	ASI	2_6	4350	1	U	15	K2	2	9	8
505	Ασπρούδι	ASI	2_6	4350	1	U	15	K2	2	9	8
500	Ασπρούδι	ASI	2_6	4350	1	U	14	K3	3	9	8
501	Ασπρούδι	ASI	2_6	4350	1	U	14	K3	3	9	8
502	Ασπρούδι	ASI	2_6	4350	1	U	14	K3	3	9	8