



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ  
ΕΡΓΑΣΤΗΡΙΟ ΟΠΤΟΗΛΕΚΤΡΟΝΙΚΗΣ, LASER ΚΑΙ ΕΦΑΡΜΟΓΩΝ ΤΟΥΣ

---

«Στατιστική ανάλυση ανακτηθέντων φασμάτων  
φθορισμού καταστρεπτικών βιολογικών  
επικαθίσεων σε επιφάνειες μνημείων με χρήση  
Επαγόμενου Φθορισμού με LASER (LIF)»

---

Διπλωματική Εργασία  
της  
Χριστίνας Γιαννακουδάκη  
Α.Μ. ge15053

**Επιβλέπων:** Δρ. Αλέξανδρος Παπαγιάννης, Καθηγητής

**Τριμελής Εξεταστική Επιτροπή:**

Α. Παπαγιάννης                      Χ. Καρώνη                      Β. Παπανικολάου  
Καθηγητής ΣΕΜΦΕ    Καθηγήτρια ΣΕΜΦΕ    Καθηγητής ΣΕΜΦΕ

Αθήνα, Σεπτέμβριος 2021

*Στους γονείς μου, Γιάννη και Ελένη  
και στον αδελφό μου, Γιώργη*

.....  
Χριστίνα Γιαννακουδάκη

© (2021) Εθνικό Μετσόβιο Πολυτεχνείο. All rights Reserved. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σ' αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## Ευχαριστίες

Η εκπόνηση της παρούσας διπλωματικής εργασίας, η οποία υλοποιήθηκε κατά το ακαδημαϊκό έτος 2020-2021 στη Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών, ήταν εφικτή εξαιτίας της συμβολής πολλών και σημαντικών ανθρώπων. Για τον λόγο αυτό θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε καθέναν από αυτούς.

Αρχικά θα ήθελα να ευχαριστήσω τον Ιωάννη Θυρή από την εταιρεία Raymetrics, ο οποίος τόσο με τις θεωρητικές όσο και με τις πρακτικές γνώσεις του με βοήθησε σε κάθε τομέα αυτής της εργασίας. Τόσο το κείμενο όσο και η ανάκτηση και η ανάλυση των δεδομένων στηρίζονται στις συμβουλές και στις διορθώσεις του, καθώς επίσης και στις πολλές συζητήσεις σχετικά με την προσέγγιση του αντικειμένου κατά την διάρκεια της εκπόνησης.

Στη συνέχεια, θα ήθελα να ευχαριστήσω την Δρ Ιωάννα Πυρρή από το Τμήμα Βιολογίας του Ε.Κ.Π.Α, η οποία πραγματοποίησε την μελέτη των δειγμάτων από την βιολογική σκοπιά. Η συμβολή της με τις γνώσεις και τις συμβουλές της σε κάθε στάδιο αυτής τη διπλωματικής ήταν καθοριστική για την εξέλιξη της έρευνας. Επιπλέον, οι ευχαριστίες μου επεκτείνονται και για κάθε άλλη βιολογική μελέτη την οποία πραγματοποίησε για εμένα μέχρι να καταλήξω τελικά στο συγκεκριμένο θέμα που μελέτησα.

Επιπλέον, θα ήθελα να ευχαριστήσω την Καθηγήτρια Δρ. Χρυσής Καρώνη, η οποία με καθοδήγησε συστηματικά στον άγνωστο για εμένα επιστημονικό τομέα της Ανάλυσης Δεδομένων. Τα αποτελέσματα αυτής της εργασίας στηρίζονται σε βασική βιβλιογραφία καθώς και σε συμβουλές με τα οποία με τροφοδοτούσε κατά την διάρκεια της εκπόνησης.

Έπειτα, θα ήθελα να ευχαριστήσω τον Στέφανο Ρίτσαρντσον, διδακτορικό φοιτητή του Πανεπιστημίου του Southampton, για τις επιστημονικές συμβουλές του κατά την διάρκεια της μελέτης, όπως επίσης και για την γνωριμία που μου παρείχε με τον τομέα των LIDAR. Επιπλέον, θέλω να ευχαριστήσω την Παναγιώτα-Ισμήνη Χάρκεν-Αλεξίου, απόφοιτη της ΣΕΜΦΕ, η οποία εκτός από φίλη μου, αποτέλεσε μια πολύτιμη βοήθεια στην στατιστική επεξεργασία των δεδομένων.

Δεν θα μπορούσα να παραλείψω του ανθρώπους που με στήριξαν ψυχολογικά κατά την διάρκεια αυτού του ταξιδιού. Αρχικά, ευχαριστώ τον Αντώνη Τζωρτζόπουλο για την στήριξη και την βοήθεια σε ολόκληρο το χρονικό διάστημα αυτής της εργασίας. Ευχαριστώ την οικογένεια μου, τους φίλους μου Μαριάνα, Χριστιάνα, Παναγιώτη, Παναγιώτα, Έλενα, Κατερίνα και Δέσποινα που χωρίς την βοήθεια τους, δεν θα ήταν τίποτα εφικτό. Επίσης ευχαριστώ πολύ την κυρία Αντωνίου και την ομάδα για την στήριξή τους στις δύσκολες στιγμές.

Για το τέλος κράτησα τις πιο θερμές και εγκάρδιες ευχαριστίες για τον επιβλέποντα Καθηγητή μου, Δρ. Αλέξανδρο Παπαγιάννη. Η εμπιστοσύνη που μου έδειξε στα τρία χρόνια της γνωριμίας μας, η δυνατότητα να υλοποιήσω την κάθε επιστημονική ιδέα που είχα, η αμέριστη στήριξή του σε κάθε εμπόδιο που ανέκυπτε και οι συμβουλές του σε πολλαπλά θέματα που με απασχολούσαν, με διαμόρφωσε στον άνθρωπο που είμαι σήμερα. Επιπλέον, τον ευχαριστώ για τις ευκαιρίες που μου έδωσε ώστε να εξερευνήσω τον υπέροχο κόσμο των LASER και των εφαρμογών τους.

# Περιεχόμενα

Περίληψη	ix
Abstract	x
<b>1 Κεφάλαιο 1: Εισαγωγικά Στοιχεία</b>	<b>1</b>
1.1 Φυσικές Διεργασίες	2
1.1.1 Κβαντομηχανικές Αρχές Αλληλεπίδρασης Ακτινοβολίας με Ύλη	2
1.1.2 Αρχή Franck-Condon - Θεωρητικό Υπόβαθρο Απορρόφησης	5
1.1.3 Χαρακτηριστικά Μεγέθη Απορρόφησης	6
1.1.4 Εσωτερική Μετατροπή	9
1.1.5 Φθορισμός	10
1.1.6 Διασυστημική Διέλευση και Φωσφορισμός	13
1.2 Διάβρωση Έργων Πολιτιστικής Κληρονομίας	15
1.2.1 Τύποι και Ιδιότητες Υποστρώματος	15
1.2.2 Περιβαλλοντικοί Παράγοντες Διάβρωσης	16
1.2.3 Βιολογικοί Παράγοντες Διάβρωσης - Οργανισμοί	18
1.2.4 Βιολογικοί Παράγοντες Διάβρωσης - Μικροοργανισμοί	18
<b>2 Κεφάλαιο 2: Μαθηματικό Υπόβαθρο</b>	<b>25</b>
2.1 Στατιστική μέθοδος ανάλυσης Principal Component Analysis (PCA)	26
2.1.1 Ανάλυση πίνακα σε ιδιάζουσες τιμές (Singular Value Decomposition-SVD)	26
2.1.2 Ανάλυση μεθόδου Principal Component Analysis (PCA)	28
2.2 Ιεραρχική Ομαδοποίηση κατά Συστάδες	33
2.3 Γραμμική Διακριτική Ανάλυση - LDA	35
2.3.1 Θεώρημα ταξινόμησης του Bayes	35
2.3.2 LDA με χρήση μιας μεταβλητής πρόβλεψης	35
2.3.3 LDA σε πολυμεταβλητά δεδομένα	36
2.3.4 Γεωμετρική Ερμηνεία LDA	37
2.4 Αποδοτικότητα Μοντέλου	40
<b>3 Κεφάλαιο 3: Μελέτη Βιοδιαβρωμένων Επιφανειών με την Τεχνική LIF</b>	<b>42</b>
3.1 Βιβλιογραφική Ανασκόπηση	43
3.2 Περιγραφή Πειράματος και Πειραματική Διάταξη	46
3.3 Βιολογική και Φασματική Ανάλυση Δειγμάτων Μελέτης	51
3.3.1 CALLOS CER-3	52
3.3.2 CALLOS MAR-2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1	56
3.3.3 CALLOS MAR-2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2	60
<b>4 Κεφάλαιο 4: Στατιστική Επεξεργασία Φασματικών Δεδομένων</b>	<b>64</b>
4.1 CALLOS CER 3	66
4.2 CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1	75
4.3 CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2	80
4.3.1 Βιολογική Κατανομή Κλάσεων	80
4.3.2 Φασματική Κατανομή Κλάσεων	85
4.4 Συμπεράσματα	87
<b>5 Επίλογος</b>	<b>89</b>

6 Βιβλιογραφία	90
A Παράρτημα: Στοιχεία Στατιστικής Πολυμεταβλητών Δεδομένων	95

## Λίστα Εικόνων

1	<i>Ενεργειακή κατανομή μοριακών τροχιακών φορμαλδεύδης[1].</i> . . . . .	4
2	<i>Κλασική (a) και κβαντομηχανική (b) περιγραφή του δυναμικού του μορίου κατά την μετάβαση ενός ηλεκτρονίου σύμφωνα με την αρχή Franck-Condon. Το μήκος του δεσμού <math>R_e</math> αυξάνεται κατά την διέγερση καθώς τα ηλεκτρόνια διεγείρονται σε αντιδεσμικά τροχιακά[2].</i> . . . . .	6
3	<i>Τυπικό φάσμα απορρόφησης το οποίο αντικατοπτρίζει τις ενεργειακές αποστάσεις των δονητικών επιπέδων μια μοριακής ηλεκτρονιακής κατάστασης[1].</i> . . . . .	7
4	<i>Ηλεκτρονιακή πυκνότητα κατά μήκος του μεγάλου άξονα των μορίων χλωροφύλλης a και b αντίστοιχα (a) και το διάγραμμα ροπής μετάβασης κάθετα στον δακτύλιο πορφυρίνης των δύο μορίων (b) [3]</i> . . . . .	9
5	<i>Φάσματα απορρόφησης της χλωροφύλλης a (αριστερά) και της χλωροφύλλης b (δεξιά) σύμφωνα με προσομοίωση TDDFT (Time Dependent Density Function Theory) σε κενό (μαύρη γραμμή με παύλες) και σε διάλυμα αιθανόλης (κόκκινη γραμμή)[3].</i> . . . .	9
6	<i>Διάγραμμα Jablonski[4].</i> . . . . .	9
7	<i>Πειραματικό Φάσμα Απορρόφησης και Φθορισμού της Χλωροφύλλης a[5].</i> . . . . .	11
8	<i>Αλληλεπικάλυψη φασμάτων απορρόφησης και φθορισμού σε θερμοκρασία περιβάλλοντος[1].</i>	11
9	<i>Χλωροφύκη</i> . . . . .	16
10	<i>Κυανοβακτήρια</i> . . . . .	16
11	<i>Αριθμός ειδών χλωροφυκών (αριστερά) και κυανοβακτηρίων (δεξιά) που ανιχνεύθηκαν σε μνημεία πολιτιστικής κληρονομιάς στην ευρύτερη περιοχή της Μεσογείου σε έξι διαφορετικά υποστρώματα[6]</i> . . . . .	16
12	<i>Η προέλευση και η κατάληξη των ρύπων επί των μνημείων[7]</i> . . . . .	17
13	<i>Γεωμετρική και μαθηματική περιγραφή της τυποποίησης δεδομένων. Μετά την διαδικασία της τυποποίησης τα δεδομένα έχουν μέση τιμή το 0 και τυπική απόκλιση ίση με 1.</i> . . .	28
14	<i>Γεωμετρική Αναπαράσταση αξόνων principal components (<a href="https://www.cheric.org/files/education/cyberlecture/d201401/d201401-401.pdf">https://www.cheric.org/files/education/cyberlecture/d201401/d201401-401.pdf</a>). . . . .</i>	30
15	<i>Σχηματική αναπαράσταση εξαγωγής στοιχείων της PCA μέσω της ανάλυσης του πίνακα τυποποιημένων δεδομένων με SVD.</i> . . . . .	32
16	<i>Απεικόνιση συναρτήσεων σύνδεσης, οι οποίες προσδιορίζουν τους τρεις βασικούς τρόπους εύρεσης της απόστασης μεταξύ των υποσυνόλων: ελάχιστη απόσταση μεταξύ των σημείων των συστάδων (αριστερά), μέγιστη απόσταση μεταξύ των συστάδων - διάμετρος (κέντρο), μέση τιμή απόστασης μεταξύ όλων των σημείων των συστάδων (δεξιά)[8]</i> . .	34
17	<i>Γραφική απεικόνιση της ιεραρχικής ομαδοποίησης κατά συστάδες ανάλογα με τον τρόπο σύνδεσης των υποσυνόλων σε κάθε βήμα της ταξινόμησης[9]</i> . . . . .	34
18	<i>Προβολές δεδομένων με μέσο m που ανήκουν σε τρεις κλάσεις με μέσους <math>m_1</math>, <math>m_2</math> και <math>m_3</math> και προβολές μέσων <math>\mu</math>, <math>\mu_1</math>, <math>\mu_2</math> και <math>\mu_3</math> αντίστοιχα στην βέλτιστη κατεύθυνση [10].</i> . . . .	37
19	<i>Οπτική περιγραφή συστήματος Hyperspectral Fluorescence LIDAR [11]</i> . . . . .	45
20	<i>Σημειακή Πειραματική Διάταξη LIF</i> . . . . .	47
21	<i>Σχηματικό χαρτογράφησης δειγμάτων. Οι λήψεις των φασμάτων πραγματοποιήθηκαν στις κορυφές των ορθογωνίων με κατάλληλο βήμα <math>\Delta x</math> κατά μήκος και <math>\Delta y</math> καθ' ύψος σε κάθε δείγμα.</i> . . . . .	49
22	<i>Σχηματικό Τυπικής Διάταξης Fluorescence LiDAR σταθερού στόχου [12]</i> . . . . .	50
23	<i>Βρύο γένους Tortula</i> . . . . .	52
24	<i>Ύλην σπορίων μυκήτων και πράσινα στίγματα που υποδεικνύουν την ύπαρξη χλωροφυκών</i>	52
25	<i>CALLOS CER-3</i> . . . . .	53
26	<i>Περιοχές αναφοράς φασμάτων φθορισμού δείγματος CALLOS CER-3</i> . . . . .	53

27	<i>Η ομάδα μηδέν των φασμάτων αναφοράς του δείγματος CALLOS CER-3 αντιστοιχεί σε έντονη βιολογική επιβάρυνση από φωτο-αυτότροφους μικροοργανισμούς</i>	55
28	<i>Η ομάδα ένα των φασμάτων αναφοράς του δείγματος CALLOS CER-3 αντιστοιχεί σε έντονη βιολογική επιβάρυνση από φωτο-αυτότροφους μικροοργανισμούς όπως επίσης και από μικροοργανισμούς με ευρύ φάσμα στην περιοχή 440-600 nm όπως σπόρια μυκήτων</i>	55
29	<i>Η ομάδα δύο των φασμάτων αναφοράς του δείγματος CALLOS CER-3 αντιστοιχεί σε ασθενή βιολογική επιβάρυνση από φωτο-αυτότροφους μικροοργανισμούς</i>	55
30	<i>Βιολογική δειγματοληψία στην επιφάνεια του δείγματος CALLOS MAR-2. Τα σφαιρικά κίτρινα στίγματα είναι σπόρια μυκήτων, οι πράσινες και κίτρινες περιοχές καταδυναμώνουν ύπαρξη χλωροφυκών και πιθανώς κυανοβακτηρίων και οι μαύρες αποικίες υποδηλώνουν ανάπτυξη μαύρων μεριστευματικών μυκήτων.</i>	56
31	<i>Περιοχές αναφοράς φασμάτων φθορισμού δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i>	56
32	<i>Η ομάδα μηδέν των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i>	58
33	<i>Η ομάδα ένα των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i>	58
34	<i>Η ομάδα δύο των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i>	59
35	<i>Η ομάδα τρία των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i>	59
36	<i>Η ομάδα τέσσερα των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i>	59
37	<i>Περιοχές αναφοράς φασμάτων φθορισμού δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i>	61
38	<i>Η ομάδα μηδέν των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i>	61
39	<i>Η ομάδα ένα των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i>	62
40	<i>Η ομάδα δύο των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i>	62
41	<i>Η ομάδα τρία των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i>	62
42	<i>Οι ομάδες των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 σύμφωνα με την φασματική ομοιότητα</i>	63
43	<i>Διαγραμμα Ροής Στατιστικής Επεξεργασίας Δεδομένων</i>	65
44	<i>Κρυμνογράφημα τυποποιημένων και μη τυποποιημένων δεδομένων</i>	67
45	<i>Τα φορτία των κύριων συνιστωσών τα οποία προκύπτουν από τον πίνακα δεδομένων του δείγματος CALLOS CER-3</i>	68
46	<i>Score plots των παρατηρήσεων που αντιστοιχούν στο δείγμα CALLOS CER 3. Η χρωματική κλίμακα αντιστοιχεί στην περιεκτικότητα χλωροφύλλης του σημείου μέτρησης. Από το ανοιχτόχρωμο πράσινο προς το σκουρόχρωμο, η τιμή του ολοκληρώματος της συνάρτησης της έντασης φθορισμού στο εύρος 650-800 nm αυξάνεται</i>	69
47	<i>Δενδροδιάγραμμα περιοχών αναφοράς δείγματος CALLOS CER-3. Οι περιοχές A-M αντιστοιχούν στους αριθμούς 0-12</i>	70
48	<i>Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS CER-3</i>	71

49	<i>Confusion Matrix</i> μεταξύ προβλεπόμενης και "ψευδοπαρατηρούμενης" κλάσης των φασμάτων της χαρτογράφησης του CALLOS CER-3 . . . . .	72
50	Score Plots δεδομένων δείγματος CALLOS CER-3 στα οποία η χρωματική κλίμακα αντιστοιχεί στις κλάσεις που προβλέφθηκαν από τον αλγόριθμο της LDA. Στις διαγωνίους παρουσιάζεται η συνάρτηση πυκνότητας της κάθε κλάσης σύμφωνα με τις τιμές του component της αντίστοιχης στήλης . . . . .	74
51	<i>Κρυμνογράφημα τυποποιημένων δεδομένων του δείγματος CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i> . . . . .	75
52	<i>Τα φορτία των κύριων συνιστωσών τα οποία προκύπτουν από τον πίνακα δεδομένων του δείγματος CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i> . . . . .	76
53	<i>Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1</i> . . . . .	77
54	<i>Confusion Matrix</i> μεταξύ προβλεπόμενης και πραγματικής κλάσης των φασμάτων της αναφοράς του CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 με χρήση της μεθόδου <i>Leave-One-Out Cross Validation (LOO)</i> . . . . .	78
55	Score Plots δεδομένων δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 στα οποία η χρωματική κλίμακα αντιστοιχεί στις κλάσεις που προβλέφθηκαν από τον αλγόριθμο της LDA. Στις διαγωνίους παρουσιάζεται η συνάρτηση πυκνότητας της κάθε κλάσης σύμφωνα με τις τιμές του component της αντίστοιχης στήλης . . . . .	79
56	<i>Κρυμνόγραφημα τυποποιημένων δεδομένων του δείγματος CALLOS MAR 2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i> . . . . .	80
57	<i>Τα φορτία των κύριων συνιστωσών τα οποία προκύπτουν από τον πίνακα δεδομένων του δείγματος CALLOS MAR 2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i> . . . . .	81
58	<i>Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2</i> . . . . .	82
59	<i>Confusion Matrix</i> μεταξύ προβλεπόμενης και πραγματικής κλάσης των φασμάτων της αναφοράς του CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 με χρήση της μεθόδου <i>Leave-One-Out Cross Validation (LOO)</i> . . . . .	83
60	Score Plots δεδομένων δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 στα οποία η χρωματική κλίμακα αντιστοιχεί στις κλάσεις που προβλέφθηκαν από τον αλγόριθμο της LDA. Στις διαγωνίους παρουσιάζεται η συνάρτηση πυκνότητας της κάθε κλάσης σύμφωνα με τις τιμές του component της αντίστοιχης στήλης . . . . .	84
61	<i>Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2(Φασματική Κατανομή Κλάσεων)</i> . . . . .	85



## Λίστα Πινάκων

1	Σημαντικοί ατμοσφαιρικοί ρύποι, οι φυσικές και οι τεχνητές πηγές τους καθώς και η επίδραση τους στα μνημεία πολιτιστικής κληρονομιάς[7]	17
2	Περιοχές Αναφοράς Μετρήσεων CALLOS CER-3 και αντίστοιχη οπτική περιγραφή	54
3	Περιοχές Αναφοράς Μετρήσεων CALLOS MAR-2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 και αντίστοιχη οπτική περιγραφή	57
4	Περιοχές Αναφοράς Μετρήσεων CALLOS MAR-2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 και αντίστοιχη οπτική περιγραφή	60
5	Επεξηγούμενη διασπορά για κάθε component για τυποποιημένα και μη τυποποιημένα δεδομένα	66
6	Πίνακας τιμών - scores των principal components στις περιοχές αναφοράς του δείγματος CALLOS CER-3	68
7	Περιοχές Αναφοράς δείγματος CALLOS CER-3, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης	70
8	Περιοχές αναφοράς, οι αντίστοιχοι λόγοι μέγιστης έντασης φθορισμού στα εύρη 430-600nm/412nm και 650-800nm/412nm και η διαφορά τους, καθώς και η κλάση στην οποία ανήκουν	71
9	Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση των μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS CER-3	73
10	Περιοχές Αναφοράς δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης	76
11	Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση το μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1	78
12	Περιοχές Αναφοράς δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης	81
13	Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση των μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2	83
14	Περιοχές Αναφοράς δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 με βάση την φασματική τους ομοιότητα, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης	85
15	Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση των μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2	86

## Περίληψη

Η παρούσα διπλωματική εργασία έχει ως βασικό αντικείμενο μελέτης την εφαρμογή στατιστικών μεθόδων επεξεργασίας δεδομένων επάνω στην φασματοσκοπία Φθορισμού με Laser. Κύριος στόχος της συγκεκριμένης μελέτης αποτέλεσε η ανάκτηση φασμάτων φθορισμού με χρήση της Τεχνικής Επαγόμενου Φθορισμού με Laser (LIF) και στη συνέχεια η επεξεργασία τους με κατάλληλους αλγόριθμους από τον κλάδο της Επιστήμης των Δεδομένων.

Τα προαναφερόμενα φάσματα προέρχονται από βιολογικά επιβαρυμένες επιφάνειες μνημείων πολιτιστικής κληρονομιάς. Η στατιστική τους επεξεργασία σε συνδυασμό με κατάλληλη βιολογική πληροφορία, στοχεύει στον διαχωρισμό της βιολογικής τους επιβάρυνσης σε ξεχωριστές ομάδες. Με βάση αυτή την οπτική κατηγοριοποίηση της επιβάρυνσης, η εργασία του συντηρητή αρχαιοτήτων αποδεσμεύεται από την ανάγκη πολλαπλής δειγματοληψίας από το σύνολο του μνημείου. Η χρήση του Fluorescence LIDAR, δηλαδή της επέκτασης της LIF στην τηλεπισκόπηση, δίνει την δυνατότητα στους συντηρητές να χαρτογραφήσουν το μνημείο και να ομαδοποιήσουν τις επιβαρύνσεις του ώστε να τις αντιμετωπίσουν πιο αποτελεσματικά.

Στο κεφάλαιο 1 αναλύεται το φυσικό υπόβαθρο της αλληλεπίδρασης της ακτινοβολίας με την ύλη και πραγματοποιείται η κατάλληλη εμβάθυνση στο φαινόμενο του Φθορισμού. Επιπλέον, επεξηγούνται οι βασικοί παράγοντες διάβρωσης των επιφανειών των μνημείων καθώς επίσης και τα αποτελέσματα που επιφέρουν.

Στο κεφάλαιο 2 παρουσιάζεται αναλυτικά το μαθηματικό υπόβαθρο των στατιστικών μεθόδων που χρησιμοποιήθηκαν. Ειδικότερα, αναλύονται οι αλγόριθμοι της Principal Component Analysis (PCA), της Hierarchical Agglomerative Clustering (HAC) και της Linear Discriminant Analysis (LDA). Επιπλέον, περιγράφεται και μια μέθοδος αξιολόγησης μοντέλων πρόβλεψης κλάσης, η μέθοδος Leave-One-Out Cross Validation (LOO).

Στο κεφάλαιο 3 καταγράφεται η βασική βιβλιογραφική ανασκόπηση επάνω στο συγκεκριμένο αντικείμενο και παρουσιάζεται η πειραματική διάταξη, η οποία χρησιμοποιήθηκε. Επίσης, παρουσιάζεται η ανάλυση των φασμάτων φθορισμού με βάση την βιολογική πληροφορία του σημείου μέτρησης. Επιπλέον, πραγματοποιείται η ομαδοποίηση ορισμένων χαρακτηριστικών φασμάτων σε κατάλληλες ομάδες επιβάρυνσης. Αυτές θα αποτελέσουν την βάση για την δημιουργία ενός μοντέλου πρόβλεψης του τύπου επιβάρυνσης σε κάθε σημείο μέτρησης.

Στο κεφάλαιο 4, περιγράφεται βήμα προς βήμα η στατιστική ανάλυση των φασμάτων φθορισμού και η κατηγοριοποίηση τους σε κατάλληλες κλάσεις βιολογικής επιβάρυνσης. Η κατηγοριοποίηση προκύπτει με βάση την δημιουργία ενός μοντέλου πρόβλεψης, το οποίο στη συνέχεια αξιολογείται. Τέλος, παρουσιάζονται τα συμπεράσματα που ανακύπτουν από την συγκεκριμένη μελέτη καθώς επίσης και οι δυνατότητες για μελλοντική έρευνα.

## Abstract

The main object of this thesis is the application of statistical data processing methods on Laser Fluorescence Spectroscopy. The aim of this study was the recovery of fluorescence spectra using the Laser-Induced Fluorescence Technique (LIF) and then the spectrum analysis with appropriate algorithms from the field of Data Science.

The spectra mentioned above come from biologically loaded surfaces of cultural heritage monuments. Their statistical processing in combination with appropriate biological information aims to separate their biological burden into distinctive groups. Based on this categorization of the burden, the conservator's work disengages from the need for multiple sampling from the entire monument. The use of Fluorescence LIDAR, namely the extension of LIF to remote sensing, enables the conservators to map the monument and group its biological burdens to deal with them more effectively.

The 1st Chapter is an introduction to the physical background of the interaction of radiation with matter with an emphasis on the fluorescence phenomenon. In addition, the critical factors of monument's surface erosion as well as its effects are introduced.

The 2nd Chapter presents in detail the mathematical background of the statistical methods that are used in this thesis. In particular, the algorithms of Principal Component Analysis (PCA), Hierarchical Agglomerative Clustering (HAC) and Linear Discriminant Analysis (LDA) are analyzed. In addition, the chapter includes the description of a method for evaluating class predictive models, the Leave- One-Out Cross Validation (LOO) method.

The 3rd Chapter begins with a basic literature review on the application of LIF on monuments and continues with the experimental setup which was used. Moreover, the analysis of fluorescence spectra based on the biological information of the measuring point is presented and certain characteristic spectra are grouped into appropriate classes. These classes are the basis for the creation of a predictive model about the biological burden type at each measuring point.

The 4th Chapter describes step-by-step the statistical analysis of fluorescence spectra and their categorization into appropriate biological burden classes. This categorization is based on the creation of the aforementioned model, which is then evaluated with the LOO method. Finally, the conclusions that emerge from this study are presented as well as opportunities for future research.

# 1 Κεφάλαιο 1: Εισαγωγικά Στοιχεία

Η μελέτη των μνημείων πολιτιστικής κληρονομιάς από την σκοπιά της φασματοσκοπίας φθορισμού απαιτεί την θεμελίωση του κατάλληλου θεωρητικού υποβάθρου σχετικά με την αλληλεπίδραση της ακτινοβολίας με την ύλη. Βασικά φαινόμενα όπως η απορρόφηση, η εσωτερική μετατροπή, ο φθορισμός, η διασυστημική διέλευση και ο φωσφορισμός συντελούνται κατά την διάρκεια αυτής της αλληλεπίδρασης. Συνεπώς, η πρώτη ενότητα του παρόντος κεφαλαίου αφιερώνεται στην θεωρητική περιγραφή των φυσικών διεργασιών που διέπουν την φασματοσκοπία φθορισμού.

Η δεύτερη ενότητα του κεφαλαίου 1 στοχεύει στην ανάλυση των παραγόντων διάβρωσης, οι οποίοι επάγουν την διάβρωση των επιφανειών στα μνημεία πολιτιστικής κληρονομιάς. Ειδικότερα, οι περιβαλλοντικές και κλιματολογικές συνθήκες στις οποίες υφίσταται το μνημείο αποτελούν πρωτεύουσας σημασίας παράγοντα, καθώς οι έντονα μολυσμένες αστικές περιοχές συντρέχουν στην ταχύτερη καταστροφή των επιφανειών. Επιπλέον, η βιολογική επιβάρυνση της επιφάνειας, η οποία αποτελεί και το αντικείμενο μελέτης της παρούσας διπλωματικής εργασίας, επιφέρεται από ποικίλους οργανισμούς και μικροοργανισμούς οι οποίοι είναι είτε αερομεταφερόμενοι είτε υπάρχουν στο έδαφος περιμετρικά του μνημείου. Τόσο η καταγραφή αυτών των παραγόντων όσο και η καταστροφική τους δράση αποτελεί το αντικείμενο της δεύτερης ενότητας.

## 1.1 Φυσικές Διεργασίες

### 1.1.1 Κβαντομηχανικές Αρχές Αλληλεπίδρασης Ακτινοβολίας με Ύλη

Η ύλη σε ατομικό και μοριακό επίπεδο διέπεται από τους νόμους της Κβαντομηχανικής οι οποίοι απορρέουν αυστηρά από την επίλυση της εξίσωσης του Schrödinger για κάθε επιμέρους σύστημα που μελετάται, η οποία είναι:

$$i\hbar \frac{\partial |\Psi\rangle}{\partial t} = \hat{H}|\Psi\rangle \quad (1)$$

με την αντίστοιχη εξίσωση ιδιοτιμών ενέργειας:

$$\hat{H}|\Psi(r_1, r_2, r_3, \dots, n_1, n_2, n_3, \dots)\rangle = E|\Psi(r_1, r_2, r_3, \dots, n_1, n_2, n_3, \dots)\rangle \quad (2)$$

όπου  $E$  η ενέργεια της ενεργειακής κατάστασης η οποία περιγράφεται από την κυματοσυνάρτηση  $|\Psi\rangle$ . Η γνώση της  $|\Psi\rangle$  εξασφαλίζει μια πλήρη περιγραφή της κατάστασης του συστήματος. Τα ορίσματα  $r_i$  και  $n_i$  αντιστοιχούν στις θέσεις των ηλεκτρονίων και των πυρήνων του συστήματος. Το αντίστοιχο ατομικό τροχιακό προκύπτει από την επίλυση της εξίσωσης 1 για την χαμιλτονιανή  $\hat{H}$  με τύπο:

$$\hat{H} = -\frac{\hbar^2}{2m_e} \sum_i^n \nabla_i^2 - \sum_i^n \sum_I^N \frac{Z_I e^2}{4\pi\epsilon_0 r_{Ii}} + \frac{1}{2} \sum_{ij}^n \frac{e^2}{4\pi\epsilon_0 r_{ij}} \quad (3)$$

όπου σύμφωνα με την προσέγγιση Born-Oppenheimer [2] η αμελητέα μάζα των ηλεκτρονίων συγκριτικά με την μάζα του πυρήνα προκαλεί μια στιγμιαία αντίδραση των ηλεκτρονίων στην οποιαδήποτε μετατόπιση των πυρήνων. Συνεπώς η επίλυση της εξίσωσης 2 θεωρεί ως παραδοχή την σταθερή θέση των πυρήνων και επομένως το δυναμικό Coulomb θεωρείται στατικό. Για τον λόγο αυτό, στη σχέση 3 ο πρώτος όρος αντιστοιχεί στην κινητική ενέργεια των ηλεκτρονίων, ο δεύτερος στην έλξη Coulomb από τους πυρήνες και ο τρίτος στην άπωση Coulomb μεταξύ των ηλεκτρονίων.

Καθώς το φαινόμενο της αλληλεπίδρασης της ηλεκτρομαγνητικής ακτινοβολίας με την ύλη επηρεάζει την κινητικότητα των ηλεκτρονίων μεταξύ των ατομικών ενεργειακών επιπέδων με ενέργεια  $E$ , τότε είναι απαραίτητο να είναι γνωστές οι κυματοσυναρτήσεις των αντίστοιχων τροχιακών. Επιπρόσθετα, η κυματοσυνάρτηση  $|\Psi\rangle$  της κατάστασης του συστήματος πρέπει να είναι γνωστή καθώς προσδιορίζει την πυκνότητα πιθανότητας εύρεσης ηλεκτρονίων σε μια συγκεκριμένη θέση του χώρου:

$$\rho_e = \langle \Psi | \Psi \rangle \quad (4)$$

Η διεργασία εκκίνησης της αλληλεπίδρασης της ακτινοβολίας με την ύλη αντιστοιχεί στην απορρόφηση των φωτονίων από τα μόρια. Κατά την διαδικασία της απορρόφησης, ένα φωτόνιο κατάλληλης ενέργειας διεγείρει ένα ηλεκτρόνιο ενός μοριακού τροχιακού, το οποίο στη συνέχεια αξιοποιεί την προσφερόμενη ενέργεια και πραγματοποιεί μετάβαση σε ένα υψηλότερο ενεργειακά μοριακό τροχιακό. Τα βασικά μοριακά τροχιακά, τα οποία κατέχουν τον καθοριστικό ρόλο τόσο στην προαναφερόμενη διαδικασία όσο και στην αντίστοιχη διαδικασία εκπομπής είναι το Υψηλότερο Κατειλημμένο Μοριακό Τροχιακό (HOMO-Highest Occupied Molecular Orbital) και το Χαμηλότερο Μη Κατειλημμένο Μοριακό Τροχιακό (LUMO-Lowest Unoccupied Molecular Orbital), τα οποία δημιουργούνται κατά την σύνθεση του δεσμού μεταξύ των ατόμων που αποτελούν το μόριο. Οι δεσμοί αυτοί δύναται να είναι είτε τύπου  $\sigma$  είτε τύπου  $\pi$ , οι οποίοι θα περιγραφούν στη συνέχεια. Συνεπώς, απαιτείται η διεξοδική ανάλυση τόσο των τύπων των μοριακών τροχιακών όσο και των κανόνων που διέπουν την μετακίνηση ηλεκτρονίων μεταξύ των ενεργειακών επιπέδων.

Σύμφωνα με την θεωρία μοριακών τροχιακών, τα ηλεκτρόνια σε ένα μόριο είναι απεντοπισμένα από τον πυρήνα που αντιστοιχούν και συνεπώς η ηλεκτρονιακή πυκνότητα (σχέση 4) είναι κατανομημένη σε ολόκληρη την έκταση του μορίου. Οι κυματοσυναρτήσεις των μοριακών τροχιακών, οι οποίες προσδιορίζουν και την πιθανότητα εύρεσης ενός ηλεκτρονίου σε μια συγκεκριμένη θέση, σύμφωνα με την προσέγγιση LCAO (Linear Combination of Atomic Orbitals) [2], αποτελούν έναν γραμμικό συνδυασμό των επιμέρους ατομικών τροχιακών των ατόμων που συνθέτουν το μόριο. Τα ατομικά τροχιακά προκύπτουν από την επίλυση της εξίσωσης Schrödinger για κάθε ένα άτομο ξεχωριστά. Τα μοριακά τροχιακά διακρίνονται σε  $\sigma$  και  $\pi$  τροχιακά ανάλογα με τον άξονα αλληλεπικάλυψης των ατομικών τροχιακών που τα απαρτίζουν. Ειδικότερα, ένα  $\sigma$  τροχιακό δύναται να σχηματιστεί μέσω τριών διαφορετικών συνδυασμών, δηλαδή είτε από δύο  $s$  ατομικά τροχιακά, είτε από ένα  $s$  ατομικό τροχιακό και ένα  $p$  ατομικό τροχιακό, είτε από δύο  $p$  τροχιακά με γραμμικό άξονα συμμετρίας. Ο αντίστοιχος δεσμός ονομάζεται  $\sigma$  δεσμός και διαφέρει ως προς τον  $\pi$  δεσμό αναφορικά με την υψηλότερη ενέργεια που τον χαρακτηρίζει. Ο  $\pi$  δεσμός, ο οποίος προσδιορίζεται από το αντίστοιχο  $\pi$  μοριακό τροχιακό σχηματίζεται από την πλευρική αλληλεπικάλυψη των  $p$  τροχιακών. Στη συνέχεια, θα περιγραφούν οι προαναφερόμενοι δεσμοί για ένα διατομικό μόριο, για λόγους απλούστευσης και κατανόησης.

## Μοντέλο διατομικού μορίου

### Δεσμός $\sigma$

Κατά την δημιουργία οποιουδήποτε δεσμού δημιουργούνται δύο σύνολα τροχιακών, τα δεσμικά και τα αντιδεσμικά, η κατάληψη ή μη των οποίων προσδιορίζει και την σταθερότητα του μορίου. Πιο συγκεκριμένα, σε έναν  $\sigma$  δεσμό, δημιουργείται το  $\sigma$  δεσμικό τροχιακό και το  $\sigma^*$  αντιδεσμικό τροχιακό, των οποίων η ενεργειακή διαφορά προσδιορίζει το κέρδος ενέργειας από την δημιουργία του δεσμού. Όταν ένα δεσμικό τροχιακό είναι συμπληρωμένο τότε το σύστημα επέρχεται σε χαμηλότερη ενέργεια προ της δημιουργίας του δεσμού, ενώ αν συμπληρωθεί και το αντιδεσμικό τροχιακό ή ηλεκτρόνια από το δεσμικό ανέλθουν στο αντιδεσμικό, τότε το σύστημα βρίσκεται σε υψηλότερη ενέργεια συγκριτικά με την απουσία του δεσμού και άρα προκύπτει μια λιγότερο σταθερή κατάσταση.

Στο  $\sigma$  μοριακό τροχιακό, το μέγιστο της ηλεκτρονιακής πυκνότητας βρίσκεται μεταξύ των δύο ατόμων που συνθέτουν το μόριο, κατά μήκος του δεσμικού τους άξονα, διάταξη η οποία ελαχιστοποιεί την ενέργεια του συστήματος καθώς τα ηλεκτρόνια προασπίζουν τους επιμέρους πυρήνες και μειώνουν την άπωση μεταξύ τους. Επομένως, στο συγκεκριμένο σημείο, οι κυματοσυναρτήσεις των ατομικών τροχιακών συμβάλλουν ενισχυτικά. Αντίθετα, στο  $\sigma^*$  μοριακό τροχιακό, οι αντίστοιχες κυματοσυναρτήσεις συμβάλλουν αναιρετικά μεταξύ των δύο πυρήνων και επομένως δημιουργείται ένα σημείο μηδενικής πιθανότητας εύρεσης των ηλεκτρονίων. Συνεπώς, η ενέργεια του συστήματος αυξάνεται καθώς ενισχύεται η άπωση μεταξύ των πυρήνων και επακόλουθα η αστάθεια του μορίου. Ο  $\sigma$  δεσμός είναι πάντοτε ο πρώτος δεσμός που σχηματίζεται σε ένα μόριο.

### Δεσμός $\pi$

Ο δεσμός  $\pi$  προκύπτει όταν σε ένα μόριο εμφανίζεται επιπρόσθετη αλληλεπικάλυψη σε ατομικά τροχιακά και συνεπώς εμφανίζεται είτε διπλός είτε τριπλός δεσμός. Δημιουργείται από πλευρική επικάλυψη  $p$  ατομικών τροχιακών και κατ' αντιστοιχία με τον  $\sigma$  δεσμό, απαρτίζεται από δύο σύνολα δεσμικών και αντιδεσμικών τροχιακών. Στο  $\pi$  δεσμικό τροχιακό η ηλεκτρονιακή πυκνότητα κατανέμεται εκατέρωθεν του δεσμικού άξονα και οι αντίστοιχες κυματοσυναρτήσεις συμβάλλουν ενισχυτικά και στους δύο ημιχώρους. Καθώς τα  $p$  ατομικά τροχιακά δεν εμφανίζουν ηλεκτρονιακή πυκνότητα

επί του πυρήνα, τότε αντίστοιχα και το π δεσμικό τροχιακό παρουσιάζει κόμβο πιθανότητας στον δεσμικό άξονα. Καθώς όμως τα ηλεκτρόνια περιβάλλουν τους πυρήνες σε μορφή λοβών, τότε η ενισχυτική συμβολή των κυματοσυναρτήσεων τοποθετεί ένα νέφος ηλεκτρονίων στον θετικό ημιχώρο με αποτέλεσμα να δημιουργούνται ελκτικές δυνάμεις με τους πυρήνες και το σύστημα να μειώνει τελικά την ενέργεια του.

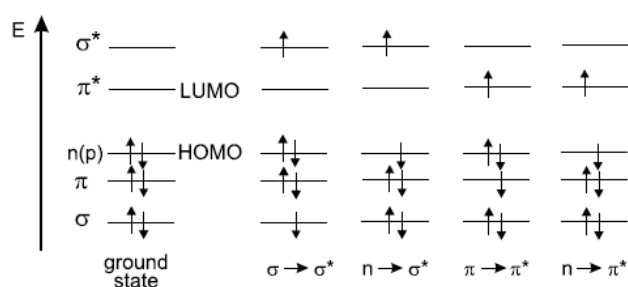
Αντίθετα, στο π\* αντιδεσμικό τροχιακό οι κυματοσυναρτήσεις των επιμέρους ατομικών τροχιακών είναι εκτός φάσης και συνεπώς συμβάλλουν αναιρετικά και στους δύο ημιχώρους. Αυτό έχει ως αποτέλεσμα την δημιουργία ενός δευτερεύοντος κόμβου πιθανότητας ύπαρξης ηλεκτρονίων μεταξύ των ατόμων και συνεπώς την ενίσχυση των απώσεων μεταξύ των πυρήνων. Με αυτήν την ηλεκτρονιακή διαμόρφωση, η ενέργεια του συστήματος αυξάνεται και συνεπώς η σταθερότητα του μορίου ελαττώνεται. Η κατανομή των ηλεκτρονίων στα μοριακά τροχιακά, είτε αυτά είναι σ είτε π, δημιουργούν τα LUMO και HOMO μοριακά τροχιακά, όπως αυτά παρουσιάζονται στο Σχήμα 1. Η διέγερση ενός ηλεκτρονίου από ένα σ τροχιακό απαιτεί σημαντικά υψηλότερη ενέργεια φωτονίων σε σχέση με την διέγερση ενός ηλεκτρονίου από π τροχιακό[1].

### Ηλεκτρονιακές Μεταβάσεις σε Πολυατομικά Μόρια

Σε ένα πολυατομικό μόριο, εκτός από τα σ και π τροχιακά, δύναται να υπάρχουν και μη δεσμικά ηλεκτρόνια από άλλα άτομα όπως το οξυγόνο και το άζωτο. Τα αντίστοιχα μοριακά τροχιακά ονομάζονται n και πλέον οι δυνατές μεταβάσεις ηλεκτρονίων με αύξουσα σειρά ενεργειακής απαίτησης [1] θα είναι:

$$n \rightarrow \pi^* < \pi \rightarrow \pi^* < n \rightarrow \sigma^* < \sigma \rightarrow \pi^* < \sigma \rightarrow \sigma^* \quad (5)$$

Στο διάγραμμα του Σχήματος 1 παρουσιάζονται τα μοριακά τροχιακά του μορίου της φορμαλδεΐδης τόσο στην βασική τους κατάσταση όσο και στην πρώτη διεγερμένη σε κάθε περίπτωση μετάβασης. Παρατηρείται πως η ενεργειακή διαφορά μεταξύ του αρχικού και του τελικού μοριακού τροχιακού κατά την διέγερση ενός ηλεκτρονίου υπακούει στη σχέση 1. Σε κάθε διαφορετικό μόριο, τα LUMO και HOMO μπορεί να είναι διαφορετικά τροχιακά από το π\* και n στην φορμαλδεΐδη, ανάλογα με την κατανομή των ηλεκτρονίων.



Σχήμα 1: Ενεργειακή κατανομή μοριακών τροχιακών φορμαλδεΐδης[1].

Ένας επιπρόσθετος παράγοντας ο οποίος έχει σημαντική επιρροή στις ηλεκτρονιακές μεταβάσεις είναι η πολλαπλότητα του χβαντικού αριθμού του spin S που δύναται να έχει το ηλεκτρόνιο ανάλογα με το τροχιακό στο οποίο βρίσκεται. Ειδικότερα, σύμφωνα με τους κανόνες επιλογής, οι οποίοι προκύπτουν από την διαδικασία επίλυσης της εξίσωσης Schrödinger, θα πρέπει η πολλαπλότητα  $M=2S+1$  του ηλεκτρονιακού spin μεταξύ δύο επιπέδων μετάβασης να παραμείνει αμετάβλητη [2], διότι στη σχέση 3 δεν εμφανίζεται ο τελεστής του spin. Επομένως, ορίζονται οι καταστάσεις singlet στις οποίες ο ολικός αριθμός spin είναι  $S=(-1/2)+(1/2)=0$  για αντιπαράλληλα spins, η πολλαπλότητα  $M = 1$ , και

συμβολίζονται ως  $S_0$  για την βασική κατάσταση,  $S_1$  για την πρώτη διεγερμένη κλπ. Οι καταστάσεις αυτές ορίζουν και τις singlet-singlet μεταβάσεις. Στην περίπτωση των παράλληλων spins, ορίζονται οι triplet καταστάσεις με συνολικό αριθμό spin  $S=(1/2)+(1/2)=1$  και άρα πολλαπλότητα  $M = 3$ . Οι μεταβάσεις που αντιστοιχούν σε αυτά τα επίπεδα ονομάζονται triplet-triplet μεταβάσεις. Για την διατήρηση επομένως της πολλαπλότητας οι singlet-triplet μεταβάσεις θεωρούνται απαγορευμένες και επιπλέον σύμφωνα με τον Κανόνα του Hund μια singlet κατάσταση έχει μικρότερη ενέργεια από μια αντίστοιχη triplet.

### 1.1.2 Αρχή Franck-Condon - Θεωρητικό Υπόβαθρο Απορρόφησης

Η κατανομή των ηλεκτρονίων στα μοριακά τροχιακά προσδίδει ενέργεια ηλεκτρομαγνητικής αλληλεπίδρασής στο σύστημα, η οποία μεταφράζεται ως ελκτική ή απωστική ενέργεια Coulomb. Όμως η βαθμιά ελευθερίας κινήσεως του μορίου και των περιλαμβανομένων ηλεκτρονίων αντιστοιχούν σε επιπρόσθετες μορφές κινητικής ενέργειας, οι οποίες αντιστοιχούν σε υποεπίπεδα των προαναφερόμενων ενεργειακών επιπέδων. Επεξηγηματικά, η διάσπαση των ενεργειακών επιπέδων τα οποία αντιστοιχούν στον κύριο κβαντικό αριθμό και στον κβαντικό αριθμό της στροφορμής προκύπτει εξαιτίας:

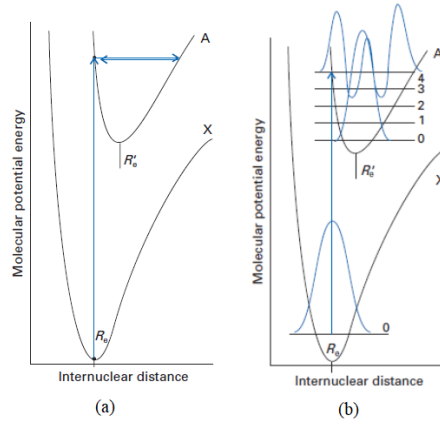
1. της περιστροφικής ενέργειας των ηλεκτρονίων λόγω τροχιακής γωνιακής στροφορμής
2. της δονητικής ενέργειας των ατόμων εντός του μορίου εξαιτίας της περιοδικής μετατόπισης από την θέση ισορροπίας
3. του προσανατολισμού του spin κατά την αλληλεπίδραση με ηλεκτρομαγνητικό πεδίο [2, 13].

Όταν πραγματοποιείται μια ηλεκτρονική μετάβαση σε ένα υψηλότερο ενεργειακό επίπεδο του μορίου, οι πυρήνες αντιλαμβάνονται μια αλλαγή στην ηλεκτρομαγνητική δύναμη την οποία δέχονται εξαιτίας της ανακατανομής του φορτίου στον χώρο. Επομένως, κατά την διάρκεια της μετάβασης διαφοροποιείται η δυναμική ενέργεια του συστήματος και οι πυρήνες τείνουν να εξαλείψουν την αύξηση που επιβλήθηκε μέσω έντονης αύξησης της ταλαντωτικής τους κίνησης γύρω από την θέση ισορροπίας τους. Συνεπώς, το φάσμα απορρόφησης ενός μορίου παρουσιάζει μια δομή χαρακτηριστική των δονητικών επιπέδων της κάθε κατάστασης που μπορεί να βρεθεί το σύστημα.

Η μορφή των κυματοσυνάρτησεων των δονητικών καταστάσεων υπακούει στην αρχή Franck-Condon, σύμφωνα με την οποία μια ηλεκτρονική μετάβαση πραγματοποιείται σε ένα ακίνητο πυρηνικό σύστημα αναφοράς εξαιτίας της αμελητέας μάζας των ηλεκτρονίων συγκριτικά με την μάζα των πυρήνων. Η συγκεκριμένη προσέγγιση δύναται να λάβει τόσο κλασσική όσο και κβαντομηχανική επεξήγηση. Σύμφωνα με το Σχήμα 2, από κλασσικής σκοπιάς η μετάβαση μπορεί να πραγματοποιηθεί όταν οι αποστάσεις των πυρήνων ισοδυναμούν με το μήκος ισορροπίας δεσμού  $R_e$  του μορίου στην βασική του κατάσταση. Όσο πραγματοποιείται η μετάβαση από το πρώτο δονητικό επίπεδο της βασικής κατάστασης, η θέση του πυρήνα δεν μεταβάλλεται και δεν επιταχύνεται καθώς η διάρκεια της μετάβασης είναι της τάξης του  $10^{-15}$  s [4] ενώ η πυρηνική ταλάντωση έχει περίοδο της τάξης του  $10^{-10}$  –  $10^{-12}$  s [1]. Η μετάβαση ολοκληρώνεται όταν η ενέργεια του ηλεκτρονίου για την ίδια θέση του πυρήνα ταυτιστεί με την δυναμική ενέργεια της πρώτης διεγερμένης κατάστασης. Μετά το πέρας της μετάβασης, το ηλεκτρόνιο ταλαντώνεται με ενέργεια που αντιστοιχεί στην τομή της κάθετης γραμμής με την καμπύλη του δυναμικού της διεγερμένης κατάστασης (Σχήμα 2a).

Υπό την κβαντομηχανική σκοπιά, η συνολική κυματοσυνάρτηση του συστήματος μπορεί να περιγραφεί από το γινόμενο της ηλεκτρονιακής κυματοσυνάρτησης  $|\Psi\rangle$  και της δονητικής συνιστώσας  $\chi$ , θεωρώντας ασήμαντη την περιστροφική συνιστώσα καθώς η περιστροφική ενέργεια είναι αμελητέα συγκριτικά με τις δυο προαναφερόμενες [13]. Η πιθανότητα της μετάβασης μεταξύ της αρχικής κατάστασης  $\Psi_1\chi_a$  και της τελικής κατάστασης  $\Psi_2\chi_b$  είναι ανάλογη του  $|\langle\Psi_1\chi_a|\mathbf{M}|\Psi_2\chi_b\rangle|^2$  όπου  $\mathbf{M}$  ο τελεστής της





Σχήμα 2: Κλασική (a) και κβαντομηχανική (b) περιγραφή του δυναμικού του μορίου κατά την μετάβαση ενός ηλεκτρονίου σύμφωνα με την αρχή Franck-Condon. Το μήκος του δεσμού  $R_e$  αυξάνεται κατά την διέγερση καθώς τα ηλεκτρόνια διεγείρονται σε αντιδεσμικά τροχιακά[2].

διπολικής ροπής του μορίου. Η εμφάνιση της διπολικής ροπής σε αυτό το σημείο αντιστοιχεί στην αλληλεπίδραση της ηλεκτρονιακής πυκνότητας του μορίου με το ηλεκτρομαγνητικό πεδίο του φωτονίου του οποίου η απορρόφηση προκαλεί την μετάβαση. Επιπλέον, καθώς το  $\mathbf{M}$  εξαρτάται μόνο από τις συντεταγμένες των ηλεκτρονίων, ισχύει ότι:

$$|\langle \Psi_1 \chi_a | \mathbf{M} | \Psi_2 \chi_b \rangle|^2 = |\langle \Psi_1 | \mathbf{M} | \Psi_2 \rangle|^2 |\langle \chi_a | \chi_b \rangle|^2 \quad (6)$$

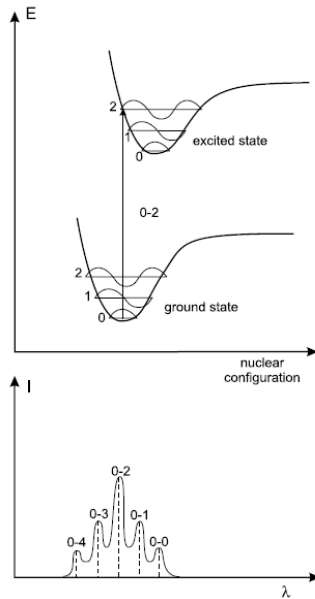
όπου ο δεύτερος παράγοντας ονομάζεται παράγοντας Franck-Condon. Η μετάβαση πραγματοποιείται όταν ο παράγοντας αυτός μεγιστοποιείται, δηλαδή όταν η κυματοσυνάρτηση της διεγερμένης κατάστασης είναι σχεδόν πανομοιότυπη με την κυματοσυνάρτηση της βασικής κατάστασης. Με αυτόν το τρόπο η κυματοσυνάρτηση υφίσταται την ελάχιστη αλλαγή, η οποία αντιστοιχεί στην διατήρηση της δυναμικής κατάστασης του πυρήνα, όπως προδιαγράφει η αρχή Franck-Condon [2]. Καθώς όμως από την σχέση 6 προκύπτει ότι η πιθανότητα μετάβασης και σε άλλες δονητικές καταστάσεις είναι μη μηδενική, τότε στο φάσμα απορρόφησης δεν προκύπτει μόνο μια γκαουσιανή συνάρτηση έντασης σε ένα μήκος κύματος. Αντιθέτως εμφανίζονται περισσότερες με μικρότερη ένταση ανάλογα με την τιμή του παράγοντα Franck-Condon. Συνεπώς τα φάσματα απορρόφησης των μορίων έχουν την μορφή του Σχήματος 3.

Παρόλα αυτά όμως η διακριτότητα στις δονητικές γραμμές του φάσματος απορρόφησης είναι θεωρητικά αποδεδειγμένη αλλά στην πράξη είναι δύσκολα ανιχνεύσιμη. Αυτό συμβαίνει διότι οι δονητικές ενεργειακές στάθμες απέχουν ελάχιστα η μία από την άλλη και συνεπώς η διακριτική ικανότητα τυπικών φασματομέτρων δεν είναι ικανή να τις διαχωρήσει [1].

### 1.1.3 Χαρακτηριστικά Μεγέθη Απορρόφησης

Το φαινόμενο της απορρόφησης περιγράφεται με μεγάλη ακρίβεια από τον νόμο Beer-Lambert, ο οποίος συσχετίζει την απορρόφηση  $A(\lambda)$  με την ένταση  $I_\lambda^0$  της προσπίπτουσας δέσμης φωτός μήκους κύματος  $\lambda$  σε ένα δείγμα και την ένταση  $I_\lambda$  της αντίστοιχης εξερχόμενης δέσμης και εκφράζεται μέσω της σχέσης:

$$A(\lambda) = \log \frac{I_\lambda^0}{I_\lambda} = \varepsilon(\lambda)lc \quad (7)$$



Σχήμα 3: Τυπικό φάσμα απορρόφησης το οποίο αντικατοπτρίζει τις ενεργειακές αποστάσεις των δονητικών επιπέδων μια μοριακής ηλεκτρονιακής κατάστασης[1].

όπου  $\epsilon(\lambda)$  είναι ο μοριακός συντελεστής απορρόφησης σε  $Lmol^{-1}cm^{-1}$ . Ο συγκεκριμένος συντελεστής εκφράζει την δυνατότητα ενός μορίου, που βρίσκεται εντός συγκεκριμένου διαλύματος, να απορροφήσει ακτινοβολία μήκους κύματος  $\lambda$ . Επίσης,  $c$  είναι η συγκέντρωση του δείγματος που απορροφά σε  $molL^{-1}$  και  $l$  το πάχος του. Παρόλα αυτά όμως, ένας βολικότερος τρόπος έκφρασης σχετικά με την απορρόφηση προκύπτει από την έκφραση του νόμου Beer-Lambert σε εκθετική μορφή μέσω του Νεπέριου συντελεστή απορρόφησης:

$$\alpha(\lambda) = \frac{1}{l} \ln \frac{I_{\lambda}^0}{I_{\lambda}} = \frac{1}{l} \log \frac{I_{\lambda}^0}{I_{\lambda}} \ln 10 \quad (8)$$

και συνεπώς ο νόμος Beer-Lambert μετασχηματίζεται στη σχέση:

$$I_{\lambda} = I_{\lambda}^0 e^{-\alpha(\lambda)l} \quad (9)$$

Επιπρόσθετη σημαντική ποσότητα για την πειραματική μελέτη του φαινομένου της απορρόφησης είναι η μοριακή ενεργός διατομή απορρόφησης  $\sigma(\lambda)$  δηλαδή η ενεργός επιφάνεια περιφερειακά του μορίου στην οποία όταν προσπίπτει ένα φωτόνιο κατάλληλης ενέργειας, δύναται να απορροφηθεί από το μόριο. ο συντελεστής αυτός συνδέεται με τον νεπέριο συντελεστή απορρόφησης και τον αριθμό των μορίων  $N$  μέσω της σχέσης:

$$\sigma(\lambda) = \frac{\alpha(\lambda)}{N} \quad (10)$$

Η αποδοτικότητα απορρόφησης εξαρτάται από την δυνατότητα ταλάντωσης του νέφους ηλεκτρονίων του μορίου σύμφωνα με την συχνότητα και την πόλωση του προσπίπτοντος ηλεκτρομαγνητικού πεδίου. Η δυνατότητα αυτή εκφράζεται μέσω της ηλεκτρικής διπολικής ροπής του νέφους ηλεκτρονίων του μορίου, η οποία κλασικά περιγράφεται από την *Δύναμη Ταλαντωτή*  $f$  και κβαντομηχανικά από την *Ροπή Μετάβασης*.

Σύμφωνα με την κλασική θεωρία, η απορρόφηση φωτός από ένα μόριο περιγράφεται αποτελεσματικά εάν θεωρηθεί ότι τα ηλεκτρόνια του δημιουργούν ένα ταλαντούμενο δίπολο, το οποίο περιγράφεται από την *δύναμη ταλαντωτή*  $f$  μέσω της σχέσης:

$$f = 2303 \frac{mc_0^2}{N_a \pi e^2 n} \int \varepsilon(\bar{\nu}) d\bar{\nu} = \frac{4.32 \times 10^{-9}}{n} \int \varepsilon(\bar{\nu}) d\bar{\nu} \quad (11)$$

όπου  $m$  και  $e$  η μάζα και το φορτίο του ηλεκτρονίου αντίστοιχα,  $c_0$  η ταχύτητα του φωτός,  $n$  ο δείκτης διάθλασης και  $\bar{\nu}$  ο κυματαριθμός σε  $cm^{-1}$  [1]. Η φυσική σημασία του συγκεκριμένου μεγέθους αναδεικνύεται ως η πιθανότητα απορρόφησης φωτονίων στο εύρος κυματαριθμών του ολοκληρώματος της σχέσης 11.

Η *Ροπή Μετάβασης*, η οποία περιγράφει το ίδιο φαινόμενο από κβαντομηχανικής σκοπιάς, αντιστοιχεί στο στιγμιαίο δίπολο που δημιουργείται από την μετατόπιση των φορτίων κατά την διάρκεια της μετάβασης από την αρχική στην τελική κατάσταση. Επομένως δεν είναι αυστηρά μια διπολική ροπή. Η σημαντικότητα αυτού του μεγέθους προκύπτει ειδικά όταν μελετώνται φαινόμενα απορρόφησης με χρήση πολωμένου φωτός [14]. Η τιμή του προκύπτει μέσω της σχέσης:

$$\mathbf{M}_{ab} = \langle \psi_b | (q\mathbf{r}) | \psi_a \rangle = q \int \psi_b^*(\mathbf{r}) \mathbf{r} \psi_a(\mathbf{r}) d^3\mathbf{r} \quad (12)$$

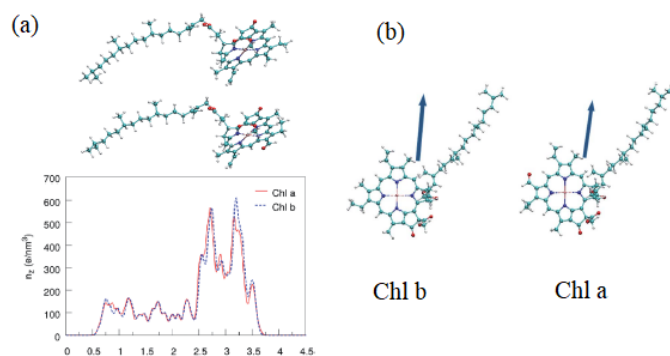
όπου  $|\psi_a\rangle$  η κυματοσυνάρτηση της αρχικής κατάστασης και αντίστοιχα  $|\psi_b\rangle$  της τελικής,  $q$  το συνολικό φορτίο και  $\mathbf{r}$  ο τελεστής θέσης των ηλεκτρονίων.

## Παράδειγμα Χλωροφύλλης

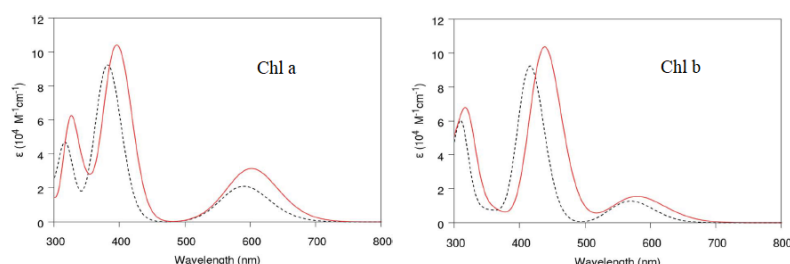
Ένα παράδειγμα της σημαντικότητας της διπολικής ροπής μετάβασης στα φάσματα απορρόφησης είναι η περίπτωση του μορίου της χλωροφύλλης. Γενικότερα οι χλωροφύλλες βρίσκονται στα φωτοσυστήματα των χλωροπλαστών των φυτικών κυττάρων και απορροφούν φωτόνια κατάλληλης ενέργειας. Την ενέργεια αυτήν την μεταφέρουν αλυσιδωτά σε περιφερειακά μόρια χλωροφύλλης με την μορφή εξιτονίων μέσω του μηχανισμού μεταφοράς ενέργειας Förster [15].

Κάθε μόριο χλωροφύλλης αποτελείται από έναν δακτύλιο πορφυρίνης και μια υδροφοβική ουρά η οποία καλύπτεται από το θυλακοειδές τοίχωμα. Ο συγκεκριμένος δακτύλιος, ο οποίος περιλαμβάνει ένα κεντρικό άτομο μαγνησίου  $Mg$  και τέσσερα άτομα αζώτου  $N$  περιμετρικά [5], είναι υπεύθυνος για την απορρόφηση και την μεταφορά ενέργειας, καθώς όπως παρουσιάζεται και στο σχήμα, η πυκνότητα ηλεκτρονίων κατά μήκος του μεγαλύτερου άξονα είναι σημαντικά ισχυρότερη στο εύρος του δακτυλίου. Αυτό έχει ως αποτέλεσμα την επαγωγή της ροπής μετάβασης κατακόρυφα στον άξονα του δακτυλίου [3], όπως φαίνεται στο Σχήμα 4, η οποία ενισχύει την απορρόφηση και την μεταφορά ενέργειας εξαιτίας του προσανατολισμού των μορίων χλωροφύλλης εντός των θυλακοειδών μεμβρανών στα φωτοσυστήματα [16].

Επιπλέον, στο Σχήμα 5 παρουσιάζονται τα φάσματα απορρόφησης της χλωροφύλλης  $a$  και  $b$  με χημικούς τύπους  $C_{55}H_{72}MgN_4O_5$  και  $C_{55}H_{70}MgN_4O_6$  αντίστοιχα. Η υψηλή ένταση σε συγκεκριμένα μήκη κύματος δεν εξαρτάται μόνο από την ενεργειακή διαφορά των επιπέδων HOMO και LUMO αλλά και από την διηλεκτρική σταθερά του περιβάλλοντος του μορίου, η οποία δημιουργεί αλλαγές στον προσανατολισμό των ροπών μετάβασης των μορίων [3]. Ενδεικτικά, αναφέρεται ότι η 1η και η 7η singlet μεταβάσεις ( $S_1$  και  $S_7$ ) συνεισφέρουν στις κορυφές που παρουσιάζονται στην μπλε και κόκκινη περιοχή του φάσματος. Ειδικότερα για την  $S_1$  μετάβαση, τα τροχιακά HOMO και LUMO συνεισφέρουν κατά 84% ενώ για την  $S_7$  τα  $HOMO - 1$  και  $LUMO + 1$  συνεισφέρουν κατά 72% [3].



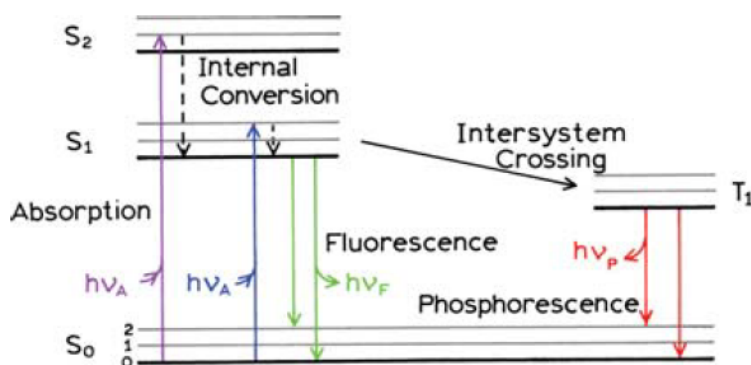
Σχήμα 4: Ηλεκτρονική πυκνότητα κατά μήκος του μεγάλου άξονα των μορίων χλωροφύλλης a και b αντίστοιχα (a) και το διάνυσμα ροπής μετάβασης κάθετα στον δακτύλιο πορφυρίνης των δύο μορίων (b) [3]



Σχήμα 5: Φάσματα απορρόφησης της χλωροφύλλης a (αριστερά) και της χλωροφύλλης b (δεξιά) σύμφωνα με προσομοίωση TDDFT (Time Dependent Density Function Theory) σε κενό (μαύρη γραμμή με παύλες) και σε διάλυμα αιθανόλης (κόκκινη γραμμή)[3].

#### 1.1.4 Εσωτερική Μετατροπή

Ακολουθώντας της απορρόφησης, η ενέργεια η οποία απορροφήθηκε από τα ηλεκτρόνια του μορίου ακολουθεί ποικίλα μονοπάτια επιστροφής στο περιβάλλον και επίτευξης εκ νέου της ισορροπίας του συστήματος. Τα μονοπάτια αυτά εκφράζονται σχηματικά στο διάγραμμα Jablonski, όπως αυτό παρουσιάζεται στο Σχήμα 6. Το πρώτο βήμα των συγκεκριμένων διαδρομών ανταποκρίνεται στην διαδικασία της εσωτερικής μετατροπής, η οποία είναι μη ακτινοβολητική και συμβαίνει μεταξύ επιπέδων με ίδια πολλαπλότητα spin.



Σχήμα 6: Διάγραμμα Jablonski[4].

Όταν ένα μόριο διεγείρεται σε ένα ενεργειακό επίπεδο υψηλότερα από το χαμηλότερο δονητικό επίπεδο της  $S_1$ , τότε το ηλεκτρόνιο χάνει ενέργεια μέσω των συγκρούσεων του μορίου με άλλα μόρια του διαλύματος στο οποίο βρίσκεται. Η αποδοτικότητα της μεταφοράς ενέργειας ενισχύεται όταν τα ενεργειακά διάκενα του μορίου σχεδόν ταυτίζονται με τα αντίστοιχα διάκενα των μορίων του διαλύματος. Καθώς τα συστήματα επικοινωνούν, συνοριακά οι καταστάσεις τους αναμειγνύονται [2]. Το αποτέλεσμα της διαδικασίας αυτής είναι να αποδιεγερθεί το ηλεκτρόνιο έως το χαμηλότερο ενεργειακό δονητικό επίπεδο της  $S_1$ . Το φαινόμενο αυτό διαρκεί περίπου  $10^{-12}$  s και επομένως η προσέγγιση συνεχίζει να υπακούει στην Αρχή Franck-Condon [4]. Επιπλέον, η εσωτερική μετατροπή μεταξύ  $S_1$  και  $S_0$  είναι εφικτή αλλά λιγότερη πιθανή. Αυτό συμβαίνει διότι το ενεργειακό διάκενο είναι αρκετά μεγαλύτερο από την διαφορά των  $S_2$  και  $S_1$  και επομένως μεγαλώνει η πιθανότητα για ακτινοβολητική αποδιέγερση (φθορισμός) ή διασυστημικής διέλευσης (intersystem crossing) σε triplet κατάσταση και έπειτα ακτινοβολητική αποδιέγερση (φωσφορισμός)[1]. Τόσο ο φθορισμός όσο και ο φωσφορισμός πραγματοποιούνται αφού το σύστημα έχει επέλθει σε θερμική ισορροπία.

### 1.1.5 Φθορισμός

Όταν το μόριο υποστεί την πλήρη θερμική εξισορρόπηση με το περιβάλλον του, δηλαδή τα διεγερμένα ηλεκτρόνια μέσω της δονητικής χαλάρωσης καταλήξουν στο χαμηλότερο δονητικό επίπεδο της  $S_1$ , έχουν την δυνατότητα να αποδιεγερθούν στην  $S_0$  είτε μη ακτινοβολητικά μέσω εσωτερικής μετατροπής, είτε ακτινοβολητικά μέσω του φαινομένου του φθορισμού (*Fluorescence*). Και οι δύο τρόποι αποδιέγερσης είναι επιτρεπτοί καθώς διατηρείται η πολλαπλότητα του spin. Παρόλα αυτά όμως, ειδικά στην αποδιέγερση  $S_1 \rightarrow S_0$ , ο φθορισμός είναι ενισχυμένος συγκριτικά με την εσωτερική μετατροπή διότι τα ενεργειακά επίπεδα των μορίων του διαλύματος που βρίσκεται το φθορίζον μόριο εμφανίζουν σημαντικές ενεργειακές διαφορές με τα αντίστοιχα του φθορίζοντος μορίου [2]. Το γεγονός αυτό δυσκολεύει την μεταφορά ενέργειας μέσω συγκρούσεων γιατί δεν υπάρχουν διαθέσιμα επίπεδα με την διαφορά  $E_{S_1} - E_{S_0}$ , ώστε η ενέργεια να μεταφερθεί κβαντωμένη.

Επιπλέον, οι ενεργειακές διαφορές μεταξύ των δονητικών επιπέδων είναι ίδιες για κάθε ηλεκτρονιακό ενεργειακό επίπεδο. Επομένως τα φάσματα φθορισμού είναι κατοπτρικές εικόνες των φασμάτων απορρόφησης (mirror image rule) [1]. Αυτό όμως ισχύει μόνο για τα δονητικά επίπεδα που συμμετέχουν στην μετάβαση  $S_1 \rightarrow S_0$  καθώς η απορρόφηση μπορεί να καταλήξει σε υψηλότερες ενεργειακά singlet καταστάσεις ενώ ο φθορισμός πραγματοποιείται μόνο από την  $S_1$  στην  $S_0$ . Επίσης, ο παράγοντας Franck-Condon στη σχέση 6 είναι ίδιος και για την μετάβαση  $b \rightarrow a$  και επομένως όσο πιθανή είναι μια μετάβαση κατά την απορρόφηση, εξίσου πιθανή είναι και για την εκπομπή [4]. Παρόλα αυτά όμως, η ένταση των φασμάτων εκπομπής παρουσιάζεται μειωμένη σε σχέση με τα φάσματα απορρόφησης λόγω ποικίλων φαινομένων τα οποία συμπυκνώνονται στο φαινόμενο του "σβήσιματος" φθορισμού (fluorescence quenching) και θα περιγραφούν στη συνέχεια.

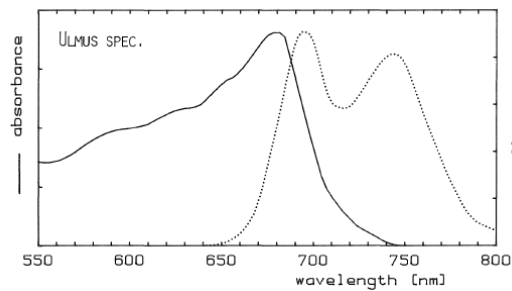
### Μετατόπιση Stokes vs Κατανομή Boltzmann

Καθώς ο φθορισμός συντελείται αυστηρά από την  $S_1$  στην  $S_0$  τότε το φάσμα εκπομπής του δεν εξαρτάται από το μήκος κύματος διέγερσης, φαινόμενο το οποίο στην βιβλιογραφία συναντάται ως νόμος του Kasha[4]. Επιπροσθέτως, τα μήκη κύματος εκπομπής του φθορισμού πρέπει πάντα να είναι υψηλότερα από τα μήκη κύματος απορρόφησης, καθώς υφίστανται απώλειες ενέργειας λόγω δονητικής χαλάρωσης. Ειδικότερα, η διαφορά μεταξύ του μεγίστου της έντασης για το φάσμα απορρόφησης στην μετάβαση  $S_0 \rightarrow S_1$  (πρώτη ζώνη απορρόφησησης) και του μεγίστου του φάσματος φθορισμού, ονομάζεται μετατόπιση Stokes και συνήθως εκφράζεται σε μονάδες κυματαριθμών[1].

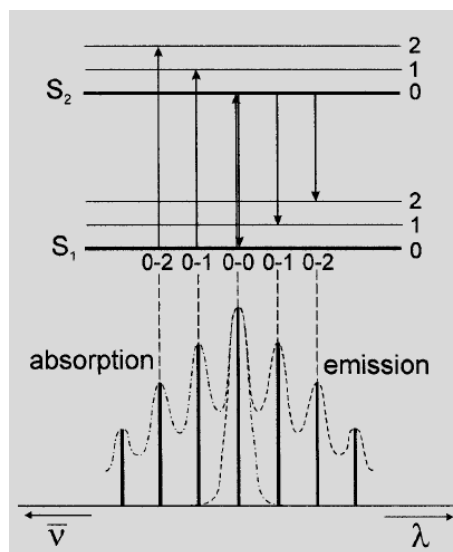
Ωστόσο, όταν πραγματοποιούνται πειραματικές καταγραφές φασμάτων απορρόφησης και εκπομπής, παρατηρείται πως τα δύο αυτά φάσματα εμφανίζουν ένα ποσοστό αλληλεπικάλυψης, όπως παρουσιάζεται παραδείγματος χάριν στα αντίστοιχα φάσματα της χλωροφύλλης του Σχήματος 7. Το γεγονός αυτό φαινομενικά παραβιάζει την αρχή διατήρησης της ενέργειας καθώς εκπέμπονται φωτόνια χαμηλότερου μήκους κύματος από ορισμένα πιθανά μήκη κύματος απορρόφησης. Παρόλα αυτά η ερμηνεία αυτή είναι λανθασμένη, διότι δεν λήφθηκε υπόψη πως το μόριο δεν βρίσκεται στους 0 K, αλλά σε μια οποιαδήποτε άλλη θερμοκρασία. Ειδικότερα, η κατανομή των μορίων στα ηλεκτρονικά ενεργειακά επίπεδα ακολουθεί την κατανομή Boltzmann:

$$\frac{N_1}{N_0} = \exp \frac{-(E_1 - E_0)}{kT} \quad (13)$$

όπου  $N_1$  και  $N_0$  οι πληθυσμοί δύο ενεργειακών επιπέδων και  $E_1$ ,  $E_0$  οι αντίστοιχες ενέργειες,  $k$  η σταθερά Boltzmann και  $T$  η απόλυτη θερμοκρασία του περιβάλλοντος. Όταν η ενεργειακή διαφορά μεταξύ των δύο επιπέδων είναι εξαιρετικά μικρή, όπως συμβαίνει μεταξύ δονητικών ενεργειακών επιπέδων τότε ο λόγος  $N_1/N_0$  αρχίζει να γίνεται σημαντικός. Συνεπώς, η απορρόφηση δύναται να πραγματοποιηθεί από κάποιο διεγερμένο δονητικό επίπεδο της  $S_0$  αλλά η εκπομπή φθορισμού να καταλήξει σε ένα χαμηλότερο ενεργειακό επίπεδο της  $S_0$ , όπως φαίνεται στο Σχήμα 8. Η περιγραφή εξηγεί αποτελεσματικά τον φαινόμενο της αλληλεπικάλυψης και προτάθηκε πρώτη φορά από τον Einstein[1].



Σχήμα 7: Πειραματικό Φάσμα Απορρόφησης και Φθορισμού της Χλωροφύλλης  $a$ [5].



Σχήμα 8: Αλληλεπικάλυψη φασμάτων απορρόφησης και φθορισμού σε θερμοκρασία περιβάλλοντος[1].

## Χρόνος ζωής διεγερμένης κατάστασης - Κβαντική Απόδοση Φθορισμού

Ένα ιδιαίτερα σημαντικό χαρακτηριστικό του φαινομένου του φθορισμού είναι ο χρόνος ζωής της διεγερμένης κατάστασης  $S_1$ ,  $\tau_S$ , ο οποίος δύναται να υπολογιστεί όταν το φθορίζον δείγμα ακτινοβοληθεί με έναν βραχύ παλμό φωτός. Η σημαντικότητα του έγκειται στο γεγονός ότι προσδιορίζει το χρονικό παράθυρο που συντελούνται όλα τα δυναμικά φαινόμενα, με ενδεικτικό εύρος  $[\frac{\tau_S}{100}, 10\tau_S]$  [1]. Ειδικότερα, εάν ακτινοβοληθεί ένα διάλυμα το οποίο περιέχει συγκέντρωση  $[A]$  από ένα συγκεκριμένο είδος φθορίζοντων μορίων, τότε ορισμένα από τα μόρια θα διεγερθούν και σύμφωνα με τις προαναφερόμενες διαδικασίες θα καταλήξουν στη πρώτη διεγερμένη κατάσταση  $S_1$ . Τα μόρια αυτά θα επιστρέψουν στην  $S_0$  είτε ακτινοβολητικά με ρυθμό μείωσης  $k_r^S$ , είτε μη ακτινοβολητικά με ρυθμό μείωσης  $k_{nr}^S$ . Επομένως, η συγκέντρωση των μορίων που βρίσκονται στην πρώτη διεγερμένη κατάσταση  $[^1 A^*]$  θα περιγράφεται από την διαφορική εξίσωση και την αντίστοιχη λύση της:

$$-\frac{d[^1 A^*]}{dt} = (k_r^S + k_{nr}^S) [^1 A^*] \Rightarrow [^1 A^*] = [^1 A^*]_0 \exp\left(-\frac{t}{\tau_S}\right) \quad (14)$$

Επομένως, παρατηρείται πως ο πληθυσμός των μορίων στην πρώτη διεγερμένη κατάσταση μειώνεται εκθετικά με σταθερά χρόνου  $\tau_S$  ή αλλιώς *χρόνο ζωής της  $S_1$*  [1]:

$$\tau_S = \frac{1}{k_r^S + k_{nr}^S} \quad (15)$$

και το αντίστοιχο ποσοστό των μορίων που αποδιεγείρεται ακτινοβολητικά ή αλλιώς η *Κβαντική Απόδοση Φθορισμού (Quantum Yield)* [4] θα είναι:

$$Q = \frac{k_r^S}{k_r^S + k_{nr}^S} \quad (16)$$

Η ένταση φθορισμού που επάγεται από έναν βραχύ παλμό θα δίνεται από την σχέση:

$$i_F(t) = k_r^S [^1 A^*] = k_r^S [^1 A^*]_0 \exp\left(-\frac{t}{\tau_S}\right) \quad (17)$$

### Φθορισμός σταθερής κατάστασης

Αντίστοιχη ανάλυση πραγματοποιείται και όταν η ακτινοβολήση του δείγματος είναι σταθερή με τον χρόνο [1]. Πιο συγκεκριμένα, ο πληθυσμός της διεγερμένης κατάστασης παραμένει σταθερός με τον χρόνο, επομένως ο ρυθμός μείωσης της συγκέντρωσης θα είναι μηδενικός, άρα:

$$\frac{d[^1 A^*]}{dt} = 0 = k_a \alpha N_0 - (k_r^S + k_{nr}^S) [^1 A^*] \Rightarrow [^1 A^*] = \frac{\alpha I_0}{k_r^S + k_{nr}^S} \quad (18)$$

όπου ο όρος  $k_a \alpha N_0$  αντιστοιχεί στο ποσοστό των απορροφούμενων φωτονίων ανά μονάδα όγκου ανά μονάδα χρόνου και άρα  $I_0$  θα είναι η ένταση της προσπίπτουσας δέσμης ( $mol/Ls$ ). Επομένως, η ένταση φθορισμού σταθερής κατάστασης σύνδεεται με την κβαντική απόδοση μέσω της σχέσης:

$$i_F = k_r^S [^1 A^*] = \alpha I_0 \frac{k_r^S}{k_r^S + k_{nr}^S} = \alpha I_0 Q_F \quad (19)$$

## Fluorescence Quenching

Καθώς τα φθορίζοντα μόρια δεν βρίσκονται απομονωμένα από το περιβάλλον αλλά συνήθως συνυπάρχουν τόσο με μόρια διαλυτών όσο και με μόρια φθορίζοντων και μη φθορίζοντων ουσιών, θα πρέπει να ληφθεί υπόψη η αλληλεπίδραση τους, η οποία μειώνει την ένταση του φθορισμού που δύναται να καταγραφεί. Οι βασικοί παράγοντες μείωσης της έντασης του φθορισμού, φαινόμενο το οποίο στην βιβλιογραφία παρουσιάζεται ως *Fluorescence Quenching*, και οι αντίστοιχες επεξηγήσεις τους είναι:

1. Τα φθορίζοντα μόρια, αφότου έχουν περιέλθει στην  $S_1$  διεγερμένη κατάσταση, αντί να αποδιεγερθούν ακτινοβολητικά μέσω του φθορισμού, μεταλαμπαδεύουν την ενέργεια τους σε μόρια του περιβάλλοντος τους (quenchers) μέσω συγκρούσεων (collisional quenching). Η ένταση του φαινομένου ενισχύεται ανάλογα με την συγκέντρωση των quenchers σύμφωνα με την σχέση [4]:

$$\frac{F_0}{F} = 1 + k_q \tau_0 [Q] \quad (20)$$

όπου  $F_0$  η αρχικά εκπεμπόμενη ένταση φθορισμού,  $F$  η τελική ένταση φθορισμού,  $k_q$  η διαμοριακή σταθερά quenching, ο μέσος χρόνος μεταξύ δύο συγκρούσεων και  $[Q]$  η συγκέντρωση των quenchers. Συνήθως τα μόρια που επάγουν το φαινόμενο αυτό είναι ισχυρά ηλεκτραρνητικά, καθώς ο μηχανισμός που αντλούν την ενέργεια από τα φθορίζοντα μόρια είναι η μεταφορά των διεγερμένων ηλεκτρονίων στα δικά τους τροχιακά[4].

2. Ενισχύεται το φαινόμενο του φωσφορισμού, όπως αυτό περιγράφεται στην επόμενη ενότητα, όταν φθορίζον μόριο περιλαμβάνει ένα άτομο με υψηλό ατομικό αριθμό (internal heavy atom effect)[1].
3. Συντελείται το φαινόμενο του δυναμικού αναπροσανατολισμού των διπόλων που υπάρχουν εντός του διαλύματος που βρίσκεται το φθορίζον μόριο. Αυτό προκύπτει εξαιτίας της στιγμιαίας ύπαρξης της ροπής μετάβασης κατά την διάρκεια της απορρόφησης, η οποία επάγει τον αναπροσανατολισμό των διπόλων και συνεπώς την τοπική αλλαγή της συγκέντρωσης των quenchers (σχέση 20).
4. Το φαινόμενο του φθορισμού συντελείται από φθορίζοντες ουσίες, σημειακά επί του μορίου. Εάν το μόριο αυτό βρίσκεται σε διάλυμα με συγκεκριμένη οπτική πυκνότητα, η οποία αντιστοιχεί σε απορρόφηση σε συγκεκριμένο μήκος κύματος, τότε τόσο η ακτινοβολία διέγερσης όσο και η ακτινοβολία εκπομπής θα υφίστανται εξασθένιση (inner filter effect). Επομένως, λαμβάνοντας υπόψη τα φάσματα απορρόφησης των μορίων του διαλύματος καθώς επίσης και τον νόμο Beer-Lambert (σχέση 9), προκύπτει πώς η μετρούμενη σε ιδανικές συνθήκες  $F_{\text{corr}}$  ένταση φθορισμού θα είναι:

$$F_{\text{corr}} = F_{\text{rec}} * 10^{[OD(\text{em})+OD(\text{ex})]/2} \quad (21)$$

όπου  $F_{\text{rec}}$  η αρχική ένταση και  $OD_{(em)}, OD_{(ex)}$  η οπτική πυκνότητα του διαλύματος στο μήκος κύματος εκπομπής και διέγερσης αντίστοιχα [13].

### 1.1.6 Διασυστημική Διέλευση και Φωσφορισμός

Εκτός από τον φθορισμό, ο οποίος δεν παραβιάζει την αρχή διατήρησης της πολλαπλότητας του spin, παρατηρείται από το διάγραμμα Jablonski ότι το σύστημα δύναται να περιέλθει σε μετάβαση με αλλαγή στην πολλαπλότητα του spin. Η διαδικασία αυτή είναι φαινομενικά απαγορευμένη, εάν θεωρηθεί πως η τροχιακή στροφορμή του ηλεκτρονίου δεν συζεύγνυται με την στροφορμή λόγω spin. Όμως, αυτό αποτελεί αυθαιρεσία μιας και η ενεργειακή σύζευξη spin-τροχιακής στροφορμής (LS coupling) υπάρχει και ενίοτε είναι αρκετά ισχυρή, όπως στην περίπτωση που το μόριο περιέχει ένα "βαρύ" άτομο (internal



heavy atom effect), όπου η ενεργειακή σύζευξη αυξάνεται σύμφωνα με την τέταρτη δύναμη του ατομικού αριθμού [1]. Η διαδικασία αυτή ονομάζεται *Διασυστημική Διέλευση (intersystem crossing)* και προκύπτει καθώς στην μοριακή χαμιλτονιανή της σχέσης (3) προστίθεται ο όρος:

$$H_{so} = \sum_i \xi_i l_i \cdot s_i \quad (22)$$

όπου  $\mathbf{l}$  ο τελεστής της τροχιακής στροφορμής,  $\mathbf{s}$  ο τελεστής του spin και το άθροισμα γίνεται επάνω σε όλα τα ηλεκτρόνια του μορίου. Η πρόσθεση αυτού του όρου αλλάζει την αρχική χαμιλτονιανή (σχέση 3) η οποία επέβαλλε την διατήρηση της πολλαπλότητας του spin και πλέον οι μεταβάσεις singlet-triplet εμφανίζουν μη μηδενική πιθανότητα. Όσο μεγαλύτερη η σύζευξη, τόσο περισσότερες καταστάσεις των συστημάτων singlet και triplet αναμυγνύονται και η προαναφερόμενη πιθανότητα αρχίζει να εμφανίζει σημαντικό πλάτος [2]. Συνεπώς, πάλι υπό την αρχή Franck-Condon, το ηλεκτρόνιο θα αποδιεγερθεί σε εκείνο το δονητικό επίπεδο της  $T_1$  του οποίου η κυματοσυνάρτηση θα είναι παρόμοια με την αντίστοιχη της  $S_1$  και στη συνέχεια θα υποστεί και πάλι εσωτερική μετατροπή στο χαμηλότερο δονητικό επίπεδο της  $T_1$ . Από την  $T_1$  στην  $S_0$ , η μετάβαση δεν είναι πια απαγορευμένη λόγω της προαναφερόμενης σύζευξης και έτσι προκύπτει η ακτινοβολητική αποδιέγερση του *Φωσφορισμού* με χρονική διάρκεια της τάξης του  $10^{-6} - 1$  s [1], η οποία είναι σημαντικά μεγαλύτερη και ετεροχρονισμένη του φθορισμού. Χαρακτηριστικό μέγεθος και σε αυτές τις δύο διεργασίες είναι η χβαντική τους απόδοση με  $Q_{isc}$  για την διασυστημική διέλευση και  $Q_P$  για τον φωσφορισμό:

$$Q_{isc} = \frac{k_{isc}}{k_r^S + k_{nr}^S} = k_{isc} \tau_S \quad (23)$$

$$Q_P = \frac{k_r^T}{k_r^T + k_{nr}^T} \Phi_{isc} \quad (24)$$

όπου  $k_{isc}$  και  $k_r^T$  οι σταθερές χρόνου για την διασυστημική διέλευση και την αποδιέγερση φθορισμού από την triplet κατάσταση.

## 1.2 Διάβρωση Έργων Πολιτιστικής Κληρονομίας

### 1.2.1 Τύποι και Ιδιότητες Υποστρώματος

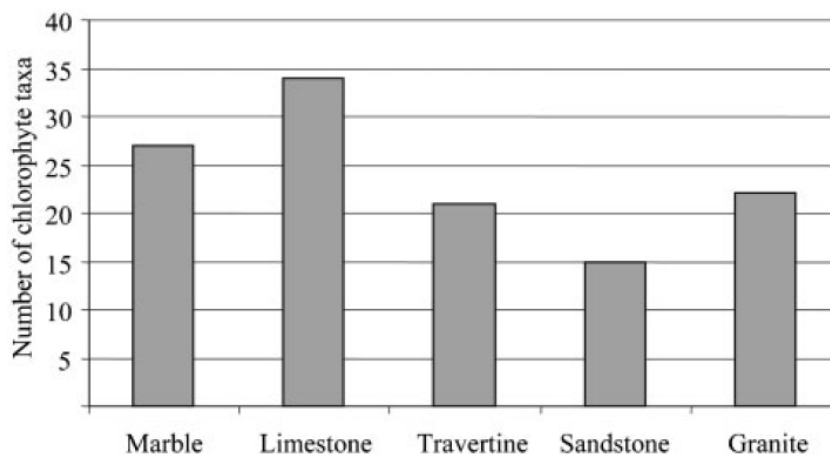
Η διάβρωση των επιφανειών είναι μια διαδικασία η οποία εξαρτάται άμεσα από το υλικό κατασκευής. Τα έργα και τα μνημεία πολιτιστικής κληρονομίας είναι κατασκευασμένα τόσο από φυσικά πετρώματα όπως το μάρμαρο (marble), ο ασβεστόλιθος (limestone), ο ψαμμίτης (sandstone) και ο γρανίτης (granite) καθώς επίσης και από τεχνητά πετρώματα όπως τα κονιάματα (mortars), οι πλίνθοι (bricks), τα κονιάματα (concretes) και τα κεραμικά (ceramics)[17]. Ειδικότερα, τα μνημεία της Ακρόπολης των Αθηνών τα οποία εξετάζονται στην παρούσα εργασία αποτελούνται κυρίως από Πεντελικό Μάρμαρο ως βασικό δομικό στοιχείο και τα θεμέλια τους εδράζονται σε ασβεστόλιθο[18].

Οι λίθινες επιφάνειες ταξινομούνται σε κατηγορίες ανάλογα με τον τύπο της πέτρας σε ασβεστολιθικές και πυριτικές. Οι ασβεστολιθικές πέτρες περιλαμβάνουν τον ασβεστόλιθο και το μάρμαρο με βασικό τους συστατικό είναι το ανθρακικό ασβέστιο, το οποίο τα καθιστά ιδιαίτερα ευαίσθητα σε υπερβολικά όξινο ή αλκαλικό περιβάλλον. Για τον λόγο αυτό δύναται να διαλυθούν χωρίς να αφήσουν υπολείμματα[7]. Οι πυριτικές πέτρες, που περιλαμβάνουν τον ψαμμίτη και τον γρανίτη, αποτελούνται κυρίως από διοξείδιο του πυριτίου και χαλαζία. Παρότι οι πυριτικές πέτρες είναι περισσότερο ανθεκτικές από της ασβεστολιθικές, ωστόσο ο ασβεστόλιθος και το μάρμαρο χρησιμοποιούνται σε μεγάλη έκταση στα έργα πολιτιστικής κληρονομίας.

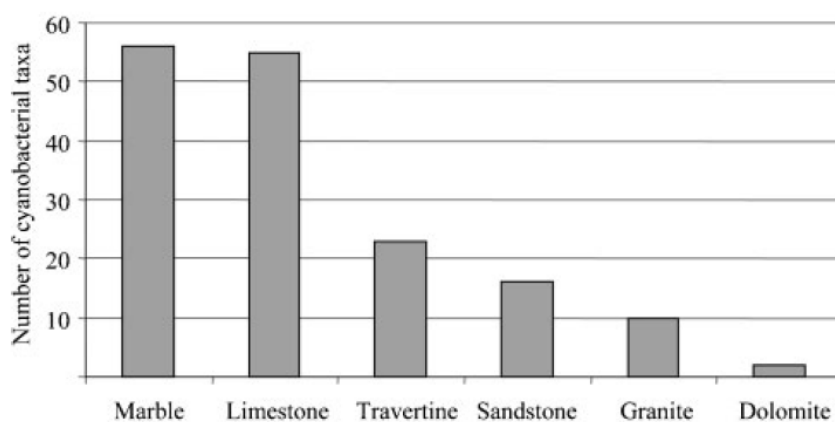
Τόσο τα φυσικά όσο και τα τεχνητά πετρώματα διαφέρουν μεταξύ τους στην χημική σύσταση, στην επιφανειακή υφή, στην σκληρότητα, στο pH και στο πόσο πορώδεις είναι *-Porosity*. Αυτό έχει ως αποτέλεσμα την διαφορετική επιδεκτικότητά τους στην ανάπτυξη βιολογικών επικαθίσεων και την επαγωγή Βιοδιάβρωσης. Το μέτρο της ικανότητας ενός πετρώματος να δεχθεί και να διατηρήσει την ανάπτυξη βιολογικών επικαθίσεων στην επιφάνεια ή σε κοιλότητες του ονομάζεται *Βιοδεκτικότητα - Bioreceptivity* και προτάθηκε από τον Guillite το 1995 [19]. Η ιδιότητα αυτή διακρίνεται σε:

1. Πρωτογενής Βιοδεκτικότητα, η οποία προσδιορίζει το δυναμικό ενός καθαρού και πλήρως μη διαβρωμένου υλικού να ενοικιστεί από βιολογικούς παράγοντες
2. Δευτερογενής Βιοδεκτικότητα, η οποία πιστοποιεί το αποτέλεσμα της αλλοίωσης από άβιους και έμβιους παράγοντες
3. Τριτογενής Βιοδεκτικότητα, η οποία επάγεται από τα θρεπτικά συστατικά που δύναται να υπάρχουν επί της επιφάνειας, όπως νεκρή βιομάζα, σωματίδια σκόνης, ζωικά κόπρανα, απωθητικά υγρασίας και βιοκτόνα.

Οι παράγοντες που επηρεάζουν την Βιοδεκτικότητα εστιάζονται στην πορώδη έκταση, στην επιφανειακή τραχύτητα και στην χημική σύσταση του υποστρώματος. Γενικότερα, τα υλικά τα οποία είναι αρκετά πορώδη είναι πιο επιδεκτικά στην *Βιοαποίκηση (Biocolonization)*, δηλαδή την εποίκηση μια επιφάνειας από οργανισμούς, εξαιτίας της δυνατότητάς τους να απορροφούν περισσότερο νερό από το περιβάλλον και να το διατηρούν για μεγαλύτερα χρονικά διαστήματα. Ειδικά όταν οι πόροι είναι εμφανής στην επιφάνεια του υλικού δημιουργείται ένα δίκτυο διαδρομών για την μετακίνηση του νερού κατά μήκος του συνόλου της κατασκευής[17]. Επιπλέον, τα υποστρώματα που είναι αρκετά τραχιά εμφανίζουν στην επιφάνεια τους μικρές εγχοπές, οι οποίες προσφέρουν μια προστατευμένη περιοχή για την ανάπτυξη μικροοργανισμών. Το παράδειγμα των κεραμικών, το οποίο σαν δείγμα θα μελετηθεί εκτενώς στην παρούσα εργασία, αποτελεί κυρίαρχο παράδειγμα στην επίπτωση της υψηλής ποροσιμότητας και της επιφανειακής τραχύτητας στην βιοδεκτικότητα. Ειδικότερα, οι φωτοαυτότροφοι μικροοργανισμοί οι οποίοι έχουν μικρή ανθεκτικότητα σε συνθήκες ξηρασίας, ενοικίζουν ισχυρά τα κεραμικά εξαιτίας της ικανότητάς τους να διατηρούν υγρασία [20]. Όσον αφορά την χημική σύσταση του υποστρώματος, το



Σχήμα 9: Χλωροφύκη



Σχήμα 10: Κυανοβακτήρια

Σχήμα 11: Αριθμός ειδών χλωροφυκών (αριστερά) και κυανοβακτηρίων (δεξιά) που ανιχνεύθηκαν σε μνημεία πολιτιστικής κληρονομιάς στην ευρύτερη περιοχή της Μεσογείου σε έξι διαφορετικά υποστρώματα[6]

pH καθορίζει σε σημαντικό βαθμό το είδος των μικροοργανισμών που αναπτύσσεται καθώς τα ιόντα  $H^+$  και  $OH^-$  λειτουργούν καταστρεπτικά ανάλογα με το είδος των εξωκυτταρικών πολυμερικών ουσιών (EPS) που εκκρίνονται από τους μικροοργανισμούς σε αντίζοες συνθήκες[6]. Παραδείγματος χάριν, ορισμένοι μύκητες προτιμούν ελαφρώς όξινα περιβάλλοντα όπως τον γρανίτη και ορισμένους ψαμμίτες ενώ ορισμένα κυανοβακτήρια αναπτύσσονται σε πιο αλκαλικά περιβάλλοντα όπως ο ασβεστόλιθος και το μάρμαρο [6, 17]. Αντίθετα, τα χλωροφύκη δύναται να αποικήσουν οποιοδήποτε υπόστρωμα, όσον αφορά την χημική του σύσταση αλλά η ανάπτυξή τους ευνοείται από την ποροσιμότητα και την επιφανειακή τραχύτητα εξαιτίας της υψηλής υγρασίας[6]. Οι παρατηρήσεις αυτές εμφανίζονται και στο Σχήμα 11.

### 1.2.2 Περιβαλλοντικοί Παράγοντες Διάβρωσης

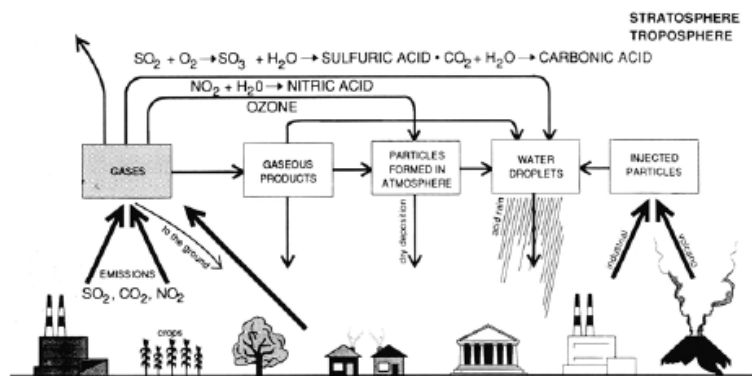
Τα έργα πολιτιστικής κληρονομιάς αλλοιώνονται εξαιτίας της Βιοαποικοδόμησης, η οποία συνδέεται με την αποίκηση οργανισμών επί των επιφανειών τους, όπως επίσης και εξαιτίας περιβαλλοντικών παραγόντων που συνδέονται κυρίως με την ρύπανση της ατμόσφαιρας. Στη συνέχεια παρουσιάζονται οι σημαντικότεροι περιβαλλοντικοί περιβαλλοντικοί παράγοντες διάβρωσης των επιφανειών των μνημείων

πολιτιστικής κληρονομιάς καθώς επίσης και η αντίστοιχη επίπτωση που προκαλούν. Οι βιολογικοί παράγοντες επιβάρυνσης περιγράφονται στις υποενότητες 1.2.3 και 1.2.4. Σε αστικές και βιομηχανικές περιοχές, χημικές ενώσεις και στοιχεία όπως  $CO_2$ ,  $SO_2$ ,  $SO_3$ ,  $NO_x$ ,  $O_3$  και  $Cl_2$  [7] υπάρχουν σε υψηλή περιεκτικότητα σε μορφή αεροζόλ, τα οποία δύναται να επικαθήσουν σε ξηρή μορφή χωρίς να πυροδοτήσουν κάποιον βιοδιαβρωτικό μηχανισμό. Όμως κατά την διάρκεια βροχόπτωσης, τα θειικά, τα νιτρικά και τα ανθρακικά οξείδια διαλύονται στο νερό και μειώνουν κατακόρυφα το pH του. Επομένως, η όξινη βροχόπτωση όταν έρχεται σε επαφή με τους ανθρακικούς λίθους όπως ο ασβεστόλιθος και το μάρμαρο αντιδρά και προκαλεί όξινη διάβρωση[21]. Στον Πίνακα 1 παρουσιάζονται οι σημαντικότεροι ρύποι της ατμόσφαιρας, η φυσική και τεχνητή πηγή τους καθώς και η επίπτωση που έχουν στην λίθινες επιφάνειες. Επιπλέον στο Σχήμα 12 παρουσιάζεται η προέλευση και η κατάληξη των ρύπων επί των μνημείων.

Ρύπος	Φυσική Πηγή	Τεχνητή Πηγή	Επίπτωση
Σωματίδια	Ηφαιστειακή Τέφρα, Πυρκαγιές Δασών	Βιομηχανικές Εξατμίσεις	Κάλυμμα αιθάλης
$CO_2$	Ηφαιστεια, Βλάστηση	Καύση Υδρογονανθράκων	Όξινη Βροχή
$SO_2$	Ηφαιστεια, Ερημική Σκόνη	Καύση Υδρογονανθράκων	Όξινη Βροχή
$SO_3$	-	Όξειδωση του $SO_2$	Όξινη Βροχή
$O_x$	Ηφαιστεια, βλάστηση, βακτήρια	Καύση Υδρογονανθράκων	Όξινη Βροχή, Φωτοοξείδωση
$O_3$	Μικροποσότητες στην Στρατόσφαιρα	Όξειδωση $O_2$ λόγω άλλων ρύπων	Όξινη Βροχή, Μεταλλική Διάβρωση
$Cl$	Ηφαιστεια, Ερημική Σκόνη	Καύση	Όξινη Βροχή

Πίνακας 1: Σημαντικοί ατμοσφαιρικοί ρύποι, οι φυσικές και οι τεχνητές πηγές τους καθώς και η επίδραση τους στα μνημεία πολιτιστικής κληρονομιάς[7]

Η επίπτωση όμως της ατμοσφαιρικής ρύπανσης εξαρτάται και από το εύρος έκθεσης του μνημείου στους ρυπογόνους παράγοντες. Ειδικότερα εάν κάποιο σημείο δεν είναι άμεσα εκτεθειμένο στην όξινη βροχή, τότε η υγρασία συσσωρεύεται εκεί και καθώς εξατμίζεται με την πάροδο του χρόνου δημιουργείται *Μαύρη Κρούστα*. Ποικίλες μελέτες που έχουν πραγματοποιηθεί, έδειξαν πως η κρούστα προέρχεται συγκεκριμένα από την ρύπανση των ανθρακικών σωματιδίων από την καύση υδρογονανθράκων. Αυτό συμβαίνει καθώς περιέχουν μεταλλικά οξείδια τα οποία καταλύουν την οξείδωση του  $SO_2$  και προωθούν την δημιουργία κρούστας. Η διάβρωση που προκαλείται από την μαύρη κρούστα είναι διπλής φύσης καθώς κάτω από την μαύρη επικάλυψη επί του ασβεστολιθικού πετρώματος δημιουργείται επίσης ένα στρώμα καθαρού γύψου[21].



Σχήμα 12: Η προέλευση και η κατάληξη των ρύπων επί των μνημείων[7]

Ωστόσο το περιβάλλον της μαύρης κρούστας εξαιτίας της υψηλής οργανικής επιβάρυνσης, δύναται να αυξήσει το μέτρο της δευτερογενούς Βιοδεκτικότητας και συνεπώς να φιλοξενήσει επιπρόσθετα και ορισμένους μικροοργανισμούς. Πιο συγκεκριμένα, μερικές ομάδες κυανοβαντηρίων όπως επίσης και κάποια νιτροποιούντα βακτήρια ευδοκούν στις μαύρες κρούστες καθώς σταδιακά διαλύουν τα επικαθήμενα θειικά άλατα[21].

### 1.2.3 Βιολογικοί Παράγοντες Διάβρωσης - Οργανισμοί

Τα μνημεία πολιτιστικής κληρονομιάς, καθώς είναι εκτεθειμένα στο φυσικό περιβάλλον, είναι δυνατόν να επηρεαστούν σημαντικά από τους οργανισμούς που ζουν σε αυτό. Ιδιαίτερα συχνό είναι το φαινόμενο της ύπαρξης μεγάλων φυτών και δέντρων όπου με τις ρίζες τους προκαλούν υψηλή μηχανική καταπόνηση στο μνημείο και μπορούν να προκαλέσουν μέχρι και την ολική καταστροφή του. Καθώς όμως στη συγκεκριμένη εργασία εξετάζονται οι παράγοντες βιοδιάβρωσης σε μικρότερη κλίμακα, θα πρέπει να αναφερθούν τα Βρυόφυτα ή Βρύα, τα οποία εμφανίζονται σε πληθώρα στις επιφάνειες μνημείων όπου η υγρασία είναι εξαιρετικά υψηλή και η ίδια η επιφάνεια έχει τη δυνατότητα να την συγκρατήσει, όπως είναι δηλαδή τα κεραμικά.

Ο όρος Βρυόφυτα είναι μια γενικότερη ονομασία για φυτά χωρίς αγωγό ιστό των οποίων ο κύκλος ζωής χαρακτηρίζεται από την αλληλουχία απλοειδών και διπλοειδών φάσεων με ένα κεντρικό γαμετόφυτο ή σπορόφυτο αντίστοιχα [22]. Το γαμετόφυτο, κατά την απλοειδή φάση, παράγει τους γαμέτες του βρυόφυτου οι οποίοι θα συντηχθούν έτσι ώστε να δημιουργηθεί στη συνέχεια το ζυγωτό. Στην διπλοειδή φάση της ζωής του βρυόφυτου, το σπορόφυτο παράγει με την διαδικασία της μείωσης απλοειδή σπόρια τα οποία διασπείρονται με τον άνεμο.

Τα σπόρια όταν βρεθούν σε κατάλληλο υπόστρωμα φυτρώνουν και σχηματίζουν ένα πολυδιακλαδιζόμενο δίκτυο υφών το πρωτόνυμα. Όταν το υπόστρωμα έχει υψηλά επίπεδα υγρασίας, όπως ο πορόλιθος, τότε το πρωτόνυμα αναπτύσσεται και σχηματίζονται τα γαμετόφυτα. Η βιοδιαβρωση από τα βρύα είναι πιο έμμεση και αφορά κυρίως την μηχανική καταπόνηση που δέχεται το υπόστρωμα και την διατήρηση της υγρασίας περιφερειακά του υποστρώματος.

### 1.2.4 Βιολογικοί Παράγοντες Διάβρωσης - Μικροοργανισμοί

Οι μικροοργανισμοί δύναται να αναπτύξουν ποικίλες σχέσεις με το υπόστρωμα που εποικίζουν. Αρχικά, έχουν την δυνατότητα να ευδοκιμήσουν επιλιθικά, αναπτυσσόμενοι στην επιφάνεια του υποστρώματος, όπως επίσης και ενδολιθικά διεισδύοντας στο εσωτερικό της πέτρας. Οι μικροοργανισμοί που αναπτύσσονται στην επιφάνεια ονομάζονται επιλιθικοί ενώ οι μικροοργανισμοί που έχουν την ικανότητα να διεισδύουν στο υπόστρωμα και να αναπτύσσονται στο εσωτερικό λέγονται ενδολιθικοί. Οι ενδολιθικοί μικροοργανισμοί διαχωρίζονται σε χασμοενδολιθικούς, οι οποίοι εποικούν προϋπάρχοντες σχισμές και κοιλότητες, σε κρυπτοενδολιθικούς οι οποίοι εποικούν δομικές κοιλότητες σε πορώδη πετρώματα και σε ευενδολιθικούς, οι οποίοι διαλύουν την πέτρα και διεισδύουν σε αυτή [23]. Ορισμένα είδη είναι αποκλειστικά ενδολιθικά ενώ άλλα προσαρμόζονται σε κάθε είδος μορφολογίας. Τα χημικά και δομικά χαρακτηριστικά των υποστρωμάτων όπως η επιφανειακή τραχύτητα, η πορώδης επιφάνεια και η κατάσταση συντήρησης όπως επίσης και οι κλιματολογικές συνθήκες διαμορφώνουν την ποικιλότητα και την ανάπτυξη διάφορων τύπων μικροοργανισμών, όπως θα περιγραφούν στη συνέχεια. Ένας λιθότυπος με υψηλή βιοδεκτικότητα (bioreceptivity) [24] σε συνδυασμό με κατάλληλες περιβαλλοντικές συνθήκες δύναται να επάγει έντονο βιολογικό αποικισμό εντός μερικών χρόνων. Η διάκριση των βιοδιαβρωτικών μικροοργανισμών ανάλογα με την τον τρόπο κάλυψης των ενεργειακών τους αναγκών τους διακρίνει στις κατηγορίες των *Φωτοαντότροφων* στους οποίους ανήκουν τα *Χλωροφύκη*, τα *Κυανοβακτήρια* και οι *Λειχήνες*, στους *Ετερότροφους* στους οποίους ανήκουν οι *Μύκητες* και στους

Χημειολιθοτροφικούς στους οποίους ανήκουν ορισμένα βακτήρια που αξιοποιούν τα ανόργανα υλικά του υποστρώματος.

## 1. Φωτοαυτότροφοι Μικροοργανισμοί

### (a) Χλωροφύκη

Αρχικά, μια σημαντική κατηγορία μικροοργανισμών που εποικούν στην επιφάνεια των μνημείων είναι τα χλωροφύκη, τα οποία ανήκουν βασίλειο των Πρωτίστων. Οι τάξεις τους υπάγονται τόσο στους ευκαρυωτικούς, πολυφυλετικούς οργανισμούς με φωτοσυνθετική λειτουργία. Εμφανίζονται είτε ως μονοκύτταροι σχηματίζοντας αποικίες, είτε ως πολυκύτταροι[25], ωστόσο διακρίνονται από την κατηγορία των φυτών καθώς δεν αναπτύσσουν εξωτερικά χαρακτηριστικά όπως ρίζες, φύλλα ή αγωγούς ιστούς. Επιπλέον, εμφανίζονται σε ένα εύρος διαστάσεων έως και τα 60m [25].

Η πλειονότητα των χλωροφυκών τα οποία απαντούν σε συστήματα βιοδιάβρωσης μνημειακών επιφανειών ανήκουν στους προκαρυωτικούς οργανισμούς, στην ομάδα των Κυανοφύτων[26]. Περιέχουν χρωμοφόρες ουσίες όπως η χλωροφύλλη α, φυκοβιλλιπρωτεΐνες όπως η c-φυκοερυθρίνη και c-φυκοκυανίνη και καροτενοΐδη όπως το β-καροτένιο[25]. Όταν οι συνθήκες του περιβάλλοντος γίνουν αντίξοες, τότε τα χλωροφύκη παράγουν τα "έμμονα κύτταρα", τα οποία παρουσιάζουν χαμηλούς μεταβολικούς ρυθμούς και παχιά τοιχώματα προκειμένου να επιβιώσουν[25]. Η ανάπτυξή τους συνήθως πραγματοποιείται επιλιθικά ώστε να αξιοποιούν την ηλιακή ακτινοβολία στην διαδικασία της φωτοσύνθεσης[26]. Σημαντικό τμήμα ευκαρυωτικών χλωροφυκών που αποτελεί παράγοντα βιοδιάβρωσης είναι τα Χλωρόφυτα[6] και ειδικότερα η κλάση Chlorophyceae. Ο τρόπος αναπαραγωγής τους είναι όμοιος με τα Κυανόφυτα αλλά διαφέρουν στα φθορίζοντα μόρια καθώς περιέχουν χλωροφύλλη α και β, απουσία όμως φυκοβιλλιπρωτεϊνών[25].

### (b) Βακτήρια - Κυανοβακτήρια

Πρωταρχικό παράγοντα βιοδιάβρωσης αποτελούν τα βακτήρια, τα οποία ανήκουν στους προκαρυωτικούς μικροοργανισμούς και έχουν την δυνατότητα επιβίωσης και αναπαραγωγής στα πιο αντίξοα περιβάλλοντα. Κύριο χαρακτηριστικό τους, με βάση το οποίο διαχωρίζονται και σε δύο κατηγορίες, Gram-Positive και Gram-Negative [27], αποτελεί το κυτταρικό τους τοίχωμα το οποίο συντίθεται κυρίως από πεπτιδογλυκάνη που τους παρέχει δομική στήριξη και προστασία. Η αλληλουχία σακχάρων και αμινοξέων της πεπτιδογλυκάνης προσδίδει επιπλέον στα βακτήρια την δυνατότητα να επιβιώσουν σε περιβάλλοντα ακραίου pH, αλατότητας και θερμοκρασίας παράλληλα με την ικανότητα τους να μετατρέπονται σε ενδοσπόρια. Ειδικά τα αυτότροφα βακτήρια αποτελούν έναν εξαιρετικά αυτόνομο μικροοργανισμό καθώς μετατρέπουν την ανόργανη ένωση του  $CO_2$  σε χρήσιμες για αυτά οργανικές ενώσεις όπως υδατάνθρακες, λιπίδια και πρωτεΐνες. Η βιοχημική διαδικασία που διεκπεραιώνεται για την παραγωγή αυτών των οργανικών ενώσεων μπορεί να στηρίζεται στην αξιοποίηση ενέργειας μέσω ακτινοβολίας -Φωτοαυτότροφα Βακτήρια- είτε στην οξείδωση αζώτου, θείου ή άλλων στοιχείων -Χημειοαυτότροφα, τα οποία όμως θα περιγραφούν στη συνέχεια. Όταν αυτή η διαδικασία απουσιάζει και τα βακτήρια λαμβάνουν από το περιβάλλον τους τα θρεπτικά συστατικά που έχουν ανάγκη, τότε χαρακτηρίζονται ως *Ετερότροφα*.

Το Βασίλειο των βακτηρίων διακρίνεται σε Φύλα, ορισμένα από τα οποία κυριαρχούν στον τομέα της βιοδιάβρωσης. Πιο συγκεκριμένα, τα ετερότροφα φύλα τα οποία βιοδιαβρώνουν κυρίως το μάρμαρο είναι τα Actinobacteria, τα Proteobacteria και τα Firmicutes [28], ενώ στα φωτοαυτότροφα βακτήρια το κύριαρχο βιοδιαβρωτικό φύλο είναι τα Κυανοβακτήρια τα οποία παράγουν θρεπτικά συστατικά και οξυγόνο μέσω φωτοσύνθεσης. Ειδικότερα τα λιθοφιλικά κυανοβακτήρια εποικίζουν τις πέτρινες επιφάνειες είτε επιλιθικά είτε ενδολιθικά του υποστρώματος σε βάθος μερικών χιλιοστών [29]. Η πολυπλοκότητα και η ποικιλία στα είδη των κυανοβακτηρίων εξαρτάται άμεσα από την διαθεσιμότητα του νερού στο περιβάλλον του υποστρώματος [30]. Συνήθως η ανάπτυξή τους συνδέεται με την δημιουργία του "βιοτάπητα", μια επιλιθική συμβίωση πολλαπλών τύπων μικροοργανισμών, η οποία θα περιγραφεί στην επόμενη ενότητα. Τα κυανοβακτήρια προσδιορίζονται από ένα χαρακτηριστικό μπλε-πράσινο χρώμα εξαιτίας των χρωμοφόρων μορίων που περιέχουν, δηλαδή την χλωροφύλλη α (πράσινο), την φυκοβιλλιπρωτεΐνη φυκοκυανίνη (μπλε) και τα καροτενοειδή (κίτρινο-πορτοκαλί). Παρόλα αυτά όμως, ορισμένα είδη κυανοβακτηρίων είναι κόκκινα ή ρόζ εξαιτίας της φυκοερυθρίνης, η οποία επίσης είναι φυκοβιλλιπρωτεΐνη [31].

### (c) Λειχήνες (Λειχηνοποιημένοι μύκητες)

Οι λειχήνες, ενώ αποτελούν συμβιωτικά συστήματα τα οποία αποτελούνται από μύκητες (συνήθως Ascomycetes), ευκαρυωτικά χλωροφύκη και κυανοβακτήρια, δεν παρουσιάζουν τις ιδιότητες των συμβιωτών αλλά πλήρως διαφορετικές [17]. Ο μύκητας, ο οποίος συνήθως αναφέρεται ως μυκητοβιότης, φιλοξενεί τα χλωροφύκη ή τα κυανοβακτήρια, δηλαδή τον φωτοβιότη, δημιουργώντας ένα κατάλληλο και αυστηρά ελεγχόμενο οικότοπο ο οποίος προσομοιάζει ένα θερμοκήπιο. Ειδικότερα, ο μυκητοβιότης οργανώνει και συγκεντρώνει τον φωτοβιότη σε ένα διαχωρισμένο στρώμα εντός της μυκητιακής δομής, ρυθμίζοντας την ποιότητα και την ποσότητα του φωτός, το νερό και τα θρεπτικά συστατικά που καταφθάνουν στον φωτοβιότη. Επιπλέον, προωθεί την διανταλλαγή αερίων, ελέγχει το χημικό περιβάλλον και εξουδετερώνει αποτελεσματικά ορισμένες ομάδες δυνητικά τοξικών αερομεταφερόμενων στοιχείων τα οποία μπορεί να συσσωρεύσει η λειχήνα. Σε αντάλλαγμα ο φωτοβιότη παρέχει τις οργανικές ενώσεις μέσω φωτοσύνθεσης που ο μυκητοβιότης έχει ανάγκη για να καλύψει τις ενεργειακές του ανάγκες [32].

Οι λειχήνες έχουν την δυνατότητα να επιβιώσουν στα πιο ακραία περιβάλλοντα της Γης όπως η αρκτική τούντρα, οι έρημοι και οι πετρώδεις ακτές εξαιτίας της συνδυαστικής ανθεκτικότητας των συμβιωτών. Το χαρακτηριστικό αυτό συνδέεται με την ιδιότητα τους να αποξηραίνονται εξαιρετικά εύκολα όταν οι συνθήκες γίνουν αντίξοες, κατά την διάρκεια των οποίων σταματά και η διαδικασία της φωτοσύνθεσης [17]. Τα χαρακτηριστικά αυτά τις αναδεικνύουν σημαντικό παράγοντα βιοδιάβρωσης πέτρινων επιφανειών καθώς το περιβλήμα τόσο των επιλιθικών όσο και των ενδολιθικών λειχήνων διειδύει στα υποστρώματα, τα οποία διαβρώνει προκειμένου να προσκολληθεί [23].

## Επιπτώσεις Βιοαποίκησης φωτοαυτότροφων μικροοργανισμών - Βιοτάπητες

Οι περιβαλλοντικές συνθήκες, ειδικά στην περίπτωση των φωτοαυτότροφων μικροοργανισμών ευνοούν εξαιρετικά την επιλιθική ανάπτυξη τους. Ανάλογα με τον τύπο του υποστρώματος η αποίκηση είτε από κυανοβακτήρια είτε από χλωροφύκη ποικίλει. Η καταστροφή που προκαλούν στο υπόστρωμα που εποικίζουν προέρχεται συνήθως από την συμβίωση τους στους *Βιοτάπητες*, στα οποία όμως συμμετέχουν και ετερότροφοι μικροοργανισμοί όπως μύκητες και βακτήρια[29]. Εξετάζοντας συγκεκριμένα όμως μόνο τα χλωροφύκη, η καταστρεπτική τους δράση εστιάζεται στο ότι αποτελούν κατάλληλο υπόστρωμα θρεπτικών συστατικών για την δημιουργία αποικιών ετερότροφων μικροοργανισμών [17].

Γενικότερα οι μικροοργανισμοί και σε ακόμη μεγαλύτερο βαθμό τα κυανοβακτήρια, όταν οι συνθήκες μετατραπούν σε αντίξοες παράγουν έναν εξωκυτταρικό περίβλημα από πολυμερείς ουσίες -*Extracellular Polymeric Substances (EPS)*- όπως πολυσακχαρίτες, λιποπολυσακχαρίτες, πρωτεΐνες, γλυκοπρωτεΐνες, λιπίδια, γλυκολιπίδια, λιπαρά οξέα και ένζυμα[6]. Η κολλώδης υφή του επιτρέπει τον εγκλωβισμό υγρασίας ενώ παράλληλα προσκολλάται ισχυρά στο υπόστρωμα[30]. Η δημιουργία του βιοτάπητα βασίζεται στην ύπαρξη των EPS ως συγκολλητική ουσία των δευτερογενών αποικιών, η οποία αυξάνεται σημαντικά όταν αυξάνεται και η τραχύτητα της επιφάνειας. Αυτό συμβαίνει διότι μειώνονται σημαντικά οι διατμητικές δυνάμεις λόγω του αέρα και επομένως η συγκόλληση τόσο των EPS στο υπόστρωμα όσο και τον ετερότροφων μικροοργανισμών στα EPS γίνεται αποτελεσματικότερη. Οι βιοτάπητες προκαλούν αλλοίωση στις επιφάνειες που αναπτύσσονται μέσω δύο διαφορετικών κατηγοριών από μηχανισμούς, βιογεωφυσικούς και τους βιογεωχημικούς:

### (a) Βιογεωφυσική Αλλοίωση

Με τον όρο Βιογεωφυσική Αλλοίωση προσδιορίζεται η μηχανική καταπόνηση η οποία επάγεται εξαιτίας της βιολογικής ανάπτυξης, η οποία καταλήγει σε επιφανειακές αποκολλήσεις (Biopitting) ή σε διείδυση στο υπόστρωμα και αύξηση της ποροσιότητάς του. Οι βιοτάπητες επάγουν μηχανικό στρες στην επιφάνεια που προσκολλούνται εξαιτίας της συστολής και της διαστολής τους ανάλογα με το αν το περιβάλλον είναι υγρό ή ξηρό αντίστοιχα. Αυτό μπορεί να έχει ως αποτέλεσμα την ανακατανομή των πόρων και επομένως της ενίσχυσης της ρήσης του νερού επί της επιφάνειας [6].

### (b) Βιογεωχημική Αλλοίωση

Η βιογεωχημική αλλοίωση προκύπτει εξαιτίας της άμεσης επαφής του υποστρώματος με προϊόντα του μεταβολισμού των μικροοργανισμών. Συνηθέστερη πηγή βιοδιαβρωσης είναι τα οργανικά οξέα που παράγονται από τα κύτταρα, τα οποία διαλυτοποιούν την ασβεστολιθική συνήθως επιφάνεια. Επιπλέον, οι μικροοργανισμοί αποβάλλουν οργανικούς χηλικούς παράγοντες που δεσμεύουν μεταλλικά κατιόντα από πέτρα όπως επίσης μετατρέπουν ανόργανες ουσίες από αντιδράσεις οξειδοαναγωγής, η οποίες στη συνέχεια σχηματίζουν οξέα που χαράζουν πέτρα και συμβάλλουν στο σχηματισμό αλάτων[6].

Επιπρόσθετα, η βιογεωχημική αλλοίωση συνδέεται με την ύπαρξη αποχρωματισμών, δυσχρωματισμών και πατίνας. Ειδικότερα, η κολλώδης επιφάνεια των βιοτάπητων ενισχύει την επικόλληση αερομεταφερόμενων σωματιδίων όπως σκόνη, γύρη, σπόρια, σωματίδια από καύση υδρογονανθράκων, γεγονός το οποίο ενισχύει την δημιουργία κρούστας [6]. Επίσης, οι επιλιθικοί βιοτάπητες, οι οποίοι σχηματίζονται από κυανοβακτήρια, δημιουργούν χρωματισμούς στο υπόστρωμα λόγω των χρωμοφόρων ουσιών που περιέχουν δηλαδή της χλωροφύλλης, των φυκοβιλλιπρωτεϊνών και των καροτενοειδών καθώς και ουσιών προστασίας από την UV ακτινοβολία[29].



Στην περίπτωση των Λειχήνων, η κύρια πηγή βιοαλλοίωσης του υποστρώματος προκύπτει εξαιτίας του μυκητοβιότη, που έρχεται σε άμεση επαφή με το υπόστρωμα. Το βιογεωφυσικό στρες προκύπτει και σε αυτήν περίπτωση από την διαδοχική συστολή και διαστολή του θαλλού, καθώς επίσης και από την ευενδολιθική δράση των υφών των μυκήτων, δομή η οποία θα περιγραφεί στη συνέχεια, οι οποίες είναι ικανές να διεισδύσουν στο υπόστρωμα έως και 20 mm ανάλογα με το είδος του λειχήνα που ανήκουν [23].

Η βιογεωχημική δράση των λειχήνων επάγει την νεομορφοποίηση των υποστρωμάτων λόγω της αλληλεπίδρασης του με το παραγόμενο οξαλικό οξύ, καθώς και άλλως ενώσεων με χηλική ικανότητα. Εξαρτάται τόσο από το είδος του υποστρώματος όσο και από το είδος του λειχήνα. Συνήθως μετά την αλληλεπίδραση με ασβεστολιθικά πετρώματα, εμφανίζεται οξαλικό ασβέστιο καθώς υπάρχει αφθονία ανόργανων αλάτων ασβεστίου. Το οξαλικό ασβέστιο δύναται να εναποτεθεί σε διαφορετικά σημεία του λειχήνα ανάλογα με το είδος και επίσης υπάρχει η πιθανότητα να εισχωρήσει και στο εσωτερικό του βράχου έως και το βάθος που διεισδύουν και οι υφές. Αντίστοιχα, μπορούν να δημιουργηθούν οξαλικό μαγνήσιο, χαλκό, σίδηρο και μαγγάνιο ανάλογα με το είδος του υποστρώματος[23].

## 2. Χημειολιθοτροφικοί Μικροοργανισμοί

Ως χημειολιθοτροφικούς οργανισμούς συνήθως θεωρούνται ορισμένα ετερότροφα βακτήρια που μετατρέπουν το θείο σε θειικό άλας, το οποίο μπορεί να βιοδιαβρώσει το υπόστρωμα εάν έρθει σε επαφή με όξινη βροχή και συνεπώς η δράση τους είναι κυρίως βιογεωχημική. Παρόλα αυτά όμως, η δράση τους μπορεί να διασταυρωθεί με την δράση της όξινης βροχής καθώς το θειικό οξύ που παράγεται μπορεί να προέρχεται απευθείας από την βροχόπτωση. Επιπλέον, εκκρίνουν θειικό ασβέστιο το οποίο διασπά το πέτρωμα και επίσης όταν αυτά τα βακτήρια βρίσκονται στο χώμα γύρω από το περιβάλλον του μνημείου, τότε μειώνουν τα θειικά άλατα του εδάφους και παράγουν υδρόθειο ως προϊόν του μεταβολισμού τους. Στη συνέχεια, το υδρόθειο μεταφέρεται στην πέτρα και αυξάνει την ποροσιμότητά της.

Η δεύτερη κατηγορία χημειολιθοτροφικών βακτηρίων είναι τα βακτήρια νιτροποίησης. Ειδικότερα, σε μια φθαρμένη πέτρα, όπου τα επίπεδα θειικού άλατος είναι ήδη χαμηλά, οξειδώνονται με την αμμωνία από τον αέρα και σχηματίζουν νιτρικά οξέα αντιδρώντας με το ανθρακικό ασβέστιο ασβεστολιθικού βράχου. Αυτό έχει ως αποτέλεσμα να σχηματιστεί διαλυτό νιτρικό ασβέστιο το οποίο, όταν εκπλυθεί με νερό βροχής, αφήνει πίσω σκόνη διοξειδίου του πυριτίου, αποσυνθέτοντας περαιτέρω την πέτρα. Επιπλέον, τα βακτήρια νιτροποίησης οξειδώνουν το ανόργανο άζωτο, όπως νιτρικά άλατα, για ενεργειακούς λόγους και το χρησιμοποιούν για να παράγουν όξινα προϊόντα όπως το νιτρώδες και το νιτρικό οξύ. Επομένως, η αποικία και η δράση τους είναι ενισχυμένη μόνο εάν το υπόστρωμα είναι πλούσιο σε αμμωνία και νιτρικά άλατα. Η αμμωνία μπορεί να στερεωθεί στην πέτρα με τη μορφή σκόνης αλάτων αμμωνίου και το νιτρικό άλας οφείλεται στους ρύπους των αυτοκινήτων, του εδάφους ή των εργοστασίων.

### 3. Ετερότροφοι Μικροοργανισμοί - Μύκητες

Οι μύκητες ως ετερότροφοι και συγκεκριμένα χημειοργανότροφοι [33] ευκαρυωτικοί μικροοργανισμοί, αξιοποιούν για την επιβίωση τους οργανικές ενώσεις η οποίες έχουν παραχθεί από άλλους οργανισμούς. Το εύρος της οργάνωσής τους ποικίλει από μονοκύτταρους έως εξαιρετικά πολύπλοκους πολυκύτταρους. Περιβάλλονται από ένα ανθεκτικό κυτταρικό τοίχωμα το οποίο αποτελείται από χιτίνη (chitin), γλυκάνες και πρωτεΐνες. Η μεταβολική ευελιξία αυτών των μικροοργανισμών τους επιτρέπει να αποικούν ποικίλα υποστρώματα όπως ξύλο, γυαλί, πέτρα και χαρτί.

Συνήθως εμφανίζονται ως διακλαδισμένα νημάτια, τις υφές, οι οποίες προέρχονται από την εκβλάστηση των σπορίων, μέσω των οποίων πραγματοποιείται η αναπαραγωγή και διασπορά των μυκήτων στο περιβάλλον τους. Η διασπορά αυτή επηρεάζεται από τα φυσικά χαρακτηριστικά των σπορίων όπως το μέγεθος, η πυκνότητα και το σχήμα[34]. Επιπλέον, τα σπόρια χαρακτηρίζονται από ένα εξαιρετικά σκληρό περίβλημα που επιτρέπει στους μύκητες να επιβιώνουν σε ακραία περιβάλλοντα[35]. Πιο συγκεκριμένα, οι υφές εμφανίζονται ως κυλινδρικές δομές με διάμετρο από 2 έως 10 μm και απεριόριστο μήκος, όσο υπάρχουν διαθέσιμα θρεπτικά συστατικά[17]. Αναπτύσσονται στην άκρη τους και διαιρούνται συνεχώς κατά μήκος δημιουργώντας ένα πολυδιακλαδισμένο δίκτυο υφών, τα μυκήλια.Στις περισσότερες περιπτώσεις τα σπόρια των μυκήτων δημιουργούν υφές όταν οι περιβαλλοντικές συνθήκες το επιτρέπουν. Όμως αρκετές φορές η ενεργοποίηση τους απαιτεί κατάλληλο χημικό περιβάλλον, γεγονός το οποίο εντείνει την βιοδιάβρωση αντί να την αντιμετωπίσει όταν οι διαδικασίες συντήρησης δεν διεκπεραιωθούν σωστά[35]. Οι βασικές κατηγορίες μυκήτων που εποικίζουν την πέτρα και γενικότερα τα υποστρώματα με δομικά υλικά είναι οι *Υφομύκητες - (Hyphomycetes)* και οι *Μαύροι Μεριστεματικοί Μύκητες - Black Meristematic Fungi (BMF)*.

Οι μύκητες που ανήκουν στην ομάδα των Υφομύκητων, παραδείγματα των οποίων είναι οι *Penicillium, Fusarium, Aspergillus - Hyaline Hyphomycetes-* καθώς και *Alternaria, Ulocladium, Cladosporium - Dematiaceous Hyphomycetes-* είναι κυρίως αερομεταφερόμενοι αλλά σε κάποιες περιπτώσεις υπάρχουν και στο περιβάλλον του μνημείου[35]. Αναπτύσσονται όταν οι κλιματικές συνθήκες είναι ήπιες με μέτρια επίπεδα υγρασίας και σε περιοχές με πλούσιο οργανικό απόθεμα, παρόλα αυτά όμως επιβιώνουν σε μεγάλες θερμοκρασιακές μεταβολές και ανωμαλίες στα επίπεδα υγρασίας. Περιλαμβάνουν είδη τα οποία παράγουν λευκές ή συνήθως έντονα χρωματισμένες αποικίες και ανήκουν στην πρώτη κατηγορία, ενώ στην δεύτερη κατηγορία οι αποικίες παρουσιάζουν καφέ, σκούρο πράσινο ή μαύρο χρώμα. Βιοδιαβρώνουν τις επιφάνειες παράγοντας οργανικά οξέα τα οποία είναι διαλύτες των ανθρακικών αλάτων ενώ σπανιότερα παράγουν μαύρες χρωστικές όπως μελανίνες[23].

Η δεύτερη κατηγορία μυκήτων και συνήθως η πιο βιοδιαβρωτική[23] είναι οι Μαύροι Μεριστεματικοί Μύκητες (BMF), στους οποίους τα μορφολογικά και φυσιολογικά χαρακτηριστικά τους επιτρέπουν την επιβίωση σε ακραίες συνθήκες θερμοκρασιών, ακτινοβολίας και αποθέματος τροφής. Ενώ ο τύπος μεριστεματικής ανάπτυξης είναι συνήθως ένα χαρακτηριστικό των μυκήτων το οποίο εμφανίζεται όταν οι συνθήκες γίνουν αντίξοες, αποτελεί βασικό χαρακτηριστικό των BMF. Η δυνατότητα τους αναπτύσσονται ακτινωτά, να σχηματίζουν κυκλικές αποικίες και να διογκώνονται ισοδιαμετρικά μειώνει τον αναλογία όγκου/επιφάνειας που εφάπτεται του υποστρώματος και έτσι ελαχιστοποιείται η έκθεση σε εξωτερικούς στρεσογόνους παράγοντες [36]. Επιπλέον, είναι εξαιρετικά ολιγοτροφικοί, το οποίο τους καθιστά ικανούς να τρέφονται με αραιά, αερομεταφερόμενα θρεπτικά συστατικά τα οποία είναι διαθέσιμα σε μορφή σκόνης. Ειδικότερα, παρουσιάζουν αξιοσημείωτες πεπτικές ικανότητες καθώς μπορούν να αφομοιώσουν τους ρύπους ως αρωματικές ενώσεις [24], γεγονός το οποίο αυξάνει την ποικιλομορφία τους σε περιβάλλοντα με υψηλό

οργανικό ρυπογόνο φορτίο όπως τα αστικά [37]. Επιπροσθέτως, οι BMF εκτίθενται σε υψηλά επίπεδα εσωτερικής ωσμωτικής πίεσης εξαιτίας της υψηλής αλατότητας και έλλειψής νερού στις πετρώδεις επιφάνειες, συνθήκη την οποία αντιμετωπίζουν ενδοκυτταρικά παράγοντας κατάλληλες ενώσεις για την εξισορρόπηση των πιέσεων. Μια τέτοια ένωση είναι η μελανίνη, η οποία επιπρόσθετα προσφέρει στους BMF προστασία από την διαρκή θερμική καταπόνηση, τα ακραία pH και την πολυχρωματική UV ακτινοβολία[24], ενώ παράλληλα τους προσδίδει το σκούρο πράσινο ή μαύρο χρώμα που χαρακτηρίζει τις αποικίες τους.

## **Βιοαλλοιωτική Δράση Μυκήτων**

Η συγκεκριμένη ομάδα μικροοργανισμών παρουσιάζει εξαιρετική προσαρμοστικότητα σε ακραία περιβάλλοντα όπως σε υψηλή μεταλλική τοξικότητα, UV ακτινοβολία, υψηλό ή χαμηλό pH, αλατότητα και αποξήρανση και αναπτύσσει μια ποικιλία τρόπων ανάπτυξης, μεταβολικής διεργασίας και συμβίωσης με άλλους μικροοργανισμούς για να επιβιώσει. Η νηματοειδής ανάπτυξη τους δίνει την δυνατότητα να αξιοποιήσουν τις περιοχές του υποστρώματος οι οποίες είναι καταλληλότερες για την ανάπτυξή τους. Όπως και στις προηγούμενες κατηγορίες βιοαλλοιωτικών μικροοργανισμών, έτσι και στην περίπτωση των μυκήτων, η αλλοίωση του υποστρώματος επί του οποίου αναπτύσσονται διαχωρίζεται σε βιογεωφυσική και βιογεωχημική καταλήγοντας στην βιοδιάβρωση των επιφανειών και την δημιουργία πατίνας [38]. Αντίστοιχα με τους λειχήνες, το μηχανικό στρες που επάγεται στο υπόστρωμα, δηλαδή η βιογεωφυσική συνιστώσα αλλοίωσης, προκαλείται από την διείσδυση των υφών εντός του υποστρώματος [33]. Πιο συγκεκριμένα, όταν οι περιβαλλοντικές συνθήκες γίνουν εξαιρετικά αντίξοες στην περίπτωση των BMF, τότε η αποικία εξαφανίζεται από την επιφάνεια και διεισδύει στο εσωτερικό του λίθου ακολουθώντας την μικροσκοπική κρυσταλλική δομή του ειδικά στην περίπτωση ασβεστολιθικών πετρωμάτων [38]. Επομένως η δημιουργία σχισμών και ρωγμών είναι αναπόφευκτη.

Οι καταστροφές που προκαλούνται εξαιτίας της βιογεωχημικής δράσης των μυκήτων οφείλονται αφενώς σε ουσίες που διαθέτουν οι μύκητες για την προστασία τους όπως η μελανίνη στους BMF και αφετέρου σε ουσίες που παράγονται κατά τον μεταβολισμό τους και οξειδώνουν τα ανόργανα υλικά των υποστρωμάτων όπου αναπτύσσονται. Ειδικότερα, οι μύκητες παράγουν οξέα όπως το οξαλικό και το κιτρικό, τα οποία είναι σημαντικοί βιοδιαβρωτές ειδικά των πυριτικών πετρωμάτων όπως ο ψαμμίτης [33]. Επιπλέον, μια ενδιαφέρουσα μεταβολή που προκαλούν είναι η δημιουργία πιο όξινου περιβάλλοντος εκεί όπου δημιουργείται η αποικία τους. Αυτό συμβαίνει εξαιτίας της μεταφοράς πρωτονίων από την πλασματική τους μεμβράνη καθώς μετατοπίζεται η ATPase προκειμένου να γίνει η μεταφορά θρεπτικών συστατικών [33]. Τέλος, λαμβάνοντας υπόψη τις υψηλές περιεκτικότητες καροτενοϊδών και μελανίνης που εμφανίζεται σε αρκετούς μύκητες και ειδικά στους BMF, τα φαινόμενα χρωματισμού και σχηματισμού χρούστας επί των επιφανειών που εποικίζουν οι μύκητες είναι αρκετά συνηθισμένα [24]. Πιο συγκεκριμένα, δημιουργούν διαχρωματισμούς στο υπόστρωμα όπως γκρι, μαύρες και πορτοκαλί πατίνες [23].

## 2 Κεφάλαιο 2: Μαθηματικό Υπόβαθρο

Η επεξεργασία δεδομένων με υψηλό αριθμό μεταβλητών και παρατηρήσεων απαιτεί την χρήση κατάλληλων αλγορίθμων, οι οποίοι θα είναι ικανοί να διακρίνουν μοτίβα που υφίστανται μεταξύ των μεταβλητών, τα οποία εξαιτίας του μεγάλου όγκου της πληροφορίας, είναι αδύνατο να παρατηρηθούν χωρίς την βοήθεια υπολογιστή. Στην συγκεκριμένη διπλωματική εργασία, τα δεδομένα που μελετήθηκαν περιείχαν 1688 μεταβλητές δηλαδή τα μήκη κύματος των φασμάτων φθορισμού. Επομένως, η αποτελεσματικότερη εξόρυξη της πληροφορίας επετεύχθη με χρήση αλγορίθμων μείωσης διαστάσεων όπως η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis - PCA), ομαδοποίησης όπως η Ομαδοποίηση κατά Συστάδες (Hierarchical Agglomerative Clustering - HAC) και πρόβλεψης όπως η Γραμμική Διακριτική Ανάλυση (Linear Discriminant Analysis - LDA).

Τα μοτίβα που προκύπτουν μεταξύ των μεταβλητών που προσδιορίζουν τα δεδομένα προσφέρουν την δυνατότητα ταξινόμησης τους σε κλάσεις. Εάν ο αλγόριθμος δεν λαμβάνει καμία πληροφορία σχετικά με τις μεταβλητές που προσδιορίζουν κάποια κλάση, τότε η αναζήτηση των μοτίβων στηρίζεται σε εγγενή χαρακτηριστικά των δεδομένων και ονομάζεται *Μη επιβλεπόμενη - Unsupervised*. Αυτό πρακτικά σημαίνει πως πραγματοποιείται κατάλληλος έλεγχος εάν τα δεδομένα ομαδοποιούνται χωρίς κάποια συγκεκριμένη επεξεργασία. Στην κατηγορία αυτών των αλγορίθμων ανήκει η PCA και η HAC. Σύμφωνα με την PCA, όταν ο αριθμός των μεταβλητών του δείγματος μειωθεί σε σημαντικά μικρότερο όπως πχ 2 ή 3, γίνεται σαφώς ευκολότερη η γραφική τους απεικόνιση και εξερεύνηση ύπαρξης εγγενών κλάσεων. Εναλλακτικά, ο αλγόριθμος της HAC υπολογίζει τις γεωμετρικές αποστάσεις μεταξύ των τιμών των μεταβλητών για κάθε παρατήρηση και στη συνέχεια ομαδοποιεί εκείνες τις παρατηρήσεις που βρίσκονται πιο "κοντά" μεταξύ τους.

Ωστόσο, οι σχέσεις μεταξύ των μεταβλητών είναι συνήθως εξαιρετικά πολύπλοκες και οι unsupervised αλγόριθμοι δεν είναι αρκετοί για την ταξινόμηση των δεδομένων σε ομάδες. Η λύση σε αυτό το πρόβλημα αναδύεται από τον κλάδο της Μηχανικής Μάθησης μέσω της αξιοποίησης στατιστικών μοντέλων μάθησης και πρόβλεψης. Πλέον, η ομαδοποίηση των δεδομένων μετατρέπεται σε *Επιβλεπόμενη - Supervised*, καθώς πριν την ταξινόμηση των παρατηρήσεων σε κλάσεις, έχει δημιουργηθεί ένα μοντέλο από παρατηρήσεις αναφοράς. Οι κλάσεις αυτών των παρατηρήσεων είναι εκ των προτέρων γνωστές και τα στατιστικά χαρακτηριστικά τους προσδιορίζουν την πιθανότητα μια νέα παρατήρηση να ανήκει σε μία από αυτές. Μια χαρακτηριστική μέθοδος ταξινόμησης σε αυτή την κατηγορία είναι και η LDA. Στην μέθοδο αυτή, όταν δίνεται η πληροφορία για τις κλάσεις των παρατηρήσεων αναφοράς, διαχωρίζει τις κλάσεις αυτές με χρήση γραμμικών ορίων (linear boundaries).

Στο κεφάλαιο αυτό θα παρουσιαστεί αναλυτικά το μαθηματικό υπόβαθρο της μεθόδου PCA με χρήση της μεθόδου Ανάλυσης πίνακα σε ιδιάζουσες τιμές (Singular Value Decomposition-SVD). Επίσης θα θεμελιωθούν οι μέθοδοι HAC και LDA. Επιπλέον, θα παρουσιαστεί ένα τρόπος αξιολόγησης της απόδοσης του μοντέλου μιας supervised ταξινόμησης με χρήση της μεθόδου Leave-One-Out Cross Validation και κατάλληλων εργαλείων ελέγχου αλήθειας των προβλέψεων όπως ο Confusion Matrix. Επίσης, στο Παράρτημα A παρατίθενται τα βασικά στοιχεία στατιστικής, τα οποία είναι απαραίτητα για την κατανόηση των μεθόδων.

## 2.1 Στατιστική μέθοδος ανάλυσης Principal Component Analysis (PCA)

### 2.1.1 Ανάλυση πίνακα σε ιδιάζουσες τιμές (Singular Value Decomposition-SVD)

Ένας πίνακας, τετραγωνικός ή μη, έχει τη δυνατότητα να αναλυθεί σε τρεις επιμέρους πίνακες εφαρμόζοντας την μέθοδο ανάλυσης ιδιάζουσών τιμών. Έστω ένας πίνακας  $X$  με διαστάσεις  $n \times p$ . Η ανάλυση πραγματοποιείται μέσω της παραγοντοποίησης:

$$X = U\Sigma V^T \quad (25)$$

όπου ο πίνακας  $U$  έχει διάσταση  $n \times n$ , ο πίνακας  $\Sigma$  έχει διάσταση  $n \times p$ , και ο πίνακας  $V$  έχει διάσταση  $p \times p$ . Οι πίνακες  $U$  και  $V$  είναι πραγματικοί μοναδιαίοι, δηλαδή ισχύει ότι:

$$\begin{aligned} U^T U &= U U^T = I \\ V^T V &= V V^T = I \end{aligned} \quad (26)$$

Ο πίνακας  $\Sigma$  έχει τη μορφή:

$$\Sigma = \begin{pmatrix} \Sigma_+ & 0_{r \times (p-r)} \\ 0_{(n-r) \times r} & 0_{(n-r) \times (p-r)} \end{pmatrix} \mu\epsilon \quad \Sigma_+ = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_r \end{pmatrix} \quad (27)$$

όπου  $r$  ο βαθμός του πίνακα  $X$  και  $\sigma_1, \dots, \sigma_r$ , οι ιδιάζουσες τιμές του  $X$ .

Για την εύρεση των πινάκων  $U$ ,  $\Sigma$  και  $V$  θα υπολογιστούν οι ιδιοτιμές και τα ιδιοδιανύσματα των πινάκων  $X^T X$  και  $X X^T$ . Ισχύει ότι:

$$X^T X = (V \Sigma^T U^T) (U \Sigma V^T) = V \Sigma^T \Sigma V^T \quad (28)$$

$$X X^T = (U \Sigma V^T) (V \Sigma^T U^T) = U \Sigma \Sigma^T U^T \quad (29)$$

όπου:

$$\Sigma^T \Sigma = \begin{pmatrix} \Sigma_+^2 & 0_{r \times (p-r)} \\ 0_{(p-r) \times r} & 0_{(p-r) \times (p-r)} \end{pmatrix} \quad (30)$$

με:

$$\Sigma_+^2 = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_r^2 \end{pmatrix} \quad (31)$$

όπου οι στήλες του  $V$  περιέχουν τα ιδιοδιανύσματα του  $X^T X$  και ο διαγώνιος πίνακας  $\Sigma^T \Sigma$  περιέχει τις ιδιοτιμές του. Αντίστοιχα ο  $U$  και ο  $\Sigma \Sigma^T$  για τον  $X X^T$ . Παρατηρείται πως ο  $V$  και ο  $\Sigma^T \Sigma$  περιέχουν τα ιδιοδιανύσματα και τις ιδιοτιμές (απουσία του πολλαπλασιαστικού παράγοντα  $n-1$ ) του πίνακα συνδιακύμανσης πολυμεταβλητών τυποποιημένων δεδομένων σύμφωνα με τη σχέση A-6. Καθώς όμως ο  $\Sigma$  έχει μηδενικές  $(p-r)$  μηδενικές στήλες και  $(p-r)$  μηδενικές γραμμές, στις σχέσεις 28 και 29 θα αφαιρεθούν αυτά τα μηδενικά στοιχεία από τον  $\Sigma$ , καθώς και οι στήλες που αυτά θα μηδένιζαν

στους  $V$  και  $U$ . Επομένως θα ισχύει ότι:

$$\left\{ \begin{array}{l} X^T X = V \Sigma^T \Sigma V^T = V_{(p \times r)} \Sigma_+^2 V^T \\ X X^T = U \Sigma \Sigma^T U^T = U_{(n \times r)} \Sigma_+^2 U^T \end{array} \right\} \Rightarrow X = U_{(n \times r)} \Sigma_+ V_{(p \times r)}^T \quad (32)$$

Οι στήλες  $u_i$  και  $v_i$  όπου αντιστοιχούν στον  $U$  και  $V$  θα δίνονται από τις σχέσεις:

$$\left\{ \begin{array}{l} X V_{(p \times r)} = U_{(n \times r)} \Sigma_+ \\ X^T U_{(n \times r)} = V_{(p \times r)} \Sigma_+ \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} u_i = \frac{1}{\sigma_i} X v_i \\ v_i = \frac{1}{\sigma_i} X^T u_i \end{array} \right\} \quad (33)$$

Τα ιδιοδιανύσματα στον ορθογώνιο πίνακα  $U$  ονομάζονται αριστερά ιδιάζοντα διανύσματα του  $X$  και τα αντίστοιχα στον ορθογώνιο πίνακα  $V$  ονομάζονται δεξιά ιδιάζοντα διανύσματα του  $X$ . Η αντιστοιχία που τα προσδιορίζει δίνεται από τη σχέση:

$$u_i^T X v_i = v_i^T X^T u_i = \sigma_i \quad (34)$$

Με την ανάλυση σε ιδιάζουσες τιμές έχουν ταυτόχρονα προσδιορισθεί ορθογώνιες βάσεις για τους τέσσερις θεμελιώδεις υπόχωρους του πίνακα  $A$ . Τα διανύσματα του  $U$  που αντιστοιχούν στις μη μηδενικές ιδιάζουσες τιμές αποτελούν βάση για το χώρο των στηλών του  $A$ , ενώ τα υπόλοιπα είναι βάση για τον αριστερό μηδενόχωρο του  $A$ . Τα διανύσματα του  $V$  που αντιστοιχούν στις μη μηδενικές ιδιάζουσες τιμές αποτελούν βάση για το χώρο των γραμμών του  $A$ , ενώ τα υπόλοιπα είναι βάση για το μηδενόχωρο του  $A$  [39].

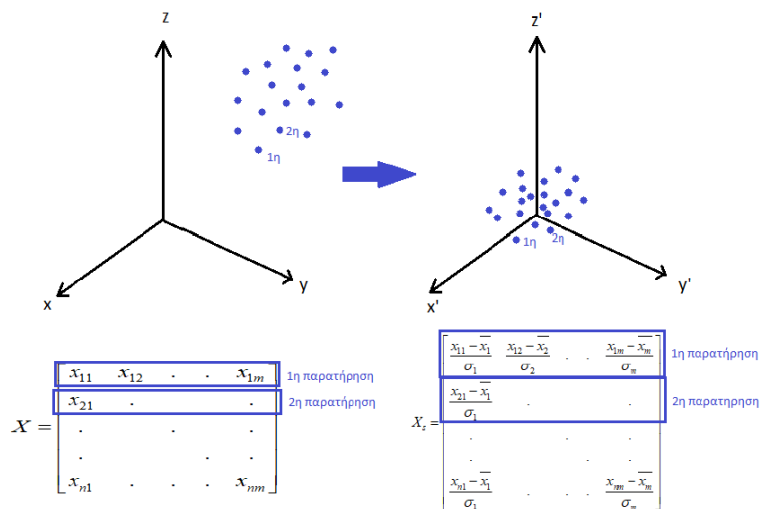
### 2.1.2 Ανάλυση μεθόδου Principal Component Analysis (PCA)

Η στατιστική μέθοδος Principal Component Analysis βρίσκεται μεταξύ των παλαιότερων και εκτενέστερα ερευνημένων στατιστικών μεθόδων ανάλυσης δεδομένων που στοχεύουν στην μείωση των μεταβλητών που συμμετέχουν σε ένα πείραμα. Η βασική ιδέα της μεθόδου είναι να περιγράψει την διακύμανση που παρουσιάζουν οι διαφορετικές παρατηρήσεις ενός συνόλου πολυμεταβλητών δεδομένων χρησιμοποιώντας ένα νέο σύνολο μεταβλητών, τα Principal Components. Οι νέες αυτές μεταβλητές εξάγονται σε φθίνουσα σειρά σημαντικότητας, δηλαδή το πρώτο principal component περιγράφει την όσο το δυνατόν μεγαλύτερη διακύμανση στα αρχικά δεδομένα. Στη συνέχεια, το δεύτερο principal component επιλέγεται ώστε να περιγράψει την όσο το δυνατόν μεγαλύτερη διακύμανση από την αντίστοιχη υπολειπόμενη, με την προϋπόθεση πως είναι ασυσχέτιστο με το πρώτο principal component. Η διαδικασία συνεχίζεται και για τα υπόλοιπα principal components. Το αντικείμενο της παρούσας ανάλυσης εστιάζει στο αν τα πρώτα ελάχιστα components μπορούν να περιγράψουν σχεδόν τη συνολική διακύμανση των δεδομένων, με ελάχιστη απώλεια πληροφορίας, μειώνοντας τις διαστάσεις του αρχικού χώρου στις διαστάσεις του χώρου των components. Η ανάλυση θα πραγματοποιηθεί με χρήση της μεθόδου SVD.

#### Μεθοδολογία

##### 1. Τυποποίηση Δεδομένων - Standardization:

Έστω ο πίνακας πολυμεταβλητών δεδομένων  $X$  με διαστάσεις  $n \times p$ , ο οποίος παρουσιάζεται παρακάτω στο Σχήμα 13. Η διαδικασία της τυποποίησης περιλαμβάνει την αφαίρεση της μέσης τιμής της κάθε μεταβλητής στις  $n$  παρατηρήσεις από τις αντίστοιχες τιμές και στη συνέχεια τη διαίρεση των τιμών αυτών με την τυπική απόκλιση της αντίστοιχης μεταβλητής, δηλαδή με την τετραγωνική ρίζα της διακύμανσης. Μέσω αυτών των πράξεων παράγονται τα τυποποιημένα δεδομένα, δηλαδή ο πίνακας  $X_s$  όπως φαίνεται στο Σχήμα 13. Η γεωμετρική ερμηνεία της διαδικασίας αυτής παρουσιάζεται και πάλι στο Σχήμα 13 και προσδιορίζεται ως μεταφορά των δεδομένων στην αρχή των αξόνων και επιβολή τυπικής απόκλισης ίσης με 1. Ο συγκεκριμένος



Σχήμα 13: Γεωμετρική και μαθηματική περιγραφή της τυποποίησης δεδομένων. Μετά την διαδικασία της τυποποίησης τα δεδομένα έχουν μέση τιμή το 0 και τυπική απόκλιση ίση με 1.

μετασχηματισμός δεν παραλλάσει τα δεδομένα, αλλά τα μεταφέρει όλα στην ίδια κλίμακα. Η ανάγκη τυποποίησης προκύπτει καθώς στη συνέχεια θα χρησιμοποιηθεί ο πίνακας συνδιακύμανσης, ο οποίος είναι εξαρτημένος από τις διαστάσεις των δεδομένων, γεγονός το οποίο επηρεάζει

έντονα την ανάλυση που ακολουθεί. Ο μετασχηματισμός αυτός ισοδυναμεί με την μετατροπή του πίνακα συνδυαζόμενης σε πίνακα συσχέτισης και συνεπώς τα δεδομένα μπορούν πλέον να αναλυθούν συλλογικά [40].

## 2. Ορισμός των Principal Components

- Το πρώτο principal component των παρατηρήσεων ορίζεται ως ο γραμμικός συνδυασμός  $Y_1$  των αρχικών μεταβλητών, του οποίου η δειγματική διακύμανση είναι η μεγαλύτερη από κάθε άλλο γραμμικό συνδυασμό.

$$Y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p = \mathbf{a}_1 \mathbf{x}^T \quad (35)$$

όπου  $\mathbf{x} = [x_1, x_2, \dots, x_p]$  οι μεταβλητές που αντιστοιχούν στις στήλες του πίνακα δεδομένων. Καθώς η διακύμανση μπορεί να αυξάνεται επ'άπειρον απλώς αυξάνοντας τους συντελεστές  $\mathbf{a}_1 = [a_{11}, a_{12}, \dots, a_{1p}]$ , οι οποίοι ονομάζονται βάρη, θα πρέπει να επιβληθεί κατάλληλος περιορισμός ο οποίος ικανοποιείται από την κανονικοποίηση του  $Y_1$ , δηλαδή:

$$\mathbf{a}_1^T \mathbf{a}_1 = 1 \quad (36)$$

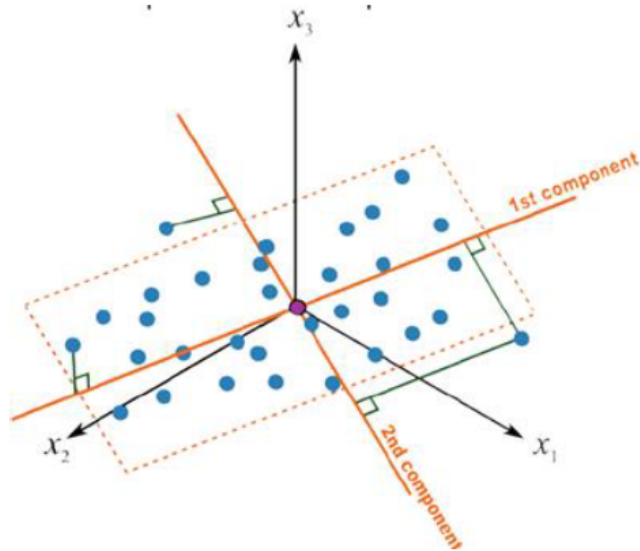
- Το δεύτερο principal component είναι ο γραμμικός συνδυασμός  $Y_2$  με την δεύτερη μεγαλύτερη δειγματική διακύμανση και του οποίου οι συντελεστές  $\mathbf{a}_2^T = [a_{21}, a_{22}, \dots, a_{2p}]$  θα πρέπει να υπακούν τους παρακάτω περιορισμούς:

$$\left\{ \begin{array}{l} \mathbf{a}_2^T \mathbf{a}_2 = 1 \\ \mathbf{a}_2^T \mathbf{a}_1 = 0 \end{array} \right\} \quad (37)$$

Ο δεύτερος περιορισμός, ο οποίος δηλώνει την ορθογωνιότητα των διανυσμάτων  $\mathbf{a}_1$  και  $\mathbf{a}_2$ , διαβεβαιώνει πως τα principal components είναι κάθετα μεταξύ τους.

- Η διαδικασία αυτή μπορεί να συνεχιστεί μέχρι να κατασκευαστούν  $p$  principal components, όσα δηλαδή και οι αρχικές μεταβλητές, τα οποία σχηματίζουν μια νέα ορθοκανονική βάση στον χώρο των  $p$  μεταβλητών.
- Αφού προσδιοριστούν οι νέοι άξονες των principal components, τότε κάθε παρατήρηση λαμβάνει μια νέα συντεταγμένη σε κάθε άξονα, η οποία ονομάζεται score της παρατήρησης στο αντίστοιχο principal component. Καθώς όμως η διακύμανση των δεδομένων είναι πεπερασμένη, συνήθως τα πρώτα δύο ή τρία components είναι αρκετά για να τη συμπεριλάβουν με μικρές απώλειες. Συνεπώς δεν είναι αναγκαίο να υπολογιστούν και τα  $p$  components, αλλά μονάχα τα πρώτα δύο ή τρία. Αυτό παρουσιάζεται γραφικά στο Σχήμα 14 και θα προσδιοριστεί και στη συνέχεια με μαθηματική ακρίβεια.





Σχήμα 14: Γεωμετρική Αναπαράσταση αξόνων *principal components*  
 (<https://www.cheric.org/files/education/cyberlecture/d201401/d201401-401.pdf>).

### 3. Υπολογισμός των Principal Components με χρήση της μεθόδου SVD

- Καθώς το πρώτο principal component αντιστοιχεί στον γραμμικό συνδυασμό των αρχικών μεταβλητών, ο οποίος θα εκφράζει την μέγιστη δειγματική διακύμανση, θα πρέπει να ισχύει ότι:

$$\text{Var}(Y_1) = \text{Var}(\mathbf{a}_1 \mathbf{x}^T) = \max \quad (38)$$

όμως ισχύει ότι:

$$\text{Var}(Y_1) = \text{Var}(\mathbf{a}_1 \mathbf{x}^T) = \mathbf{a}_1 \mathbf{x}^T \mathbf{x} \mathbf{a}_1^T \quad (39)$$

Και επειδή το  $x$  αναφέρεται σε κάθε πιθανή τιμή που έλαβαν οι μεταβλητές στις  $n$  παρατηρήσεις, τότε το γινόμενο αντιστοιχεί στον πίνακα συνδιακύμανσης των αρχικών δεδομένων σύμφωνα με τη σχέση A-6, δηλαδή:

$$\mathbf{a}_1 \mathbf{x}^T \mathbf{x} \mathbf{a}_1^T = (n-1) \mathbf{a}_1^T \Sigma \mathbf{a}_1 \quad (40)$$

Επομένως σύμφωνα με τη σχέση 40 και τη σχέση 32, τα βάρη του πρώτου principal component θα πρέπει να αντιστοιχούν στη μέγιστη διακύμανση, δηλαδή:

$$\mathbf{a}_1 \mathbf{x}^T \mathbf{x} \mathbf{a}_1^T = \max(VX^T XV^T) = \max(\Sigma^T \Sigma) = \max(\Sigma_+^2) \quad (41)$$

όπου ο  $\Sigma_+^2$  περιέχει τα τετράγωνα των μη μηδενικών ιδιζουσών τιμών του πίνακα  $X$  των δεδομένων (SVD). Επομένως το ιδιοδιάνυσμα που αντιστοιχεί στην μέγιστη από τις ιδιοτιμές  $\{\lambda_1, \dots, \lambda_p\}$  του πίνακα συνδιακύμανσης των δεδομένων θα προσδιορίζει τα βάρη του πρώτου principal component. Αντίστοιχα, από τη δεύτερη μεγαλύτερη ιδιοτιμή προκύπτουν τα βάρη του δεύτερου principal component κ.ο.κ. Τα principal components που είναι απαραίτητα για την περιγραφή του μεγαλύτερου ποσοστού δειγματικής διακύμανσης προσδιορίζονται από τον αριθμό των μη μηδενικών ιδιζουσών τιμών του πίνακα  $X$ . Όσες από αυτές είναι περίπου μηδέν, απορρίπτονται επίσης με κατάλληλα κριτήρια [41].

- Το ποσοστό της εξηγούμενης διακύμανσης (explained variance ratio) από το  $i$ -οστό component, δίνεται από τη σχέση:

$$\frac{\lambda_i}{\sum_{s=1}^p \lambda_s} \quad (42)$$

όπου  $\lambda_i$  η ιδιοτιμή του πίνακα συνδιακύμανσης που αντιστοιχεί στο  $i$ -οστό component. Το ποσοστό της συσσωρευμένης διακύμανσης που εκφράζουν τα  $j$  πρώτα principal component (cumulative explained variance ratio) θα δίνεται από τη σχέση:

$$\frac{\sum_{i=1}^j \lambda_i}{\sum_{s=1}^p \lambda_s} \quad (43)$$

- Τα scores των principal components για κάθε παρατήρηση θα δίνονται από τη σχέση:

$$T = US \quad (44)$$

όπου οι στήλες του πίνακα  $U$  περιέχουν τα αριστερά ιδιάζοντα διανύσματα του πίνακα  $X$  των δεδομένων και ο πίνακας  $\Sigma$  τις ιδιάζουσες τιμές του  $X$ , δηλαδή τις τετραγωνικές ρίζες των ιδιοτιμών του πίνακα συνδιακύμανσης [42]. Στο Σχήμα 15, συνοψίζεται η μέθοδος PCA μέσα από την ανάλυση του τυποποιημένου πίνακα  $X$  των δεδομένων με την μέθοδο SVD.

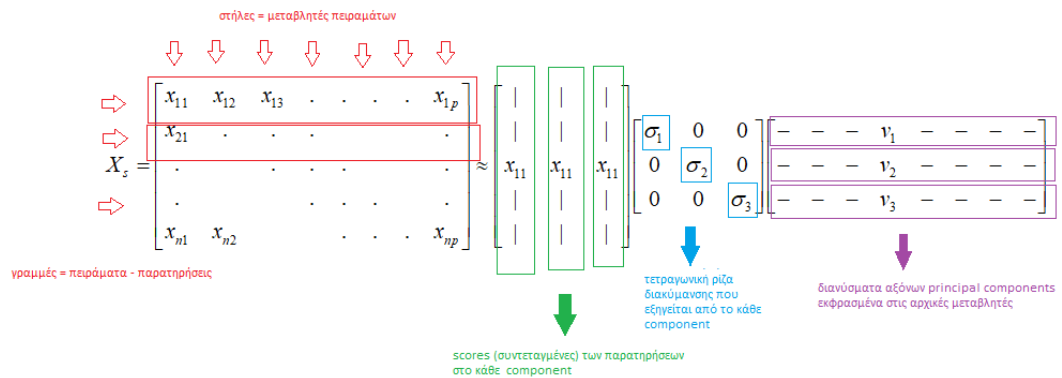
- Τα φορτία των principal components (loadings) θα δίνονται από τη σχέση:

$$r_{ij} = \frac{v_i \sqrt{\lambda_i}}{\sigma_{ii}^2} \quad (45)$$

όπου  $v_i$  το ιδιοδιάνυσμα του  $i$ -οστού component όπως δίνεται από τη σχέση 33,  $\lambda_i$  η ιδιοτιμή του πίνακα συνδιακύμανσης που αντιστοιχεί στο  $i$ -οστό component και  $\sigma_{ii}^2$  η διακύμανση της  $i$ -οστής μεταβλητής ( $i$ -οστό διαγώνιο στοιχείο του πίνακα συνδιακύμανσης - σχέση A-6). Το φορτίο ενός component εκφράζει τη συσχέτιση του με την κάθε παλαιά μεταβλητή των αρχικών δεδομένων.

- Το κριτήριο επιλογής του αριθμού των principal components βασίζεται στις ιδιοτιμές του πίνακα συνδιακύμανσης των δεδομένων [43]. Ειδικότερα, αξιοποιείται το κριτήριο του "αγκώνα" σύμφωνα με το οποίο διατηρούνται τα πρώτα components για τα οποία ο λόγος της σχέσης 42 συνεχίζει να μεταβάλλεται με γρήγορο ρυθμό. Η ονομασία "αγκώνας" προκύπτει καθώς εάν δημιουργηθεί ένα γράφημα της σχέσης 42 συναρτήσει του αριθμού των components, διατηρούνται τα εκείνα τα components πριν το γράφημα παρουσιάσει καμπή - "αγκώνα". Τα γραφήματα αυτά ονομάζονται κρημνογραφήματα και θα χρησιμοποιηθούν εκτενώς στην παρούσα εργασία.
- Όταν τα δεδομένα προβληθούν στους άξονες των principal components, όπως στο Σχήμα 14, δημιουργούνται τα score plots καθώς αντιστοιχούν στο γράφημα των τιμών ενός principal components συναρτήσει ενός άλλου component. Στα γραφήματα αυτά είναι δυνατό να παρατηρηθούν ομαδοποιήσεις των δεδομένων σε συγκεκριμένες κλάσεις, οι οποίες προκύπτουν από εγγενή μοτίβα των τιμών των μεταβλητών. Η δημιουργία αυτών των κλάσεων ανήκει στην Μη Επιβλεπόμενη (unsupervised) ταξινόμηση και δεν λαμβάνει κάποια επιπρόσθετη πληροφορία για τα δεδομένα. Η αναζήτηση αυτών των κλάσεων θα αποτελέσει κυρίαρχο στόχο της παρούσας διπλωματικής εργασίας.

Συνοψίζοντας την διαδικασία εύρεσης των principal components καθώς και των scores τους για κάθε παρατήρηση, μέσω της εφαρμογής της μεθόδου SVD στον πίνακα των δεδομένων  $X$ , παρατίθεται το Σχήμα 15. Η μέθοδος SVD χρησιμοποιείται στην PCA για την ανάλυση πινάκων οι οποίοι περιέχουν σημαντικά περισσότερες μεταβλητές συγκριτικά με τις παρατηρήσεις. Εάν ο πίνακας δεδομένων περιέχει σημαντικά λιγότερες μεταβλητές αναφορικά με τα δείγματα τότε οι μέθοδοι που χρησιμοποιούνται είναι διαφορετικές, όπως για παράδειγμα η μέθοδος πολλαπλασιαστών Lagrange με εφαρμογή της στη σχέση 38 συνδυαστικά με τους περιορισμούς της σχέσης 37 [40].



Σχήμα 15: Σχηματική αναπαράσταση εξαγωγής στοιχείων της PCA μέσω της ανάλυσης του πίνακα τυποποιημένων δεδομένων με SVD.

## 2.2 Ιεραρχική Ομαδοποίηση κατά Συστάδες

Οι αλγόριθμοι ιεραρχικής ομαδοποίησης και ειδικότερα η ιεραρχική ομαδοποίηση κατά συστάδες - Hierarchical Agglomerative Clustering (HAC) είναι ένα unsupervised εργαλείο οργάνωσης των δεδομένων σε ευρύτερες ομάδες, οι οποίες δίνουν την δυνατότητα εξερεύνησης μοτίβων που υπάρχουν στα δεδομένα σε σχέση με μια ποικιλία μεταβλητών. Στην περίπτωση της Ομαδοποίησης κατά Συστάδες, η οποία αποτελεί μια μέθοδο ταξινόμησης "από κάτω προς τα πάνω" - *Bottom-Up*, μια συλλογή μεμονωμένων δεδομένων διαχωρίζεται σε υποσύνολα με κριτήριο την μεταξύ τους απόσταση στον φυσικό χώρο των μεταβλητών τους. Τα δεδομένα αυτών των υποσυνόλων παρουσιάζουν την μέγιστη δυνατή ομοιογένεια, ενώ τα δεδομένα εκτός των υποσυνόλων θεωρούνται μεταξύ τους ανομοιογενή. Η ιεραρχική οργάνωση των υποσυνόλων-συστάδων αντιστοιχεί στην διαδοχική ομαδοποίηση τους, έτσι ώστε σε κάθε βήμα ταξινόμησης, οι συστάδες που ανήκουν στην ίδια ομάδα να είναι πιο όμοιες μεταξύ τους από αυτές που ανήκουν σε άλλες ομάδες.

Κάθε επίπεδο της ιεραρχίας αντιπροσωπεύει μια συγκεκριμένη ομαδοποίηση των δεδομένων σε ασύνδετες ομάδες παρατηρήσεων και επομένως ολόκληρη η ιεραρχία αντιπροσωπεύει μια διατεταγμένη σειρά τέτοιων ομαδοποιήσεων. Εξαρτάται από τον χρήστη του αλγορίθμου να αποφασίσει ποιο επίπεδο (εάν υπάρχει) αντιπροσωπεύει πραγματικά μια «φυσική» ομαδοποίηση με την έννοια ότι οι παρατηρήσεις σε κάθε μία από τις ομάδες της είναι αρκετά ομοιότερες μεταξύ τους σε σχέση με τις παρατηρήσεις που έχουν ανατεθεί σε διαφορετικές ομάδες σε αυτό το επίπεδο. Οι συγκεκριμένοι αλγόριθμοι, επομένως, ξεκινούν με την κάθε μεμονωμένη παρατήρηση ενός δείγματος με  $N$  παρατηρήσεις να αποτελεί μια μεμονωμένη ομάδα - *Singleton Cluster*. Σε καθένα από τα  $N-1$  βήματα ταξινόμησης, οι δύο ομοιότερες παρατηρήσεις συνδυάζονται σε μια νέα ομάδα - *Cluster*, παράγοντας έτσι μια λιγότερη ομάδα σε κάθε υψηλότερο επίπεδο[9]. Συνεπώς, πρέπει να οριστεί ένα μέτρο για τον προσδιορισμό της ομοιότητας μεταξύ των ομάδων. Κατάλληλη επιλογή μέτρου αποτελεί η γεωμετρική απόσταση μεταξύ των σημείων που αποτελούν τις ομάδες. Καθώς όμως περισσότερα από ένα στοιχεία υπάρχουν εντός των ομάδων, θα πρέπει επιπλέον να οριστεί η κατάλληλη γεωμετρική απόσταση που θα προσδιορίζει ολόκληρο το cluster και όχι μεμονωμένα σημεία.

Έστω ένα σύνολο δεδομένων  $X$  το οποίο περιέχει τις παρατηρήσεις  $x_i, i = 1, \dots, N$ . Ορίζεται η στοιχειώδης απόσταση μεταξύ δύο στοιχείων του  $X$  ως  $D(x_i, x_j)$ , η οποία συνήθως αντιστοιχεί στη ευκλείδεια απόσταση. Προκειμένου σε κάθε βήμα να επιλεγούν τα δύο ομοιότερα υποσύνολα, ορίζεται η απόσταση σύνδεσης - *Linkage Distance*,  $\Delta(X_i, X_j)$ , όπου  $X_i$  και  $X_j$  δύο τυχαία υποσύνολα που δημιουργήθηκαν σε ένα στάδιο της ταξινόμησης. Όταν δύο υποσύνολα αποτελούν μονοσύνολα ισχύει ότι  $\Delta(X_i, X_j) = D(x_i, x_j)$ . Οι συνήθεις τρόποι δημιουργίας αποστάσεων σύνδεσης είναι τρεις, οι τύποι των οποίων παρουσιάζονται παρακάτω[8] και η γεωμετρική τους απεικόνιση στο σχήμα:

- **Απλή Σύνδεση - Single Linkage**

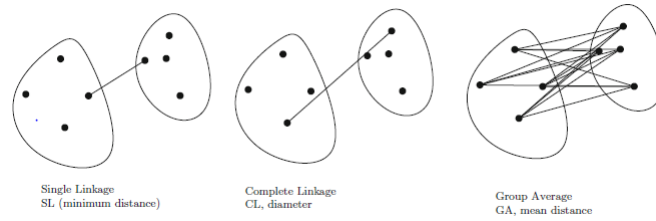
$$\Delta(X_i, X_j) = \min_{x_i \in X_i, x_j \in X_j} D(x_i, x_j) \quad (46)$$

- **Πλήρης Σύνδεση μέσω Διαμέτρου - Complete Linkage (Diameter)**

$$\Delta(X_i, X_j) = \max_{x_i \in X_i, x_j \in X_j} D(x_i, x_j) \quad (47)$$

- **Μέση Σύνδεση - Group Average Linkage**

$$\Delta(X_i, X_j) = \frac{1}{|X_i| |X_j|} \sum_{x_i \in X_i} \sum_{x_j \in X_j} D(x_i, x_j) \quad (48)$$

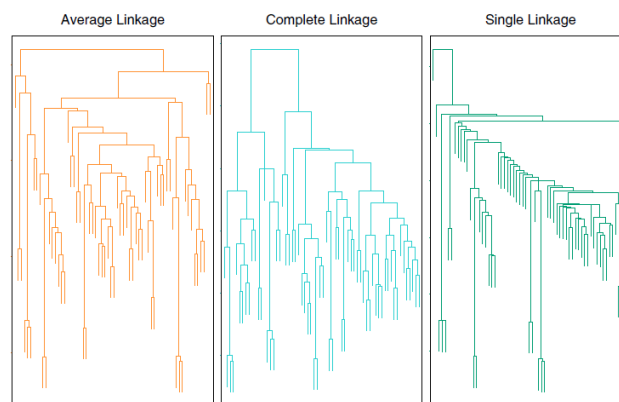


Σχήμα 16: Απεικόνιση συναρτήσεων σύνδεσης, οι οποίες προσδιορίζουν τους τρεις βασικούς τρόπους εύρεσης της απόστασης μεταξύ των υποσυνόλων: ελάχιστη απόσταση μεταξύ των σημείων των συστάδων (αριστερά), μέγιστη απόσταση μεταξύ των συστάδων - διάμετρος (κέντρο), μέση τιμή απόστασης μεταξύ όλων των σημείων των συστάδων (δεξιά)[8]

Ο κορμός του αλγορίθμου της Ομαδοποίησης κατά Συστάδες συμπυκνώνεται στα εξής βήματα:

1. Αρχικοποίηση για κάθε  $x_i \in X$  την αντίστοιχη ομάδα του -singleton-  $G_i = x_i$  σε μια λίστα.
2. Όσο παραμένουν τουλάχιστον δύο στοιχεία στην λίστα:
  - Επιλογή  $G_i$  και  $G_j$  έτσι ώστε η συνάρτηση  $\Delta(G_i, G_j)$  να ελαχιστοποιείται μεταξύ όλων των ζευγών.
  - Ενοποίηση των  $G_{i,j} = G_i \cup G_j$  και πρόσθεση του  $G_{i,j}$  στη λίστα με παράλληλη αφαίρεση των  $G_i$  και  $G_j$ .
3. Δημιουργία δενδρογράμματος, η σχηματική απεικόνιση της διαδικασίας ομαδοποίησης, με ρίζα το τελευταίο στοιχείο που απομένει στη λίστα  $G_{root} = X$ .

Η γραφική αναπαράσταση της ανωτέρω διαδικασίας ονομάζεται δενδρόγραμμα. Η σχεδίαση του ξεκινά με έναν εσωτερικό κόμβο  $s(X')$ , ο οποίος περιέχει το υποσύνολο  $X' \subseteq X$  σε ύψος  $h(X') = |X'|$  όπου το σύμβολο της απόλυτης τιμής δηλώνει τον πληθύνισμο του  $X'$ . Στη συνέχεια δημιουργούνται οριζόντιες ακμές από τον κόμβο  $s(X')$  προς τους δύο εγγύτερους κόμβους  $s(X_1)$  και  $s(X_2)$  με  $X' = X_1 \cup X_2$  και  $X_1 \cap X_2 = \emptyset$ . Ανάλογα με τον τύπο της σύνδεσης τα δενδρογράμματα μπορεί να διαφέρουν καθώς οι κόμβοι δημιουργούνται σε διαφορετικά σημεία. Χαρακτηριστικό παράδειγμα αποτελεί το σχήμα στο οποίο παρουσιάζεται ο διαφορετικός τρόπος δημιουργίας των συστάδων ανάλογα με την σύνδεσή τους.



Σχήμα 17: Γραφική απεικόνιση της ιεραρχικής ομαδοποίησης κατά συστάδες ανάλογα με τον τρόπο σύνδεσης των υποσυνόλων σε κάθε βήμα της ταξινόμησης[9]

## 2.3 Γραμμική Διακριτική Ανάλυση - LDA

Η Γραμμική Διακριτική Ανάλυση - *Linear Discriminant Analysis (LDA)* αποτελεί μια επιβλεπόμενη (supervised) μέθοδο ταξινόμησης των δεδομένων σε συγκεκριμένες κλάσεις. Οι κλάσεις προσδιορίζονται από τα δεδομένα εκπαίδευσης - *training set*, για τα οποία η πληροφορία της κλάσης που ανήκουν είναι εκ των προτέρων γνωστή. Για κάθε κλάση που δημιουργείται, ορίζονται τα στατιστικά μεγέθη των παρατηρήσεων της, όπως ο δειγματικός μέσος  $\mu$  και η διακύμανση  $\sigma$ . Με βάση αυτά τα μεγέθη μπορεί να οριστεί και κατάλληλη συνάρτηση πυκνότητας πιθανότητας για κάθε κλάση. Μια παρατήρηση μπορεί να ταξινομηθεί σε μία από τις κλάσεις του σετ εκπαίδευσης με χρήση του Θεωρήματος Ταξινόμησης του Bayes, όταν αυτό εφαρμοστεί στην προαναφερόμενη συνάρτηση πυκνότητας πιθανότητας. Γεωμετρικά, η LDA αναζητεί την εκείνη την κατεύθυνση επί της οποίας θα προβληθούν τα δεδομένα και θα εμφανίζουν τον αποτελεσματικότερο διαχωρισμό μεταξύ των κλάσεων τους. Η μαθηματική θεμελίωση των προαναφερόμενων εννοιών παρουσιάζεται στη συνέχεια αυτής της ενότητας.

### 2.3.1 Θεώρημα ταξινόμησης του Bayes

Έστω μια παρατήρηση με διάνυσμα μεταβλητών  $X=x$  η οποία πρέπει να ταξινομηθεί σε μια από τις  $K$  κλάσεις δεδομένων που υπάρχουν διαθέσιμες. Έστω επίσης η μεταβλητή  $Y$ , η οποία ονομάζεται μεταβλητή απόκρισης -*Response Variable*- και είναι αυτή που θα προσδιορίσει τον αριθμό της κλάσης που ανήκει η παρατήρηση με διάνυσμα  $X=x$ . Επιπλέον, τίθεται ως  $\pi_k$  η συνολική πιθανότητα -*Prior Probability*- μια τυχαία επιλεγμένη παρατήρηση από ένα δείγμα με  $n$  παρατηρήσεις να ανήκει στην κλάση  $k$ . Εκτός κι αν δίνεται σαν εξωτερική πληροφορία, η πιθανότητα αυτή δίνεται από την απλή σχέση:

$$\hat{\pi}_k = \frac{n_k}{n} \quad (49)$$

όπου  $n_k$  ο αριθμός των παρατηρήσεων του δείγματος που ανήκουν στην κλάση  $k$ . Στη συνέχεια ορίζεται η συνάρτηση πυκνότητας πιθανότητας για την κλάση  $k$ :

$$f_k(x) \equiv \Pr(X = x | Y = k) \quad (50)$$

η οποία προσδιορίζει την πιθανότητα -*Posterior Probability*- η μεταβλητή  $X$  να πάρει την τιμή  $X=x$  εάν είναι γνωστό ότι ανήκει στην κλάση  $k$ , δηλαδή εάν  $Y = k$ . Εάν η  $f$  είναι αεκετά υψηλή τότε υπάρχει σημαντική πιθανότητα μια παρατήρηση στην  $k$  κλάση να έχει  $X \approx x$

Το θεώρημα του Bayes [44] αντιστρέφει την προηγούμενη σχέση, δηλώνοντας πώς η πιθανότητα μια παρατήρηση να ανήκει στην κλάση  $k$  αν είναι γνωστό ότι  $X=x$ , δίνεται από τη σχέση:

$$p_k(X) = \Pr(Y = k | X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^K \pi_l f_l(x)} \quad (51)$$

Επομένως, για να ταξινομηθεί η παρατήρηση με  $X=x$  θα πρέπει να προσδιοριστεί η συνάρτηση πυκνότητας πιθανότητας για την κάθε κλάση.

### 2.3.2 LDA με χρήση μιας μεταβλητής πρόβλεψης

Έστω μια παρατήρηση η οποία προσδιορίζεται από μια μόνο μεταβλητή  $x$  για τη οποία γίνεται η παραδοχή ότι ακολουθεί γκαουσιανή κατανομή. Συνεπώς, η συνάρτηση πυκνότητας πιθανότητας για την κλάση  $k$  θα δίνεται από την σχέση [44]:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right) \quad (52)$$

όπου  $\mu_k$  και  $\sigma_k^2$  είναι ο πληθυσμιακός μέσος και η διασπορά της k-οστής κλάσης. Επιπλέον, θεωρείται πως όλες οι κλάσεις έχουν την ίδια διασπορά δηλαδή  $\sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ . Εφαρμόζοντας επομένως το θεώρημα του Bayes προκύπτει ότι:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_k)^2\right)}{\sum_{l=1}^K \pi_l \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu_l)^2\right)} \quad (53)$$

Επομένως, το θεώρημα ταξινόμησης του Bayes μπορεί ταξινομεί μια παρατήρηση με μεταβλητή πρόβλεψης  $X=x$  στην κλάση για την οποία η  $p_k(x)$  μεγιστοποιείται. Σύμφωνα με τον αλγόριθμο της γραμμικής διακριτικής ανάλυσης, τόσο ο μέσος της κάθε κλάσης όσο και η διασπορά της προκύπτει από ένα γνωστό σετ δεδομένων, το σετ εκπαίδευσης (training set), με  $n$  παρατηρήσεις που είναι ήδη ταξινομημένο στις κλάσεις και δίνεται από τις σχέσεις:

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i:y_i=k} x_i \quad (54)$$

$$\hat{\sigma}^2 = \frac{1}{n - K} \sum_{k=1}^K \sum_{i:y_i=k} (x_i - \hat{\mu}_k)^2 \quad (55)$$

όπου  $n_k$  ο αριθμός παρατηρήσεων της k-οστής τάξης και  $K$  ο αριθμός όλων των τάξεων. Συνεπώς, λογαριθμίζοντας την σχέση και αντικαθιστώντας της τιμές από τις σχέσεις, προκύπτει η Γραμμική Διακριτική Συνάρτηση - *Linear Discriminant Function* [44]:

$$\hat{\delta}_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k) \quad (56)$$

Επομένως, η παρατήρηση με  $X=x$  ταξινομείται στην κλάση όπου η γραμμική διακριτική συνάρτηση μεγιστοποιείται. Ο όρος "Γραμμική" προκύπτει καθώς η διακριτική συνάρτηση είναι γραμμική συνάρτηση της μεταβλητής  $x$ .

### 2.3.3 LDA σε πολυμεταβλητά δεδομένα

Ο αλγόριθμος της Γραμμικής Διακριτικής Ανάλυσης επεκτείνεται και σε πολυμεταβλητά δεδομένα, εάν τόσο η παρατήρηση προς ταξινόμηση όσο και οι παρατηρήσεις του δείγματος εκπαίδευσης περιγράφονται από ένα σύνολο  $p$  μεταβλητών με  $X = (X_1, X_2, \dots, X_p)$ , οι οποίες ακολουθούν η καθεμία μια μονοδιάστατη γκαουσιανή κατανομή και συνολικά μια πολυδιάστατη γκαουσιανή κατανομή. Η αντίστοιχη μαθηματική έκφραση είναι  $X \sim N(\mu, \Sigma)$  όπου  $\mu$  είναι ο μέσος του  $X$  και  $\Sigma = Cov(X)$  ο πίνακας συνδιακύμανσης. Συνήθως όμως, σε αυτήν περίπτωση οι μεταβλητές δεν είναι όλες ανεξάρτητες μεταξύ τους, αλλά εμφανίζονται συσχετίσεις και επίσης οι διασπορές τους δεν είναι ίσες. Αυτό μπορεί να δημιουργήσει διαταραχές στην καμπανοειδή μορφή της πολυδιάστατης γκαουσιανής έτσι η ταξινόμηση να μην είναι αξιόπιστη. Το πρόβλημα αυτό μπορεί να επιλυθεί εφαρμόζοντας τυποποίηση των δεδομένων, όπως περιγράφηκε στην προηγούμενη ενότητα, ώστε η διασπορά όλων των μεταβλητών να γίνει ίση με την μονάδα και στη συνέχεια να εφαρμοστεί PCA ώστε να χρησιμοποιηθούν εγγενείς μεταβλητές του δείγματος, οι οποίες όμως είναι ασυσχέτιστες μεταξύ τους.

Για να εφαρμοστεί και σε αυτή την περίπτωση το θεώρημα ταξινόμησης του Bayes θα πρέπει να οριστεί η πολυμεταβλητή κατανομή πυκνότητας πιθανότητας για τις μεταβλητές των δειγμάτων της κάθε κλάσης [44], η οποία θα είναι γκαουσιανή και δίνεται από τη σχέση:

$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)\right) \quad (57)$$

όπου  $\mu_k$  και  $\Sigma_k$  το διάνυσμα - μέσος για τις παρατηρήσεις της  $k$ -οστής κλάσης και ο αντίστοιχος πίνακας συνδιακύμανσης των μεταβλητών. Αντίστοιχα, η πολυμεταβλητή διακριτική συνάρτηση [44] θα δίνεται από τη σχέση:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k \quad (58)$$

όπου ο πίνακας συνδιακύμανσης  $\Sigma$  των δειγμάτων που είναι κατανεμημένα σε  $K$  κλάσεις δίνεται από τη σχέση:

$$\Sigma = \frac{1}{n - k} \sum_{j=1}^k \sum_{x \in C_j} (x - \mu_j) (x - \mu_j)^T \quad (59)$$

Τα όρια απόφασης του θεωρήματος Bayes [44], δηλαδή οι τιμές των μεταβλητών του διανύσματος  $X = (X_1, X_2, \dots, X_p)$  οι οποίες διαχωρίζουν τις κλάσεις μεταξύ τους προκύπτουν από την επίλυση της παρακάτω εξίσωσης για δύο διαφορετικές κλάσεις  $k$  και  $l$ :

$$\delta_k(x) = \delta_l(x) \implies x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k = x^T \Sigma^{-1} \mu_l - \frac{1}{2} \mu_l^T \Sigma^{-1} \mu_l \quad (60)$$

Η λύση της παραπάνω εξίσωσης προσδιορίζει ένα υπερεπίπεδο στον χώρο των μεταβλητών  $(X_1, X_2, \dots, X_p)$  με διάνυσμα:

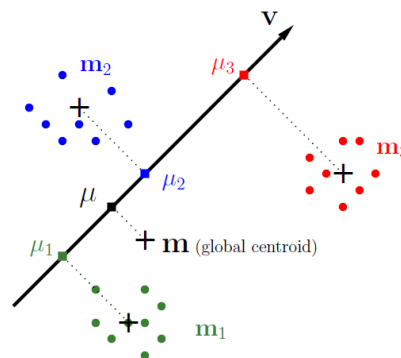
$$v = \Sigma^{-1} (\mu_k - \mu_l) \quad (61)$$

### 2.3.4 Γεωμετρική Ερμηνεία LDA

Το σετ δεδομένων εκπαίδευσης το οποίο προσδίδει τις παραμέτρους που χρειάζεται η πολυμεταβλητή κανονική κατανομή, αξιοποιείται για την γεωμετρική απεικόνιση των δεδομένων σε κλάσεις στον φυσικό χώρο των μεταβλητών τους. Σύμφωνα με το Σχήμα 18, αναζητείται η βέλτιστη κατεύθυνση  $v$  στην οποία αν προβληθούν τα δεδομένα, θα παρουσιάζεται ο καλύτερος δυνατός διαχωρισμός των κλάσεων τους. Αυτό επιτυγχάνεται σε εκείνη την κατεύθυνση  $v$  όπου:

1. η προβολή της κάθε κλάσης είναι όσο το δυνατόν πιο συμπτυγμένη
2. οι προβολές των κεντροειδών τους είναι όσο το δυνατόν πιο απομακρυσμένες η μία από την άλλη.

Ο συνδυασμός αυτών αναφέρεται στις διασπορές των δεδομένων επί της επιλεγόμενης κατεύθυνσης.



Σχήμα 18: Προβολές δεδομένων με μέσο  $m$  που ανήκουν σε τρεις κλάσεις με μέσους  $m_1$ ,  $m_2$  και  $m_3$  και προβολές μέσων  $\mu$ ,  $\mu_1$ ,  $\mu_2$  και  $\mu_3$  αντίστοιχα στην βέλτιστη κατεύθυνση [10].



Η πρώτη προϋπόθεση εκφράζεται μέσω της ελαχιστοποίησης της συνολικής προβαλλόμενης διασποράς όλων των δεδομένων εντός των κλάσεων τους. Η διασπορά για τα δεδομένα της κλάσης  $j$  στην κατεύθυνση  $\mathbf{v}$ , θα είναι [10]:

$$\begin{aligned}
s_j^2 &= \sum_{\mathbf{x}_i \in C_j} (a_i - \mu_j)^2 = \sum_{\mathbf{x}_i \in C_j} (\mathbf{v}^T \mathbf{x}_i - \mathbf{v}^T \mathbf{m}_j)^2 \\
&= \sum_{\mathbf{x}_i \in C_j} \mathbf{v}^T (\mathbf{x}_i - \mathbf{m}_j) (\mathbf{x}_i - \mathbf{m}_j)^T \mathbf{v} \\
&= \mathbf{v}^T \left[ \sum_{\mathbf{x}_i \in C_j} (\mathbf{x}_i - \mathbf{m}_j) (\mathbf{x}_i - \mathbf{m}_j)^T \right] \mathbf{v} \\
&= \mathbf{v}^T \mathbf{S}_j \mathbf{v},
\end{aligned} \tag{62}$$

όπου  $a_i$  οι συντεταγμένες των δεδομένων στην κατεύθυνση  $\mathbf{v}$ ,  $x_i$  οι συντεταγμένες δεδομένων στον φυσικό τους χώρο και  $m_j$  το κεντροειδές της κάθε κλάσης. Ο πίνακας:

$$\mathbf{S}_j = \sum_{\mathbf{x}_i \in C_j} (\mathbf{x}_i - \mathbf{m}_j) (\mathbf{x}_i - \mathbf{m}_j)^T \in R_{p \times p} \tag{63}$$

ονομάζεται *Πίνακας Διασποράς Εντός της Κλάσης  $j$*  [10]. Επομένως για  $K$  κλάσεις ισχύει ότι:

$$\sum_{j=1}^K s_j^2 = \sum \mathbf{v}^T \mathbf{S}_j \mathbf{v} = \mathbf{v}^T \left( \sum \mathbf{S}_j \right) \mathbf{v} = \mathbf{v}^T \mathbf{S}_w \mathbf{v} \tag{64}$$

όπου  $\mathbf{S}_w$  ο *Πίνακας Συνολικής Διασποράς Εντός Όλων των Κλάσεων*. Με βάση τον ορισμό του, ο συγκεκριμένος πίνακας είναι τετραγωνικός, συμμετρικός και θετικά ημιορισμένος.

Η δεύτερη προϋπόθεση εκφράζεται μέσω της μέγιστης δυνατής απόστασης των προβολών των κεντροειδών της κάθε κλάσης. Το τετράγωνο της απόστασης των προβαλλόμενων μέσων για δύο κλάσεις  $j$  και  $k$  θα είναι:

$$\begin{aligned}
(\mu_j - \mu_k)^2 &= (\mathbf{v}^T \mathbf{m}_j - \mathbf{v}^T \mathbf{m}_k)^2 = (\mathbf{v}^T (\mathbf{m}_j - \mathbf{m}_k))^2 \\
&= \mathbf{v}^T (\mathbf{m}_j - \mathbf{m}_k) \cdot (\mathbf{m}_j - \mathbf{m}_k)^T \mathbf{v} \\
&= \mathbf{v}^T \mathbf{S}_b \mathbf{v},
\end{aligned} \tag{65}$$

όπου ο πίνακας:

$$\mathbf{S}_b = (\mathbf{m}_1 - \mathbf{m}_2) (\mathbf{m}_1 - \mathbf{m}_2)^T \in R^{d \times d} \tag{66}$$

ονομάζεται *Πίνακας Διασποράς Μεταξύ των Κλάσεων* [10]. Για να υπολογιστεί η συνολική διασπορά μεταξύ όλων των κλάσεων θα πρέπει να χρησιμοποιηθεί ο σταθμισμένος μέσος  $\mu$ , καθώς η κάθε κλάση έχει διαφορετικό αριθμό παρατηρήσεων. Έστω ότι η κλάση  $j$  περιέχει  $n_j$  παρατηρήσεις από τις  $n$  του συνολικού δείγματος και άρα για  $K$  κλάσεις ισχύει ότι:

$$\mu = \frac{1}{n} \sum_{j=1}^K n_j \mu_j = \mathbf{v}^T \left( \frac{1}{n} \sum_{j=1}^K n_j \mathbf{m}_j \right) = \mathbf{v}^T \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \right) = \mathbf{v}^T \mathbf{m} \tag{67}$$

όπου  $m_j$  τα κεντροειδή των κλάσεων και  $\mathbf{m}$  το συνολικό κεντροειδές του δείγματος.

Άρα η συνολική σταθμισμένη διασπορά μεταξύ των κλάσεων θα είναι:

$$\begin{aligned}
\sum_{j=1}^c n_j (\mu_j - \mu)^2 &= \sum n_j (\mathbf{v}^T (\mathbf{m}_j - \mathbf{m}))^2 \\
&= \sum n_j \mathbf{v}^T (\mathbf{m}_j - \mathbf{m}) (\mathbf{m}_j - \mathbf{m})^T \mathbf{v} \\
&= \mathbf{v}^T \left( \sum n_j (\mathbf{m}_j - \mathbf{m}) (\mathbf{m}_j - \mathbf{m})^T \right) \mathbf{v} \\
&= \mathbf{v}^T \mathbf{S}_b \mathbf{v}
\end{aligned} \tag{68}$$

Συνεπώς, οι δύο προϋποθέσεις που αναφέρθηκαν παραπάνω συμπεκνώνονται στην επιλογή της κατεύθυνσης  $\mathbf{v}$ , στην οποία μεγιστοποιείται η απόσταση των προβλλόμενων κεντροειδών, ενώ παράλληλα ελαχιστοποιείται η συνολική διασπορά εντός των κλάσεων. Επομένως, συντελείται η μεγιστοποίηση του κλάσματος [10]:

$$\max_{\mathbf{v}: \|\mathbf{v}\|=1} \frac{\mathbf{v}^T \mathbf{S}_b \mathbf{v}}{\mathbf{v}^T \mathbf{S}_w \mathbf{v}} \leftarrow \frac{\sum n_j (\mu_j - \mu)^2}{\sum s_j^2} \tag{69}$$

Η λύση αυτού του προβλήματος, εάν υποθέσουμε ότι ο  $\mathbf{S}_w$  είναι μη-ιδιάζων, δίνεται από την εύρεση του ιδιοδιανύσματος  $\mathbf{v}$  με την υψηλότερη ιδιοτιμή, από το πρόβλημα ιδιοτιμών:

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{v}_1 = \lambda_1 \mathbf{v}_1 \tag{70}$$

Σε αρκετές περιπτώσεις όμως, ειδικά όταν ο πίνακας δεδομένων  $X \in \mathbb{R}_{n \times p}$  είναι υψηλών διαστάσεων, για τον  $\mathbf{S}_w \in \mathbb{R}_{p \times p}$  ισχύει ότι  $rank(X) = rank(\mathbf{S}_w) < p$ . Αυτό σημαίνει πως ο  $\mathbf{S}_w$  είναι ιδιάζων [10]. Για να επιλυθεί αυτό το ζήτημα υπάρχουν ποικίλες μορφές επεξεργασίας των δεδομένων. στην παρούσα εργασία όμως θα εφαρμοστεί PCA στον πίνακα δεδομένων και στη συνέχεια θα τροφοδοτηθεί ο πίνακας μειωμένων διαστάσεων στον αλγόριθμο της LDA [10].

### Παρατήρηση

Σύμφωνα με την γεωμετρική ερμηνεία της LDA, ο αλγόριθμός της λειτουργεί παρόμοια με τον αλγόριθμο της PCA καθώς μειώνει τις διαστάσεις των δεδομένων, προβάλλοντας σε μια συγκεκριμένη κατεύθυνση, η οποία στην περίπτωση δύο κλάσεων απλοποιείται στο διάνυσμα:

$$\mathbf{v} = \mathbf{S}_w^{-1} (\mathbf{m}_1 - \mathbf{m}_2) \tag{71}$$

Σύμφωνα με την εξίσωση, ο αλγόριθμος της LDA σύμφωνα με το θεώρημα ταξινόμησης του Bayes επιβάλλει ως όριο διαχωρισμού των κλάσεων το υπερεπίπεδο με διάνυσμα:

$$\mathbf{v} = \mathbf{\Sigma}^{-1} (\mu_k - \mu_l) \tag{72}$$

Παρατηρείται πως τα δύο διανύσματα είναι πανομοιότυπα [45] καθώς:

$$\begin{aligned}
\mathbf{S}_w &= \sum_{j=1}^2 \sum_{\mathbf{x} \in C_j} (\mathbf{x} - \mathbf{m}_j) (\mathbf{x} - \mathbf{m}_j)^T = (n - 2) \mathbf{\Sigma} \\
\mathbf{m}_1 &= \mu_j \\
\mathbf{m}_2 &= \mu_k
\end{aligned} \tag{73}$$

Συμπερασματικά, η LDA διαχωρίζει τις κλάσεις των δεδομένων με επίπεδα στα οποία, εάν προβληθούν τα δεδομένα, θα διαχωρίζονται οι κλάσεις με τον πιο αποτελεσματικό τρόπο. Για μια οποιαδήποτε νέα παρατήρηση, η μεγιστοποίηση της γραμμικής διακριτικής συνάρτησης θα υποδηλώνει τον υποχώρο - κλάση που ανήκει η παρατήρηση με την μεγαλύτερη δυνατή πιθανότητα.

## 2.4 Αποδοτικότητα Μοντέλου

Η χρήση μοντέλων ταξινόμησης δεδομένων σε κατάλληλες κλάσεις, όπως το μοντέλο της LDA, απαιτεί την αξιολόγηση της απόδοσης του αλγορίθμου που πραγματοποιεί την πρόβλεψη. Η αξιολόγηση αυτή πραγματοποιείται με βάση ένα σετ δεδομένων του οποίου οι κλάσεις των παρατηρήσεων είναι εκ των προτέρων γνωστές. Συνήθως, το σετ που πληρεί αυτή την προϋπόθεση είναι το σετ εκπαίδευσης του οποίου οι κλάσεις είναι γνωστές και συνεπώς είναι δυνατή η σύγκριση μεταξύ της πραγματικής και της προβλεπόμενης κλάσης. Επιπλέον, η εξαγωγή κατάλληλων δεικτών όπως η *Ακρίβεια-Accuracy*, η *Εναισθησία-Sensitivity* και η *Ανάκληση-Recall*, προσδίδουν τον κατάλληλο ποσοτικό χαρακτήρα στην αξιολόγηση της ταξινόμησης. Στη συνέχεια, θα παρουσιαστεί η μέθοδος αξιολόγησης Leave-One-Out Cross Validation (LOO) και θα αναλυθεί και ο τρόπος λειτουργίας του Confusion Matrix με βάση τον οποίο υπολογίζονται οι δείκτες της Ακρίβειας, της Εναισθησίας και της Ανάκλησης.

### Leave-One-Out Cross Validation - LOO

Ο έλεγχος απόδοσης του αλγορίθμου πραγματοποιείται συνήθως διαχωρίζοντας με τυχαίο τρόπο το σετ εκπαίδευσης, του οποίου οι κλάσεις είναι εκ των προτέρων γνωστές, σε δύο υποσύνολα, το training set (80%) και το test set (20%) [46]. Εναλλακτικά, σύμφωνα με τη μέθοδο *Leave-One-Out Cross Validation (LOO)* η οποία θα αξιοποιηθεί και στην παρούσα εργασία, το training set αποτελείται από ολόκληρο το σετ εκπαίδευσης, εκτός από μια παρατήρηση η οποία αντιστοιχεί στο test set [47]. Το training set χρησιμοποιείται για την δημιουργία του μοντέλου και το test set αξιοποιείται ως σετ πρόβλεψης. Η διαδικασία "Δημιουργία Μοντέλου μέσω training set - Πρόβλεψη μέσω test set" πραγματοποιείται συνολικά  $N$  φορές, όπου  $N$  το πλήθος των παρατηρήσεων του αρχικού σετ εκπαίδευσης, έτσι ώστε κάθε μια παρατήρηση να έχει αναχθεί ακριβώς μια φορά ως test set. Μετά τις  $N$  προβλέψεις, κάθε παρατήρηση του σετ εκπαίδευσης χαρακτηρίζεται από την πραγματική κλάση της (*Actual Class*) και από την προβλεπόμενη (*Predicted Class*), στοιχεία τα οποία απεικονίζονται γραφικά στον Confusion Matrix, ο οποίος θα περιγραφεί στη συνέχεια.

### Confusion Matrix

Ο Confusion Matrix δηλώνει τον αριθμό των σωστών και λανθασμένων προβλέψεων για κάθε κλάση. Ειδικότερα, κάθε γραμμή αντιστοιχεί στην προβλεπόμενη από το μοντέλο κλάση και κάθε στήλη στην πραγματική κλάση. Επομένως κάθε στοιχείο του πίνακα δηλώνει τον αριθμό των προβλέψεων που έκανε το μοντέλο για την κλάση της συγκεκριμένης γραμμής ενώ η πατήρηση ανήκει στην κλάση της συγκεκριμένης στήλης. Με αυτό τον τρόπο μπορεί να υπολογιστούν για κάθε κλάση τα παρακάτω στοιχεία [48]:

1. Αληθώς Θετικές προβλέψεις για κάθε κλάση (True Positive - TP), οι οποίες βρίσκονται στην διαγώνιο του Confusion Matrix
2. Αληθώς Αρνητικές προβλέψεις για κάθε κλάση (True Negative - TN), οι οποίες είναι το άθροισμα των προβλέψεων που ανήκουν εκτός της γραμμής και τις στήλης της συγκεκριμένης κλάσης.
3. Ψευδώς Θετικές προβλέψεις για κάθε κλάση (False Positive - FP), οι οποίες είναι το άθροισμα των προβλέψεων στην γραμμή της κλάσης, εκτός από το στοιχείο της διαγωνίου.
4. Ψευδώς Αρνητικές προβλέψεις για κάθε κλάση (False Negative - FN), οι οποίες είναι το άθροισμα των προβλέψεων στην στήλη της κλάσης, εκτός από το στοιχείο της διαγωνίου.

Με βάση αυτά τα στοιχεία, υπολογίζονται για κάθε κλάση τα παρακάτω στοιχεία, τα οποία όσο πιο υψηλά είναι τόσο πιο αποδοτικό είναι το μοντέλο που αξιολογείται [48]:

1. Ακρίβεια (Accuracy) πρόβλεψης της συγκεκριμένης κλάσης, η οποία εκφράζει το ποσοστό των σωστών προβλέψεων της κλάσης από το μοντέλο:

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FN} + \text{FP} + \text{TP}} = \frac{\text{TN} + \text{TP}}{N} \quad (74)$$

2. Ευαισθησία (Sensitivity) πρόβλεψης της συγκεκριμένης κλάσης, η οποία εκφράζει το ποσοστό των αληθώς θετικών προβλέψεων ως προς τον αριθμό των συνολικών θετικών προβλέψεων του μοντέλου για τη συγκεκριμένη κλάση:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (75)$$

3. Ανάκληση (Recall) πρόβλεψης της συγκεκριμένης κλάσης, η οποία εκφράζει το ποσοστό των αληθώς θετικών προβλέψεων ως προς τον πραγματικό αριθμό των παρατηρήσεων της κλάσης:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (76)$$

### 3 Κεφάλαιο 3: Μελέτη Βιοδιαβρωμένων Επιφανειών με την Τεχνική LIF

Η πορεία της διπλωματικής εργασίας μέχρι και το προηγούμενο κεφάλαιο επικεντρώθηκε στην θεμελίωση του βασικού θεωρητικού υποβάθρου, το οποίο είναι απαραίτητο για να αποκτήσει κανείς μια σφαιρική κατανόηση του αντικείμενου που εξετάζεται. Στο Κεφάλαιο 1 αναλύθηκε το φυσικό περιεχόμενο του φαινομένου του φθορισμού, στο οποίο στηρίζεται η μελέτη, καθώς επίσης και οι βασικοί παράγοντες διάβρωσης των επιφανειών των μνημείων πολιτιστικής κληρονομιάς. Στο Κεφάλαιο 2 περιγράφηκαν οι βασικές στατιστικές μέθοδοι ανάλυσης των δεδομένων που θα ανακτηθούν σε αυτή την εργασία, ενώ παράλληλα παρουσιάστηκε και ένα τρόπος αξιολόγησης των μοντέλων πρόβλεψης που θα χρησιμοποιηθούν. Στο παρόν κεφάλαιο θα περιγραφεί η φασματοσκοπική μέθοδος μελέτης των διαβρωμένων επιφανειών των μνημείων και θα καταγραφεί η βασική βιβλιογραφία που πλαισιώνει αντικείμενο μελέτης. Επιπλέον, θα παρουσιαστούν τα βασικά φάσματα φθορισμού που λήφθηκαν καθώς επίσης και τα βασικά φυσικά και βιολογικά χαρακτηριστικά τους.

Ειδικότερα, στην πρώτη ενότητα αναλύονται οι σημαντικότερες μελέτες που έχουν πραγματοποιηθεί έως τώρα αναφορικά με τον φθορισμό διαβρωμένων επιφανειών σε συνδυασμό με την στατιστική ανάλυση των δεδομένων, όπως επίσης και τα συμπεράσματα που έχουν προκύψει. Η δεύτερη ενότητα αφιερώνεται στην περιγραφή της μεθόδου που θα μελετηθούν οι διαβρωμένες επιφάνειες, δηλαδή στον *Επαγόμενο Φθορισμό με Laser - Laser Induced Fluorescence (LIF)*. Στη συνέχεια παρουσιάζεται η πειραματική μεθοδολογία και η πειραματική διάταξη που χρησιμοποιήθηκε για την μελέτη τριών βιολογικά επιβαρυσμένων επιφανειών. Επιπλέον, περιγράφεται η επέκταση της LIF στην χρήση συστημάτων Fluorescence LIDAR (Light Detection and Ranging) με στόχο την μελέτη φθορισμού των επιφανειών απευθείας επάνω στο μνημείο, ώστε να αποφευχθεί η μεταφορά θραύσματος στο εργαστήριο και να δοθεί η δυνατότητα χαρτογράφησης φθορισμού του μνημείου.

Η τρίτη ενότητα αφιερώνεται στην αναλυτική περιγραφή των φασμάτων που λήφθηκαν από τις τρεις βιοδιαβρωμένες επιφάνειες που μελετήθηκαν. Ειδικότερα, μετά από κατάλληλη βιολογική δειγματοληψία προσδιορίζονται και παρουσιάζονται τα βασικά είδη βιοδιαβρωτικών παραγόντων που επιβαρύνουν την κάθε επιφάνεια. Στη συνέχεια, λαμβάνονται φάσματα φθορισμού από συγκεκριμένες περιοχές των επιφανειών οι οποίες, μετά από οπτική παρατήρηση, συνδυάστηκαν με τους συγκεκριμένους βιολογικούς παράγοντες που προσδιορίστηκαν. Τα φάσματα παρουσιάζονται κανονικοποιημένα στην μέγιστη τιμή τους. Τελικά, τόσο με βάση την φασματική όσο και με την βιολογική πληροφορία, τα φάσματα των περιοχών ομαδοποιούνται σε κλάσεις που προσδιορίζουν έναν συγκεκριμένο τύπο βιοδιάβρωσης όπως επιβάρυνση από φωτοαυτότροφους μικροοργανισμούς, από συμβίωση ετερότροφων με φωτοαυτότροφους ή καθαρές περιοχές. Τα φάσματα αυτών των περιοχών καθώς και η πληροφορία της κλάσης που ανήκουν θα αποτελέσουν το σετ εκπαίδευσης του αλγορίθμου της LDA, η οποία με τη σειρά της θα δημιουργήσει ένα μοντέλο πρόβλεψης του τύπου βιοδιάβρωσης σε κάθε σημείο που πραγματοποιείται μέτρηση φθορισμού.

### 3.1 Βιβλιογραφική Ανασκόπηση

Η τεχνική *LIF - Laser Induced Fluorescence* και η επέκτασή της σε μετρήσεις μεγάλης κλίμακας μέσω της τεχνικής *Fluorescence LIDAR - Light Detection and Ranging*, χρησιμοποιείται ευρέως σε ποικίλες φασματοσκοπικές εφαρμογές για αρκετές δεκαετίες. Η ενδεδειγμένη ανάλυσή τους παρουσιάζεται στην ενότητα 3.2. Η δημιουργία του πρώτου Fluorescence LIDAR και η εκπόνηση συστηματικών πειραμάτων πραγματοποιήθηκε από τους Measures et al., 1971 [49]. Στα πειράματα αυτά μετρήθηκε ο φθορισμός ποικίλων υγρών καυσίμων τα οποία επικαθόνταν σε λεπτά υμένα επάνω σε νερό καθώς επίσης εξετάστηκε και η αξιοποίηση της διάταξης για την ανίχνευση χλωροφύλλης, τον φασματικό χαρακτηρισμό πετρωμάτων και την καταγραφή των εστιών μόλυνσης του θαλάσσιου περιβάλλοντος.

Η συστηματική αξιοποίηση του Fluorescence LIDAR στην ανίχνευση σημαντικών βιομορίων όπως η χλωροφύλλη και την φασματική μελέτη της βλάστησης και των ιδιοτήτων της πραγματοποιήθηκε από τους Edner et al., 1992 [50] και 1994 [51] και Valentini et al., 1994 [52]. Στις μελέτες αυτές εφαρμόστηκαν τρεις τρόποι λήψης σήματος φθορισμού, σημειακά, με φορητή συσκευή LIDAR και με τοποθέτηση του συστήματος LIDAR σε ειδικό όχημα από το οποίο εξέρχεται η δέσμη του laser προς τον υπό μελέτη στόχο. Επιπλέον, μελετήθηκε η συσχέτιση της έντασης του φθορισμού της χλωροφύλλης με την κατάσταση της υγείας των φυτικών κυττάρων από τα οποία προέρχεται.

Η έναρξη της μελέτης των μνημείων πολιτιστικής κληρονομιάς με χρήση LIF συντελείται με τα πειράματα των Cecchi et al., 1996 [53], με τα οποία μελετάται ο αυτοφθορισμός των πετρωμάτων που χρησιμοποιούνται ως υλικά δόμησης των μνημείων. Τα επόμενα χρόνια θα ακολουθήσουν και επιπρόσθετα πειράματα φθορισμού των διαφόρων χρησιμοποιούμενων λιθότυπων είτε σε εργαστηριακό περιβάλλον είτε επί του μνημείου. Ειδικότερα, οι Raimondi et al., 1998 [54] ανίχνευσαν τον φθορισμό του ασβεστόλιθου (limestone) και παρατήρησαν την μείωση της έντασης του όταν περιέχει κατιόντα  $Fe^{2+}$  (quenching). Επιπλέον συσχέτισαν τον λιθότυπο του μνημείου με την περιοχή από την οποία εξορύχτηκε. Στη συνέχεια, οι Pantani et al., 2003 [55] διαχώρισαν ποικίλους λιθότυπους όπως ασβεστόλιθο, αμμόλιθο, μάρμαρο και ασβεστοκονίαμα με βάση τα φάσματα φθορισμού τους αξιοποιώντας διάφορα μήκη κύματος διέγερσης. Τόσο στον διαχωρισμό των λιθότυπων όσο και παρατήρηση της μεταβολής τους φάσματος φθορισμού από σημείο σε σημείο για τον κάθε λιθότυπο αξιοποιήθηκε και η στατιστική μέθοδος PCA. Επίσης, οι Palombi et al., 2008 [56] ανίχνευσαν μέσω φθορισμού και εφαρμογής της PCA τις περιοχές όπου υφίσταται σκουριά από μεταλλικές στηρίξεις εντός της δομής του μνημείου. Χαρακτηριστικό της ύπαρξης σκουριάς αποτέλεσε η κορυφή φθορισμού στα 380 nm. Επιπλέον, μελέτησαν την ανομοιομορφή κατανομή των προσμίξεων σε ένα πέτρωμα μέσω της εφαρμογής της PCA σε φάσματα φθορισμού από διάφορα σημεία της επιφάνειάς του.

Οι ανάγκες για οπτική μελέτη της επιβάρυνσης των επιφανειών των μνημείων, χωρίς την χρήση σκαλωσιάς για την μελέτη σε μεγάλο ύψος, ικανοποιούνται μέσω των μετρήσεων φθορισμού με χρήση Fluorescence LIDAR. Ο διαχωρισμός του είδους της βιολογικής επιβάρυνσης της επιφάνειας με βάση τον φθορισμό έχει πραγματοποιηθεί για την διάκριση της επιβάρυνσης των φωτοαυτότροφων μικροοργανισμών, δηλαδή από κυανοβακτήρια ή από χλωροφύκη από τους Cecchi et al., 1996 [57], Lognioli et al., 1999 [58], Pantani et al., 2000 [59] και Lognioli et al., 2002 [60]. Και τα δύο είδη παρουσιάζουν την κορυφή της χλωροφύλλης περίπου στα 690 nm ως φωτοαυτότροφοι, αλλά η χαρακτηριστική κορυφή της φυκοκυανίνης στα 660 nm καθώς επίσης και κατά περιπτώσεις της φυκοερινίνης στα 560-612 nm [31] αποτελούν ενδείξεις ύπαρξης κυανοβακτηρίων. Η επιβάρυνση από ετερότροφους μικροοργανισμούς αποδεικνύεται δυσκολότερη και έως τώρα έχει πραγματοποιηθεί μόνο σε μονοκαλλιέργειες μυκήτων και βακτηρίων με χρήση της PCA στα φάσματα φθορισμού τους από τους Raimondi et al., 2007 [61] και 2008 [62].

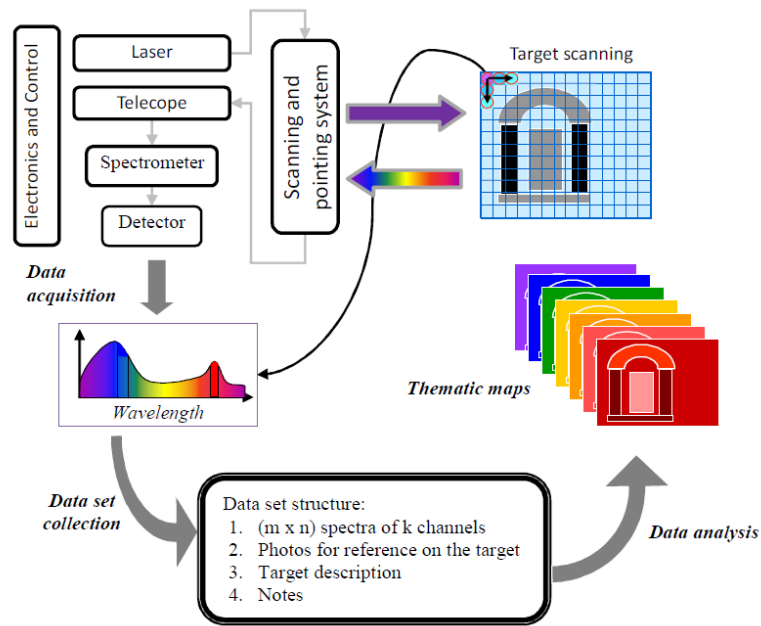
Η αντιμετώπιση της βιολογικής επιβάρυνσης πραγματοποιείται με χρήση κατάλληλων βιοκτόνων ανάλογα με το είδος του μικροοργανισμού που αναπτύσσεται στην επιφάνεια. Η αποτελεσματικότητα των θεραπειών μπορεί να ανιχνευθεί με χρήση του Fluorescence LIDAR σε συνδυασμό με την PCA, όπως πραγματοποίησαν οι Lognioli et al.,1999 [58], Pantani et al.,2000 [59] και Lognioli et al.,2003 [63]. λαμβάνοντας φάσματα πριν και μετά της εφαρμογής τους. Επίσης, μπορεί να καταγραφεί η χρονική εξέλιξη του φθορισμού της επιφάνειας ώστε να παρατηρηθεί ο χρόνος δράσης και η αποτελεσματικότητα των βιοκτόνων, όπως πραγματοποίησαν οι Lognioli et al.,2002 [60] με χρήση επίσης της PCA. Επιπλέον, είναι δυνατό να διαχωριστούν οι επιφάνειες με διαφορετικά είδη βιολογικών θεραπειών μέσω του φάσματος φθορισμού τους και την χρήση της PCA, μελέτη την οποία πραγματοποίησαν οι Bracci et al.,2001 [64].

Οι δυνατότητες συνεχίζουν να επεκτείνονται στην ανίχνευση χρωστικών σε τοιχογραφίες. Ειδικότερα, οι Raimondi et al.,2013 [65] συνδύασαν την χρήση του Fluorescence LIDAR με φασματοσκοπικές τεχνικές όπως FT-IR και φασματοσκοπία Raman και ανίχνευσαν την χρωστική κινναβάρη (cinnabar) και την κίτρινη ώχρα. Το κινναβάρη παρουσιάζει κορυφή στα 615 nm και ενώ η κίτρινη ώχρα φθορίζει περίπου στα 450 nm. Επιπλέον, οι Raimondi et al.,2014 [66] εφάρμοσαν την PCA σε φάσματα φθορισμού τοιχογραφιών με στόχο την εύρεση περιοχών όπου χρησιμοποιήθηκαν ίδιες χρωστικές.

Οι προαναφερόμενες εφαρμογές αναδεικνύουν το εύρος της πληροφορίας που δύναται να παρέχουν όταν εφαρμόζονται σε μεγάλο εύρος της επιφάνειας του μνημείου. Αυτό είναι εφικτό όταν το σύστημα του Fluorescence LIDAR συνδυάζεται με κατάλληλο σύστημα scanning της επιφάνειας και πραγματοποίηση χαρτογράφησης φθορισμού. Με αυτό τον τρόπο είναι εφικτή η δημιουργία κατάλληλων θεματικών χαρτών επάνω στην ψηφιακά φωτογραφημένη και χαρτογραφημένη επιφάνεια. Ειδικότερα, η υπό μελέτη επιφάνεια διαχωρίζεται σε υποπεριοχές όπου για την κάθε μια λαμβάνεται ένα φάσμα φθορισμού. Η χωρική διακριτική ικανότητα μπορεί να φτάσει μέχρι και το 1 cm [65]. Μετά το πέρας της διαδικασίας, η οποία παρουσιάζεται στο Σχήμα 19, προκύπτει μια "εικόνα" της οποίας το κάθε pixel-περιοχή χαρακτηρίζεται από ένα φάσμα φθορισμού.

Από αυτά τα φάσματα και την εφαρμογή στατιστικών μεθόδων όπως η PCA και η Ομαδοποίηση κατά Συστάδες - Cluster Analysis (στην διπλωματική εργασία ονομάζεται με τη συντομογραφία HAC), προκύπτουν ποικίλες πληροφορίες όπως η ταξινόμηση των περιοχών του μνημείου με βάση το εύρος και το είδος της βιολογικής επιβάρυνσης, το είδος του υποστρώματος και την αποτελεσματική ή μη εφαρμογή βιολογικών θεραπειών. Η εξέλιξη του συστήματος ονομάζεται Hyperspectral Fluorescence LIDAR και παρουσιάζεται αναλυτικά όπως επίσης και οι εφαρμογές του από τους Weibring et al., 2001 [67], Weibring et al., 2003 [68], Hällstrom et al., 2008 [69], Hällstrom et al. [70], 2009, Palombi et al.,2008 [56], Raimondi et al.,2009 [12], Raimondi et al.,2013 [65], και Raimondi et al.,2008 [62]. Επιπλέον, η δυνατότητες του συστήματος επεκτείνονται καθώς δύναται να καταγραφεί η χρονική εξέλιξη του φθορισμού (Time Resolved System) με χρονική διακριτική ικανότητα έως και 40 ps [66]. Η εύρεση του χρόνου ζωής φθορισμού για κάθε διαφορετικό φάσμα παρέχει σημαντικές πληροφορίες για την ταξινόμηση των περιοχών του μνημείου όπως για παράδειγμα την κατανομή των προσμίξεων στις πέτρινες επιφάνειες που μελετώνται. Το συγκεκριμένο σύστημα καθώς και οι αντίστοιχες εφαρμογές εξετάζονται αναλυτικά από τους Palombi et al.,2013 [71] και Raimondi et al.,2014 [66].

Στην παρούσα διπλωματική εργασία λήφθηκαν φάσματα φθορισμού από τα υπό μελέτη δείγματα, τα οποία θεωρήθηκαν ως γνωστά φάσματα αναφοράς. Τα φάσματα αυτά ταξινομήθηκαν σε κατάλληλες



Σχήμα 19: Οπτική περιγραφή συστήματος *Hyperspectral Fluorescence LIDAR* [11]

κλάσεις με βάση την βιολογική τους πληροφορία, όπως θα περιγραφεί στην ενότητα 3.3. Στη συνέχεια, πραγματοποιήθηκε χαρτογράφηση του δείγματος με στόχο την ταξινόμηση των φασμάτων χαρτογράφησης στις κλάσεις αναφοράς. Η ταξινόμηση αυτή συντελέστηκε με χρήση κατάλληλων μοντέλων πρόβλεψης κλάσεων όπως το μοντέλο της LDA (ενότητα 2.3). Η μεθοδολογία που εφαρμόστηκε προέκυψε από την μελέτη δημοσιεύσεων που αφορούν την ταξινόμηση διαφόρων τύπων τσαγιού με βάση το φάσμα φθορισμού τους [46, 47].



### 3.2 Περιγραφή Πειράματος και Πειραματική Διάταξη

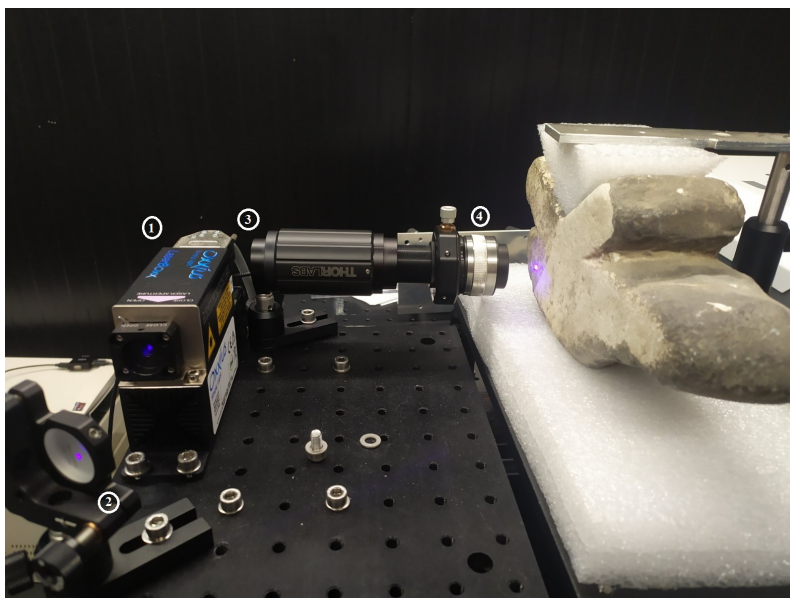
Η φασματική μελέτη των βιολογικά επιβαρυσμένων επιφανειών θα πραγματοποιηθεί με χρήση της μεθόδου *Επαγόμενου Φθορισμού με Laser - Laser Induced Fluorescence (LIF)*. Όπως εξηγήθηκε στην υποενότητα 1.1.5, η ακτινοβολία φθορισμού αποτελεί την φασματική υπογραφή των χρωμοφόρων μορίων που υφίστανται στο δείγμα το οποίο διεγείρεται. Συνεπώς, κάθε διαφορετικό δείγμα εάν ακτινοβοληθεί με laser κατάλληλου μήκους κύματος, αναμένεται να εκπέμψει ένα μοναδικό φάσμα που θα το ξεχωρίζει σε σχέση με τα υπόλοιπα. Ωστόσο, τα χρωμοφόρα μόρια που διεγείρονται, συνήθως είναι στην πλειοψηφία τους κοινά όταν ανήκουν στην ίδια ομάδα δειγμάτων. Συνακόλουθα, αυτό υποδηλώνεται και με αντίστοιχη ομοιομορφία στα φάσματα φθορισμού της συγκεκριμένης ομάδας. Επομένως, τα φάσματα τα οποία παρουσιάζουν μεταξύ τους τις μεγαλύτερες ομοιότητες, δύναται να ταξινομηθούν στην ίδια κατηγορία και ως εκ τούτου να προκύψει μια ομαδοποίηση των δειγμάτων βασισμένη στην ακτινοβολία φθορισμού. Η μέθοδος LIF χρησιμοποιείται ευρέως για πάνω από μισό αιώνα σε πληθώρα εφαρμογών επάνω σε τομείς όπως η χημεία τροφίμων [72], οι ιατρικές διαγνώσεις [73], η μικροσκοπία φθορισμού [4] και η τηλεπισκόπηση [74].

Τα μνημεία πολιτιστικής κληρονομιάς, όπως αναλύθηκε διεξοδικά στην ενότητα 1.2, καταστρέφονται επιφανειακά εξαιτίας της βιολογικής επιβάρυνσης που δέχονται από ποικίλους μικροοργανισμούς. Η μελέτη των επιφανειών των μνημείων με χρήση της μεθόδου LIF μπορεί να αναδείξει τις περιοχές έντονης βιολογικής επιβάρυνσης καθώς επίσης και το είδος των μικροοργανισμών που επάγουν την βιολογική καταστροφή της επιφάνειας. Με βάση την ομοιομορφία των φασμάτων υπάρχει η δυνατότητα ομαδοποίησης των περιοχών ανάλογα με τον τύπο της επιβάρυνσης, όπως παραδείγματος χάριν επιβάρυνση από φωτοαυτότροφους μικροοργανισμούς ή από συμβίωση φωτοαυτότροφων με ετερότροφους. Η λήψη των φασμάτων δύναται να πραγματοποιηθεί είτε εργαστηριακά, είτε στο σημείο του μνημείου με χρήση της τεχνικής Fluorescence LiDAR σταθερού στόχου. Στην παρούσα εργασία οι μετρήσεις πραγματοποιήθηκαν εργαστηριακά, στις εγκαταστάσεις της εταιρείας *Raymetrics S.A.* και η πειραματική διάταξη θα παρουσιαστεί ενδελεχώς στη συνέχεια. Επίσης, θα παρουσιαστεί η βασική διάταξη και αρχή λειτουργίας της τεχνικής Fluorescence LiDAR καθώς αποτελεί τον βασικό τρόπο μελέτης φθορισμού των μνημείων, όπως παρουσιάζεται βιβλιογραφικά. Στην επόμενη ενότητα θα περιγραφούν αναλυτικά οι βασικές μελέτες που έχουν πραγματοποιηθεί τόσο με εργαστηριακή εφαρμογή της LIF όσο και με την τεχνική Fluorescence LiDAR.

Η πειραματική διάταξη η οποία χρησιμοποιήθηκε στην διπλωματική εργασία σχεδιάστηκε και υλοποιήθηκε από την εταιρεία *Raymetrics S.A.* και παρουσιάζεται στην Εικόνα 20. Στη συνέχεια θα αναλυθούν οι βασικές αρχές λειτουργίας της πειραματικής διάταξης. Αρχικά, το δείγμα μελέτης στηρίζεται σε κάθετη θέση και διεγείρεται μέσω της ακτινοβολίας του διοδικού laser συνεχούς λειτουργίας (CW) - σημείο (1) της φωτογραφίας. Στο laser υπάρχει η δυνατότητα κατάλληλης ρύθμισης της ισχύος εξόδου στο εύρος 0-300 mW. Για τα πειράματα που πραγματοποιήθηκαν, η ισχύς εξόδου σταθεροποιήθηκε στα 25 mW, τα οποία αφενός δεν καταστρέφουν την βιολογική επικάλυψη και αφετέρου την διεγείρουν αποτελεσματικά ώστε να παράγεται ικανοποιητικό σήμα φθορισμού. Η δέσμη του laser, έπειτα από ανάκλαση σε κάτοπτρο (σημείο (2) της φωτογραφίας), προσπίπτει υπό γωνία περίπου 45° στο υπό μελέτη σημείο.

Τόσο το σήμα φθορισμού όσο και η ελαστική σκέδαση του laser συλλέγονται από ένα σύστημα φακού-οπτικής ίνας (σημείο (4) της φωτογραφίας). Για την μεγιστοποίηση του σήματος σε κάθε μέτρηση, υπάρχει δυνατότητα αλλαγής της απόστασης μεταξύ του συστήματος φακού - οπτικής ίνας και δείγματος. Η ρύθμιση αυτή είναι απαραίτητη για την ευθυγράμμιση του συστήματος σε κάθε μέτρηση καθώς η επιφάνεια είναι ανομοιομορφη και οι σχεδιάσεις σε ποικίλες κατευθύνσεις είναι αναπόφευκτες.

Επιπλέον, στο σύστημα συλλογής τοποθετείται κατάλληλο φίλτρο αποκοπής υψηλών συχνοτήτων (longpass filter) στα 407 nm ώστε να αποκοπεί η ελαστική σκέδαση του laser. Ωστόσο, το συγκεκριμένο laser το οποίο χρησιμοποιείται, παράγει και επιπρόσθετες κορυφές σε υψηλότερα μήκη κύματος σε σχέση με την κεντρική γραμμή εκπομπής του. Επομένως, τμήμα του φάσματος του laser διαπερνά το οπτικό φίλτρο και καταγράφεται από τον φασματογράφο. Ειδικότερα, η ονομαστική τιμή κορυφής του laser είναι στα 405 nm, αλλά πειραματικά ανιχνεύεται η πραγματική του κορυφή στα 402 nm. Οι υπόλοιπες κορυφές, των οποίων όμως η ένταση είναι μικρότερη του 1/1000 της κεντρικής γραμμής, εμφανίζονται στο εύρος 405-420 nm. Κατόπιν της τοποθέτησης του φίλτρου αποκοπής, η μοναδική κορυφή εμφανίζεται στα 412 nm, η οποία όμως είναι συγκρίσιμη σε ένταση με το φάσμα φθορισμού και εμφανίζεται σε μικρότερα μήκη κύματος από αυτό. Το σύνολο των οπτικών στοιχείων που περιγράφηκαν βρίσκεται τοποθετημένο σε πλατφόρμα με ρυθμιζόμενο ύψος, το οποίο έχει την δυνατότητα παράλληλης μετακίνησης επάνω σε ράγα.



Σχήμα 20: Σημειακή Πειραματική Διάταξη LIF

Το σήμα που συλλέγεται, μεταφέρεται από οπτική ίνα (σημείο (3) της φωτογραφίας) στον φασματογράφο. Με χρήση κατάλληλου φράγματος περίθλασης, ο φασματογράφος αναλύει το λαμβανόμενο σήμα στις φασματικές του συνιστώσες, οι οποίες καταγράφονται από έναν ανιχνευτή CCD. Παρακάτω, δίνονται τα βασικά χαρακτηριστικά των μερών της διάταξης:

1. Φασματογράφος με εστιακή απόσταση 37.5 mm,  $NA=0.22$ , πλάτος σχισμής εισόδου 200  $\mu\text{m}$  και διακριτική ικανότητα 6.8 nm (FWHM).
2. Διοδικό laser με ονομαστική κεντρική γραμμή εκπομπής στα 405 nm, εύρος ισχύος εξόδου 0-300mW και διάμετρος δέσμης στην έξοδο του laser 0.72 mm.
3. Αχρωματικός φακός συλλογής, εστιακής απόστασης 10 mm και διάμετρο 8 mm.

4. Οπτική Ίνα VIS/NIR με διάμετρο πυρήνα 200  $\mu\text{m}$  και  $\text{NA} = 0.22$ .
5. Οπτικό φίλτρο αποκοπής υψηλών συχνοτήτων με κατώφλι στα 407 nm και 93% ποσοστό διαπερατότητας στα 412.3 nm.
6. Κάτοπτρο με 99.5% ανακλαστικότητα στα 405 nm.

Η βασική πειραματική μεθοδολογία που ακολουθήθηκε περιλαμβάνει την δημιουργία δύο συνόλων φασμάτων για κάθε επιφάνεια που θα μελετηθεί. Το πρώτο σύνολο περιλαμβάνει τα φάσματα περιοχών αναφοράς, δηλαδή τα φάσματα των περιοχών όπου οπτικά ταυτίζονται με συγκεκριμένη βιολογική πληροφορία, όπως θα περιγραφεί στην ενότητα 3.3. Το δεύτερο σύνολο περιλαμβάνει τα φάσματα χαρτογράφησης της επιφάνειας. Στόχος της μελέτης αποτελεί η ταύτιση του κάθε σημείου χαρτογράφησης με μια ομάδα φασμάτων από τις περιοχές αναφοράς με χρήση στατιστικών μεθόδων ανάλυσης, όπως θα περιγραφεί στο κεφάλαιο 4. Η λήψη των φασμάτων πραγματοποιήθηκε με το λογισμικό του φασματογράφου και περιλαμβάνει τα εξής βήματα:

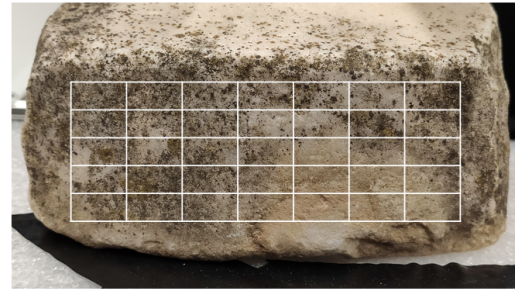
1. Στήριξη του δείγματος σε κάθετη θέση μπροστά από το σύστημα ανίχνευσης .
2. Ακτινοβολήση της περιοχής μέτρησης με τη δέσμη του διοδικού laser σε ισχύ 25 mW.
3. Όταν πραγματοποιηθεί η ευθυγράμμιση με στόχο την μεγιστοποίηση του σήματος, ορίζεται μέσω του λογισμικού ο χρόνος καταγραφής του σήματος στην κάθε μέτρηση (integration time). Εάν το σήμα είναι ασθενές, ο χρόνος καταγραφής αυξάνεται ώστε να συλλεχθούν περισσότερα φωτόνια φθορισμού ενώ αν το σήμα είναι ισχυρό, ο αντίστοιχος χρόνος μειώνεται ώστε να αποφευχθούν φαινόμενα κορεσμού του ανιχνευτή.
4. Κάθε φάσμα που λαμβάνεται προκύπτει ως ο μέσος όρος 5 μετρήσεων. Επίσης, για κάθε νέο χρόνο καταγραφής λαμβάνεται μέτρηση υποβάθρου (dark spectrum) απουσία του laser, η οποία αφαιρείται αυτόματα από την μέτρηση του φθορισμού σε κάθε λήψη.
5. Τα φάσματα αναφοράς λαμβάνονται από συγκεκριμένες περιοχές οι οποίες αντιπροσωπεύουν την πλειοψηφία των τύπων βιολογικής επιβάρυνσης του κάθε δείγματος. Οι περιοχές αυτές καθώς και η περιγραφή τους παρουσιάζονται αναλυτικά στην ενότητα 3.3. Τα φάσματα χαρτογράφησης κάθε δείγματος λαμβάνονται αφότου οριστεί ένα σημείο εκκίνησης και προσδιοριστεί το βήμα της μέτρησης κατά μήκος και καθ' ύψος του δείγματος. Ειδικότερα, όπως παρουσιάζεται στην εικόνα 21, τα φάσματα χαρτογράφησης λήφθηκαν στις κορυφές των ορθογωνίων του κάθε πλαισίου. Το βήμα της μέτρησης αλλάζει σε κάθε δείγμα καθώς και το ύψος που αλλάζει η σειρά των μετρήσεων. Για κάθε δείγμα τα βήματα μέτρησης περιγράφονται επίσης στην εικόνα 21. Επίσης, τα βιολογικά και γεωλογικά χαρακτηριστικά του κάθε δείγματος παρουσιάζονται στην ενότητα 3.3.

### Εφαρμογή της LIF σε Fluorescence LiDAR

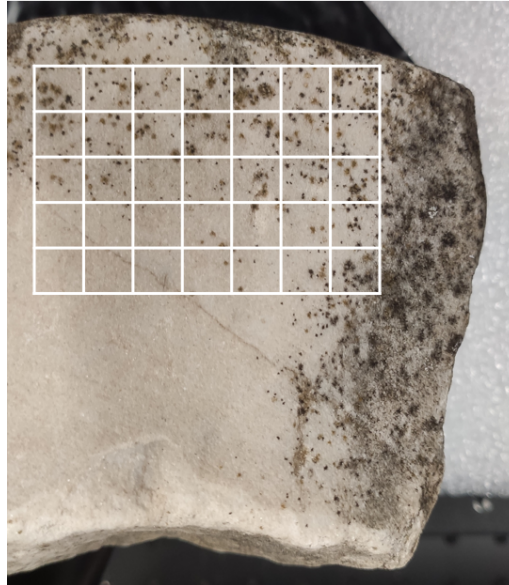
Η επέκταση της LIF σε μετρήσεις φθορισμού μεγαλύτερης κλίμακας σε εξωτερικό περιβάλλον μπορεί να πραγματοποιηθεί με τη χρήση της τεχνικής *Fluorescence LiDAR (Light Detection and Ranging)*. Ειδικότερα, η τεχνική αυτή αξιοποιείται όταν η απόσταση μεταξύ του υπό μελέτη δείγματος και του ανιχνευτή είναι μεγάλη καθώς επίσης και όταν το διάχυτο φως του περιβάλλοντος είναι ισχυρό και υπερκαλύπτει το σήμα φθορισμού [74]. Επιπλέον, η χρήση του Fluorescence LiDAR δεν εφαρμόζεται μόνο σε πειράματα που ο στόχος είναι σταθερός, όπως περιγράφηκε ως τώρα, αλλά επεκτείνεται αξιοποιώντας της δυνατότητα του ranging σε ατμοσφαιρικά [75] και υδάτινα πειράματα



(a) CALLOS CER-3:  $\Delta x=20\text{mm}$ ,  
 $\Delta y=10\text{mm}$



(b) CALLOS MAR-2 ΚΑΘΕΤΗ  
ΕΠΙΦΑΝΕΙΑ 1:  $\Delta x=20\text{mm}$ ,  
 $\Delta y=10\text{mm}$



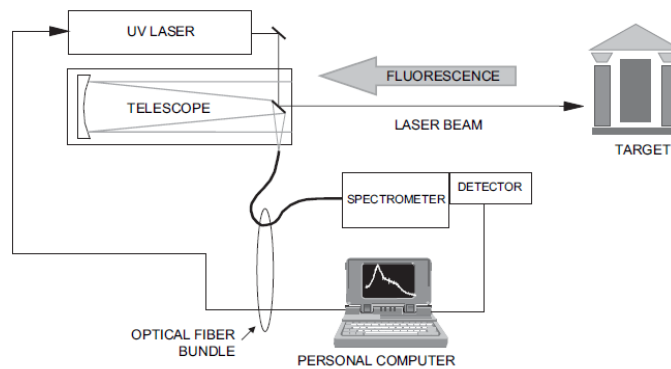
(c) CALLOS MAR-2 ΠΛΑΤΕΙΑ  
ΕΠΙΦΑΝΕΙΑ 2:  $\Delta x=5\text{mm}$ ,  $\Delta y=5\text{mm}$

Σχήμα 21: Σχηματικό χαρτογράφησης δειγμάτων. Οι λήψεις των φασμάτων πραγματοποιήθηκαν στις κορυφές των ορθογωνίων με κατάλληλο βήμα  $\Delta x$  κατά μήκος και  $\Delta y$  καθ' ύψος σε κάθε δείγμα

[76]. Το οπισθοσχεδαζόμενο σήμα, το οποίο περιέχει τόσο την ελαστική σκέδαση ενός παλμικού laser όσο και τον φθορισμό, καταγράφεται για μια σειρά σημείων κατά μήκος του ατμοσφαιρικού ή του υδάτινου στρώματος που μελετάται. Οι αποστάσεις στις οποίες λαμβάνεται το σήμα καθώς και τα αντίστοιχα φάσματα, διακρίνονται μεταξύ τους εάν είναι γνωστή η συχνότητα των παλμών του laser και ο χρόνος λήψης του σήματος σε σχέση με την εκπομπή του παλμού. Ωστόσο, στη συνέχεια θα παρουσιαστεί αναλυτικά η συγκεκριμένη τεχνική μόνο για πειράματα σταθερού στόχου, καθώς το ενδιαφέρον της παρούσας εργασίας εστιάζεται στις μετρήσεις φθορισμού από επιφάνειες μνημείων.

Σύμφωνα με την εικόνα 22, μια τυπική διάταξη Fluorescence LiDAR σταθερού στόχου [12] αποτελείται από κατάλληλο laser, συνήθως στην περιοχή του UV [74], ένα τηλεσκόπιο, ένα σύστημα ανάλυσης του φωτός στις φασματικές του συνιστώσες και έναν ανιχνευτή, ο οποίος συνδέεται με ένα σύστημα ηλεκτρονικού υπολογιστή. Η δέσμη του laser προσπίπτει στο σημείο μέτρησης επί της επιφάνειας του στόχου, όπου η ακτινοβολία αλληλεπιδρά με τα μόρια των επικαθίσεων καθώς και του υποστρώματος. Εάν η ακτινοβολία laser εμπίπτει στις ζώνες απορρόφησης των προαναφερόμενων μορίων, τότε εκπέμπεται φθορισμός ο οποίος περιέχει πληροφορίες για τα φυσικο-χημικά χαρακτηριστικά του στόχου.

Η ακτινοβολία που εκπέμπεται από τον στόχο περιλαμβάνει τόσο τον φθορισμό όσο και την ελαστική οπισθοσκέδαση του laser. Το συνολικό σήμα συλλέγεται από κατάλληλο τηλεσκόπιο, ώστε ακόμη και η αδύναμη ακτινοβολία του φθορισμού να είναι δυνατό να ανιχνευτεί. Σε περίπτωση που η ελαστική οπισθοσκέδαση δεν είναι απαραίτητη, υπάρχει η δυνατότητα τοποθέτησης κατάλληλου φίλτρου ώστε να αποκοπεί. Στη συνέχεια το σήμα καθοδηγείται με χρήση οπτικών ινών σε έναν φασματογράφο, ο οποίος το αναλύει φασματικά. Είτε εντός του φασματογράφου είτε εκτός, τοποθετείται κατάλληλο σύστημα ανίχνευσης όπως ανιχνευτής CCD ή σύστημα φωτοπολλαπλασιαστών (PMT) σε κατάλληλα εύρη μηκών κύματος ώστε να καταγραφεί το αναλυμένο σήμα φθορισμού. Η ανάκτηση των φασμάτων πραγματοποιείται από κατάλληλο λογισμικό με χρήση ηλεκτρονικού υπολογιστή.



Σχήμα 22: Σχηματικό Τυπικής Διάταξης Fluorescence LiDAR σταθερού στόχου [12]

### 3.3 Βιολογική και Φασματική Ανάλυση Δειγμάτων Μελέτης

Η ανάλυση των δειγμάτων που χρησιμοποιήθηκε στην παρούσα εργασία αποτελείται από δύο κύριες συνιστώσες. Αρχικά, πραγματοποιήθηκε βιολογική δειγματοληψία σε κάθε δείγμα ώστε να προσδιοριστούν οι βασικοί βιολογικοί παράγοντες που επιφέρουν την βιοδιάβρωση της επιφάνειας. Στη συνέχεια, λήφθηκαν τα φάσματα φθορισμού από χαρακτηριστικές περιοχές της επιφάνειας, οι οποίες προσδιορίστηκαν με οπτική παρατήρηση. Στόχος της διπλής ανάλυσης αποτελεί η προσεγγιστική συσχέτιση του ανακτιθέντος φάσματος με την βιολογική ταυτότητα του βιοδιαβρωτή. Καθώς όμως η βιολογική δειγματοληψία ήταν σε αρκετά λιγότερα σημεία συγκριτικά με τις περιοχές λήψης φασμάτων, η συσχέτιση πραγματοποιήθηκε ετεροχρονισμένα και προσεγγιστικά. Οι περιοχές που λαμβάνονται τα φάσματα θα αποτελέσουν το σύνολο δεδομένων που θα εκπαιδεύσει τον αλγόριθμο της Γραμμικής Διακριτικής Ανάλυσης.

Η βιολογική ανάλυση πραγματοποιήθηκε από την Δρ. Ιωάννα Πυρρή στον Τομέα Οικολογίας και Ταξινόμησης του τμήματος Βιολογίας στο Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών. Τα δείγματα μελετήθηκαν στις εγκαταστάσεις της εταιρείας Raymetrics S.A. στα πλαίσια του ευρωπαϊκού προγράμματος C.A.L.L.O.S - Conservation of Athens antiquities with Laser and Lidar technologies Open to Science and public - Ανοιχτό Εργαστήριο Συντήρησης των Μνημείων της Αθήνας με τεχνολογίες Λείζερ και LIDAR (κωδικός έργου Τ6ΥΒΠ-00099). Συνεπώς, η περιγραφή στηρίζεται στην αναφορά που συντάχθηκε για τον συγκεκριμένο σκοπό.

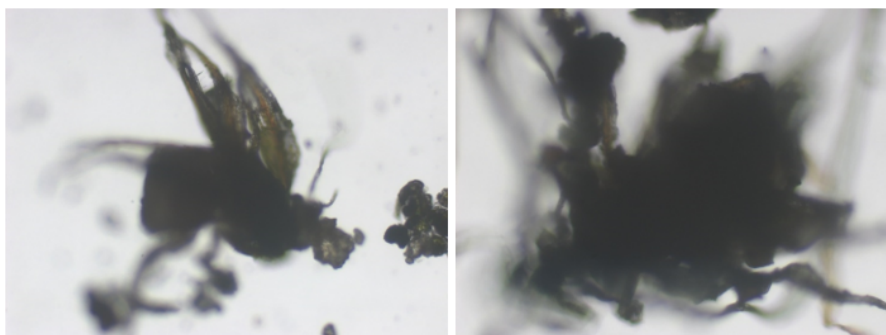
Αρχικά, η λήψη των μικροοργανισμών πραγματοποιήθηκε με χρήση κολλητικής ταινίας (adhesive tape) καθώς επίσης και με την χρήση βελόνας επί της επιφάνειας του δείγματος που μελετάται. Η καλλιέργεια των μικροοργανισμών αναπτύχθηκε σε κατάλληλο στερεό θρεπτικό υπόστρωμα και πιο συγκεκριμένα σε Water agar (WA), στο οποίο εμβολιάστηκαν τα δειγματοληπτούμενα στοιχεία. Μετά τον εμβολιασμό, τα δείγματα επώαστηκαν σε θάλαμο σταθερής θερμοκρασίας στους 25°C σε συνθήκες ημέρας-νύχτας.

Στα δείγματα με την κολλώδη ταινία προστέθηκε γλυκερόλη στο εργαστήριο και μελετήθηκαν με μικροσκόπιο Zeiss AxioImager A1 σε μεγεθύνσεις 100X, 200X και 400X. Κατά τη μικροσκοπική παρατήρηση των δειγμάτων που ελήφθησαν με κολλώδη ταινία έγινε σάρωση όλης της περιοχής του παρασκευάσματος και φωτογράφιση αντιπροσωπευτικών περιοχών. Τα τρυβλία φωτογραφήθηκαν μακροσκοπικά 2 εβδομάδες μετά τον εμβολιασμό, 1 μήνα και 2 μήνες μετά. Στα τρυβλία που αναπτύχθηκαν μικροοργανισμοί έγιναν παρασκευάσματα, παρατηρήθηκαν μικροσκοπικά και φωτογραφήθηκαν με μικρο-σκόπιο Zeiss AxioImager A1 σε μεγέθυνση 400X.

### 3.3.1 CALLOS CER-3

#### Βιολογική Περιγραφή Δείγματος

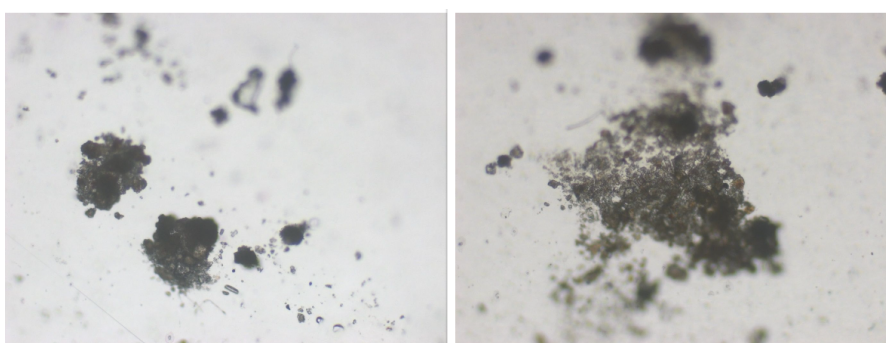
Το δείγμα αντιστοιχεί σε τμήμα κεράμου από πηλό καστανέρυθρο και προέρχεται από ανασκαφή στο μετρό Κεραμεικού (1995). Το μέγιστο σωζόμενο πλάτος είναι 8.3 cm, το αντίστοιχο μήκος είναι 28.6 cm και το αντίστοιχο πάχος είναι 5.5 cm. Ο κυρίαρχος παράγοντας βιοδιάβρωσης του συγκεκριμένου δείγματος είναι τα βρύα και ειδικότερα το βρύο γένους *Tortula*. Εξαιτίας της πορώδους επιφάνειας του κεραμικού υποστρώματος, η περιεκτικότητα υγρασίας εντός των πόρων [70] ευνοεί την ανάπτυξη βρύων, τα οποία επεκτείνονται σε μεγάλη έκταση επί του δείγματος. Η φωτογραφία του βρύου παρουσιάζεται στο Σχήμα 23.



Σχήμα 23: Βρύο γένους *Tortula*

Επιπροσθέτως στην επιφάνεια του δείγματος παρατηρήθηκαν χλωροφύκη σε κατάσταση αποσύνθεσης, με ενεργές περιοχές όμως κατά την περίοδο δειγματοληψίας. Επίσης, λόγω της ύπαρξης αερομεταφερόμενων μυκήτων, το δείγμα περιείχε σπόρια μυκήτων, ίχνη των οποίων παρουσιάζονται στο Σχήμα 24.

Η εικόνα του δείγματος όταν καταγράφηκαν οι μετρήσεις φθορισμού παρουσιάζεται στο Σχήμα 25. Στην αριστερή πλευρά του δείγματος, όπου η επιφάνεια είναι αισθητά πιο σκουρόχρωμη, εμφανίζονται τα χλωροφύκη σε αποσύνθεση ενώ στις καστανές και πράσινες ανάγλυφες περιοχές κυριαρχούν τα βρύα.



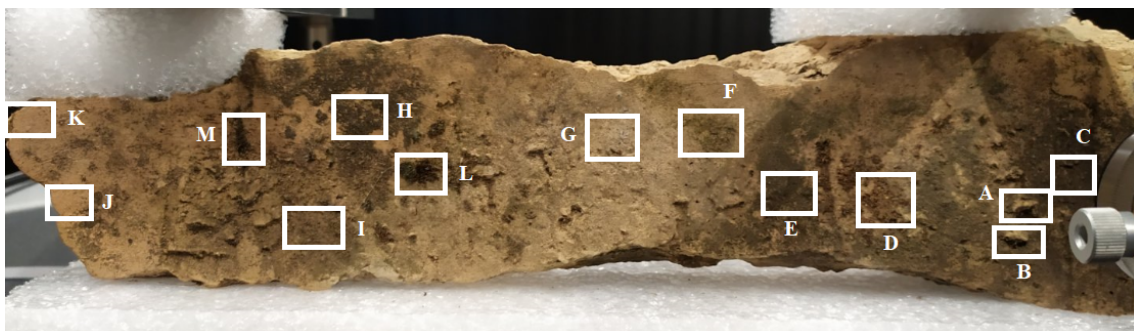
Σχήμα 24: Ίχνη σπορίων μυκήτων και πράσινα στίγματα που υποδεικνύουν την ύπαρξη χλωροφυκών



Σχήμα 25: *CALLOS CER-3*

### Φασματικές Περιοχές Αναφοράς

Στο Σχήμα 26 παρουσιάζονται οι περιοχές στις οποίες πραγματοποιήθηκε λήψη φάσματος φθορισμού. Οι περιοχές αυτές συσχετίζονται με την ύπαρξη των προαναφερόμενων βιολογικών παραγόντων βιοαποίκησης και στον Πίνακα 2 καταγράφεται η οπτική περιγραφή τους, η φασματική περιγραφή τους καθώς επίσης και η πιθανή συσχέτιση με βιολογικές χρωμοφόρες ουσίες.



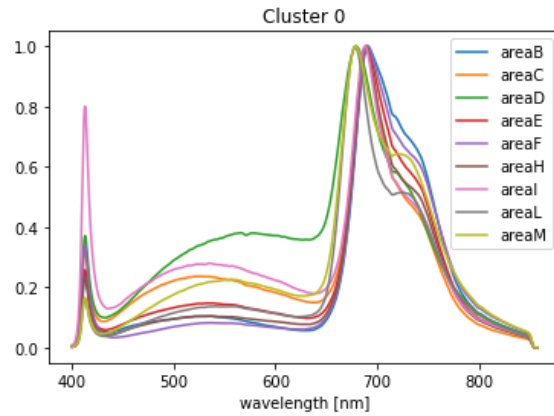
Σχήμα 26: Περιοχές αναφοράς φασμάτων φθορισμού δείγματος *CALLOS CER-3*

Τα φάσματα τα οποία προκύπτουν από τις παραπάνω περιοχές ομαδοποιούνται σε τρεις κατηγορίες ανάλογα με την κλίμακα και το είδος της βιολογικής επιβάρυνσης που προκαλούν. Ειδικότερα, η μηδενική ομάδα φασμάτων, το Cluster 0 στο Σχήμα 27, αντιστοιχεί σε περιοχές με έντονη βιολογική επιβάρυνση από φωτοαυτότροφους μικροοργανισμούς όπως βρύα και χλωροφύκη με έντονες κορυφές στο ευρος 680 - 730 nm, οι οποίες αντιστοιχούν στην χλωροφύλλη. Ο λόγος έντασης φθορισμού  $I_{690nm}/I_{735nm}$  αποτελεί χαρακτηριστικό της κατάστασης του φωτοαυτότροφου μικροοργανισμού καθώς και της περιεκτικότητας σε χλωροφύλλη[5]. Η πρώτη ομάδα φασμάτων, το Cluster 1 στο Σχήμα 28, που περιέχει όμως μόνο το φάσμα μιας περιοχής, αντιστοιχεί σε περιοχές με βιολογική επιβάρυνση φωτοαυτότροφων μικροοργανισμών αλλά και με πιθανή περιεκτικότητα σε ετερότροφους όπως μύκητες εξαιτίας του ευρέως φάσματος στην περιοχή 440-660 nm. Τέλος, η δεύτερη ομάδα φασμάτων, το Cluster 2 στο Σχήμα 29, περιλαμβάνει τα φάσματα περιοχών με μικρή βιολογική επιβάρυνση καθώς το σήμα της ελαστικής σκέδασης μιας υπολειπόμενης κορυφής του laser στα 412 nm είναι σημαντικά ισχυρότερο συγκριτικά με το σήμα φθορισμού.

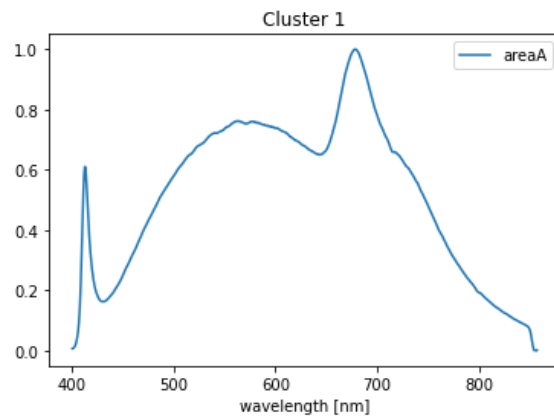


Περιοχή	Περιγραφή
A	Κοιλότητα με μαύρο χρώμα. Κορυφή χλωροφύλλης στα 678 nm, ευρύ και έντονο φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 718 nm, ώμος από τα 718 nm - 850 nm. Κορυφή του laser στα 412 nm
B	Κοιλότητα με μαύρο χρώμα. Κορυφή χλωροφύλλης στα 691 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 730 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
C	Ομοιόμορφα μαύρη περιοχή με πολύ μικρές εσοχές. Κορυφή χλωροφύλλης στα 689 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 735 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
D	Περιοχή με καστανά βρύα. Κορυφή χλωροφύλλης στα 679 nm, ευρύ μεγάλης έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 725 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
E	Πρασινό- μαύρη περιοχή, πορώδης. Κορυφή χλωροφύλλης στα 689 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 740 nm, ώμος από τα 750 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
F	Αχνή μαύρη περιοχή, εμφανίζεται έντονα το κεραμικό από πίσω. Κορυφή χλωροφύλλης στα 690 nm, ευρύ αλλά αρκετά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 730 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
G	Πιο καθαρή περιοχή, εμφανίζεται το κεραμικό και υπάρχουν διάσπαρτα μαυρα στίγματα. Κορυφή χλωροφύλλης στα 684 nm, ευρύ φάσμα στο εύρος 450-650 nm, ώμος από τα 685 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
H	Μαύρη περιοχή, σποραδική εναλλαγή με το κεραμικό σε λίγα σημεία. Κορυφή χλωροφύλλης στα 689 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 730 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
I	Γκρι κρουστώδης περιοχή, όπου εμφανίζεται και το κεραμικό. Κορυφή χλωροφύλλης στα 688 nm, ευρύ σημαντικής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 730 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
J	Καθαρή περιοχή. Κορυφή χλωροφύλλης στα 685 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 740 nm, ώμος από τα 750 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
K	Καθαρή περιοχή. Κορυφή χλωροφύλλης στα 689 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, ώμος από τα 740 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
L	Περιοχή με πράσινωπά βρύα. Κορυφή χλωροφύλλης στα 678 nm, ευρύ αλλά μικρής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 725 nm, ώμος από τα 730 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.
M	Περιοχή με πρασινωπά βρύα. Κορυφή χλωροφύλλης στα 678 nm, ευρύ, σημαντικής έντασης φάσμα στο εύρος 450-650 nm, δευτερεύουσα κορυφή χλωροφύλλης στα 725 nm, ώμος από τα 730 nm - 850 nm. Παρούσα η κορυφή στα 412 nm.

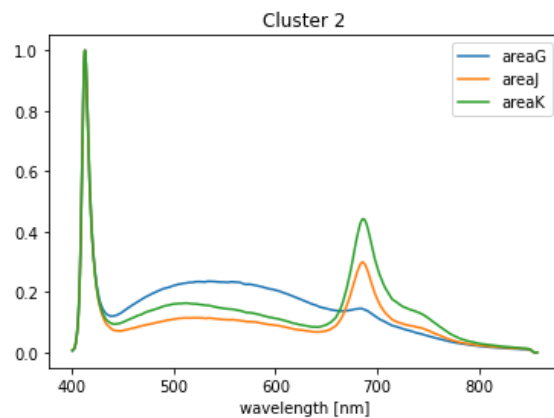
Πίνακας 2: Περιοχές Αναφοράς Μετρήσεων CALLOS CER-3 και αντίστοιχη οπτική περιγραφή



Σχήμα 27: Η ομάδα μηδέν των φασμάτων αναφοράς του δείγματος CALLOS CER-3 αντιστοιχεί σε έντονη βιολογική επιβάρυνση από φωτο-αυτότροφους μικροοργανισμούς



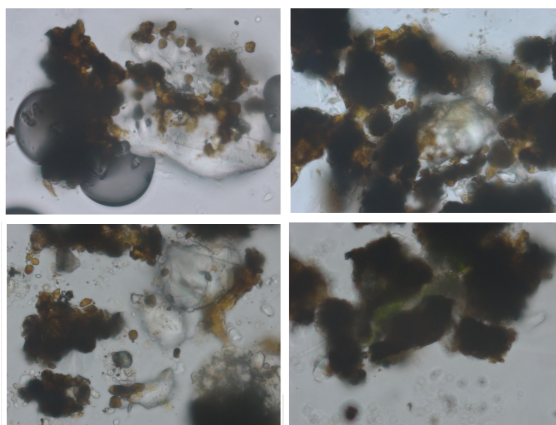
Σχήμα 28: Η ομάδα ένα των φασμάτων αναφοράς του δείγματος CALLOS CER-3 αντιστοιχεί σε έντονη βιολογική επιβάρυνση από φωτο-αυτότροφους μικροοργανισμούς όπως επίσης και από μικροοργανισμούς με ευρύ φάσμα στην περιοχή 440-600 nm όπως σπόρια μυκήτων



Σχήμα 29: Η ομάδα δύο των φασμάτων αναφοράς του δείγματος CALLOS CER-3 αντιστοιχεί σε ασθενή βιολογική επιβάρυνση από φωτο-αυτότροφους μικροοργανισμούς

### 3.3.2 CALLOS MAR-2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

Το δείγμα αυτό αποτελεί τμήμα μαρμάρινου περιρραντηρίου και εντοπίστηκε μεταξύ λιθοσωρών του αυλείου χώρου του Παλαιού Μουσείου της Ακρόπολης. Το μέγιστο σωζόμενο πλάτος είναι 10 cm, το αντίστοιχο μήκος είναι 13 cm και το αντίστοιχο πάχος είναι 8 cm. Η κύρια βιολογική επιβάρυνση του συγκεκριμένου δείγματος εστιάζεται στην ύπαρξη μαύρων μεριστεματικών μυκήτων (MCF) οι οποίοι εμφανίζονται σε μαύρες αποικίες, όπως παρουσιάζονται στο Σχήμα 30. Επιπλέον εμφανίζονται σπόρια μυκήτων είτε μαύρων μεριστεματικών είτε αερομεταφερόμενων, τα οποία εμφανίζονται ως σφαιρικά κίτρινα στίγματα στο Σχήμα 30. Επίσης, υπάρχουν μικροποσότητες χλωροφυκών ίσως και κυανοβακτηρίων, οι οποίες εμφανίζονται στις πράσινες περιοχές του σχήματος. Η συνύπαρξη των προαναφερόμενων μικροοργανισμών δημιουργεί τις συνθήκες δημιουργίας βιοφίλμ και επομένως τα ανακτηθέντα φάσματα στη συνέχεια, περιέχουν τον συνδυασμό των μικροοργανισμών σε ποικίλες περιεκτικότητες.



Σχήμα 30: Βιολογική δειγματοληψία στην επιφάνεια του δείγματος CALLOS MAR-2. Τα σφαιρικά κίτρινα στίγματα είναι σπόρια μυκήτων, οι πράσινες και κίτρινες περιοχές καταδεικνύουν ύπαρξη χλωροφυκών και πιθανώς κυανοβακτηρίων και οι μαύρες αποικίες υποδηλώνουν ανάπτυξη μαύρων μεριστεματικών μυκήτων.

Οι περιοχές του δείγματος στις οποίες πραγματοποιήθηκε η ανάκτηση φασμάτων φθορισμού παρουσιάζονται στο Σχήμα 31. Η οπτική παρατήρηση τους αντιστοιχεί στην περιγραφή που δίνεται στον Πίνακα 3, στον οποίο παρουσιάζεται και η φασματική περιγραφή των περιοχών.

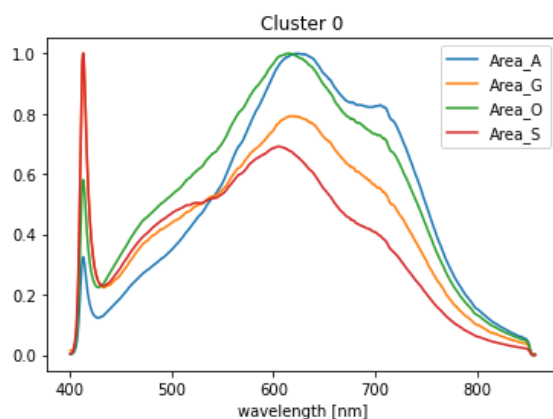


Σχήμα 31: Περιοχές αναφοράς φασμάτων φθορισμού δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

Περιοχή	Περιγραφή
A	Περιοχή με <b>κιτρινωπά στίγματα</b> . Ευρεία κορυφή στα 620 nm και μικρότερη κορυφή στα 705 nm. Κορυφή του laser στα 412 nm
B	<b>"Καθαρή" Περιοχή</b> . Ευρύ αλλά αδύναμο φάσμα φθορισμού στο εύρος 430-650 nm. Παρούσα η κορυφή στα 412 nm.
C	<b>"Καθαρή" Περιοχή</b> . Ευρεία κορυφή στα 515 nm με μειούμενη επέκταση έως τα 750 nm Παρούσα η κορυφή στα 412 nm.
D	<b>Μαύρο Στίγμα</b> . Ευρεία κορυφή στα 500 nm με μειούμενη επέκταση έως τα 650 nm. Κορυφή στα 690 nm με μειούμενη επέκταση έως τα 780 nm και τοπική αύξηση στα 730 nm. Παρούσα η κορυφή στα 412 nm.
E	<b>Διάσπαρτη μαύρη σκόνη και περιορισμένα μαύρα στίγματα</b> . Ευρύ έντονο φάσμα στα 430-660 nm με τοπική ευρεία κορυφή στα 580 nm. Κορυφή στα 685 nm με μειούμενη ένταση έως τα 780 nm. Παρούσα η κορυφή στα 412 nm.
F	<b>Ευρύ μαύρο στίγμα με διάσπαρτα λευκά</b> . Ευρύ έντονο φάσμα στα 430-660 nm με τοπική ευρεία κορυφή στα 600 nm. Κορυφή στα 690 nm με μειούμενη ένταση έως τα 780 nm. Παρούσα η κορυφή στα 412 nm.
G	Περιοχή με <b>κιτρινωπά</b> και μαύρα στίγματα. Ευρεία κορυφή στα 620 nm με μειούμενη ένταση έως τα 780 nm. Παρούσα η κορυφή στα 412 nm.
H	<b>"Καθαρή" Περιοχή</b> . Ευρύ αλλά αδύναμο φάσμα φθορισμού στο εύρος 430-650 nm. Παρούσα η κορυφή στα 412 nm.
I	<b>"Καθαρή" Περιοχή</b> με ελαφρά ίχνη χρώματος. Ευρύ αλλά αδύναμο φάσμα φθορισμού στο εύρος 430-650 nm. Παρούσα η κορυφή στα 412 nm.
J	<b>Περιοχή με ίχνη χρώματος και πιθανό σημάδι προϋπάρχουσας υγρασίας</b> . Ευρύ αλλά αδύναμο φάσμα από τα 430-660. Ασθενής κορυφή στα 685 nm. Παρούσα η κορυφή στα 412 nm.
K	<b>Μαύρο στίγμα</b> . Ευρύ αλλά αδύναμο φάσμα από τα 430-660. Ασθενής κορυφή στα 685 nm. Παρούσα η κορυφή στα 412 nm.
L	<b>Καθαρή περιοχή με αχνά μαύρα στίγματα</b> . Ευρύ φάσμα στα 430-780 nm με ευρεία κορυφή στα 600 nm. Παρούσα η κορυφή στα 412 nm.
M	Ευρεία περιοχή με μαύρα στίγματα με διάσπαρτα λευκά στίγματα. Ευρύ έντονο φάσμα στα 430-660 nm με τοπική ευρεία κορυφή στα 600 nm. Κορυφή στα 690 nm με μειούμενη ένταση έως τα 780 nm. Παρούσα η κορυφή στα 412 nm.
N	<b>Μαύρα</b> και ελαφρώς ανοιχτόχρωμα στίγματα. Ευρεία κορυφή στα 500 nm με μειούμενη επέκταση έως τα 650 nm. Κορυφή στα 690 nm με μειούμενη επέκταση έως τα 780 nm και τοπική αύξηση στα 730 nm. Παρούσα η κορυφή στα 412 nm.
O	<b>Ανοιχτόχρωμο-κιτρινωπό</b> στίγμα. Ευρεία κορυφή στα 615 nm και μικρότερη κορυφή στα 705 nm. Παρούσα η κορυφή στα 412 nm.
P	<b>Μαύρη περιοχή</b> . Ευρύ αλλά αδύναμο φάσμα από τα 430-660 με ευρεία κορυφή στα 485 nm. Ασθενής κορυφή στα 685 nm. Παρούσα η κορυφή στα 412 nm.
Q	<b>"Καθαρή" περιοχή με ανομοιόμορφη "γυψοποιημένη" επιφάνεια</b> . Ευρύ αλλά αδύναμο φάσμα από τα 430-660. Ασθενής κορυφή στα 685 nm. Παρούσα η κορυφή στα 412 nm.
R	<b>Γκρι περιοχή</b> . Ευρύ αλλά αδύναμο φάσμα από τα 430-660. Ασθενής κορυφή στα 685 nm. Παρούσα η κορυφή στα 412 nm.
S	<b>Ανοιχτόχρωμο κιτρινωπό</b> στίγμα. Ευρεία κορυφή στα 620 nm με μειούμενη ένταση έως τα 780 nm. Παρούσα η κορυφή στα 412 nm.

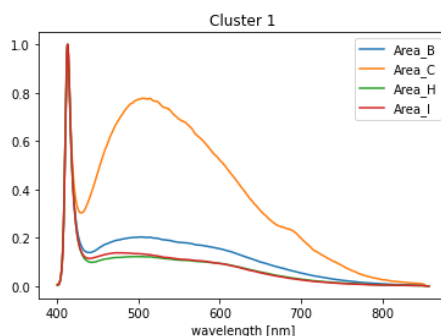
Πίνακας 3: Περιοχές Αναφοράς Μετρήσεων CALLOS MAR-2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 και αντίστοιχη οπτική περιγραφή

Τα φάσματα αναφοράς ομαδοποιούνται σε 5 κατηγορίες ανάλογα με την βιολογική τους προέλευση η οποία συνδέεται με την οπτική παρατήρηση της υπό μελέτη επιφάνειας. Η μηδενική ομάδα, δηλαδή το Cluster 0 (Σχήμα 32), συνδέεται με την ύπαρξη κίτρινων στιγμάτων τα οποία αντιστοιχούν σε σημαντική περιεκτικότητα σπορίων μυκήτων. Χαρακτηριστικό της συγκεκριμένης ομάδας αποτελεί η ευρεία κορυφή περίπου στα 620 nm, το οποίο συνδέεται πιθανώς με την κορυφή της φυκοκυανίνης[77, 31], δηλαδή με την συνύπαρξη κυανοβακτηρίων και μυκήτων. Καθώς όμως η βιολογική ανάλυση δεν υπέδειξε με βεβαιότητα την ύπαρξη κυανοβακτηρίων εξαιτίας της δυσκολίας ανάπτυξης τους στα δοκιμαστικά τριβλία, αυτό αποτελεί υπόθεση.



Σχήμα 32: Η ομάδα μηδέν των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

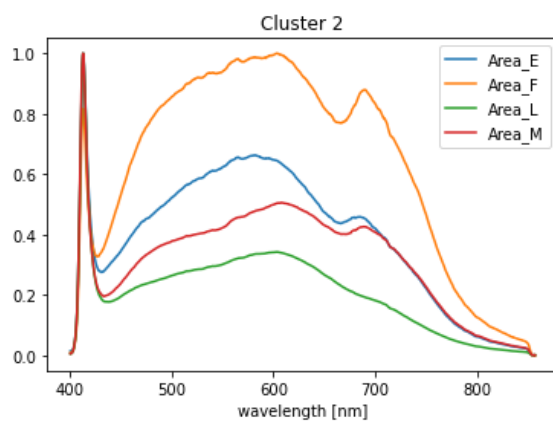
Η πρώτη ομάδα, δηλαδή το Cluster 1 (Σχήμα 33), συνδέεται με τις "καθαρότερες" περιοχές επί του μαρμάρου δηλαδή με τις περιοχές με την μικρότερη βιολογική επιβάρυνση. Το φάσμα φθορισμού είναι ευρύ στα 430-750 nm αλλά ασθενές συγκριτικά με την κορυφή ελαστικής σκέδασης του laser στα 412 nm. Εξαιρέση αποτελεί η περιοχή C στην οποία πιθανώς υπάρχει ίχνος βιολογικού στοιχείου το οποίο δεν είναι διακριτό με γυμνό μάτι ή παρουσιάζεται κάποια πρόσμιξη του υποστρώματος.



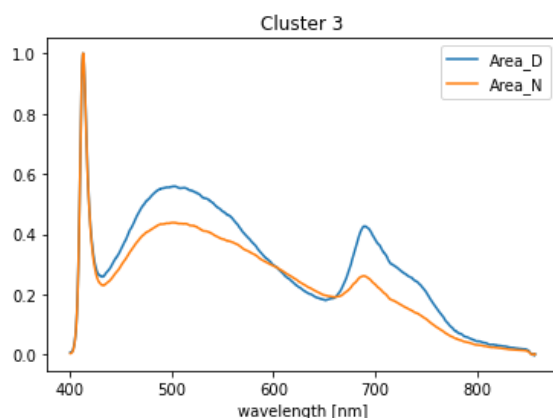
Σχήμα 33: Η ομάδα ένα των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

Η δεύτερη ομάδα, δηλαδή το Cluster 2 (Σχήμα 34), συνδέεται με τις περιοχές με μαύρα στίγματα τα οποία συνοδεύονται από διάσπαρτα λευκά στίγματα. Το φάσμα φθορισμού είναι ευρύ στα 440-750 nm με ευρείες κορυφές στα 570 - 620 nm. Επιπρόσθετη κορυφή εμφανίζεται και στα 690 nm. Χρωμοφόρα μόρια όπως η μελανίνη, η οποία όταν διεγερθεί αποτελεσματικά εκπέμπει στα 620 nm[78], τα καροτενοειδή με κορυφή περίπου στα 570 nm[78] καταδεικνύουν την ύπαρξη μυκήτων ενώ η ύπαρξη χλωροφύλλης με κορυφή στα 690 nm[79] υποδεικνύει την συνύπαρξη είτε με χλωροφύκη είτε με κυανοβακτήρια.

Η τρίτη ομάδα, δηλαδή το Cluster 3 (Σχήμα 35), συνδέεται με τις περιοχές με μαύρα και κιτρινωπά στίγματα με εντονότερη όμως την παρουσία της χλωροφύλλης, δηλαδή την ύπαρξη φωτοαυτότροφων μικροοργανισμών. Το γεγονός αυτό πιστοποιείται από την κορυφή της χλωροφύλλης στα 690 nm. Φθορίζοντα μόρια όπως το μόριο NADH στα 460 nm[78, 31], το φερουλικό οξύ στα 520 nm[70] και οι ριβοφλαβίνες στα 530 nm[78], τα οποία δημιουργούν το εύρος έντασης φθορισμού στα 430-600 nm, καταδεικνύουν την ύπαρξη μυκήτων.

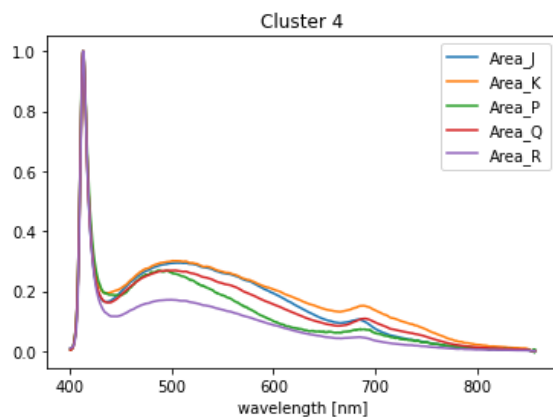


Σχήμα 34: Η ομάδα δύο των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1



Σχήμα 35: Η ομάδα τρία των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

Η τέταρτη ομάδα, δηλαδή το Cluster 4 (Σχήμα 36), περιλαμβάνει φάσματα περιοχών οι οποίες δεν εμφανίζουν μεταξύ τους οπτικές ομοιότητες αλλά παρουσιάζουν ένα ασθενές φάσμα φθορισμού συγκριτικά με την κορυφή της ελαστικής σκέδασης στα 412 nm. Το φάσμα είναι ευρύ στα 440-750 nm με ιδιαίτερα ασθενή κορυφή χλωροφύλλης περίπου στα 685 nm. Ευρεία αλλά ασθενής είναι και η κορυφή στα 460 nm πιθανώς του NADH[78, 31], το οποίο καταδεικνύει ύπαρξη μυκήτων.



Σχήμα 36: Η ομάδα τέσσερα των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

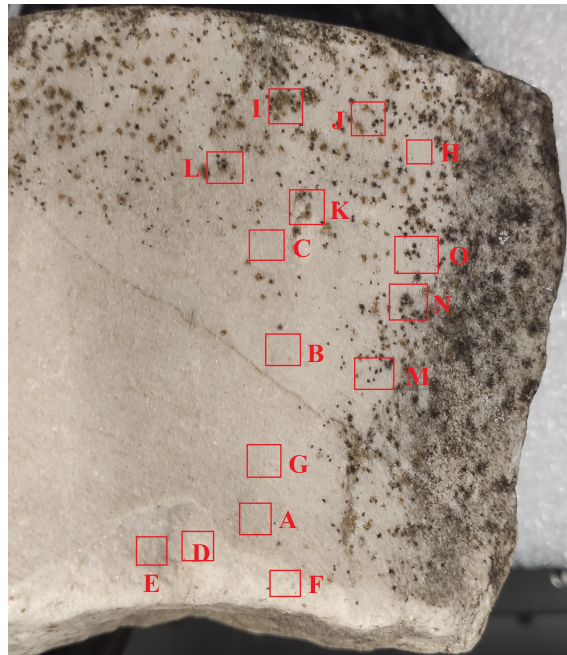
### 3.3.3 CALLOS MAR-2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

Το παρόν δείγμα αποτελεί μια δευτερεύουσα επιφάνεια του δείγματος CALLOS MAR-2 η οποία παρουσιάζει ένα εναλλακτικό ενδιαφέρον συγκριτικά με την ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1. Στην συγκεκριμένη επιφάνεια εμφανίζονται επιπροσθέτως αρκετά πράσινα στίγματα σε συνδυασμό με μαύρα, το οποίο αντικατοπτρίζεται και στα φάσματα φθορισμού που λήφθηκαν. Οι περιοχές αναφοράς στις οποίες πραγματοποιήθηκε μέτρηση φθορισμού παρουσιάζονται περιγραφικά στον Πίνακα 4 και οπτικά στο Σχήμα 37. Στην μελέτη του παρόντος δείγματος εφαρμόστηκαν δύο διαφορετικές προσεγγίσεις ομαδοποίησης φασμάτων. Στην πρώτη προσέγγιση τα φάσματα κατηγοριοποιήθηκαν με βάση την οπτική παρατήρηση των περιοχών. Σύμφωνα με την δεύτερη προσέγγιση τα φάσματα ομαδοποιήθηκαν με βάση την ομοιότητα τους.

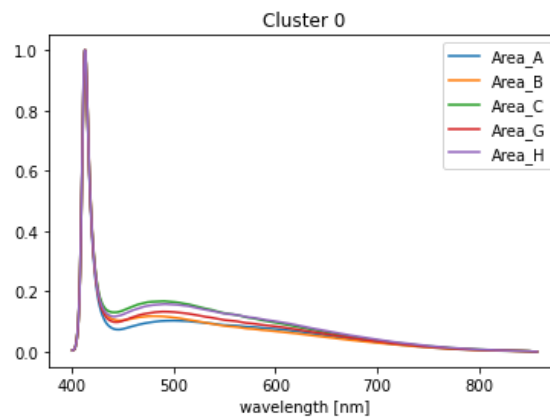
Περιοχή	Περιγραφή
A	Φαινομενικά καθαρή περιοχή μαρμάρου. Ασθενές σήμα φθορισμού στο εύρος 440-690 nm. Κορυφή του laser στα 412 nm
B	Φαινομενικά καθαρή περιοχή μαρμάρου. Ασθενές σήμα φθορισμού στο εύρος 440-690 nm. Παρούσα η κορυφή στα 412 nm.
C	Φαινομενικά καθαρή περιοχή μαρμάρου. Ασθενές σήμα φθορισμού στο εύρος 440-690 nm. Ευρεία κορυφή στα 485 nm. Παρούσα η κορυφή στα 412 nm.
D	Περιοχή που η επιφάνεια του μαρμάρου φαίνεται τραχιά, σαν να έχει αποκολληθεί ο εξωτερικός «φλοιός». Κορυφή στα 437 nm και στα 468 nm. Το φάσμα επεκτείνεται έως τα 690 nm. Παρούσα η κορυφή στα 412 nm.
E	Περιοχή που η επιφάνεια του μαρμάρου φαίνεται τραχιά, σαν να έχει αποκολληθεί ο εξωτερικός «φλοιός». Ασθενές σήμα φθορισμού στο εύρος 440-690 nm. Ευρεία κορυφή στα 485 nm. Παρούσα η κορυφή στα 412 nm.
F	Περιοχή που η επιφάνεια του μαρμάρου φαίνεται τραχιά, σαν να έχει αποκολληθεί ο εξωτερικός «φλοιός». Κορυφή στα 437 nm και στα 468 nm. Το φάσμα επεκτείνεται έως τα 690 nm. Παρούσα η κορυφή στα 412 nm.
G	Φαινομενικά καθαρή περιοχή μαρμάρου. Ασθενές σήμα φθορισμού στο εύρος 440-690 nm. Ευρεία κορυφή στα 485 nm. Παρούσα η κορυφή στα 412 nm.
H	Φαινομενικά καθαρή περιοχή μαρμάρου. Ασθενές σήμα φθορισμού στο εύρος 440-690 nm. Ευρεία κορυφή στα 485 nm. Παρούσα η κορυφή στα 412 nm.
I	Πράσινο «στίγμα»/κηλίδα. Ευρύ φάσμα από τα 440-750 nm, έντονη κορυφή στα 630 nm και επιπρόσθετες κορυφές στα 710 και 720 nm. Παρούσα η κορυφή στα 412 nm.
J	Πράσινο «στίγμα»/κηλίδα. Ευρύ φάσμα από τα 440-750 nm, έντονη κορυφή στα 630 nm και επιπρόσθετες κορυφές στα 710 και 720 nm. Παρούσα η κορυφή στα 412 nm.
K	Πράσινο «στίγμα»/κηλίδα. Ευρύ φάσμα από τα 440-750 nm, έντονη κορυφή στα 630 nm και επιπρόσθετες κορυφές στα 710 και 720 nm. Παρούσα η κορυφή στα 412 nm.
L	Πράσινο «στίγμα»/κηλίδα δίπλα σε μαύρη κηλίδα. Ευρύ φάσμα από τα 440-750 nm, έντονη κορυφή στα 630 nm και επιπρόσθετες κορυφές στα 710 και 720 nm. Παρούσα η κορυφή στα 412 nm.
M	Μαύρο «στίγμα»/κηλίδα. Κορυφή στα 440 nm και ευρεία κορυφή στα 470 nm. Ευρεία κορυφή στα 610 nm και επέκταση του φάσματος ως τα 750 nm. Παρούσα η κορυφή στα 412 nm.
N	Μαύρο «στίγμα»/κηλίδα. Ευρύ φάσμα από τα 440-750 nm, έντονη κορυφή στα 630 nm και επιπρόσθετες κορυφές στα 710 και 720 nm. Επιπρόσθετα έντονη κορυφή στα 690 nm, ένδειξη ύπαρξης χλωροφύλλης. Παρούσα η κορυφή στα 412 nm.
O	Μαύρο «στίγμα»/κηλίδα. Κορυφή στα 440 nm και ευρεία κορυφή στα 470 nm. Επέκταση του φάσματος ως τα 750 nm. Παρούσα η κορυφή στα 412 nm.

Πίνακας 4: Περιοχές Αναφοράς Μετρήσεων CALLOS MAR-2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 και αντίστοιχη οπτική περιγραφή

Σύμφωνα με την πρώτη προσέγγιση, η μηδενική ομάδα φασμάτων, δηλαδή το Cluster 0 (Σχήμα 38), αντιστοιχεί σε φαινομενικά καθαρές περιοχές μαρμάρου με ευρύ αλλά ασθενές φάσμα στα 440-690 nm, με έντονη κορυφή της ελαστικής σκέδασης του laser στα 412 nm.



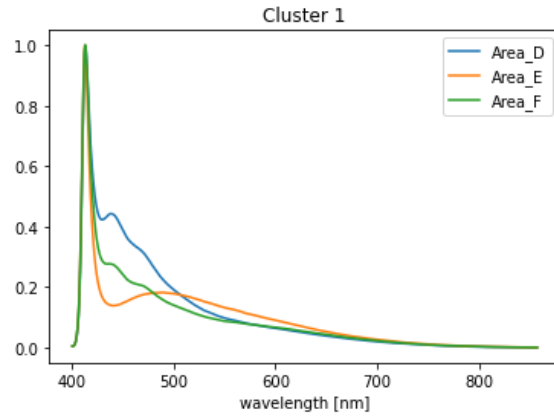
Σχήμα 37: Περιοχές αναφοράς φασμάτων φθορισμού δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2



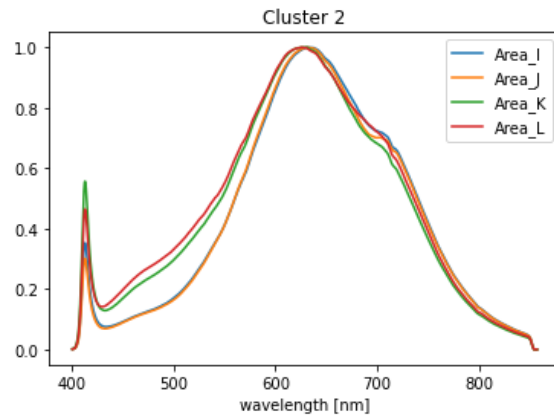
Σχήμα 38: Η ομάδα μηδέν των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

Στην πρώτη ομάδα φασμάτων, στο Cluster 1 (Σχήμα 39), ανήκουν τα φάσματα από αποφλοιωμένα σημεία της επιφανείας. Χαρακτηριστικό αυτής της ομάδας είναι οι κορυφές στα 437 nm και στα 468 nm. Η δεύτερη ομάδα φασμάτων, δηλαδή το Cluster 2 (Σχήμα 40), αντιστοιχεί στα πράσινα στίγματα τα οποία υπάρχουν επί της επιφάνειας του δείγματος. Το φάσμα είναι ευρύ και υψηλής έντασης στο εύρος 440-750 nm με έντονη και ευρεία κορυφή στα 630 nm, η οποία είναι ενδεικτική της φυκοκυανίνης. Εντός του φάσματος υπεισέρχεται και η κορυφή της χλωροφύλλης στα 690 nm. Οι δύο αυτές κορυφές είναι ενδεικτικές της ύπαρξης φωτοαυτότροφων μικροοργανισμών. Η τρίτη και τελευταία ομάδα φασμάτων, δηλαδή το Cluster 3 (Σχήμα 41), αντιστοιχεί στις περιοχές με έντονα μαύρα στίγματα με ευρείες κορυφές στα 440 nm και στα 470 nm, οι οποίες όμως δεν είναι ενδεικτικές της μελανίνης [80]. Το ένα όμως από τα τρία φάσματα παρουσιάζει κορυφή στα 620 nm και επίσης αναδεικνύεται μια κορυφή χλωροφύλλης στα 690 nm, το οποίο υποδηλώνει πιθανή συνύπαρξη των μυκήτων με φωτοαυτότροφους μικροοργανισμούς.

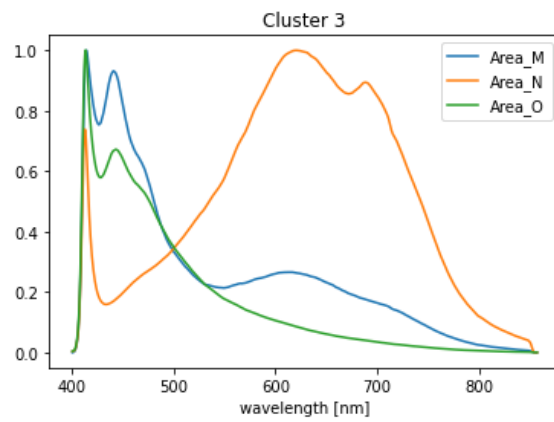




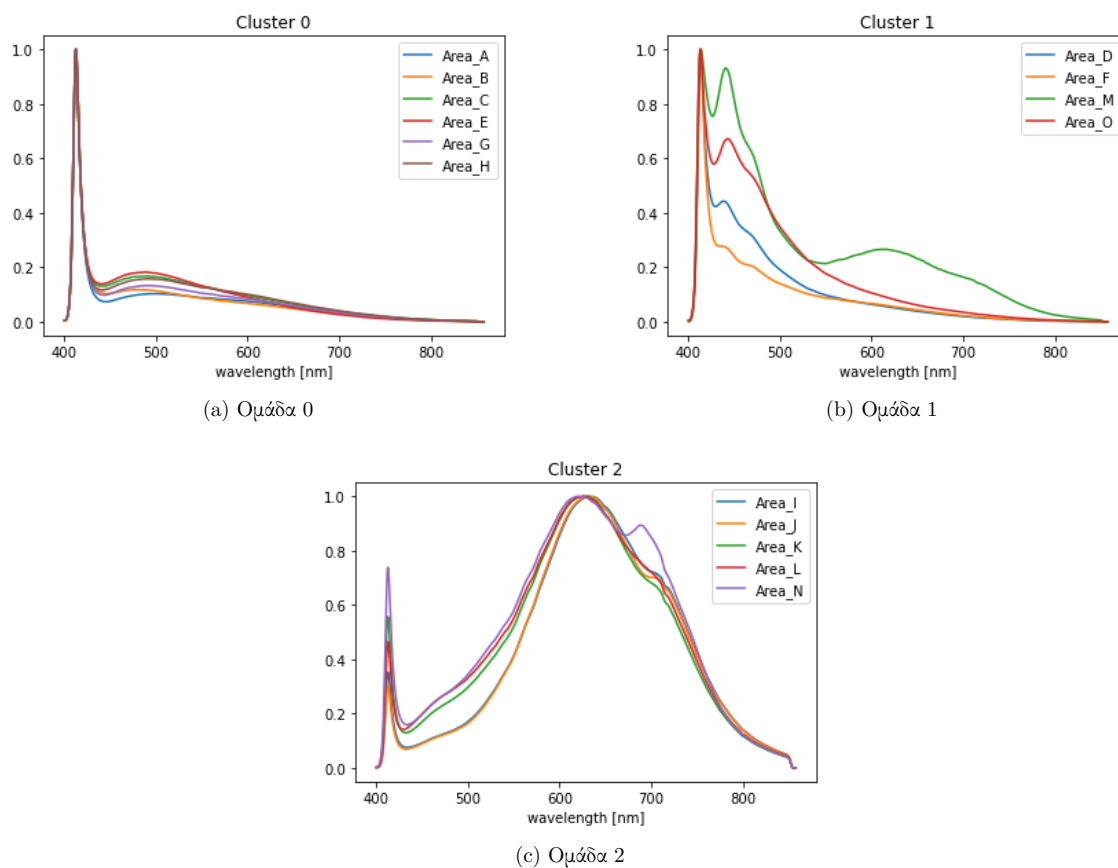
Σχήμα 39: Η ομάδα ένα των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2



Σχήμα 40: Η ομάδα δύο των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2



Σχήμα 41: Η ομάδα τρία των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2



Σχήμα 42: Οι ομάδες των φασμάτων αναφοράς του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 σύμφωνα με την φασματική ομοιότητα

Στην δεύτερη προσέγγιση τα φάσματα των περιοχών αναφοράς ομαδοποιήθηκαν σύμφωνα με την ομοιότητά τους. Οι βασικές αλλαγές εστιάζονται στην μείωση των ομάδων από 4 σε 3, οι οποίες όμως συγκεντρώνουν ενιαία φασματικά χαρακτηριστικά. Η μηδενική ομάδα, δηλαδή το Cluster 0 περιλαμβάνει τις καθαρές περιοχές μαρμάρου σε συνδυασμό με μια περιοχή μερικώς αποφλοιωμένη, η οποία όμως δεν παρουσιάζει βιολογική επιβάρυνση. Η πρώτη ομάδα, δηλαδή το Cluster 1, συγκεντρώνει αποφλοιωμένες "καθαρές" περιοχές καθώς και μαύρα στίγματα, γεγονός το οποίο δύναται να ερμηνευτεί με δύο διαφορετικούς τρόπους. Είτε στα μαύρα στίγματα δεν υφίσταται αποτελεσματική διέγερση και άρα εκπομπή της μελανίνης στα 540 nm [80] και επομένως διεγείρονται άλλα βιολογικά μόρια τα οποία υφίσταται και στην αποφλοιωμένη επιφάνεια. Είτε αποτελεί φθορισμό επιφανειακής επικάλυψης στο μάρμαρο, η οποία δεν είναι εμφανής με γυμνό μάτι. Η δεύτερη και τελευταία ομάδα, δηλαδή το Cluster 2, συσχετίζεται με τις περιοχές στις οποίες η βιολογική επιβάρυνση προκύπτει κυρίως από φωτοαυτότροφους μικροοργανισμούς με πιθανή συνύπαρξη με ετερότροφους μύκητες. Τα φάσματα των ομάδων παρουσιάζονται στο Σχήμα 42.

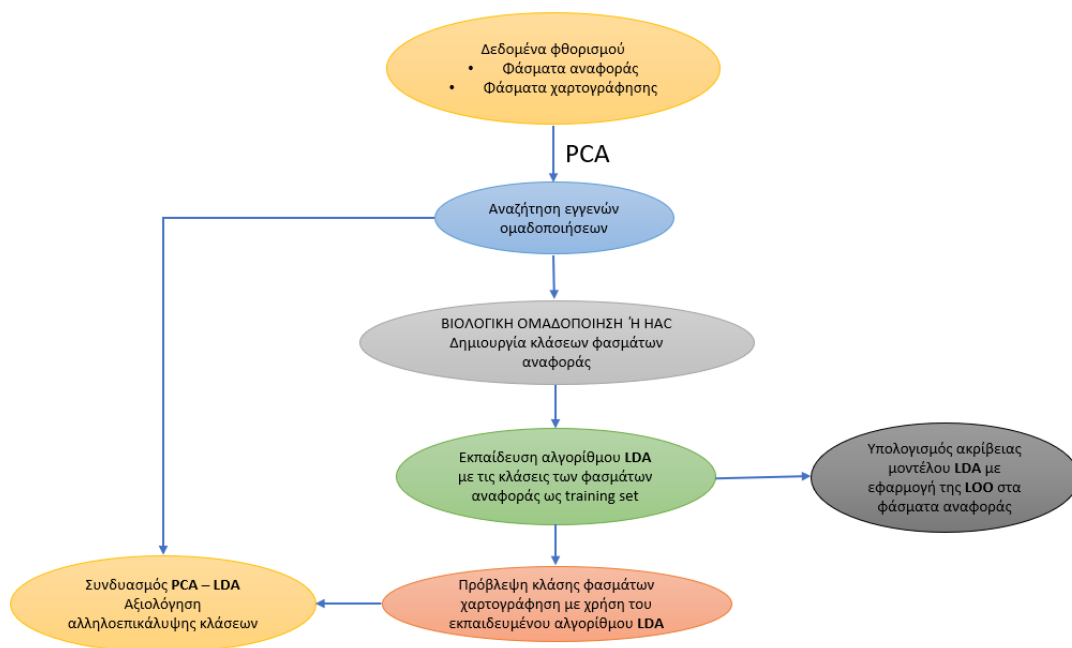
## 4 Κεφάλαιο 4: Στατιστική Επεξεργασία Φασματικών Δεδομένων

Στο συγκεκριμένο κεφάλαιο θα παρουσιαστεί αναλυτικά η διαδικασία επεξεργασίας των φασματικών δεδομένων φθορισμού, με στόχο την ταξινόμηση των περιοχών των υπό μελέτη επιφανειών σε συγκεκριμένους τύπους βιοδιάβρωσης, όπως προσδιορίστηκαν στις κλάσεις αναφοράς στην ενότητα 3.3. Η ανάλυση πραγματοποιήθηκε με χρήση της γλώσσας προγραμματισμού Python στο λογισμικό Spyder.

Τόσο τα φασματικά δεδομένα των κλάσεων αναφοράς, δηλαδή το σετ εκπαίδευσης, όσο και τα φασματικά δεδομένα χαρτογράφησης της κάθε επιφάνειας, θα υποστούν αρχικά μείωση των διαστάσεων τους με χρήση της PCA. Στα φάσματα που μελετήθηκαν, διατηρήθηκαν τα μήκη κύματος άνω των 400 nm και κανονικοποιήθηκαν στην μέγιστη τιμή τους. Επίσης, τυποποιήθηκαν σύμφωνα με την διαδικασία που περιγράφεται στην υποενότητα 2.1.2. Πιο συγκεκριμένα, τα 1688 μήκη κύματος που περιγράφουν το κάθε φάσμα-παρατήρηση θα μειωθούν σε τρεις μεταβλητές, τα principal components, που περιγράφουν πάνω από το 95% της πληροφορίας. Στη συνέχεια, μελετώντας τις συσχετίσεις μεταξύ των παλαιών και νέων μεταβλητών στο διάγραμμα των loadings, θα αναζητηθούν συνδέσεις των principal components με τις κλάσεις αναφοράς της κάθε επιφάνειας, όπως περιγράφηκαν στην ενότητα 3.3. Στόχος αυτής της αναζήτησης είναι η unsupervised ταξινόμηση των περιοχών στους συγκεκριμένους τύπους βιοδιάβρωσης. Επίσης, θα αναζητηθεί η εγγενής ομαδοποίηση των δεδομένων σε διακριτές κλάσεις με βάση τις τιμές των principal components, τα scores.

Καθώς η εφαρμογή μόνο της PCA σε κάθε επιφάνεια δεν υπέδειξε διακριτές κλάσεις βιοδιάβρωσης για τα δεδομένα, εφαρμόστηκε στη συνέχεια το μοντέλο πρόβλεψης της LDA. Τα φασματικά δεδομένα των κλάσεων αναφοράς αποτέλεσαν το σετ εκπαίδευσης του αλγορίθμου της LDA. Με αυτό τον τρόπο δημιουργήθηκε ένα μοντέλο το οποίο αποτελείται από κλάσεις με συγκεκριμένο τύπο βιοδιάβρωσης. Ειδικότερα στο δείγμα CALLOS CER-3, οι κλάσεις αναφοράς επαληθεύτηκαν και με την μέθοδο HAC. Ωστόσο, στα υπόλοιπα δείγματα δεν εφαρμόστηκε η μέθοδος αυτή, καθώς οι κλάσεις αναφοράς δημιουργήθηκαν μόνο με βάση την βιολογική πληροφορία. Στη συνέχεια, κάθε φάσμα χαρτογράφησης εισήχθη στο μοντέλο της LDA και προβλέφθηκε η κλάση αναφοράς του δηλαδή ο τύπος της βιοδιάβρωσης που το χαρακτηρίζει.

Επιπροσθέτως, εξετάστηκε η απόδοση του μοντέλου με βάση την μέθοδο Leave-One-Out Cross Validation (LOO) και προσδιορίστηκε η ακρίβεια των προβλέψεων του για κάθε τύπο βιοδιάβρωσης. Στο τέλος της ανάλυσης για την κάθε επιφάνεια, συνδυάστηκαν οι πληροφορίες που έδωσε η LDA για τον τύπο βιοδιάβρωσης των φασμάτων χαρτογράφησης με τα scores της PCA για τα ίδια φάσματα. Ο στόχος αυτού του συνδυασμού έγκειται στην αξιολόγηση της χρησιμότητας του αλγορίθμου πρόβλεψης της LDA όταν η επικάλυψη των εγγενών κλάσεων των δεδομένων, όπως προκύπτουν από την PCA, είναι σημαντική. Σε σχέση με την υπάρχουσα βιβλιογραφία, η εφαρμογή του μοντέλου πρόβλεψης του τύπου βιοδιάβρωσης με χρήση της LDA σε φάσματα φθορισμού δεν είχε έως τώρα πραγματοποιηθεί. Στην επόμενη σελίδα παρουσιάζεται γραφικά στο Σχήμα 43.



Σχήμα 43: Διαγραμμα Ροής Στατιστικής Επεξεργασίας Δεδομένων

## 4.1 CALLOS CER 3

### Principal Component Analysis

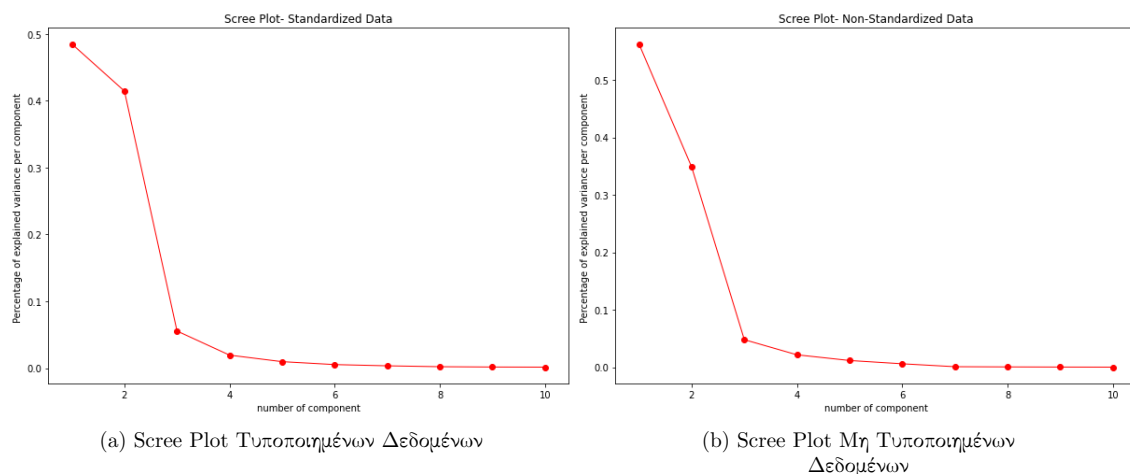
Η ανάλυση ξεκινά με την εφαρμογή της PCA και την αναζήτηση εγγενών ομαδοποιήσεων των δεδομένων. Ειδικότερα, ο πίνακας δεδομένων του δείγματος CALLOS CER 3 είναι ένας πίνακας  $52 \times 1688$  ο οποίος αποτελείται συνολικά από 52 παρατηρήσεις εκ των οποίων οι 13 αντιστοιχούν στα φάσματα των περιοχών αναφοράς και οι 39 στα σημεία της χαρτογράφησης που πραγματοποιήθηκε. Οι μεταβλητές του προβλήματος είναι 1688 και αντιστοιχούν στα μήκη κύματος που καταγράφει ο φασματογράφος από τα 400 nm και άνω. Για την εφαρμογή όμως της PCA, ο πίνακας δεδομένων προς το παρόν δεν θα διαχωριστεί. Επιπλέον, η μείωση των διαστάσεων του μέσω της PCA είναι απαραίτητη καθώς οι 1688 μεταβλητές που τα περιγράφουν δεν είναι όλες ανεξάρτητες μεταξύ τους και συνεπώς δεν είναι όλες απαραίτητες για την πλήρη περιγραφή των ιδιοτήτων των δεδομένων.

Χαρακτηριστικό παράδειγμα της προαναφερόμενης εξάρτησης αποτελεί μια από τις κορυφές της χλωροφύλλης, πχ στα 685 nm, η οποία συνήθως εμφανίζεται με συγκεκριμένο FWHM ως γκαουσιανή κορυφή [79]. Αυτό πρακτικά σημαίνει πως ορισμένες μεταβλητές που προηγούνται ή έπονται της μεταβλητής των 685 nm είναι μεταξύ τους συσχετισμένες ισχυρά καθώς αντιστοιχούν όλες στην αποδιέγερση από την  $S_1$  στην  $S_0$  για το μόριο της χλωροφύλλης. Επομένως, όσο αυξάνεται η τιμή της έντασης του φθορισμού στην μεταβλητή των 685 nm, αυξάνεται και των υπολοίπων μηκών κύματος - μεταβλητών της κορυφής. Συνεπώς, εφαρμόζοντας την PCA οι 1688 μεταβλητές μειώνονται σε σημαντικά λιγότερες ανεξάρτητες μεταξύ τους, οι οποίες περιγράφουν σχεδόν την συνολική πληροφορία που υπάρχει εντός των φασμάτων.

Τα δεδομένα, προτού τροφοδοτηθούν στον αλγόριθμο της PCA, θα πρέπει να υποστούν τυποποίηση, όπως αυτή περιγράφεται στην ενότητα 2.1. Για τον σκοπό αυτό χρησιμοποιήθηκε η κλάση StandardScaler από το πακέτο sklearn.preprocessing. Στη συνέχεια εφαρμόστηκε η μέθοδος PCA με χρήση της κλάσης PCA από το πακέτο sklearn.decomposition. Η επιλογή του αριθμού των components με βάση το ποσοστό της συνολικής επεξηγούμενης διακύμανσης προκύπτει από την ανάλυση του κρυμνογραφήματος (Scree Plot). Στο σημείο όπου το κρυμνογράφημα δημιουργεί σχηματισμό "αγκώνα", το ποσοστό της επεξηγούμενης διακύμανσης για το κάθε component μειώνεται δραστικά και γίνεται περίπου μηδενικό. Αυτός θα είναι και ο αριθμός των components που θα διατηρηθούν για την συνέχεια της μελέτης [48]. Στο Γράφημα 44 παρουσιάζονται τα κρυμνογραφήματα των τυποποιημένων και μη τυποποιημένων δεδομένων. Από τον Πίνακα 5 διαφαίνεται πως τόσο στην περίπτωση των τυποποιημένων δεδομένων όσο και στην περίπτωση των μη τυποποιημένων τα τρία πρώτα components περιλαμβάνουν σχεδόν το 95,4% της επεξηγούμενης διακύμανσης (σχέση 43). Παρόλα αυτά, θα διατηρήσουμε τα τυποποιημένα δεδομένα ώστε η ανάλυση να μην είναι ευαίσθητη σε αλλαγές στην κλίμακα των μεταβλητών [40].

	PC1	PC2	PC3	PC4	PC5
Standardized Data	0.485	0.414	0.0554	0.0194	0.0096
Non-Standardized Data	0.562	0.348	0.048	0.022	0.012

Πίνακας 5: Επεξηγούμενη διασπορά για κάθε component για τυποποιημένα και μη τυποποιημένα δεδομένα



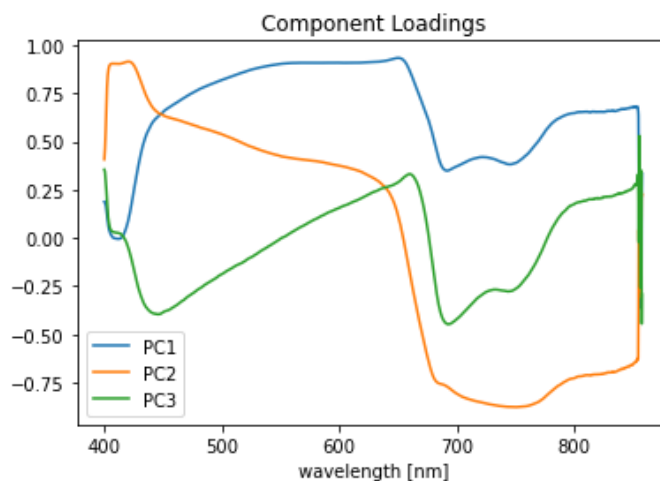
Σχήμα 44: Κρυμνογραφήματα τυποποιημένων και μη τυποποιημένων δεδομένων

Συμφωνα με το Σχήμα 44 και το κριτήριο του "αγκώνα" (υποενότητα 2.1.2), διατηρούνται τα τρία πρώτα components των τυποποιημένων δεδομένων και πλέον ο πίνακας δεδομένων είναι έχει διαστάσεις  $52 \times 3$ . Ενδεικτικά, στον Πίνακα 6 παρουσιάζονται οι τιμές των principal components για τις 13 πρώτες παρατηρήσεις, οι οποίες αντιστοιχούν στις περιοχές αναφοράς. Η συνολική επεξηγούμενη διακύμανση αντιστοιχεί στο 95,4%, το οποίο προκύπτει ως άθροισμα των ποσοστών διακύμανσης του κάθε component (σχέση 42), όπως παρουσιάζεται στον Πίνακα 5.

Τα φορτία (loadings) των components (σχέση 45), τα οποία δηλώνουν το ποσοστό συσχέτισης των παλαιών μεταβλητών (μήκη κύματος) με τις νέες (principal components), παρουσιάζονται στο Γράφημα 45. Το πρώτο principal component συσχετίζεται ισχυρά ( $> 0.5$ ) με τις μεταβλητές στο εύρος 450-650 nm περιγράφοντας έτσι τα φθορίζοντα μόρια εκτός της χλωροφύλλης και στα 750-850 nm, οι οποίες όμως δεν συνεισφέρουν σε ποιοτική πληροφορία. Το δεύτερο principal component περιγράφει κυρίως τα εύρη 400-430 nm (εύρος κορυφής ελαστικής σκέδασης στα 412 nm) και 680 - 800 nm (κορυφές χλωροφύλλης) ενώ το τρίτο συσχετίζεται με τις μεταβλητές στα εύρη 430-500 nm και 680-770 nm. Επομένως, σε πρώτη προσέγγιση κανένα από τα τρία component δεν συσχετίζεται αμιγώς με κάποια από τις ομάδες φασμάτων που ορίστηκαν στην ενότητα 3.3 στα γραφήματα 27, 28 και 29. Ωστόσο, στη συνέχεια θα μελετηθούν και οι γραφικές σχέσεις μεταξύ των principal components στα score plots, καθώς είναι πιθανό να παρατηρηθούν ομαδοποιήσεις που δεν είναι εμφανείς από το γράφημα των loadings.

Οι κανονικοποιημένες τιμές των νέων μεταβλητών, τα scores των components, παρουσιάζονται για κάθε μια παρατήρηση στα score plots του Σχήματος 46. Η κανονικοποίηση προκύπτει εάν η στήλη του score του κάθε component διαιρεθεί με την αντίστοιχη ιδιάζουσα τιμή (singular value - σχέση 27). Καθώς υπάρχουν τρεις διαφορετικές μεταβλητές που αντιστοιχούν σε κάθε παρατήρηση, τότε υπάρχουν τρεις διαφορετικοί συνδυασμοί score plot τα οποία δύναται να δώσουν πληροφορίες για πιθανές ομαδοποιήσεις των δεδομένων.

Παρατηρώντας τα score plots μπορούν να εξαχθούν ορισμένα ποιοτικά συμπεράσματα σχετικά με πιθανές ομαδοποιήσεις των δεδομένων βασισμένες σε εγγενή χαρακτηριστικά τους. Ειδικότερα, στο συγκεκριμένο δείγμα υπήρχε έντονη βιολογική επιβάρυνση από φωτοαυτότροφους μικροοργανισμούς και επομένως η περιεκτικότητα της χλωροφύλλης αποτελεί ένα μέτρο του φυτικού βιολογικού φορτίου της επιφάνειας. Επομένως μια πιθανή ομαδοποίηση θα μπορούσε να προκύψει με βάση την περιεκτικότητα

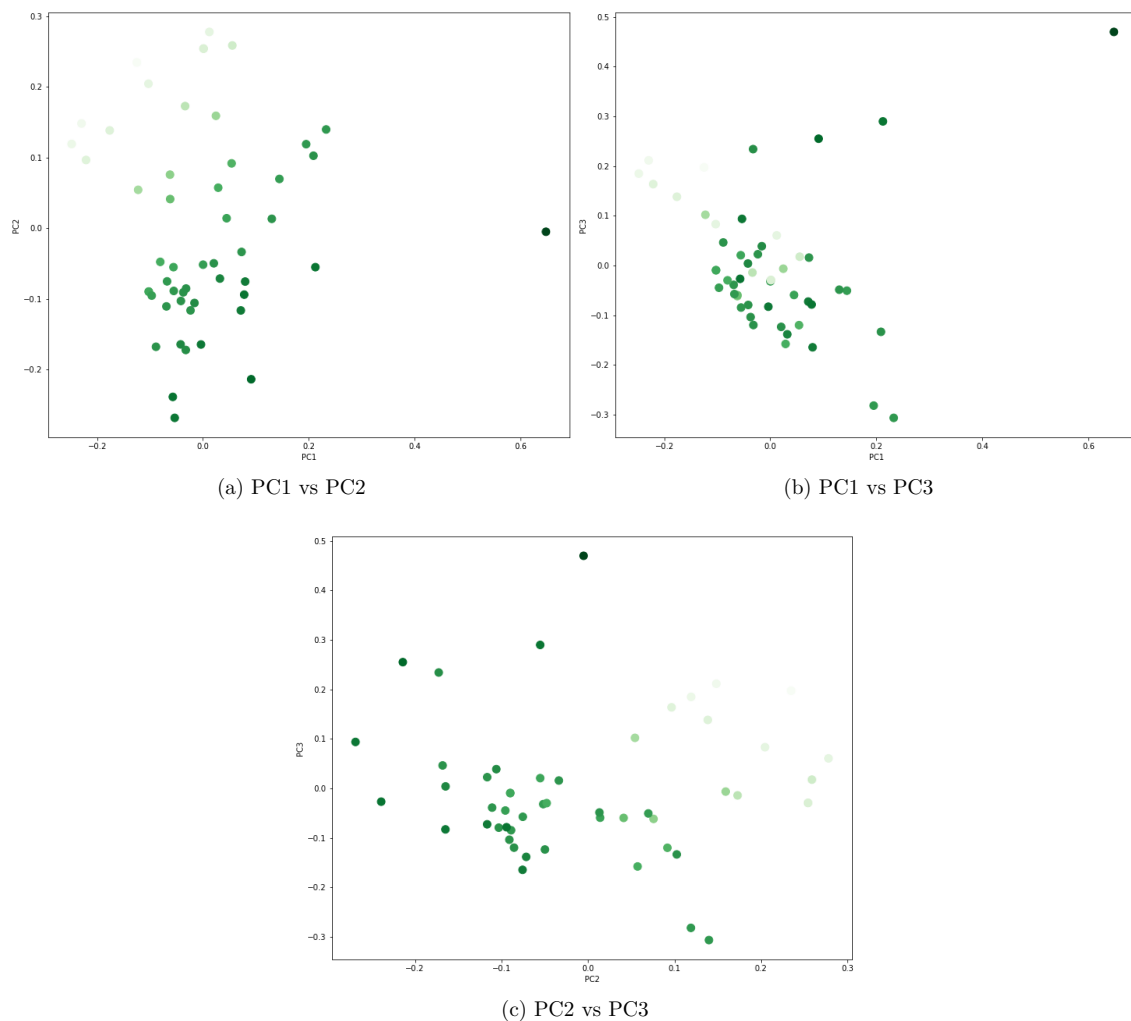


Σχήμα 45: Τα φορτία των κύριων συστατικών τα οποία προκύπτουν από τον πίνακα δεδομένων του δείγματος CALLOS CER-3

Areas	PC1	PC2	PC3
areaA	133.74	-0.96	32.79
areaB	-11.75	-45.58	-1.87
areaC	0.05	-9.86	-2.22
areaD	43.84	-10.53	20.23
areaE	-8.70	-31.39	0.29
areaF	-11.01	-51.24	6.55
areaG	-25.76	44.78	13.78
areaH	-18.31	-32.04	3.24
areaI	15.06	-6.42	1.11
areaJ	-47.35	28.29	14.76
areaK	-36.36	26.42	9.66
areaL	-6.69	-32.90	16.34
areaM	18.81	-40.81	17.80

Πίνακας 6: Πίνακας τιμών - scores των principal components στις περιοχές αναφοράς του δείγματος CALLOS CER-3

της χλωροφύλλης σε κάθε σημείο μέτρησης, η οποία υποδηλώνεται από το ολοκλήρωμα της συνάρτησης της έντασης του φθορισμού ως προς το μήκος κύματος [57] στα Γραφήματα 27, 28 και 29. Η ολοκλήρωση πραγματοποιήθηκε στο εύρος από τα 650 nm έως τα 800 nm για να συμπεριληφθούν όλες οι πιθανές κορυφές της χλωροφύλλης [79]. Ο υπολογισμός αυτός διακρίνεται στα score plots με βάση την εξής σύμβαση: Όσο πιο ανοιχτόχρωμη είναι η πράσινη απόχρωση τόσο πιο μικρή είναι η τιμή του ολοκληρώματος και άρα η περιεκτικότητα χλωροφύλλης ενώ όσο η απόχρωση εξελίσσεται σε πιο σκουρόχρωμη τόσο αυξάνεται η εν λόγω περιεκτικότητα. Παρατηρείται πως στα score plots PC1 vs PC2 και PC2 vs PC3, οι περιοχές με υψηλή περιεκτικότητα χλωροφύλλης σκοράρουν κυρίως σε αρνητικές τιμές του PC2 ενώ οι μη επιβαρυνμένες από χλωροφύλλη περιοχές σκοράρουν θετικά στο PC2. Επομένως σε έναν χρωματικό χάρτη της επιφάνειας, η διακύμανση της περιεκτικότητας χλωροφύλλης θα παρουσιάζει ανάλογη διακύμανση με το PC2. Εντούτοις, αυτή η συσχέτιση που αναδύθηκε δεν είναι ικανή να ταξινομήσει τα δεδομένα χαρτογράφησης σε κάποια από τις κλάσεις των φασμάτων αναφοράς με σαφήνεια. Αντίθετα, υποδεικνύει την ομαδοποίηση μόνο με βάση την κορυφή της χλωροφύλλης, μη λαμβάνοντας υπόψη το υπόλοιπο φάσμα. Επομένως, κρίνεται απαραίτητη η χρήση ενός μοντέλου πρόβλεψης, το οποίο θα προσδιορίζει σε ποια κλάση αναφοράς ανήκει το κάθε φάσμα.



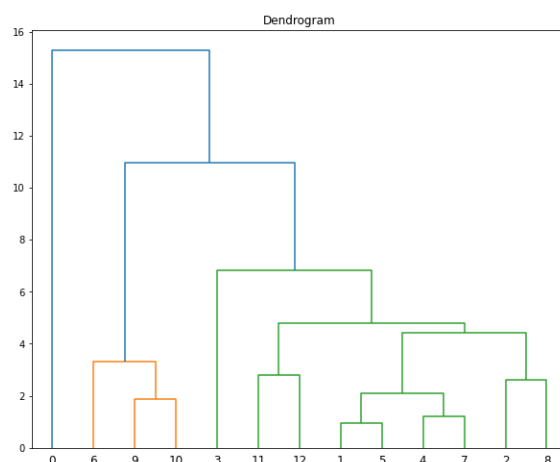
Σχήμα 46: Score plots των παρατηρήσεων που αντιστοιχούν στο δείγμα CALLOS CER 3. Η χρωματική κλίμακα αντιστοιχεί στην περιεκτικότητα χλωροφύλλης του σημείου μέτρησης. Από το ανοιχτόχρωμο πράσινο προς το σκουρόχρωμο, η τιμή του ολοκληρώματος της συνάρτησης της έντασης φθορισμού στο εύρος 650-800 nm αυξάνεται

### Linear Discriminant Analysis

Η εφαρμογή της PCA δεν ήταν αρκετή ώστε να διαχωρίσει τα δεδομένα σε κλάσεις με μη επιβλεπόμενο τρόπο (unsupervised). Επομένως, τα μειωμένων διαστάσεων δεδομένα διαχωρίζονται σε δύο σετ, το ένα χρησιμοποιείται για την δημιουργία του μοντέλου πρόβλεψης της LDA και το άλλο για την εφαρμογή του. Το πρώτο σετ, το οποίο περιέχει τα φάσματα των περιοχών αναφοράς θα διαχωριστεί σε κατάλληλες κλάσεις σύμφωνα με την βιολογική πληροφορία, όπως πραγματοποιήθηκε στην ενότητα 3.3. Στη συνέχεια, τα scores καθώς και η κλάση της κάθε περιοχής τροφοδοτούνται στον αλγόριθμο της LDA και δημιουργείται το μοντέλο πρόβλεψης όπως περιγράφεται στην ενότητα 2.3. Για τον σκοπό αυτό χρησιμοποιείται η κλάση LinearDiscriminantAnalysis από το πακέτο sklearn.discriminant-analysis. Αφού δημιουργηθεί το μοντέλο, τα scores του δεύτερου σετ δεδομένων, το οποίο αποτελείται από τα φάσματα της χαρτογράφησης του CALLOS CER-3, τροφοδοτούνται στον αλγόριθμο της LDA, ο οποίος προβλέπει τον αριθμό της κλάσης που ανήκει το κάθε φάσμα.



Στην ενότητα 3.3, οι περιοχές αναφοράς διαχωρίστηκαν σε ομάδες με βάση την βιολογική πληροφορία του σημείου μέτρησης. Οι ομάδες αυτές παρουσιάζονται στα γραφήματα 27, 28 και 29. Η επιβεβαίωση αυτής της κατηγοριοποίησης επαληθεύτηκε και αλγοριθμικά με χρήση της ομαδοποίησης κατά συστάδες (HAC), όπως περιγράφηκε στην ενότητα 2.2. Ειδικότερα, όπως πραγματοποιήθηκε και από τους Hallstrom et al, 2009 [70], τα φάσματα των περιοχών αναφοράς τροφοδοτούνται στον αλγόριθμο της ομαδοποίησης κατά συστάδες, αξιοποιώντας ως μέτρο ομοιότητας την Ευκλείδεια Απόσταση και ως απόσταση σύνδεσης την Μέση Σύνδεση της σχέσης 48. Για τον σκοπό αυτό χρησιμοποιείται η κλάση AgglomerativeClustering από το πακέτο sklearn.cluster με υπολογιζόμενο αριθμό clusters να ισούται με 3. Επίσης, για την δημιουργία του αντίστοιχου δενδρογράμματος χρησιμοποιείται η κλάση dendrogram από το πακέτο scipy.cluster.hierarchy. Στο Σχήμα 47 παρουσιάζεται το δενδροδιάγραμμα των περιοχών αναφοράς, όπου η κάθε περιοχή από το A έως το M αντιστοιχεί κατά σειρά σε έναν αριθμό από το 0 έως το 12 και στον Πίνακα 7 αναφέρεται ο αριθμός της κλάσης της κάθε περιοχής καθώς και η περιγραφή της.



Σχήμα 47: Δενδροδιάγραμμα περιοχών αναφοράς δείγματος CALLOS CER-3. Οι περιοχές A-M αντιστοιχούν στους αριθμούς 0-12

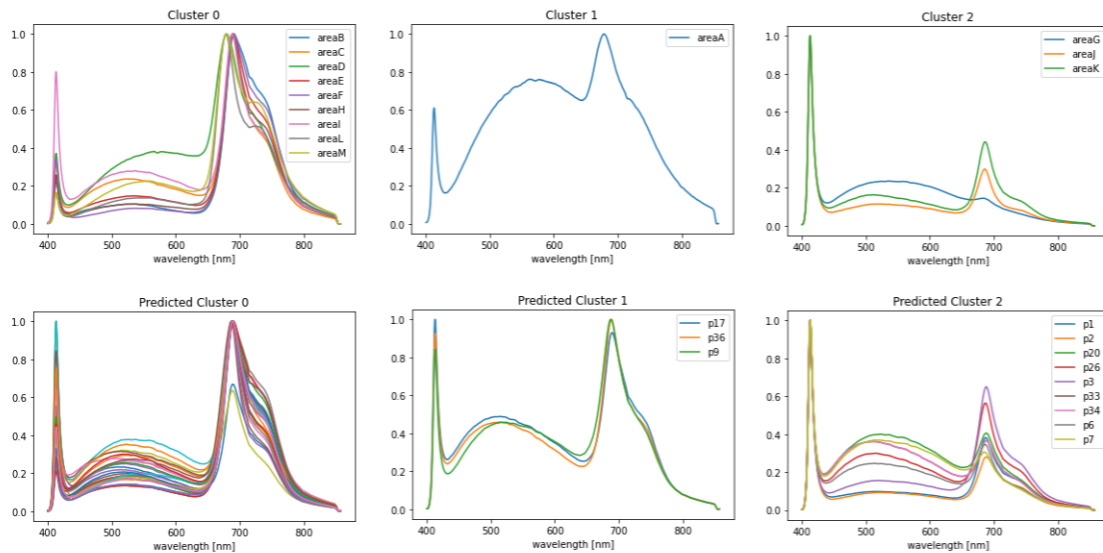
Περιοχή	Κλάση	Περιγραφή
B,C,D,E,F,H,I,L,M	0	Έντονη Βιολογική Επιβάρυνση Φωτοαυτότροφων μικροοργανισμών
A	1	Βιολογική Επιβάρυνση Φωτοαυτότροφων και Ετερότροφων μικροοργανισμών
G, J, K	2	Αμελητέα βιολογική επιβάρυνση - "Καθαρή" περιοχή

Πίνακας 7: Περιοχές Αναφοράς δείγματος CALLOS CER-3, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης

Στη συνέχεια, αφού οι κλάσεις του Πίνακα 7 τροφοδοτηθούν στην LDA, πραγματοποιείται η πρόβλεψη κλάσεων για την χαρτογράφηση. Τα αποτελέσματα παρουσιάζονται γραφικά στο Σχήμα 48. Παρατηρώντας οπτικά την ταξινόμηση, διαπιστώνεται η ομοιότητα των φασμάτων εκπαίδευσης με τα αντίστοιχα φάσματα πρόβλεψης.

### Αποδοτικότητα Μοντέλου LDA

Η αξιολόγηση της απόδοσης του μοντέλου της LDA δυνητικά θα πρέπει να πραγματοποιηθεί με χρήση της μεθόδου LOO στα δεδομένα των φασμάτων αναφοράς, όπως περιγράφηκε στην ενότητα 2.4. Καθώς όμως στο συγκεκριμένο δείγμα η κλάση 1 των δεδομένων αναφοράς περιέχει μόνο μια



Σχήμα 48: Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS CER-3

παρατήρηση, όταν αυτή αναχθεί σε test set, η μέθοδος LOO θα καταρεύσει αφού πλέον το training set θα περιέχει μόνο δυο κλάσεις και όχι τρεις. Συνεπώς, μόνο για το συγκεκριμένο δείγμα, η μέθοδος LOO εφαρμόστηκε στα δεδομένα χαρτογράφησης, προσδίδοντας σε κάθε παρατήρηση έναν "πραγματικό" αριθμό κλάσης ακολουθώντας την παρακάτω μεθοδολογία:

- Χρησιμοποιώντας τα δεδομένα του σετ εκπαίδευσης, δηλαδή τα φάσματα αναφοράς, υπολογίζονται οι λόγοι  $I_{430-600nm}/I_{412nm}$  και  $I_{650-800nm}/I_{412nm}$ , όπου  $I_{430-600nm}$  η μέγιστη ένταση φθορισμού στο εύρος 430-600 nm,  $I_{412nm}$  η ένταση της κορυφής του laser στα 412 nm και  $I_{650-800nm}$  η μέγιστη ένταση φθορισμού στο εύρος 650-800 nm. Οι λόγοι αυτοί αποτελούν χαρακτηριστικά μεγέθη της απορρόφησης του δείγματος σε σχέση με την κορυφή του laser στα 405 nm, καθώς ο λόγος οπισθοσκέδασης  $I_{405nm}/I_{412nm}$  διατηρείται σχεδόν σταθερός για κάθε μέτρηση. Επίσης, τα φάσματα αναφοράς μπορούν να διαχωριστούν σε κλάσεις και σύμφωνα με αυτούς τους λόγους, όπως περιγράφεται στη συνέχεια. Στον Πίνακα 8 παρουσιάζονται οι προαναφερόμενοι λόγοι καθώς και η διαφορά τους.

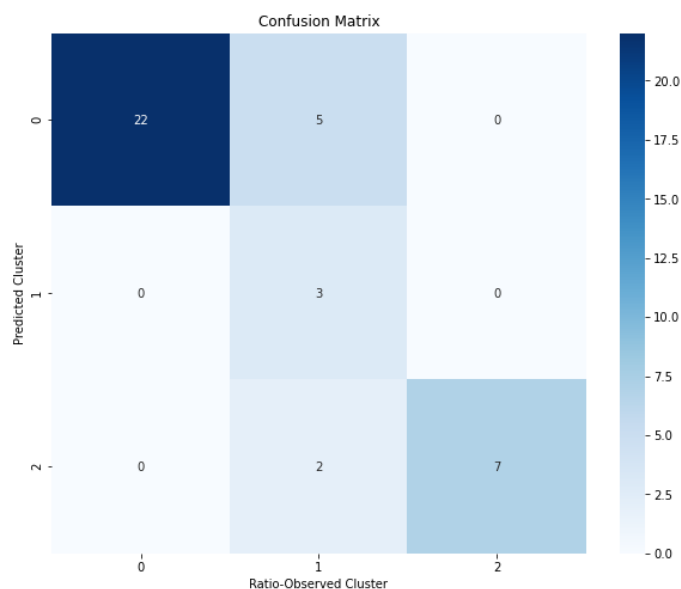
Περιοχές	$I_{430-600nm}/I_{412nm}$	$I_{650-800nm}/I_{412nm}$	Διαφορά Λόγων	Κλάση
areaA	1.25	1.64	0.39	1
areaB	0.42	4.02	3.60	0
areaC	0.74	3.11	2.37	0
areaD	1.03	2.70	1.67	0
areaE	0.57	3.89	3.32	0
areaF	0.24	2.96	2.72	0
areaG	0.24	0.15	0.09	2
areaH	0.45	4.26	3.82	0
areaI	0.35	1.25	0.90	0
areaJ	0.11	0.30	0.18	2
areaK	0.16	0.44	0.28	2
areaL	0.86	6.29	5.43	0
areaM	1.37	6.05	4.68	0

Πίνακας 8: Περιοχές αναφοράς, οι αντίστοιχοι λόγοι μέγιστης έντασης φθορισμού στα εύρη 430-600nm/412nm και 650-800nm/412nm και η διαφορά τους, καθώς και η κλάση στην οποία ανήκουν

- Από τον Πίνακα 8 παρατηρείται πως οι περιοχές οι οποίες έχουν διαφορά στους λόγους μεγαλύτερη του 0.9, τότε ανήκουν στην Κλάση 0. Εάν η διαφορά είναι μικρότερη του 0.9, τότε αν τόσο ο λόγος  $I_{430-600nm}/I_{412nm}$  όσο και ο λόγος  $I_{650-800nm}/I_{412nm}$  είναι μικρότεροι του 0.5 τότε η περιοχή ανήκει στην Κλάση 2 ενώ εάν έστω ένας από τους δύο λόγους είναι μεγαλύτερος του 0.5 τότε το φάσμα ανήκει στην Κλάση 1. Εάν αυτό το κριτήριο εφαρμοστεί στα φάσματα χαρτογράφησης, των οποίων οι κλάσεις δεν είναι γνωστές, τότε μπορεί να δημιουργηθεί μια δυνητικά πραγματική κλάση για καθένα από αυτά, η "Ψευδοπαρατηρούμενη Κλάση" του. Επομένως, κάθε σημείο χαρτογράφησης χαρακτηρίζεται πλέον από την προβλεπόμενη κλάση του μέσω του αλγορίθμου της LDA (*Predicted Class*) καθώς και από την "Ψευδοπαρατηρούμενη" κλάση του (*Ratio-Observed Class*). Με βάση αυτές τις δύο πληροφορίες για κάθε φάσμα δημιουργείται ο Confusion Matrix.

**Σχόλιο:** Συνήθως τα φάσματα είναι εξαιρετικά πιο σύνθετα και συνεπώς δυο λόγοι δεν αρκούν για την περιγραφή και την ταξινόμηση τους σε κλάσεις. Επομένως, η δημιουργία της "Ψευδοπαρατηρούμενης Κλάσης" δεν αποτελεί ένα δόκιμο τρόπο ελέγχου και γι αυτό δεν υιοθετείται και στα υπόλοιπα δείγματα. Αξιοποιήθηκε μόνο στο συγκεκριμένο δείγμα καθώς τα φάσματα αναφοράς επέτρεπαν τον διαχωρισμό σε κλάσεις με χρήση λόγων.

- Η αποδοτικότητα του μοντέλου προκύπτει με βάση τον υπολογισμό του Confusion Matrix, όπως περιγράφηκε στην ενότητα 2.4. Για το συγκεκριμένο δείγμα ο Confusion Matrix παρουσιάζεται στην εικόνα 49. Παρατηρείται πως οι περισσότερες αληθώς σωστές προβλέψεις (σκούρο μπλέ τετράγωνο) πραγματοποιήθηκαν για τον τύπο βιοδιάβρωσης από φωτοαυτότροφους μικροοργανισμούς (κλάση 0). Με χρήση των σχέσεων 74, 75, 76 και τον συγκεκριμένο πίνακα, υπολογίζονται για κάθε κλάση η ακρίβεια, η ευαισθησία και η ανάκληση και παρουσιάζονται στον Πίνακα 9. Η συνολική ακρίβεια του μοντέλου, δηλαδή οι συνολικές σωστές προβλέψεις ως προς το μέγεθος του δείγματος είναι 82%. Στη συνέχεια καταγράφονται οι παρατηρήσεις που προκύπτουν μελετώντας τον Πίνακα 9.



Σχήμα 49: *Confusion Matrix* μεταξύ προβλεπόμενης και "ψευδοπαρατηρούμενης" κλάσης των φασμάτων της χαρτογράφησης του CALLOS CER-3

Cluster	ACCURACY	PRECISION	RECALL
0	0.87	0.81	1.0
1	0.82	1.00	0.3
2	0.95	0.78	1.0

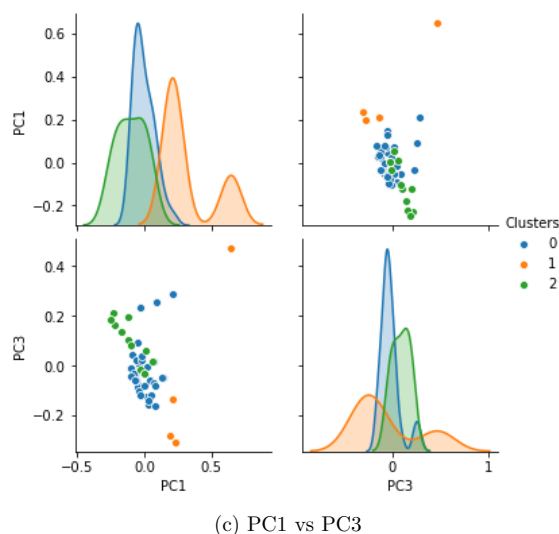
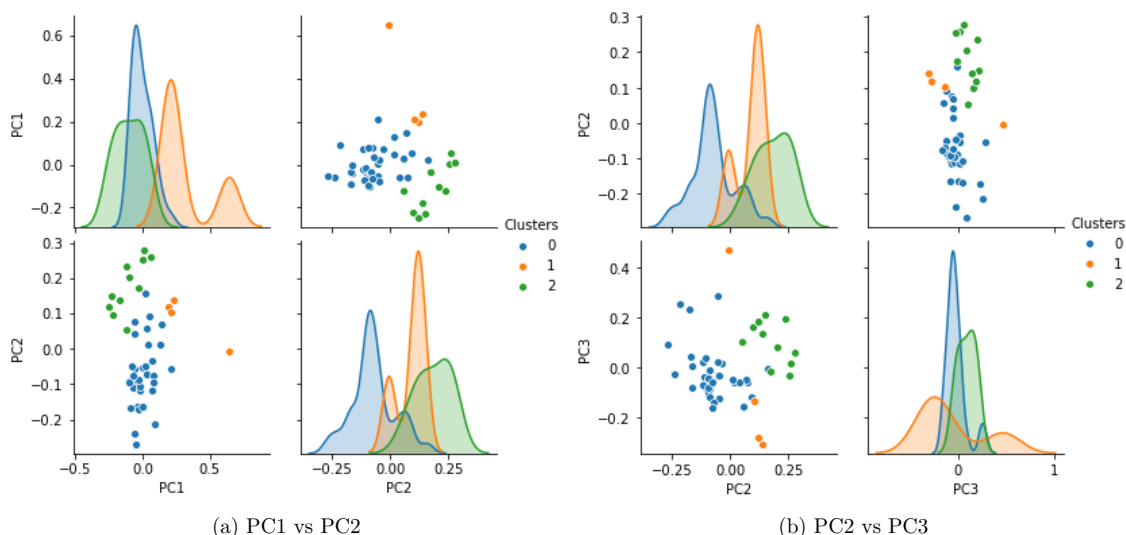
Πίνακας 9: Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση των μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS CER-3

### Παρατηρήσεις

1. Οι κλάσεις 0 και 2 παρουσιάζουν 100% ανάκληση, δηλαδή το μοντέλο προέβλεψε σωστά όλα τα φάσματα που ανήκουν πραγματικά σε αυτές τις κλάσεις. Αυτό συμβαίνει διότι και οι δύο κλάσεις δημιουργούνται από φάσματα αναφοράς με ξεκάθαρες και περιορισμένες κορυφές οι οποίες βρίσκονται και απομακρυσμένες μεταξύ τους στα 690 nm και στα 412 nm. Επίσης σε αυτές τις δύο κλάσεις είναι αρκετά υψηλή και η ακρίβεια (87% και 95%) εξαιτίας την πληθώρας των παρατηρήσεων που ανήκουν σε αυτές τις κλάσεις ως training set. Συνεπώς, οι κλάσεις 0 και 2 διαχωρίζονται αποτελεσματικά από το μοντέλο και άρα μια "καθαρή" περιοχή σε σχέση με μια βιολογικά επιβαρυσμένη από φωτοαυτότροφους μικροοργανισμούς είναι διαχωρίσιμη.
2. Η κλάση 1, της οποίας τα φάσματα αναφοράς παρουσιάζουν μεγαλύτερη πολυπλοκότητα συγκριτικά με τις άλλες δυο, παρουσιάζει σημαντικά μικρή ανάκληση καθώς το μοντέλο κατόρθωσε να προβλέψει σωστά μόνο το 30% των παρατηρήσεων που ανήκουν πραγματικά στην κλάση. Αυτό πιθανώς συμβαίνει καθώς σε αυτή την κλάση συμμετέχουν και οι δύο κορυφές στα 690 nm και στα 412 nm, οι οποίες είναι χαρακτηριστικές των κλάσεων 0 και 2. Συνεπώς, μια περιοχή με πιθανή συνύπαρξη ετερότροφων και φωτοαυτότροφων μικροοργανισμών διαχωρίζεται με μεγαλύτερη δυσκολία από μια περιοχή με την επικράτηση φωτοαυτότροφων. Παρόλα αυτά η ευαισθησία της πρόβλεψης ήταν στο 100%, δηλαδή όσα προβλέφθηκαν ότι ανήκουν στην κλάση 1, πράγματι ανήκουν στην κλάση 1. Αυτό σημαίνει πως το έντονο σήμα στο εύρος 440-650 nm αποτελεί ισχυρό στατιστικό χαρακτηριστικό της συγκεκριμένης κλάσης. Επίσης, το μοντέλο δεν πραγματοποίησε πολλές προβλέψεις για την κλάση 1 και επομένως οι σωστές προβλέψεις συγκριτικά με τις συνολικές της κλάσης αυτής δίνουν μια υψηλή ακρίβεια στο 82%.

### Συνδυαστική Μελέτη LDA - PCA

Στα score plots του Γραφήματος 46 σε συνδυασμό με τα loading plots του Γραφήματος 45 δεν ήταν ευδιακριτός ο διαχωρισμός των δεδομένων σε κλάσεις με χρήση μόνο του αλγορίθμου της PCA. Για τον λόγο αυτό, χρησιμοποιήθηκε η μέθοδος LDA, ώστε με ελάχιστη εξωγενή βιολογική πληροφορία να κατασκευαστεί ένα μοντέλο πρόβλεψης των κλάσεων που πιθανώς ανήκουν τα δεδομένα χαρτογράφησης του δείγματος. Στο Γράφημα 50 παρουσιάζονται τα score plots του Γραφήματος αξιοποιώντας όμως τις πληροφορίες για τις κλάσεις που προκύπτουν από την εφαρμογή του αλγορίθμου της LDA. Πάνω ή κάτω από κάθε score plot παρουσιάζεται η αντίστοιχη συνάρτηση πυκνότητας των κλάσεων κατά μήκος του άξονα του principal component της συγκεκριμένης στήλης.



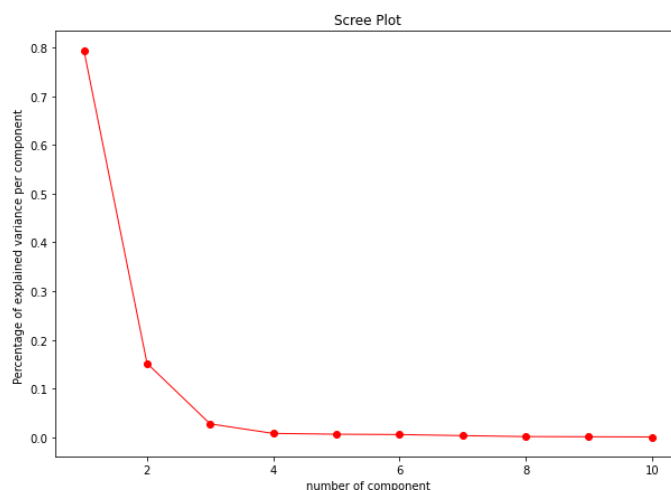
Σχήμα 50: Score Plots δεδομένων δείγματος CALLOS CER-3 στα οποία η χρωματική κλίμακα αντιστοιχεί στις κλάσεις που προβλέφθηκαν από τον αλγόριθμο της LDA. Στις διαγωνίους παρουσιάζεται η συνάρτηση πυκνότητας της κάθε κλάσης σύμφωνα με τις τιμές του component της αντίστοιχης στήλης

Τόσο από το Γράφημα 50(a) όσο και από το 50(b), διαπιστώνεται όπως και προηγουμένως, πως μόνο οι τιμές του PC2 μπορούν να διαχωρίσουν τις κλάσεις 0 και 2 με αλληλοεπικάλυψη, όμως, στο εύρος τιμών  $[-0.1, 0.25]$  του PC2. Στα υπόλοιπα principal components οι αλληλοεπικαλύψεις των κλάσεων είναι ιδιαίτερα έντονες και γι αυτόν τον λόγο η PCA δεν είναι αποτελεσματική. Με την χρήση της LDA σε συνδυασμό με την κλάσεις των φασμάτων αναφοράς είναι εφικτή η κατηγοριοποίηση των φασμάτων χαρτογράφησης και υπερπήδηση του εμποδίου της αλληλοεπικάλυψης. Το Γράφημα 50 δημιουργήθηκε με χρήση του πακέτου seaborn μέσω της συνάρτησης pairplot.

## 4.2 CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

### Principal Component Analysis

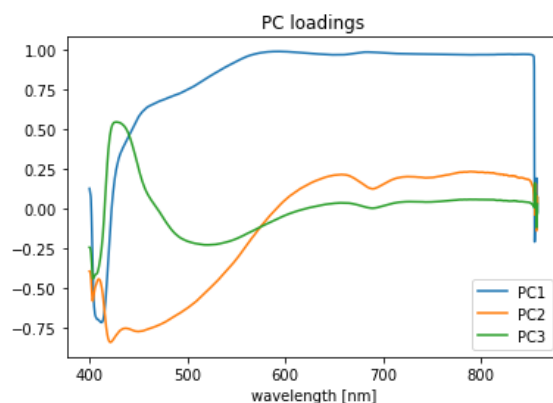
Η εφαρμογή της μεθόδου μείωσης των διαστάσεων σε αυτό το δείγμα πραγματοποιείται σε έναν πίνακα δεδομένων με διαστάσεις  $54 \times 1688$ . Ο συγκεκριμένος πίνακας περιλαμβάνει τα τυποποιημένα φάσματα των 19 περιοχών αναφοράς του δείγματος CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1, όπως αυτές περιγράφονται στην ενότητα 3.3 καθώς επίσης και τα φάσματα της αντίστοιχης χαρτογράφησης. Μέσω της PCA, οι 1688 μεταβλητές μειώνονται σε 3 principal components όπως προσδιορίζει η κάμψη του κρυμνογράφηματος στο σχήμα 51. Σε αυτά τα τρία principal components συμπυκνώνεται το 97.3% της επεξηγούμενης διακύμανσης και επομένως η απώλεια πληροφορίας είναι αμελητέα.



Σχήμα 51: Κρυμνογράφημα τυποποιημένων δεδομένων του δείγματος CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

Τα φορτία (loadings) των principal components παρουσιάζονται στο σχήμα 52 και όπως θα εξηγηθεί στη συνέχεια, κανένα από τα 3 principal components δύναται να συσχετιστεί με την παρουσία μιας συγκεκριμένης κλάσης φασμάτων αναφοράς, όπως αυτές παρουσιάζονται στα γραφήματα 32,33, 34, 35 και 36. Επιπλέον, δεν δημιουργείται κάποια εγγενής ομαδοποίηση των δεδομένων με βάση κάποιο principal component,όσον αφορά καποιον συγκεκριμένο τύπο βιοδιάβρωσης.

Ειδικότερα, το PC1 συσχετίζεται έντονα ( $> 0.75$ ) με τις μεταβλητές στα εύρη 400-420 nm, 500-850 nm, στα οποία εμφανίζεται έντονο φάσμα φθορισμού για όλες τις κλάσεις αναφοράς. Επομένως τα scores του PC1 δεν μπορούν να αντιστοιχηθούν με μια συγκεκριμένη κλάση. Το PC2 εμφανίζει σημαντική συσχέτιση ( $> 0.5$ ) με τις μεταβλητές στο εύρος 420-520 nm, το οποίο δεν αντιστοιχεί μεμονωμένα σε καμία κλάση φασμάτων. Τέλος, το PC3 συσχετίζεται με το εύρος μεταβλητών 420-480 nm, όπου αντίστοιχα με το PC2, δεν υπάρχει καμία κλάση που να εμφανίζει ένταση φθορισμού μεμονωμένα σε αυτά τα μήκη κύματος. Επομένως, όπως και στο CALLOS CER-3, η χρήση επιπρόσθετων μεθόδων ταξινόμησης των δεδομένων κρίνεται αναγκαία.



Σχήμα 52: Τα φορτία των κύριων συνιστωσών τα οποία προκύπτουν από τον πίνακα δεδομένων του δείγματος CALLOS MAR 2 - ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

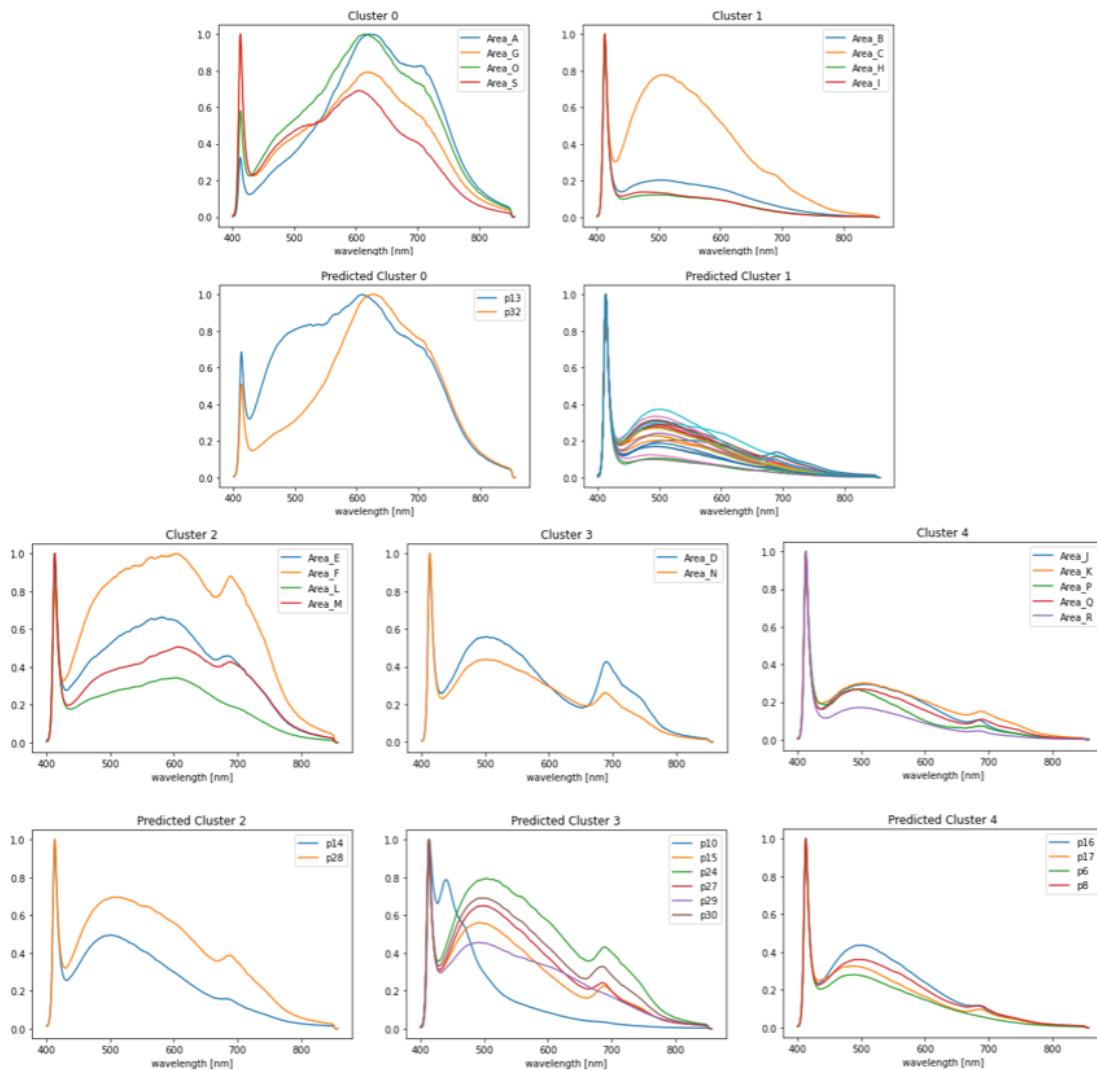
### Linear Discriminant Analysis

Η εφαρμογή του μοντέλου πρόβλεψης της LDA στα δεδομένα της χαρτογράφησης απαιτεί την εκπαίδευση του αλγορίθμου με τις κατάλληλες κλάσεις αναφοράς. Σε αυτό το δείγμα δεν αξιοποιήθηκε ο αλγόριθμος της ομαδοποίησης κατά συστάδες καθώς οι βιολογικές κλάσεις όπως παρουσιάστηκαν στην ενότητα 3.3 ήταν χαρακτηριστικές των περιοχών μέτρησης. Ειδικότερα, στον Πίνακα 10 παρουσιάζονται ομαδοποιημένες οι περιοχές αναφοράς των οποίων τα φάσματα παρατίθενται στα Γραφήματα 32, 33, 34, 35, 36. Επιπλέον προσδιορίζεται ο χαρακτηριστικός αριθμός της κλάσης καθώς και η περιγραφή της.

Περιοχή	Κλάση	Περιγραφή
A,G,O,S	0	Βιολογική Επιβάρυνση Φωτοαυτότροφων μικροοργανισμών - Πιθανή συμβίωση με ετερότροφους
B,C,H,I	1	Αμελητέα βιολογική επιβάρυνση - "Καθαρή" περιοχή
E,F,L,M	2	Μαύρα σε συνδυασμό με λευκά στίγματα
D,N	3	Μαύρα σε συνδυασμό με κιτρινωπά στίγματα - Παρουσία χλωροφύλλης
J,K,P,Q,R	4	Οπτική ανομοιομορφία περιοχών - Ασθενής βιολογική Επιβάρυνση

Πίνακας 10: Περιοχές Αναφοράς δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης

Αφότου η πληροφορία του αριθμού της κλάσης αναφοράς του κάθε φάσματος τροφοδοτηθεί στην LDA, πραγματοποιείται η πρόβλεψη του αριθμού της κλάσης του κάθε φάσματος χαρτογράφησης. Τα αποτελέσματα παρουσιάζονται γραφικά στο σχήμα 53. Παρατηρείται οπτικά πως η πιο αξιόπιστη ταξινόμηση πραγματοποιήθηκε στην κλάση 1, δηλαδή στην κλάση της καθαρής περιοχής όπως επίσης και στην κλάση 4, όπου οι περιοχές ήταν ανομοιομορφες αλλά η βιολογική επιβάρυνση ήταν επίσης αμελητέα. Επιπλέον παρατηρείται πως στη προβλεπόμενη κλάση 3, το φάσμα του p10 (γαλάζια γραμμή), το οποίο παρουσιάζει έντονη κορυφή στα 440 nm, δεν αντιστοιχεί έστω και προσεγγιστικά σε κάποιο φάσμα από τις κλάσεις αναφοράς. Επομένως, ο αλγόριθμος το ταξινόμησε στην κλάση όπου μεγιστοποιήθηκε η πιθανότητα να αντιστοιχεί, χωρίς όμως αυτό να ανταποκρίνεται στην πραγματικότητα. Παρόλα αυτά, σε κάποιες περιπτώσεις όπου η κορυφές των φασμάτων χαρτογράφησης ήταν αρκετά ισχυρές και υπήρχαν οι αντίστοιχες στα φάσματα αναφοράς, η ταξινόμηση ήταν αποτελεσματική. Χαρακτηριστικό παράδειγμα αποτελεί η κλάση 0 της πιθανής συνύπαρξης φωτοαυτότροφων με ετερότροφους και η αντίστοιχη κορυφή στα 620 nm. Επίσης, η κλάση 3 των μαύρων-κιτρινωπών σιγμάτων με τις κορυφές στα 520-530 nm και 690 nm πραγματοποίησε επαρκή ταξινόμηση.

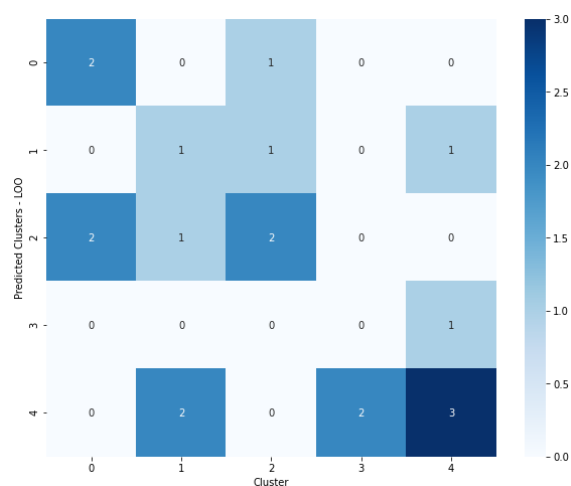


Σχήμα 53: Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

Ωστόσο τα παραπάνω συμπεράσματα είναι ποιοτικά και επομένως θα πρέπει να αξιοποιηθεί η μέθοδος LOO, όπως παρουσιάστηκε στην ενότητα 2.4 ώστε να προκύψει η ποσοτική αξιολόγηση του μοντέλου με βάση το σετ δεδομένων εκπαίδευσης. Ειδικότερα, εφαρμόζοντας την μέθοδο LOO στα φάσματα αναφοράς, των οποίων οι κλάσεις είναι γνωστές, προκύπτει η προβλεπόμενη κλάση για κάθε ένα από αυτά. Με βάση την πραγματική κλάση (Cluster) και την προβλεπόμενη (Predicted Cluster -LOO) δημιουργείται ο Confusion Matrix, όπως παρουσιάστηκε στην ενότητα 2.4. Επίσης, αξιοποιώντας τον Confusion Matrix καθώς και τις σχέσεις 74, 75, 76 υπολογίζεται η ακρίβεια, η ευαισθησία και η ανάκληση της κάθε κλάσης και παρουσιάζονται στον Πίνακα 11.

Οι συνολικές προβλέψεις του μοντέλου σύμφωνα με την μέθοδο LOO είναι 19 και αντιστοιχούν στον αριθμό των παρατηρήσεων του σετ εκπαίδευσης. Το ποσοστό των σωστών προβλέψεων ως προς τις συνολικές, δηλαδή η συνολική ακρίβεια του μοντέλου ανέρχεται στο 43%, το οποίο δεν είναι αρκετά ικανοποιητικό. Επιπλέον σε αρκετές από τις κλάσεις παρατηρούνται σημαντικά χαμηλά ποσοστά ευαισθησίας και ανάκλησης. Οι παρατηρήσεις παρουσιάζονται στη συνέχεια.





Σχήμα 54: Confusion Matrix μεταξύ προβλεπόμενης και πραγματικής κλάσης των φασμάτων της αναφοράς του CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 με χρήση της μεθόδου Leave-One-Out Cross Validation (LOO)

Cluster	ACCURACY	PRECISION	RECALL
0	0.84	0.67	0.50
1	0.74	0.33	0.25
2	0.74	0.40	0.50
3	0.84	0.00	0.00
4	0.68	0.43	0.60

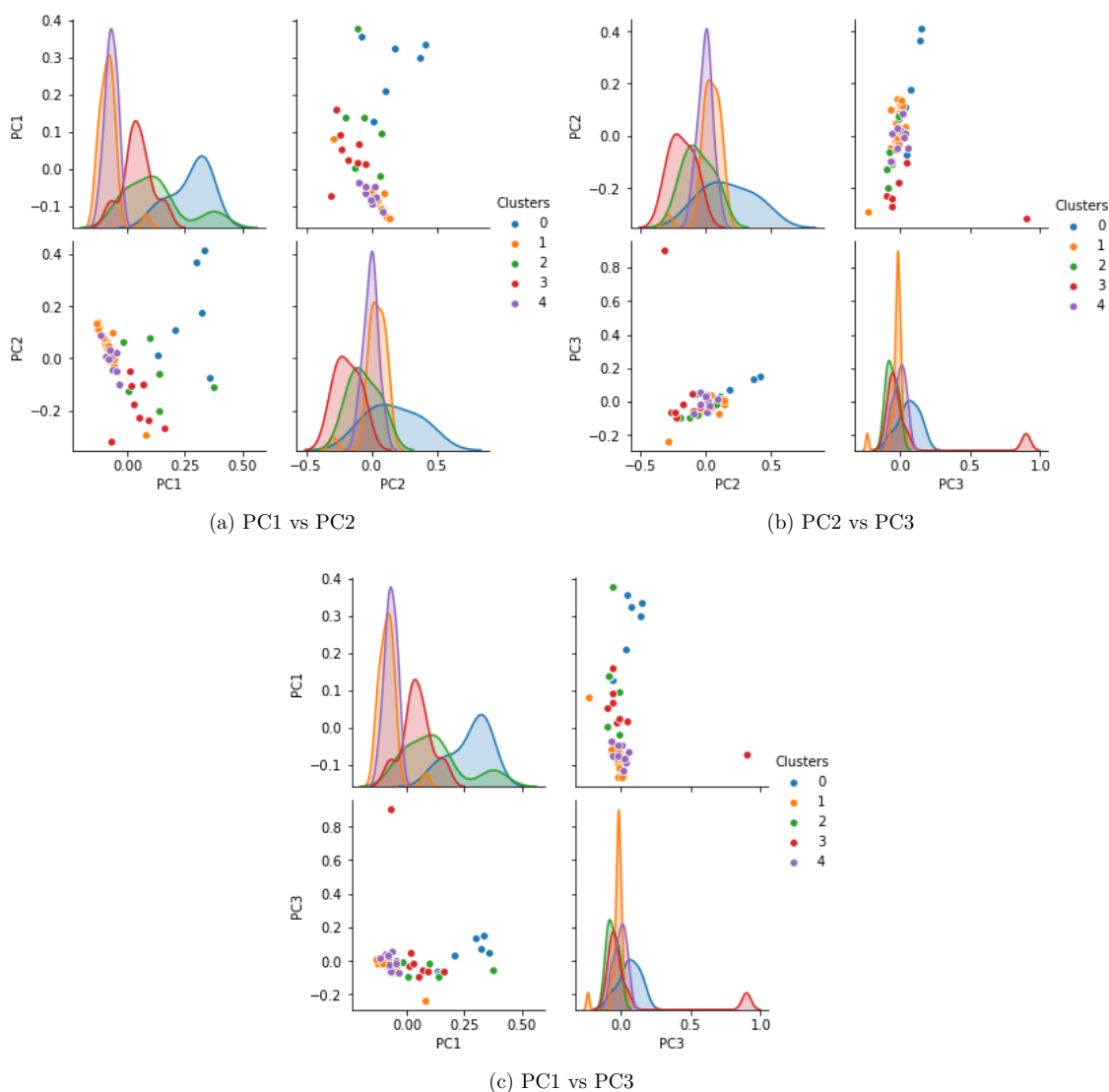
Πίνακας 11: Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση το μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1

## Παρατηρήσεις

1. Η κλάση 3, η οποία αντιστοιχεί σε μαύρα στίγματα παρουσία χλωροφύλλης, ενώ χαρακτηρίζεται από 84% ακρίβεια σε θετικές και αρνητικές προβλέψεις, τα ποσοστά ευαισθησίας και ανάκλησης ήταν μηδενικά. Πιο συγκεκριμένα, εξετάζοντας την σχέση 75, προκύπτει πως όλες οι προβλέψεις που έδωσαν τον αριθμό της κλάσης ήταν λανθασμένες (TP = 0). Επομένως ήταν αναμενόμενο και από τη σχέση 76, κανένα από τα φάσματα της κλάσης να μην επιβεβαιώθηκε ότι ανήκει πράγματι σε αυτή. Αυτό συμβαίνει διότι και τα δύο σετ χαρακτηριστικών κορυφών της κλάσης στα εύρη 460-530 nm και 680-700 nm εμφανίζονται και σε άλλες κλάσεις όπως η 2 και η 4. Επομένως, όταν η LOO αφαιρεί το ένα από τα δύο φάσματα της 3, η σωστή ταξινόμηση του άλλου καθίσταται πιο σύνθετη. Επομένως, η κλάση 3 απαιτεί περισσότερες παρατηρήσεις ώστε τα μαύρα στίγματα παρουσία χλωροφύλλης να είναι στατιστικά ευδιάκριτα.
2. Η ευαισθησία του μοντέλου στην παραγωγή σωστών προβλέψεων σε κάθε κλάση είναι μικρή, ειδικά στις περιπτώσεις των κλάσεων 1 και 4, οι οποίες αντιστοιχούν σε ασθενή βιολογική επιβάρυνση αλλά η οπτική τους ανομοιομορφία χρησιμοποιήθηκε για να τις διαχωρίσει. Πιο συγκεκριμένα, αυτό συμβαίνει διότι οι δύο κλάσεις παρουσιάζουν μια σχετική ομοιομορφία, όπως παρουσιάζεται στο Σχήμα 53 και επομένως η πιθανότητα ταξινόμησης μια παρατήρησης της κλάσης 1 στην κλάση 4 και αντιστρόφως, αυξάνεται σημαντικά. Συνεπώς, η σύμπτυξη των δύο κλάσεων θα μπορούσε φαινομενικά να επιλύσει το πρόβλημα. Για να είναι όμως σωστή η συγκεκριμένη ταξινόμηση, θα πρέπει να επιβεβαιωθεί και βιολογικά.

## Συνδυαστική Μελέτη LDA - PCA

Όπως διαπιστώθηκε κατά την μελέτη του CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 με χρήση της PCA, κανένα από τα τρία principal components δεν ήταν ικανό να διαχωρίσει αποτελεσματικά κάποια από τις κλάσεις αναφοράς. Για τον λόγο αυτό αξιοποιήθηκε το μοντέλο πρόβλεψης της LDA προκειμένου κάθε φάσμα χαρτογράφησης να αποκτήσει μια ταυτότητα ως προς την κλάση αναφοράς στην οποία ανήκει. Το Γράφημα 55 παρουσιάζει τα score plots της PCA για τον συνολικό πίνακα δεδομένων του δείγματος και η χρωματική διαφοροποίηση αντιστοιχεί στις προβλεπόμενες κλάσεις δεδομένων από την LDA. Επιβεβαιώνεται, λοιπόν πως τα φάσματα δεν θα μπορούσαν να διαχωριστούν με την PCA, καθώς ακόμη και τα δεδομένα των κλάσεων αναφοράς παρουσιάζουν ισχυρές αλληλοεπικαλύψεις εξαιτίας των πολλαπλών κοινών κορυφών φθορισμού που περιέχουν.



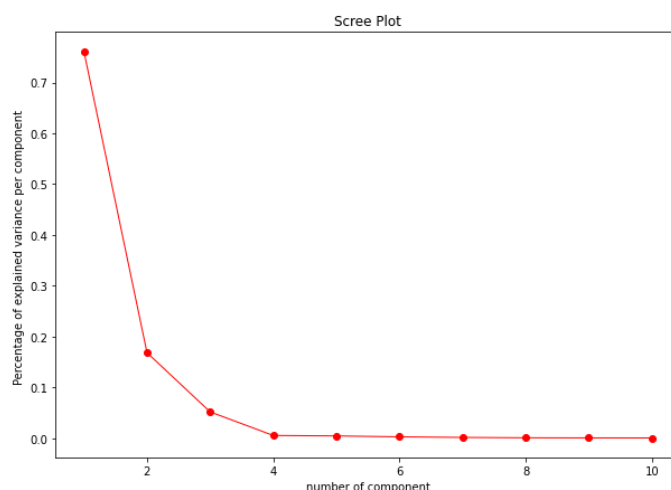
Σχήμα 55: Score Plots δεδομένων δείγματος CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 στα οποία η χρωματική κλίμακα αντιστοιχεί στις κλάσεις που προβλέφθηκαν από τον αλγόριθμο της LDA. Στις διαγωνίους παρουσιάζεται η συνάρτηση πυκνότητας της κάθε κλάσης σύμφωνα με τις τιμές του component της αντίστοιχης στήλης

## 4.3 CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

### 4.3.1 Βιολογική Κατανομή Κλάσεων

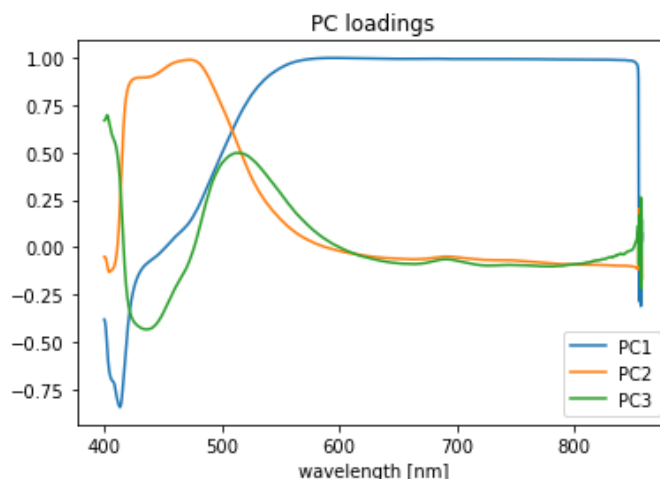
#### Principal Component Analysis

Η συγκεκριμένη επιφάνεια του δείγματος CALLOS MAR-2 αντιστοιχεί σε έναν τυποποιημένο πίνακα δεδομένων με διαστάσεις  $66 \times 1688$ , ο οποίος μετά την εφαρμογή της PCA μειώνεται σημαντικά σε  $66 \times 3$ . Οι 66 παρατηρήσεις αντιστοιχούν σε 15 φάσματα αναφοράς και σε 51 φάσματα χαρτογράφησης. Η μείωση των διαστάσεων αξιοποίησε το 98.1% της επεξηγούμενης διακύμανσης των δεδομένων και ο κατάλληλος αριθμός components είναι τα τρία, όπως παρουσιάζεται και στο κρυμνογράφημα του σχήματος 56.



Σχήμα 56: Κρυμνογράφημα τυποποιημένων δεδομένων του δείγματος CALLOS MAR 2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

Η μελέτη των φορτίων (loadings) των principal components παρέχουν την επεξήγηση της σύνδεσης των αρχικών μεταβλητών με τα principal components και δύναται να αναδείξουν πιθανές κατηγοριοποιήσεις. Πιο συγκεκριμένα, το PC1 παρουσιάζει έντονη συσχέτιση ( $> 0.5$ ) με τα εύρη μεταβλητών 400-420 nm και 500-850 nm στα οποία υφίσταται σήμα φθορισμού σε κάθε κλάση φασμάτων αναφοράς, όπως παρουσιάζονται στα γραφήματα 38, 39, 40, 41. Επομένως, το PC1 δεν ταυτίζεται με την ύπαρξη κάποιας συγκεκριμένης κλάσης φασμάτων. Το PC2 επεξηγεί κυρίως την διακύμανση των μεταβλητών στο εύρος 430-500 nm, όπου στο εύρος αυτός παρουσιάζεται έντονο σήμα φθορισμού στις κλάσεις 1 και 3, οι οποίες συνδέονται με αμελητέα βιολογική επιβάρυνση και μάυρα στίγματα ανίστοιχα. Μια πιθανή ερμηνεία της συσχέτισης αυτών των δύο κλάσεων στο συγκεκριμένο εύρος είναι η μη αποτελεσματική διέγερση των μαύρων χρωμοφόρων ουσιών των μαύρων μεριστεματικών μυκήτων, καθώς λειτουργούν φωτοπροστατευτικά [61], με αποτέλεσμα το φάσμα τους να συμπίπτει με το φάσμα του υποστρώματος. Επομένως και πάλι θα ήταν λανθασμένο να συσχετιστεί το PC2 με αυτές τις κλάσεις καθώς μεταξύ τους διαφέρουν βιολογικά. Τέλος και το PC3 αποκλείεται από την ταύτιση με κάποιο συγκεκριμένο φάσμα αναφοράς καθώς συσχετίζεται με τα εύρη 420-500 nm και 500-600 nm στα οποία εμφανίζεται σήμα φθορισμού σε κάθε γνωστό φάσμα από τις προαναφερόμενες κλάσεις. Επομένως και σε αυτό δείγμα θα πρέπει να εφαρμοστεί ένα μοντέλο πρόβλεψης της κλάσης του κάθε φάσματος χαρτογράφησης.



Σχήμα 57: Τα φορτία των κύριων συνιστωσών τα οποία προκύπτουν από τον πίνακα δεδομένων του δείγματος CALLOS MAR 2 - ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

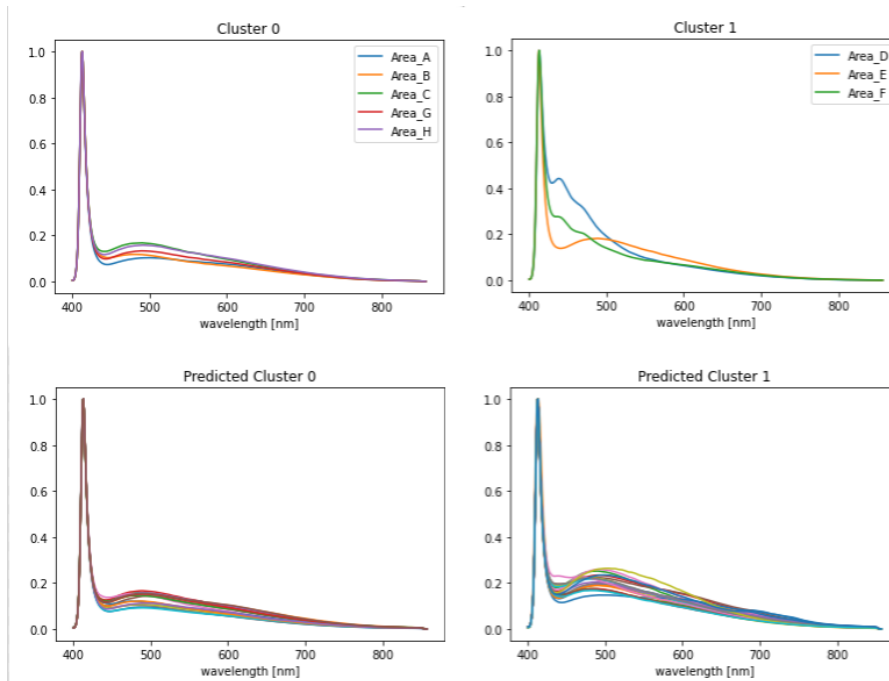
### Linear Discriminant Analysis

Το μοντέλο πρόβλεψης της κλάσης αναφοράς για το κάθε φάσμα χαρτογράφησης βασίζεται στην εκπαίδευση του αλγορίθμου της LDA μέσω των φασμάτων των κλάσεων αναφοράς, όπως αυτές παρουσιάζονται στα γραφήματα 38, 39, 40, 41. Αντίστοιχα με την ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 του δείγματος CALLOS MAR-2, δεν θα χρησιμοποιηθεί ο αλγόριθμος ομαδοποίησης κατά συστάδες καθώς η βιολογική και φασματική πληροφορία είναι συγκεντρωμένη στις κλάσεις αναφοράς, όπως αυτές δημιουργήθηκαν από την ανάλυση της ενότητας 3.3. Στον Πίνακα 12 παρουσιάζεται η ομαδοποίηση των περιοχών αναφοράς, ο αριθμός της αντίστοιχης κλάσης που τροφοδοτείται στον αλγόριθμο της LDA, καθώς επίσης και η οπτική-βιολογική περιγραφή της.

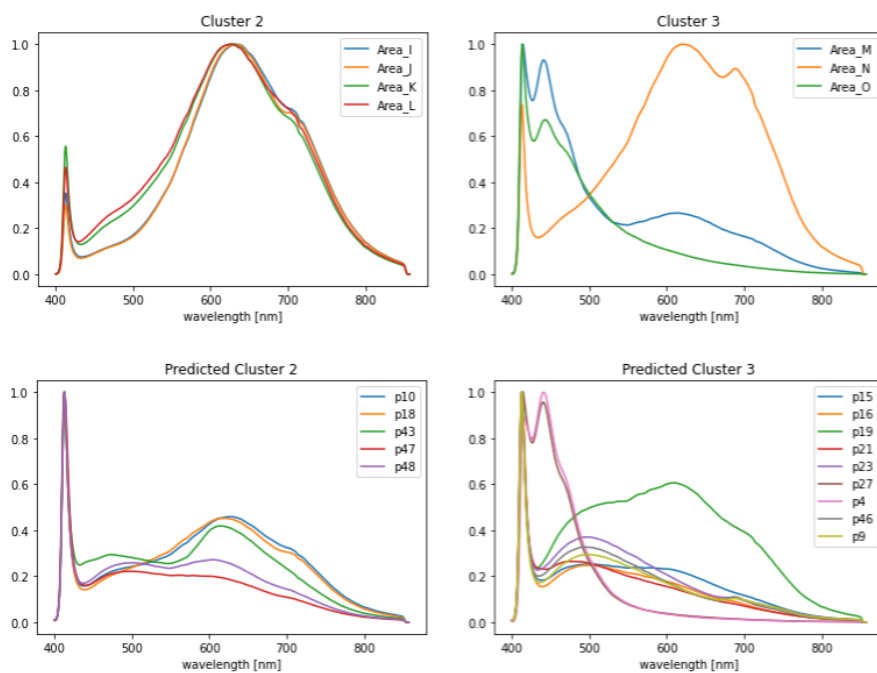
Περιοχή	Κλάση	Περιγραφή
A,B,C,G, H	0	Αμελητέα βιολογική επιβάρυνση - "Καθαρή" περιοχή
D,E,F	1	Αμελητέα βιολογική επιβάρυνση - Αποφλοιωμένη "Καθαρή" περιοχή
I,J,K,L	2	Πράσινα στίγματα
M,N,O	3	Μαύρα στίγματα - σε κάποια παρουσία χλωροφύλλης

Πίνακας 12: Περιοχές Αναφοράς δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης

Η εφαρμογή του μοντέλου στα δεδομένα χαρτογράφησης παρουσιάζεται γραφικά στο Σχήμα 58, όπου παρατίθενται συγκριτικά τα φάσματα των κλάσεων αναφοράς και τα φάσματα χαρτογράφησης στις προβλεπόμενες κλάσεις. Ορισμένα ποιοτικά συμπεράσματα προκύπτουν μέσω της οπτικής παρατήρησης της ταξινόμησης. Αντίστοιχα με το δείγμα CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1, η κλάση 0 η οποία αναπαριστά την καθαρή περιοχή, ταξινομείται αποτελεσματικά όπως επίσης και η κλάση 1 η οποία αντιστοιχεί σε αμελητέα βιολογική επιβάρυνση αποφλοιωμένης επιφάνειας. Η αποτελεσματικότητα της ταξινόμησης μειώνεται στις κλάσεις 2 και 3, οι οποίες περιγράφουν τα πράσινα και τα μαύρα στίγματα, αλλά διατηρούνται εν μέρη οι βασικές κορυφές φθορισμού στα 630 nm, 690 nm (κλάση 2), 440 nm και 470 nm (κλάση 3).



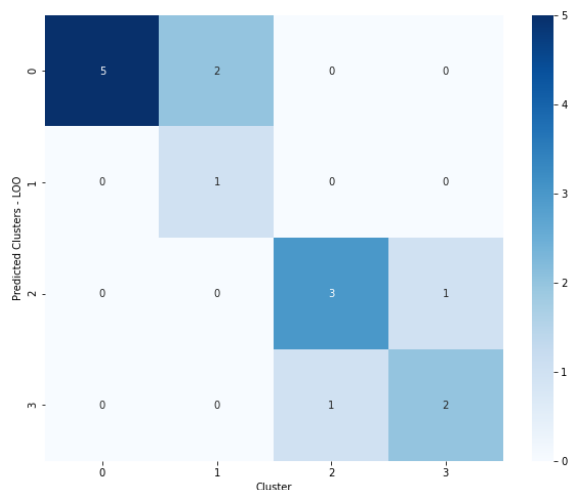
(a)



(b)

Σχήμα 58: Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

Η ποσοτική αξιολόγηση του μοντέλου της LDA για το δείγμα CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 προκύπτει μέσω της εφαρμογής της μεθόδου LOO, αντίστοιχα με το προηγούμενο δείγμα, στα φάσματα του συνόλου εκπαίδευσης. Συνεπώς, σε κάθε φάσμα των κλάσεων αναφοράς αντιστοιχίζεται η πραγματική και η προβλεπόμενη από την LDA κλάση του. Με άξονα αυτές τις προβλέψεις κατασκευάζεται ο Confusion Matrix του δείγματος, ο οποίος παρουσιάζεται στο σχήμα 59 και επιπλέον, μέσω των σχέσεων 74, 75, 76, υπολογίζεται η ακρίβεια, η ευαισθησία και η ανάκληση της κάθε κλάσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 13. Η συνολική ακρίβεια του μοντέλου ανέρχεται στα 73.3%. Εξετάζοντας τον Πίνακα 13 προκύπτουν οι παρακάτω παρατηρήσεις:



Σχήμα 59: Confusion Matrix μεταξύ προβλεπόμενης και πραγματικής κλάσης των φασμάτων της αναφοράς του CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 με χρήση της μεθόδου Leave-One-Out Cross Validation (LOO)

Cluster	ACCURACY	PRECISION	RECALL
0	0.87	0.71	1.00
1	0.87	1.00	0.33
2	0.87	0.75	0.75
3	0.87	0.67	0.67

Πίνακας 13: Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση των μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

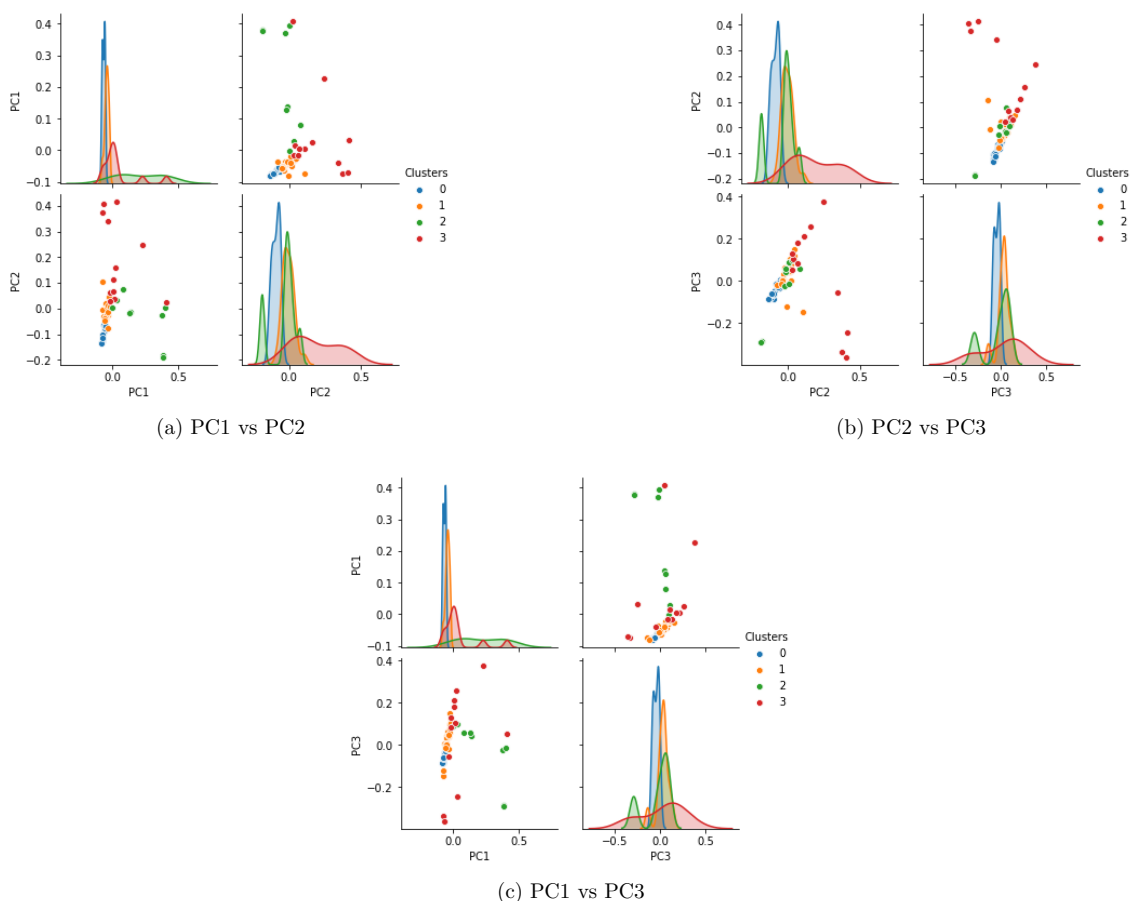
## Παρατηρήσεις

1. Οι κλάσεις 0 και 1, οι οποίες αντιστοιχούν σε περιοχές με μικρή βιολογική επιβάρυνση, παρουσιάζουν υψηλή ευαισθησία με αποτέλεσμα τα φάσματα που ταξινομούνται σε αυτές τις κλάσεις να είναι εύκολα διακριτά από το μοντέλο. Ωστόσο, η ανάκληση της κλάσης 1 είναι μικρή (33%) και επομένως μόνο ένα από τα τρία φάσματα που ανήκουν σε αυτήν επιβεβαιώθηκε. Αυτό μπορεί να βελτιωθεί αυξάνοντας τον αριθμό των παρατηρήσεων της κλάσης. Συμπερασματικά, όμως, οι καθαρές περιοχές διακρίνονται αποτελεσματικά από το μοντέλο.
2. Η κλάση 2, η οποία αντιστοιχεί στα πράσινα στίγματα, παρουσιάζει τόσο υψηλή ανάκληση όσο και ευαισθησία εξαιτίας της υψηλής ομοιομορφίας των φασμάτων της, όπως παρουσιάζεται στο σχήμα 58. Αριθμητικά, το 75% των παρατηρήσεων επιβεβαιώνεται πως πράγματι ανήκει σε αυτή την κλάση και επίσης το 75% των προβλέψεων της κλάσης είναι σωστές. Επομένως, τα πράσινα στίγματα είναι στατιστικά ευδιάκριτα.

3. Η κλάση 3, η οποία αντιστοιχίζεται στα μαύρα στίγματα, παρουσιάζει μειωμένη αποτελεσματικότητα στην πρόβλεψη συγκριτικά με τις άλλες κλάσεις. Αυτό συμβαίνει καθώς ορισμένα από τα μαύρα στίγματα παρουσιάζουν και κορυφές χλωροφύλλης με αποτέλεσμα να δημιουργούνται λανθασμένες προβλέψεις της κλάσης 3 στην 2. Αριθμητικά αυτό υποδηλώνεται στην 67% ευαισθησία, δηλαδή στην ικανότητα του μοντέλου να παράγει σωστές προβλέψεις της κλάσης 3 στο σύνολο των προβλέψεων για την ίδια κλάση. Επομένως, τα μαύρα στίγματα θα πρέπει να διαχωριστούν στους μεμονωμένους μύκητες και σε αυτούς που συνυπάρχουν με φωτοαυτότροφους μικροοργανισμούς. Ένα παράδειγμα αυτού του διαχωρισμού και της αποτελεσματικότητας που προσδίδει στο μοντέλο, παρουσιάζεται στην επόμενη ενότητα.

### Συνδυαστική Μελέτη LDA - PCA

Αντίστοιχα με τα προηγούμενα δείγματα, δημιουργείται ένα γράφημα συνδυασμού των score plots της PCA και των προβλεπόμενων κλάσεων της LDA για τα φάσματα του δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2. Όπως παρουσιάζεται στο Σχήμα 60, οι αλληλοεπικαλύψεις των κλάσεων είναι υψηλές και επομένως ούτε σε αυτό το δείγμα θα ήταν αποτελεσματική η χρήση μόνο της PCA, εάν το πρόβλημα απαιτούσε μια σαφή κατηγοριοποίηση του φάσματος σε μια συγκεκριμένη κλάση αναφοράς.



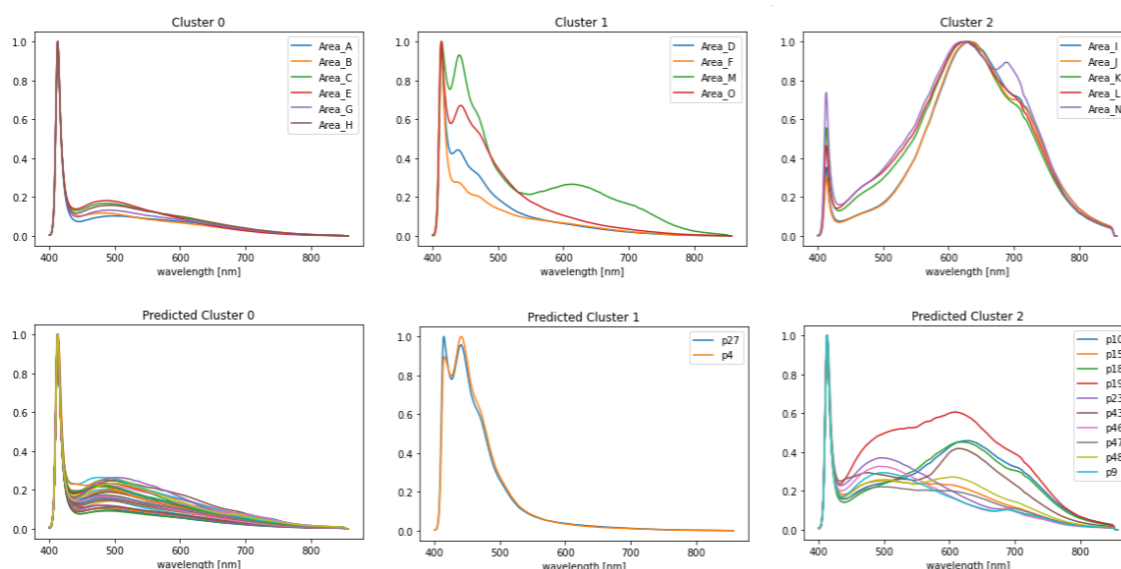
Σχήμα 60: Score Plots δεδομένων δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 στα οποία η χρωματική κλίμακα αντιστοιχεί στις κλάσεις που προβλέφθηκαν από τον αλγόριθμο της LDA. Στις διαγωνίους παρουσιάζεται η συνάρτηση πυκνότητας της κάθε κλάσης σύμφωνα με τις τιμές του component της αντίστοιχης στήλης

### 4.3.2 Φασματική Κατανομή Κλάσεων

Μια εναλλακτική κατανομή των φασμάτων αναφοράς αξιοποιείται προκειμένου να διαπιστωθεί κατά πόσο θα βελτιωθούν οι ικανότητες πρόβλεψης του μοντέλου για το δείγμα CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2. Ειδικότερα, τα φάσματα αναφοράς ταξινομήθηκαν σε 3 κλάσεις ανάλογα με την ομοιότητα που παρουσιάζουν στις κορυφές φθορισμού. Τα ομαδοποιημένα φάσματα παρουσιάζονται στο γράφημα 42. Η περιγραφή των νέων κλάσεων αναφοράς περιγράφεται στον Πίνακα 14. Αφού ο αλγόριθμος της LDA εκπαιδευτεί με τις νέες κλάσεις αναφοράς, εφαρμόζεται και πάλι το μοντέλο πρόβλεψης στα φάσματα της χαρτογράφησης και τα αποτελέσματα παρουσιάζονται γραφικά στο Σχήμα 61.

Περιοχή	Κλάση	Περιγραφή
A,B,C,E,G, H	0	Αμελητέα βιολογική επιβάρυνση - "Καθαρή" περιοχή
D,F,M,O	1	Αμελητέα βιολογική επιβάρυνση - Μαύρα Στίγματα
I,J,K,L,N	2	Μαύρα-Πρασινά στίγματα - Φωτοαυτότροφοι μικροοργανισμοί(χλωροφύλλη)

Πίνακας 14: Περιοχές Αναφοράς δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2 με βάση την φασματική τους ομοιότητα, αριθμός κλάσης στην οποία ανήκει η κάθε περιοχή καθώς και περιγραφή των ιδιοτήτων της κλάσης



Σχήμα 61: Σύγκριση Κλάσεων Φασμάτων Αναφοράς με τις Προβλεπόμενες Κλάσεις Φασμάτων Χαρτογράφησης Δείγματος CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2(Φασματική Κατανομή Κλάσεων)

Η ποσοτική αξιολόγηση του μοντέλου πραγματοποιείται και πάλι με εφαρμογή της μεθόδου LOO στα δεδομένα των νέων κλάσεων αναφοράς. Με χρήση των σχέσεων 74, 75, 76, υπολογίζεται η ακρίβεια, η ευαισθησία και η ανάκληση της κάθε κλάσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 15. Η συνολική ακρίβεια του μοντέλου αγγίζει το 93%, με εμφανή την βελτίωση από το 73.3% της βιολογικής κατανομής των κλάσεων αναφοράς. Επιπλέον, τα ποσοστά ακρίβειας, ευαισθησίας και ανάκλησης σε κάθε μεμονωμένη κλάση είναι εξαιρετικά υψηλά, αγγίζοντας κατά περιπτώσεις και το 100%. Ωστόσο η κατανομή αυτή, παρότι είναι στατιστικά πιο αποτελεσματική, πραγματοποιεί ένα άλμα πέρα από τη βιολογική ταξινόμηση που ακολουθήθηκε στις προηγούμενες αναλύσεις. Από το συγκεκριμένο άλμα όμως προκύπτουν ορισμένες ενδιαφέρουσες παρατηρήσεις σχετικά με τις ομαδοποιήσεις των δεδομένων. Πιο συγκεκριμένα:



1. Οι κλάσεις 0 και 2 παρουσιάζουν υψηλά ποσοστά ευαισθησίας και ανάκλησης και επομένως το μοντέλο είναι ικανό διακρίνει σε κλάσεις τις καθαρές περιοχές και τις βιολογικά επιβαρυνμένες από φωτοαυτότροφους μικροοργανισμούς. Ωστόσο, στην κλάση 2 συμπεριλήφθηκε και η περιοχή N η οποία αντιστοιχούσε σε μαύρο στίγμα αλλά η παρουσία της χλωροφύλλης την κατέταξε στην κατηγορία των φωτοαυτότροφων μικροοργανισμών. Επομένως, σε αυτή την περίπτωση ταξινόμησης η κλάση 2 θα πρέπει να αξιολογείται και ως κατηγορία συνύπαρξης ετερότροφων με φωτοαυτότροφους μικροοργανισμούς.
2. Η κλάση 2, η οποία περιλαμβάνει περιοχές με αμελητέα βιολογική επιβάρυνση καθώς και μαύρα στίγματα, παρουσιάζει υψηλή ανάκληση και 100% ευαισθησία, δηλαδή όσες προβλέψεις πραγματοποιήθηκαν για την κλάση αυτή ήταν και αληθείς. Εντούτοις, δεν είναι με κάποιο τρόπο διακριτό εάν το φάσμα ανήκει σε μαύρο στίγμα ή σε καθαρή επιφάνεια πιθανώς εξαιτίας της μη αποτελεσματικής διέγερσης της μελανίνης των μαύρων σιγμάτων λόγω της φωτοπροστατευτικής τους δράσης [61]. Συνεπώς, θα πρέπει να πραγματοποιηθεί επιπρόσθετη μελέτη στην καταλληλότερη διέγερση των μαύρων σιγμάτων, πιθανώς με ακτινοβολία διαφορετικού μήκους κύματος. Εναλλακτικά, δύναται να χρησιμοποιηθεί η βιολογική ταξινόμηση των φασμάτων αναφοράς με κόστος, όμως, στην ικανότητα πρόβλεψης του μοντέλου.

Cluster	ACCURACY	PRECISION	RECALL
0	0.93	0.86	1.00
1	0.93	1.00	0.75
2	1.00	1.00	1.00

Πίνακας 15: Ακρίβεια, Ευαισθησία και Ανάκληση προβλέψεων για κάθε κλάση με βάση των μοντέλο πρόβλεψης της LDA για το δείγμα CALLOS MAR-2 ΠΛΑΤΙΑ ΕΠΙΦΑΝΕΙΑ 2

#### 4.4 Συμπεράσματα

Με βάση την στατιστική επεξεργασία των φασματικών δεδομένων, τα οποία προέρχονται από την εφαρμογή της LIF σε τρεις διαφορετικές βιολογικά επιβαρυνμένες επιφάνειες, προκύπτουν τα ακόλουθα:

1. Η βιοδιάβρωση από φωτοαυτότροφους μικροοργανισμούς όπως βρύα, χλωροφύκη και κυανοβακτήρια διαχωρίζεται αποτελεσματικά από το μοντέλο καθώς και στις τρεις επιφάνειες που μελετήθηκαν η ακρίβεια ήταν υψηλή. Ειδικά στο CALLOS CER 3, η κλάση 0 των φωτοαυτότροφων παρουσίασε 100% επιβεβαίωση των παρατηρήσεων της. Στο CALLOS MAR-2 ΠΛΑΤΕΙΑ ΕΠΙΦΑΝΕΙΑ 2, η αντίστοιχη κλάση 2, επιβεβαίωσε τις παρατηρήσεις της κατά 75% ωστόσο, στο CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 η αντίστοιχη ανάκληση μειώνεται στο 50%. Αυτό συμβαίνει καθώς το φάσμα γίνεται πιο σύνθετο στο εύρος 450-600 nm, το οποίο καταδεικνύει συνύπαρξη με ετερότροφους μικροοργανισμούς.
2. Οι περιοχές με αμελητέα βιολογική επιβάρυνση διαχωρίζονται επίσης με υψηλή ακρίβεια από τους άλλους τύπους βιοδιάβρωσης. Ωστόσο όταν δημιουργούνται περισσότερες από μια κλάσεις με καθαρές περιοχές, εξαιτίας μιας αμελητέας διαφορετικής εναπόθεσης κάποιου βιοδιαβρωτή όπως έγινε στο δείγμα CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1, η ακρίβεια, η ευαισθησία και η ανάκληση του μοντέλου μειώνονται σημαντικά. Αυτό συμβαίνει καθώς ο αλγόριθμος εντοπίζει την ομοιομορφία μεταξύ των κλάσεων και μειώνει την πιθανότητα με την οποία ταξινομεί μια παρατήρηση σε μια από τις δύο κλάσεις.
3. Ο πιο απαιτητικός τύπος βιοδιάβρωσης όσον αφορά την ανίχνευση του μέσω φθορισμού είναι τα μαύρα στίγματα, δηλαδή οι Μαύροι Μερικτεματικοί Μύκητες, οι οποίοι παρουσιάστηκαν αναλυτικά στην υπόενοτητα 1.2.4. Οι κλάσεις 2 και 3 στο CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1 και η κλάση 3 CALLOS MAR-2 ΠΛΑΤΕΙΑ ΕΠΙΦΑΝΕΙΑ 2 που περιγράφουν αυτόν τον τύπο επιβάρυνσης, παρουσιάζουν εξαιρετικά μικρή απόδοση πρόβλεψης. Ειδικά στην πρώτη περίπτωση το ποσοστό των σωστών προβλέψεων ως προς τις συνολικές προβλέψεις για την κλάση ήταν μηδενικό. Αυτό πιθανώς συμβαίνει καθώς οι μύκητες αυτοί περιέχουν υψηλά ποσοστά μαύρων φωτοπροστατευτικών χρωστικών, όπως η μελανίνη, οι οποίες έχουν αμελητέο φθορισμό [81], ο οποίος είναι πιθανό να ενισχυθεί μόνο με ακτινοβολία UV [80]. Το αποτέλεσμα αυτού είναι ότι τα μαύρα στίγματα αντιπροσωπεύονται από το φθορισμό του περιβάλλοντος που βρίσκονται, δηλαδή της πέτρας ή καποια αποικίας φωτοαυτότροφων. Συνεπώς, οι κλάσεις τους περιέχουν εύρη φθορισμού που χαρακτηρίζουν και άλλες κλάσεις και έτσι εμφανίζεται σημαντική αβεβαιότητα στις συγκεκριμένες προβλέψεις.
4. Η εφαρμογή του μοντέλου στο CALLOS MAR-2 ΠΛΑΤΕΙΑ ΕΠΙΦΑΝΕΙΑ 2 για την φασματική κατανομή των κλάσεων έδωσε εξαιρετικά αποτελέσματα, γεγονός το οποίο οφείλεται στην υψηλή ομοιομορφία των φασμάτων εντός της κάθε κλάσης. Για να προκύψει όμως αυτή ομοιομορφία παραβιάστηκε η οπτική παρατήρηση με την οποία είχαν δημιουργηθεί οι κλάσεις βιολογικά. Το γεγονός αυτό καταδεικνύει την αναγκαιότητα συνεργασίας των επιστημών της εφαρμοσμένης φυσικής και βιολογίας στο πεδίο της μέτρησης. Ειδικότερα, τα φάσματα των περιοχών αναφοράς θα πρέπει να λαμβάνονται αυστηρά από τα σημεία όπου πραγματοποιήθηκε βιολογική δειγματοληψία προκειμένου να είναι γνωστό το περιεχόμενο του κάθε φάσματος. Με τον τρόπο αυτό, οι ομαδοποιήσεις των φασμάτων αναφοράς δεν θα στηρίζονται στην οπτική παρατήρηση [69, 67], αλλά στην πλήρη γνώση της βιολογικής πληροφορίας. Συνεπώς θα είναι επιστημονικά ορθότερη η ομαδοποίηση με παράλληλη ομοιομορφία στα φάσματα της κλάσης και η απόδοση του μοντέλου θα αυξηθεί.
5. Καθώς τα δεδομένα αναλύθηκαν με χρήση στατιστικών μεθόδων, η πληθώρα δειγμάτων σε κάθε κλάση αναφοράς ενισχύει την αποτελεσματικότητα του μοντέλου. Περιπτώσεις όπως η κλάση 1

του CALLOS CER-3 και η κλάση 3 του CALLOS MAR-2 ΚΑΘΕΤΗ ΕΠΙΦΑΝΕΙΑ 1, οι οποίες είχαν 1 και 2 παρατηρήσεις αντίστοιχα, αφενός δημιουργούν πολλά προβλήματα σε αλγορίθμους όπως η LOO και αφετέρου καταστρέφουν την στατιστικότητα που πρέπει να διέπει αυτόν τον τρόπο ανάλυσης.

6. Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε το μοντέλο πρόβλεψης της LDA, το οποίο εκ της κατασκευής του υποθέτει πως τα δεδομένα χωρίζονται σε κλάσεις με γραμμικά όρια. Αυτό θα μπορούσε να εισάγει σημαντικό σφάλμα στις προβλέψεις καθώς οι κλάσεις θα μπορούσαν να βρίσκονται εγκιβωτισμένες μέσα σε καμπυλοειδής επιφάνειες, στον χώρο των μεταβλητών τους. Επομένως, θα έπρεπε να εφαρμοστεί κάποιο μοντέλο που θα λάμβανε αυτή την πιθανότητα υπόψη, όπως το μοντέλο της *Quadratic Discriminant Analysis - QDA*. Σε κάθε περίπτωση όμως, γίνεται αντιληπτή η ανάγκη της εξερεύνησης της φύσης των δεδομένων φθορισμού με χρήση πολυμεταβλητών μεθόδων ανάλυσης.
7. Η ανάλυση η οποία πραγματοποιήθηκε καταδεικνύει τα ευρεία περιθώρια εφαρμογής στατιστικών μοντέλων ομαδοποίησης και πρόβλεψης στην φασματοσκοπία φθορισμού. Η αποδοτικότητα των μοντέλων ενισχύεται ανάλογα με την ποιότητα και την ποσότητα της πληροφορίας που παρέχεται στους αλγορίθμους Μηχανικής Μάθησης. Στην παρούσα διπλωματική εργασία, η πληροφορία αυτή αντιστοιχούσε στην βιολογική αναγνώριση των φασμάτων αναφοράς, τα οποία είναι σημαντικά λιγότερα από τα φάσματα χαρτογράφησης. Επομένως, υποβαθμίζεται η ανάγκη βιολογικής μελέτης του κάθε μεμονωμένου σημείου επάνω στην επιφάνεια ενός μνημείου και αναδεικνύεται η πιο άμεση και σημαντικά αποδοτική οπτική μέθοδος μελέτης. Η επέκταση στο εύρος ενός ολόκληρου μνημείου και η δημιουργία των χαρτών φθορισμού, θα μπορούσε να αποτελέσει το μέλλον της διαδικασίας συντήρησης αρχαιοτήτων.

## 5 Επίλογος

Καθώς ολοκληρώνεται η συγκεκριμένη εργασία, είναι πλέον κατανοητό πως ο συνδυασμός των επιστημών μπορεί να φέρει αρκετά ενδιαφέροντα αποτελέσματα, τα οποία μπορούν να προκύψουν για μεγάλο όγκο δεδομένων. Η εφαρμογή μοντέλων πρόβλεψης για τους τύπους της βιολογικής επιβάρυνσης των μνημείων πολιτιστικής κληρονομιάς αποτελεί μια πολλά υποσχόμενη ιδέα, καθώς ο συνδυασμός της με την LIF αναδεικνύει μια μη παρεμβατική προσέγγιση της μελέτης του μνημείου.

Η δημιουργία της χαρτογράφησης φθορισμού ενός μνημείου αποτελεί μια εξαιρετικά πολύπλοκη διαδικασία, η οποία όμως έχει πραγματοποιηθεί αποτελεσματικά στο παρελθόν. Το επόμενο βήμα της μελέτης είναι η ανάλυση του μεγάλου όγκου δεδομένων και η εξαγωγή ασφαλών συμπερασμάτων σχετικά με την κατάσταση του μνημείου. Στην παρούσα εργασία προτάθηκε ο συνδυασμός της βιολογικής μελέτης με την εφαρμογή αλγορίθμων δημιουργίας μοντέλων πρόβλεψης. Η συγκεκριμένη μεθοδολογία επέκτεινε την ήδη υπάρχουσα καθώς αξιοποίησε την τεχνική της Επιβλεπόμενης Ταξινόμησης μέσω της LDA, ενώ έως τώρα κυριαρχούσε η Μη Επιβλεπόμενη μέσω της PCA.

Εν κατακλείδι, ο στόχος της παρούσας μελέτης επετεύχθη καθώς προτάθηκαν νέες μέθοδοι προσέγγισης της επιβάρυνσης της επιφάνειας ενός μνημείου με άξονα τον συγκερασμό των επιστημών της Φυσικής, της Βιολογίας και των Δεδομένων.

## 6 Βιβλιογραφία

### References

- [1] Valeur B. *Molecular Fluorescence: Principles and Applications*. Wiley-VCH, 1st edition, 2001.
- [2] Atkins P. et al. *Molecular Quantum Mechanics*. Oxford University Press, 4th edition, 2005.
- [3] Hedayatifar et al. Optical absorption and electronic spectra of chlorophylls a and b. *The Royal Society of Chemistry*, 111(6):109778–109785, 2016.
- [4] Lakowicz J. R. *Principles of Fluorescence Spectroscopy*. Springer, 3rd edition, 2006.
- [5] Lichtenthaler et al. The role of chlorophyll fluorescence in the detection of stress conditions in plants. *Critical Reviews in Analytical Chemistry*, 19:S29–S85, 1988.
- [6] Macedo et al. Biodiversity of cyanobacteria and green algae on monuments in the mediterranean basin: an overview. *Microbiology*, 155(11):3476–3490, 2009.
- [7] Winkler E.M. *Stone in Architecture: Properties, Durability*. Springer Heidelberg New York Dordrecht London, 3rd edition, 1997.
- [8] Nielsen F. *Introduction to HPC with MPI for Data Science*. Undergraduate Topics in Computer Science. Springer, 1st edition, 2016.
- [9] Hastie T. et al. *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics. Springer, 2nd ed. 2009. corr. 3rd printing 5th printing. edition, 2009.
- [10] Chen G. Mathematical methods for data visualization: Linear discriminant analysis (lda), 2020. San Jose State University : Math 253, <https://www.sjsu.edu/faculty/guangliang.chen/Math253S20/lec11lda.pdf>.
- [11] Raimondi et al. Latest advancements in fluorescence hyperspectral lidar imaging of the cultural heritage. *Fundamentals of Laser-Assisted Micro- and Nanotechnologies 2013*, 9065, 2013.
- [12] Raimondi et al. The fluorescence lidar technique for the remote sensing of photoautotrophic biodeteriogens in the outdoor cultural heritage: A decade of in situ experiments. *International Biodeterioration Biodegradation*, 63(7):823–835, 2009.
- [13] Albani J. R. *Principles and Applications of Fluorescence Spectroscopy*. Blackwell Science, 1st edition, 2007.
- [14] Braslavsky S. E. Glossary of terms used in photochemistry. *Pure and Applied Chemistry*, 79(3):293–465, 2007.
- [15] Demmig-Adams et al. *Non-Photochemical Quenching and Energy Dissipation in Plants, Algae and Cyanobacteria*. Advances in Photosynthesis and Respiration 40. Springer Netherlands, 1st edition, 2014.
- [16] İnan İseri et al. Chlorophyll transition dipole moment orientations and pathways for flow excitation energy among the chlorophylls of the major plant antenna, lhci. *Biophysics of Structure and Mechanism*, 30(5):344–353, 2001.
- [17] Pinna D. *Coping with biological growth on stone heritage objects: methods, products, applications, and perspectives*. CRC Press, 1st edition, 2017.

- [18] Papida et al. Biodeterioration Control for Athens Acropolis Monuments: Strategy and Constraints. *Studies in Conservation*, 55(sup2):74–79, 2010.
- [19] Guillitte O. Bioreceptivity: a new concept for building ecology studies. *The Science of the Total Environment*, 167(1-3):215–220, 1995.
- [20] Coutinho et al. Biological colonization and biodeterioration of architectural ceramic materials: An overview. *Journal of Cultural Heritage*, 16(5):759–777, 2015.
- [21] Doehne E. and Price C. *Stone Conservation: An Overview of Current Research*. Getty Conservation Institut, 2nd edition, 2010.
- [22] Vanderpoorten et al. *Introduction to Bryophytes*. Cambridge University Press, 1st edition, 2009.
- [23] Salvatory et al. The role of fungi and lichens in the biodeterioration of stone monuments. *The Open Conference Proceedings Journal*, 7:39–54, 2016.
- [24] Onofri et al. Rock-inhabiting fungi and their role in deterioration of stone monuments in the Mediterranean area. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology*, 148(2):384–391, 2014.
- [25] Barsanti L. and Gualtieri P. *Algae: Anatomy, Biochemistry, and Biotechnology*. CRC Press, 1st edition, 2005.
- [26] Anagnostides et al. Epilithic and chasmolithic microflora (cyanophyta, bacillariophyta) from marbles of the parthenon (acropolis-athens, greece). *Nova Hedwigia*, 38:227–287, 1983.
- [27] Silhavy et al. The bacterial cell envelope. *Cold Spring Harbor Perspectives in Biology*, 2(5), 2010.
- [28] Savvides et al. Bacterial deterioration of marble monuments: A case study of the conservation project of acropolis monuments. *Geomicrobiology Journal*, 31(8):726–736, 2014.
- [29] Whitton B.A. *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. Springer Dordrecht Heidelberg New York London, 1st edition, 2012.
- [30] Lambrinou et al. Phenotypic and molecular biological characterization of cyanobacteria from marble surfaces of treated and untreated sites of propylaea (acropolis, athens). *Geomicrobiology Journal*, 30(4):371–378, 2014.
- [31] Persichetti et al. Spectral discrimination of planktonic cyanobacteria and microalgae based on deep uv fluorescence. *Sensors and Actuators B: Chemical*, 284:228–235, 2019.
- [32] Larry L., Clair S., and Seaward M. *Biodeterioration of Stone Surfaces Lichens and Biofilms as Weathering Agents of Rocks and Cultural Heritage*. Springer Science+Business Media Dordrecht, 1st edition, 2004.
- [33] Gadd et al. Geomicrobiology of Eukaryotic Microorganisms. *Geomicrobiology Journal*, 27:491–519, 2010.
- [34] Urzi et al. Airborne fungal spores connected with marble colonisation monitored in the terrace of Messina Museum. *Aerobiologia*, 17(1):11–17, 2001.
- [35] Misra et al. *Fungi From Different Substrates*. CRC Press, 1st edition, 2014.

- [36] Wollenzien et al. On the isolation of microcolonial fungi occurring on and in marble and other calcareous rocks. *The Science of the Total Environment*, 167(1-3):287–294, 1995.
- [37] Gorbushina et al. The fungal microcosm of Mediterranean monuments and sites-past, present and future. *Proc. 4th Int. Symp. on the Conservation of Monuments in the Mediterranean*, 4:261–270, 1997.
- [38] Gorbushina A. Life on the rocks. *Environmental Microbiology*, 9(7):1613–1631, 2007.
- [39] Nicholson W.K. *Linear algebra with applications*. Toronto [U.A.] Mcgraw Hill, 2009.
- [40] Barranco-Chamorro et al. Multivariate projection techniques to reduce dimensionality in large datasets. *Smart Data*, pages 133–160, 2019.
- [41] Everitt B. and Dunn G. *Applied multivariate data analysis*. Arnold, New York, 2001.
- [42] Brunton S. and Kutz J.N. *Data-driven science and engineering : machine learning, dynamical systems, and control*. Cambridge University Press, 2019.
- [43] Bartholomew et al. *Ανάλυση πολυμεταβλητών τεχνικών στις κοινωνικές επιστήμες*. Εκδόσεις Κλειδάριθμος, 2011.
- [44] Gareth J. et al. *An Introduction to Statistical Learning : With Applications in R*. Springer, 1st edition, 2013.
- [45] Chen G. Statistical and machine learning classification: Bayes classifiers, 2020. San Jose State University : Math 251, <https://www.sjsu.edu/faculty/guangliang.chen/Math251/lec3bayes.pdf>.
- [46] Dankowska et al. Tea types classification with data fusion of uv–vis, synchronous fluorescence and nir spectroscopies and chemometric analysis. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 211:195–202, 2019.
- [47] Mei L. et al. Tea classification and quality assessment using laser-induced fluorescence and chemometric evaluation. *Applied Optics*, 51(7):803, 2012.
- [48] Larose D. and Larose C. *Data Mining and Predictive Analytics*. Wiley Series on Methods and Applications in Data Mining. Wiley, 2 edition, 2015.
- [49] Measures R. and Bristow M. The development of a laser fluorosensor for remote environmental probing. *Joint Conference on Sensing of Environmental Pollutants*, 1971.
- [50] Edner et al. Laser-induced fluorescence monitoring of vegetation in tuscan. *EARSel Advances in Remote Sensing*, 1:119–130, 1992.
- [51] H. Edner, J. Johansson, S. Svanberg, and E. Wallinder. Fluorescence lidar multicolor imaging of vegetation. *Applied Optics*, 33:2471, 1994.
- [52] Valentini et al. Remote sensing of chlorophyll a fluorescence of vegetation canopies: 2. physiological significance of fluorescence signal in response to environmental stresses. *Remote Sensing of Environment*, 47:29–35, 1994.
- [53] Cecchi et al. Fluorescence lidar technique for the remote sensing of stony materials in ancient buildings. *Remote Sensing for Geography, Geology, Land Planning, and Cultural Heritage*, 2960, 1996.

- [54] Raimondi et al. Fluorescence lidar monitoring of historic buildings. *Applied Optics*, 37(6):1089, 1998.
- [55] Pantani et al. Lithotypes characterization with a fluorescence lidar imaging system using a multi-wavelength excitation source. *Remote Sensing for Environmental Monitoring, GIS Applications, and Geology II*, 2003.
- [56] Palombi et al. Hyperspectral fluorescence lidar imaging at the colosseum, rome: Elucidating past conservation interventions. *Optics Express*, 16(10):6794, 2008.
- [57] Cecchi et al. Fluorescence lidar technique for the monitoring of biodeteriogens in cultural heritage studies. *Remote Sensing for Geography, Geology, Land Planning, and Cultural Heritage*, Proceedings Volume 2960, 1996.
- [58] David Lognoli, Gioia Lamenti, Luca Pantani, Daniele Tirelli, and Luisa Tomaselli. Lidar remote sensing of stone cultural heritage: detection and characterization of biodeteriogens. *Remote Sensing for Earth Science, Ocean, and Sea Ice Applications*, Proceedings Volume 3868, 1999.
- [59] Pantani et al. Experiments on stony monument monitoring by laser-induced fluorescence. *Journal of Cultural Heritage*, 1:S345–S348, 2000.
- [60] Lognoli et al. Detection and characterization of biodeteriogens on stone cultural heritage by fluorescence lidar. *Applied Optics*, 41:1780, 2002.
- [61] Raimondi et al. Remote detection of laser-induced autofluorescence on pure cultures of fungal and bacterial strains and their analysis with multivariate techniques. *Optics Communications*, 273(1):219–225, 2007.
- [62] Castillejo et al. *Lasers in the conservation of artworks : Proceedings of the International Conference Lacona VII, Madrid, Spain, 17-21 September 2007*. Crc Press/Balkema, London, 2007.
- [63] Lognoli et al. Fluorescence lidar imaging of the cathedral and baptistery of parma. *Applied Physics B: Lasers and Optics*, 76:457–465, 2003.
- [64] Bracci et al. Lidar remote sensing of stone cultural heritage: detection of protective treatments. *Optical Engineering*, 40:1579, 2001.
- [65] Raimondi et al. Fluorescence lidar measurements at the archaeological site house of augustus at palatino, rome. *Earth Resources and Environmental Remote Sensing/GIS Applications IV*, Proceedings Volume 8893, 2013.
- [66] Raimondi et al. A fluorescence lidar combining spectral, lifetime and imaging capabilities for the remote sensing of cultural heritage assets. *Earth Resources and Environmental Remote Sensing/GIS Applications V*, 2014.
- [67] Weibring et al. Fluorescence lidar imaging of historical monuments. *Applied Optics*, 40(33):6111, 2001.
- [68] Weibring et al. Versatile mobile lidar system for environmental monitoring. *Applied Optics*, 42(18):3583, 2003.



- [69] Hällström et al. Remote fluorescence lidar imaging of monuments: the coliseum and the lateran baptistery in rome. In *LASERS IN THE CONSERVATION OF ARTWORKS*, pages 169–175, United Kingdom, 2008. Taylor Francis. Lacona VII - Lasers in the Conservation of Artworks ; Conference date: 17-09-2007 Through 21-09-2007.
- [70] Hällström et al. Documentation of soiled and biodeteriorated facades: A case study on the coliseum, rome, using hyperspectral imaging fluorescence lidars. *Journal of Cultural Heritage*, 10:106–115, 2009.
- [71] Palombi et al. A fluorescence lidar sensor for hyper-spectral time-resolved remote sensing and mapping. *Optics Express*, 21:14736, 2013.
- [72] Wold et al. Active photosensitizers in butter detected by fluorescence spectroscopy and multivariate curve resolution. *Journal of Agricultural and Food Chemistry*, 54:10197–10204, 2006.
- [73] Andersson-Engels et al. Time-resolved laser-induced fluorescence spectroscopy for enhanced demarcation of human atherosclerotic plaques. *Journal of Photochemistry and Photobiology B: Biology*, 4:363–369, 1990.
- [74] Fujii T. and Fukuchi T. *Laser Remote Sensing*. CRC Press, 1st edition, 2005.
- [75] Richardson et al. Bioaerosol detection over athens, greece using the laser induced fluorescence technique. *Science of The Total Environment*, 696:133–906, 2019.
- [76] Zhao et al. Inelastic hyperspectral lidar for profiling aquatic ecosystems. *Laser Photonics Reviews*, 10:807–813, 2016.
- [77] Vincent W. Fluorescence properties of the freshwater phytoplankton: Three algal classes compared. *British Phycological Journal*, 18(1):5–21, 1983.
- [78] Knauss et al. Label-free fluorescence microscopy in fungi. *Fungal Biology Reviews*, 27(2):60–66, 2013.
- [79] Donaldson L. Autofluorescence in Plants. *Molecules*, 25(10):2393, 2020.
- [80] Raimondi et al. In vivo real-time recording of uv-induced changes in the autofluorescence of a melanin-containing fungus using a micro-spectrofluorimeter and a low-cost webcam. *Optics Express*, 17:22735, 2009.
- [81] Raimondi et al. Remote detection of laser-induced autofluorescence on pure cultures of fungal and bacterial strains and their analysis with multivariate techniques. *Optics Communications*, 273(1):219–225, 2007.

## Α Παράρτημα: Στοιχεία Στατιστικής Πολυμεταβλητών Δεδομένων

Όταν συντελείται ένα πείραμα (παρατήρηση) εξαρτάται συνήθως από περισσότερες από μια μεταβλητές. Πραγματοποιώντας αυτή την απλή παραδοχή, ορίζεται ο πίνακας πολυμεταβλητών δεδομένων ως:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdots & \cdot & \cdot & x_{1p} \\ x_{21} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & x_{np} \end{bmatrix} \quad (\text{A-1})$$

με  $X \in \mathbb{R}_{n \times p}$  όπου η κάθε γραμμή αντιπροσωπεύει ένα πείραμα (παρατήρηση) και η κάθε στήλη αντιστοιχεί στην μεταβλητή όπου μετρείται για κάθε πείραμα.

Για την ανάλυση ενός πίνακα πολυμεταβλητών δεδομένων θα χρειαστούν ορισμένοι βασικοί στατιστικοί ορισμοί:

### 1. Μέση Τιμή

Θεωρώντας  $p$  μεταβλητές, το διάνυσμα του πληθυσμιακού μέσου (mean)  $\bar{x} = [\bar{x}_1 \quad \bar{x}_2 \quad \cdot \quad \cdot \quad \bar{x}_p]$  δίνεται από τη σχέση:

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_j \quad (\text{A-2})$$

όπου  $\bar{x}_i$  είναι η μέση τιμή των τιμών της μεταβλητής  $x_i$  υπολογισμένη για  $n$  παρατηρήσεις.

### 2. Διακύμανση

Θεωρώντας  $p$  μεταβλητές, η διακύμανση (variance) της κάθε μεταβλητής παρουσιάζεται στο διάνυσμα  $\bar{\sigma}^2 = [\sigma_1^2 \quad \sigma_2^2 \quad \cdot \quad \cdot \quad \sigma_p^2]$  όπου:

$$\sigma_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x}_i)^2 \quad (\text{A-3})$$

ορίζεται ως η διακύμανση των τιμών της μεταβλητής  $x_i$  υπολογισμένη για  $n$  παρατηρήσεις.

### 3. Συνδιακύμανση

Για δύο μεταβλητές  $x$  και  $y$  οι οποίες έχουν μετρηθεί σε  $n$  παρατηρήσεις και έχουν δώσει τα ζεύγη τιμών  $(x_i, y_i)$  με  $i = 1, 2, \dots, n$  ορίζεται η συνδιακύμανση (covariance) μέσω της σχέσης:

$$\text{Cov}(x, y) = \frac{1}{n} \sum_{j=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i) \quad (\text{A-4})$$

Προκύπτει ότι εάν  $x = y$ , τότε η συνδιακύμανση ταυτίζεται με την διακύμανση της  $x$  μεταβλητής. Όταν σε ένα σύνολο δεδομένων  $X$  υπάρχουν  $p$  μεταβλητές για  $n$  πειράματα τότε ορίζεται ο πίνακας συνδιακύμανσης  $S$  και κάθε στοιχείο του, το οποίο είναι η συνδιακύμανση της  $x_i$  μεταβλητής

με την  $x_j$  μεταβλητή, δίνεται από τη σχέση:

$$s_{ij} = \frac{1}{n-1} \sum_{l=1}^n (x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j) \quad (\text{A-5})$$

Προκύπτει ότι στη διαγώνιο του πίνακα συνδιακύμανσης εμφανίζεται η διακύμανση της αντίστοιχης μεταβλητής  $x_i$ . Επιπλέον, θεωρώντας έναν πίνακα πολυμεταβλητών δεδομένων  $X$  από τον οποίο έχει αφαιρεθεί το διάνυσμα του πληθυσμιακού μέσου, τότε ο πίνακας συνδιακύμανσης θα δίνεται σε κλειστή μορφή από τη σχέση:

$$\Sigma = \frac{1}{n-1} X^T X \quad (\text{A-6})$$

#### 4. Συσχέτιση

Καθώς ο πίνακας συνδιακύμανσης επηρεάζεται από τις τάξεις μεγέθους των μεταβλητών, το κατάλληλο μέτρο που δηλώνει την γραμμικότητα μεταξύ των τιμών των μεταβλητών σε ένα πίνακα δεδομένων  $X$ , περιγράφεται από τον πίνακα συσχέτισης  $R$ . Τα στοιχεία του πίνακα συσχέτισης προκύπτουν από τα στοιχεία του πίνακα συνδιακύμανσης μέσω της σχέσης :

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} \quad (\text{A-7})$$

Τα στοιχεία αυτά, τα οποία ονομάζονται συντελεστές συσχέτισης, παίρνουν τιμές από -1 έως 1, το οποίο δηλώνει την γραμμική συσχέτιση μεταξύ των μεταβλητών  $x_i$  και  $x_j$ . Εάν είναι θετική η τιμή του εκάστοτε στοιχείου, τότε οι τιμές των μεταβλητών που συμμετέχουν στον υπολογισμό τους μεγαλώνουν ή μικραίνουν μαζί. Αντίθετα αν η τιμή είναι αρνητική, τότε οι υψηλές τιμές της μίας μεταβλητής αντιστοιχούν στις χαμηλές τιμές της άλλης μεταβλητής. Ο πίνακας συσχέτισης προκύπτει από τον πίνακα συνδιακύμανσης σε κλειστή μορφή μέσω της σχέσης:

$$R = D^{-1/2} \Sigma D^{-1/2} \quad (\text{A-8})$$

όπου  $D = \text{diag} \left( 1/\sqrt{\sigma_{ii}^2} \right)$ . Προκύπτει πως ο πίνακας συσχέτισης ταυτίζεται με τον πίνακα συνδιακύμανσης, όταν ο πίνακας δεδομένων  $X$  έχει υποστεί τυποποίηση (standardization), η οποία θα περιγραφεί στη συνέχεια.