



**ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΕΠΙΣΤΗΜΗΣ ΥΛΙΚΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΧΗΜΕΙΑ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΤΩΝ ΥΛΙΚΩΝ»**

**ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ (Μ.Δ.Ε.)
ΠΑΝΑΓΟΠΟΥΛΟΣ ΔΙΟΝΥΣΗΣ ΠΑΝΑΓΙΩΤΗΣ**

*Ανάλυση μεγάλων δεδομένων
κατά τη μελέτη Θρησκευτικών Φορητών Εικόνων
με Απεικονιστική Φασματοσκοπία Φθορισμού Ακτίνων-Χ*

ΙΩΑΝΝΙΝΑ, 2025



**HELLENIC REPUBLIC
UNIVERSITY OF IOANNINA
POLYTECHNIC SCHOOL
DEPARTMENT OF MATERIAL SCIENCEE AND ENGINEERING**

**INTERDEPARTMENTAL POSTGRADUATE PROGRAM
«CHEMISTRY AND TECHNOLOGY OF MATERIALS»**

POSTGRADUATE DIPLOMA THESIS (M.S.C.)

ΠΑΝΑΓΟΠΟΥΛΟΣ ΔΙΟΝΥΣΗΣ ΠΑΝΑΓΙΩΤΗΣ

**BIG DATA ANALYSIS
IN THE STUDY OF RELIGIOUS PORTABLE ICONS
USING X-RAY FLUORESCENCE IMAGING
SPECTROSCOPY**

IOANNINA, 2025

Εσώφυλλο:

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην εξειδίκευση:

«ΧΗΜΕΙΑ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΤΩΝ ΥΛΙΚΩΝ»

που απονέμει το Τμήμα Μηχανικών Επιστήμης Υλικών του Πανεπιστημίου Ιωαννίνων.

Εγκρίθηκε την 06/10/2025 από την εξεταστική επιτροπή:

Αναγνωστόπουλος Δημήτριος,

Καθηγητής του Τμήματος Μηχανικών Επιστήμης των Υλικών,

Πολυτεχνική Σχολή του Πανεπιστημίου Ιωαννίνων, Επιβλέπων.

Παπαγεωργίου Δημήτριος,

Καθηγητής του Τμήματος Μηχανικών Επιστήμης των Υλικών,

Πολυτεχνική Σχολή του Πανεπιστημίου Ιωαννίνων.

Τζίμα Αναστασία,

Μέλος Ε.Δ.Ι.Π του Τμήματος Μηχανικών Επιστήμης των Υλικών,

Πολυτεχνική Σχολή του Πανεπιστημίου Ιωαννίνων.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

"Δηλώνω υπεύθυνα ότι η παρούσα διατριβή εκπονήθηκε κάτω από τους διεθνείς ηθικούς και ακαδημαϊκούς κανόνες δεοντολογίας και προστασίας της πνευματικής ιδιοκτησίας. Σύμφωνα με τους κανόνες αυτούς, δεν έχω προβεί σε ιδιοποίηση ξένου επιστημονικού έργου και έχω πλήρως αναφέρει τις πηγές που χρησιμοποίησα στην εργασία αυτή."

Παναγόπουλος Διονύσιος Παναγιώτης

ΠΕΡΙΛΗΨΗ

Η παρούσα μελέτη αντιμετωπίζει την πρόκληση της ανάλυσης των πολύπλοκων, υψηλής διαστατικότητας δεδομένων που προκύπτουν από τη Φασματοσκοπία Απεικόνισης Φθορισμού Ακτίνων-X (MA-XRF) σε θρησκευτικές φορητές εικόνες. Για τον σκοπό αυτό, αναπτύχθηκε ένα εξειδικευμένο λογισμικό και μια ολοκληρωμένη χημειομετρική μεθοδολογία. Η κεντρική προσέγγιση βασίζεται σε μια υβριδική στρατηγική Ανάλυσης Κύριων Συνιστωσών και Μη-Αρνητικής Παραγοντοποίησης Πινάκων (NMF-PCA) για την αποτελεσματική μείωση της διαστατικότητας και την εξαγωγή φυσικά ερμηνεύσιμων φασματικών υπογραφών. Τα αποτελέσματα αυτά τροφοδοτούν τον αλγόριθμο συσταδοποίησης K-Means (DBScan, TSNE, Spectral, GMM) για την παραγωγή θεματικών χαρτών στοιχείων υψηλής ευκρίνειας. Η μεθοδολογία εφαρμόστηκε επιτυχώς σε βυζαντινές και μεταβυζαντινές εικόνες, επιτυγχάνοντας έναν ανώτερο διαχωρισμό των υλικών της παλέτας, συμπεριλαμβανομένου του υποστρώματος gesso, των γήινων χρωστικών, του κιννάβαρι, του λευκού του μολύβδου και της επιχρύσωσης. Επιπλέον, η δυνατότητα ιεραρχικής ανάλυσης (sub-clustering) σε περιοχές ενδιαφέροντος επέτρεψε την αποκάλυψη της καλλιτεχνικής τεχνικής, αναδεικνύοντας τη λεπτομερή στρωματογραφία των υλικών (π.χ. προπλάσμος, φωτίσματα, σκιές) και τις μεμονωμένες πινελιές. Η εργασία αυτή καταδεικνύει ότι ο συνδυασμός MA-XRF με προηγμένες, μη-επιβλεπόμενες μεθόδους μηχανικής μάθησης αποτελεί ένα πανίσχυρο εργαλείο για την εις βάθος, μη-καταστροφική μελέτη της υλικής ιστορίας και τεχνολογίας των πολιτιστικών αγαθών.

***Λέξεις-κλειδιά:** MA-XRF, Ανάλυση Μεγάλων Δεδομένων, Μηχανική Μάθηση, PCA, NMF, K-Means, Συσταδοποίηση, Υπερφασματική Απεικόνιση, Πολιτιστική Κληρονομιά, Ανάλυση Εικόνων.*

ABSTRACT

This study addresses the challenge of analyzing complex, high-dimensional data generated by Macro X-ray Fluorescence (MA-XRF) imaging of religious portable icons. For this purpose, a specialized software and an integrated chemometric methodology were developed. The core approach is based on a hybrid Principal Component Analysis and Non-negative Matrix Factorization (NMF-PCA) strategy for effective dimensionality reduction and the extraction of physically interpretable spectral signatures. These results are then used as input into the K-Means (DBScan, TSNE, Spectral, GMM) clustering algorithm to generate high-definition elemental distribution maps. The methodology was successfully applied to Byzantine and post-Byzantine icons, achieving superior separation of palette materials, including the gesso ground, earth pigments, vermilion, lead white, and gilding. Furthermore, the capability of hierarchical analysis (sub-clustering) on regions of interest allowed for the revelation of the artistic technique, highlighting the detailed stratigraphy of materials (e.g., proplasmos, highlights, shadows) and individual brushstrokes. This work demonstrates that the combination of MA-XRF with advanced unsupervised machine learning methods constitutes a powerful tool for the in-depth, non-destructive study of the material history and technology of cultural heritage objects.

Keywords: *MA-XRF, Big Data Analysis, Machine Learning, PCA, NMF, K-Means, Clustering, Hyperspectral Imaging, Cultural Heritage, Icon Analysis*

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να εκφράσω τη βαθύτατη ευγνωμοσύνη μου σε όλους όσους συνέβαλαν στην ολοκλήρωση της παρούσας διπλωματικής εργασίας. Πρώτα απ' όλα, στους γονείς μου, που με στήριξαν με ατέρμονη αγάπη και θυσίες σε κάθε βήμα της ακαδημαϊκής μου πορείας. Η συνεχής τους ενθάρρυνση και η πίστη στις δυνατότητές μου υπήρξαν ο βασικός πυλώνας που με βοήθησε να φτάσω μέχρι εδώ. Οι φίλοι μου αποτέλεσαν κρίσιμο σύστημα υποστήριξης, προσφέροντας μου συναισθηματική ισορροπία και ευχάριστες στιγμές χαλάρωσης που με βοήθησαν να διατηρήσω την ψυχική μου υγεία κατά τη διάρκεια αυτής της απαιτητικής διαδικασίας. Ειδική αναγνώριση αξίζει στον αξιότιμο καθηγητή Δημήτριο Αναγνωστόπουλο, για την πολύτιμη επιστημονική καθοδήγηση, την υπομονή και τις εποικοδομητικές κριτικές του. Η ακριβής μεθοδολογία και η επιστημονική του ακεραιότητα αποτέλεσαν πρότυπο για την ανάπτυξη της εργασίας αυτής. Ιδιαίτερες ευχαριστίες οφείλω και στον διδάκτορα Θεοφάνη Γεροδήμο, για την ανεκτίμητη τεχνική του υποστήριξη στην αντιμετώπιση πληροφοριακών προκλήσεων, την αδιάλειπτη διαθεσιμότητα και την επιστημονική του αυστηρότητα που με βοήθησαν να υπερβώ δυσκολίες. Η συμβολή όλων σας υπήρξε καθοριστική για την ολοκλήρωση αυτής της εργασίας και η ευγνωμοσύνη μου θα παραμείνει αμετάβλητη με το πέρασμα του χρόνου.

Πίνακας περιεχομένων

Πίνακας περιεχομένων	9
EIKONES KEIMENOU	11
A' ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ	20
1. ΗΛΕΚΤΡΟΜΑΓΝΗΤΙΚΗ ΑΚΤΙΝΟΒΟΛΙΑ X – Ray	21
1.1 Εκπομπή συνεχούς ακτινοβολίας.....	23
1.2 Συνεχής φασματική κατανομή ακτίνων-X σε στόχους όπου προσκρούουν ηλεκτρόνια	23
1.3 Φασματικά χαρακτηριστικά.....	24
2. ΕΚΠΟΜΠΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΚΤΙΝΩΝ - X.....	25
2.1 Φασματική σειρά στις ακτίνες X	27
2.2 Απόδοση φθορισμού.....	28
2.3 Αλληλεπίδραση των φωτονίων με την ύλη	28
B' ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ	31
ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	32
Κεφάλαιο: Εισαγωγή στη Μηχανική Μάθηση για την Ανάλυση Έργων Τέχνης	32
Παραδείγματα μάθησης.....	37
Επιλογή του κατάλληλου αλγορίθμου μάθησης	39
Μέθοδοι Μείωσης Διαστατικότητας και Συσταδοποίησης για Υπερφασματικά Δεδομένα.....	42
Ανάλυση Κύριων Συνιστωσών PCA	43
Μη Αρνητική Παραγοντοποίηση Πινάκων NMF	47
Η Υβριδική Προσέγγιση NMF-PCA	50
Μέθοδοι Συσταδοποίησης	50
Γ' Πειραματική διάταξη	66
M6 JETSTREAM	67
4.1.1 Χαρακτηριστικά του φασματόμετρου M6 JETSTREAM.....	67
4.1.2 Δομή του συστήματος M6 JETSTREAM	67
4.1.3 Μέρη του φασματόμετρου	68
Δ' ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ	73
Υλικά, Μέθοδοι και Αναλυτική Ροή Εργασίας.....	74
Αντικείμενα Μελέτης: Ιστορικό και Τεχνοτροπικό Πλαίσιο.....	74
Χημειομετρική Ανάλυση με το Πλαίσιο NICE	80
Πρόγραμμα Ανάλυσης/ Αλγόριθμος.....	81
GENERATE_XRF_LIBRARY.py.....	85
NICE_HDF5_CORE.py: Θεμελίωση της Πρόσβασης και Επικύρωσης Δεδομένων	87
ANALYSIS_METHODS_4_NEW18.py	89

ANALYSIS_VISUALIZER_3_NEW18.py	93
ANALYSIS_WINDOW_6_NEW18.py	94
HDF5_Explorer_GUI.py	97
Αποτελέσματα και Συζήτηση.....	103
ΑΓΙΟΣ ΙΩΑΝΝΗΣ ΧΑΜΗΛΗΣ ΑΝΑΛΥΣΗΣ	103
ΑΓΙΟΣ ΙΩΑΝΝΗΣ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ	168
ΠΑΝΑΓΙΑ Η ΚΑΘΟΔΗΓΗΤΡΙΑ	207
ΣΥΜΠΕΡΑΣΜΑΤΑ	254
<i>Αξιολόγηση και Βελτιστοποίηση της Αναλυτικής Μεθοδολογίας</i>	254
Αποτελέσματα από την Ανάλυση των Έργων Τέχνης.....	254
Επικύρωση Αποτελεσμάτων και Όρια Μεθοδολογίας	255
BIBΛΙΟΓΡΑΦΙΑ.....	258
Case Studies – Important Papers for Analysis and Results.....	258
Bibliography – Electromagnetic Waves	258
Bibliography – Auger	259
Bibliography – Machine Learning – Supervised – Unsupervised.....	259
Bibliography – K Means	260
Bibliography – DBScan	261
Bibliography – NMF	262
Bibliography – PCA	263
Bibliography – GMM	264
Bibliography – XRF	264
Bibliography – Ma – XRF	268

ΕΙΚΟΝΕΣ ΚΕΙΜΕΝΟΥ

ΕΙΚΟΝΑ 1: ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΟΥ ΤΟΜΕΑ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	32
ΕΙΚΟΝΑ 2: ΠΡΩΤΕΡΓΑΤΕΣ ΤΗΣ ΤΕΧΝΙΤΗΣ ΝΟΗΜΟΣΥΝΗΣ.....	34
ΕΙΚΟΝΑ 3: ΙΣΤΟΡΙΚΗ ΠΟΡΕΙΑ ΤΗΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ.....	36
ΕΙΚΟΝΑ 4: ΑΠΕΙΚΟΝΙΣΗ ΜΕΘΟΔΩΝ, ΤΑΞΙΝΟΜΗΣΗΣ ΑΡΙΣΤΕΡΑ ΚΑΙ ΠΑΛΙΝΔΡΟΜΗΣΗΣ ΔΕΞΙΑ.....	37
ΕΙΚΟΝΑ 5: ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ (ΟΜΑΔΟΠΟΙΗΣΗΣ) ΔΕΔΟΜΕΝΩΝ	38
ΕΙΚΟΝΑ 6: ΠΙΝΑΚΑΣ ΑΝΑΚΕΦΑΛΑΙΩΣΗΣ ΓΙΑ ΤΗΝ ΜΕΘΟΔΟ ΤΗΣ PCA	47
ΕΙΚΟΝΑ 7: ΔΙΑΓΡΑΜΜΑ UML ΑΛΓΟΡΙΘΜΟΥ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ DBSCAN ΜΕ ΤΑ ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΤΟΥ	58
ΕΙΚΟΝΑ 8 - ΕΞΑΡΤΗΜΑΤΑ ΣΥΣΤΗΜΑΤΟΣ M6 JETSTREAM ΤΗΣ ΕΤΑΙΡΕΙΑΣ BRUKER [12].	68
ΕΙΚΟΝΑ 9: ΛΕΠΤΟΜΕΡΗΣ ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΚΕΦΑΛΗΣ ΜΕΤΡΗΣΗΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ M6 JETSTREAM. ΔΙΑΚΡΙΝΟΝΤΑΙ Ο ΣΩΛΗΝΑΣ ΑΚΤΙΝΩΝ-Χ (X-RAY TUBE HOUSING), Ο ΑΝΙΧΝΕΥΤΗΣ SDD, ΤΟ ΣΥΣΤΗΜΑ ΟΠΤΙΚΗΣ ΑΠΕΙΚΟΝΙΣΗΣ (SAMPLE ILLUMINATION) ΚΑΙ Η ΠΟΛΥΤΡΙΧΟΕΙΔΗΣ ΟΠΤΙΚΗ (POLYCAPILLARY LENS).....	69
ΕΙΚΟΝΑ 10 - ΑΝΙΧΝΕΥΤΗΣ ΟΛΙΣΘΗΣΗΣ ΠΥΡΙΤΙΟΥ ΚΑΙ MIN SVE (ΜΕΡΟΣ ΤΟΥ ΗΛΕΚΤΡΟΝΙΚΟΥ ΚΟΥΤΙΟΥ)..	70
ΕΙΚΟΝΑ 11 - ΔΙΑΤΟΜΗ ΕΝΟΣ ΑΝΙΧΝΕΥΤΗ ΑΚΤΙΝΩΝ-Χ SDD.	71
ΕΙΚΟΝΑ 12 - ΤΡΟΠΟΣ ΛΕΙΤΟΥΡΓΙΑΣ ΕΝΟΣ ΑΝΙΧΝΕΥΤΗ ΑΚΤΙΝΩΝ-Χ SDD.....	71
ΕΙΚΟΝΑ 13 - ΦΙΑΛΗ ΑΕΡΙΟΥ ΗΛΙΟΥ ΜΕ ΤΑ ΕΞΑΡΤΗΜΑΤΑ ΣΥΝΔΕΣΗΣ ΓΙΑ ΤΟΝ ΚΑΘΑΡΙΣΜΟ ΗΕ.	72
ΕΙΚΟΝΑ 14: ΟΙ ΤΡΙΕΣ ΘΡΗΣΚΕΥΤΙΚΕΣ ΦΟΡΗΤΕΣ ΕΙΚΟΝΕΣ ΠΟΥ ΑΠΟΤΕΛΕΣΑΝ ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΠΕΙΡΑΜΑΤΙΚΗΣ ΜΕΛΕΤΗΣ. (Α) Η ΠΑΝΑΓΙΑ Η ΟΔΗΓΗΤΡΙΑ, (Β) Ο ΆΓΙΟΣ ΙΩΑΝΝΗΣ Ο ΠΡΟΔΡΟΜΟΣ (ΣΕ ΚΑΤΑΣΤΑΣΗ ΦΘΟΡΑΣ), ΚΑΙ (C) Η ΠΙΣΩ ΜΕΡΙΑ ΤΗΣ ΕΙΚΟΝΑΣ (Β)	75
ΕΙΚΟΝΑ 15: ΔΙΑΓΡΑΜΜΑ ΡΟΗΣ ΤΗΣ ΥΒΡΙΔΙΚΗΣ ΜΕΘΟΔΟΥ NMF-PCA. Η PCA ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΑΡΧΙΚΑ ΓΙΑ ΤΗΝ ΑΠΟΜΑΚΡΥΝΣΗ ΤΟΥ ΘΟΡΥΒΟΥ ΚΑΙ ΤΗ ΜΕΙΩΣΗ ΤΗΣ ΔΙΑΣΤΑΤΙΚΟΤΗΤΑΣ. ΣΤΗ ΣΥΝΕΧΕΙΑ, Η NMF ΕΦΑΡΜΟΖΕΤΑΙ ΣΤΑ ΑΠΟΘΟΡΥΒΟΠΟΙΗΜΕΝΑ ΔΕΔΟΜΕΝΑ ΤΗΣ PCA, ΜΕΤΑΤΡΕΠΟΝΤΑΣ ΤΙΣ ΑΦΗΡΗΜΕΝΕΣ, ΟΡΘΟΓΩΝΙΕΣ ΣΥΝΙΣΤΩΣΕΣ ΣΕ ΜΗ-ΑΡΝΗΤΙΚΟΥΣ, ΦΥΣΙΚΑ ΕΡΜΗΝΕΥΣΙΜΟΥΣ ΧΑΡΤΕΣ ΑΦΘΟΝΙΑΣ ΥΛΙΚΩΝ.....	81
ΕΙΚΟΝΑ 16: ΔΙΑΓΡΑΜΜΑ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ ΚΑΙ ΡΟΗΣ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΠΛΑΙΣΙΟΥ NICE. Η ΚΥΡΙΑ ΕΦΑΡΜΟΓΗ (NICE EXPLORER GUI) ΔΙΑΧΕΙΡΙΖΕΤΑΙ ΤΗΝ ΠΛΟΗΓΗΣΗ ΣΤΑ ΑΡΧΕΙΑ HDF5 ΚΑΙ ΕΚΚΙΝΕΙ ΤΟ ΕΞΕΙΔΙΚΕΥΜΕΝΟ ANALYSIS WINDOW. ΤΟ ΠΑΡΑΘΥΡΟ ΑΝΑΛΥΣΗΣ ΥΛΟΠΟΙΕΙ ΜΙΑ ΣΤΡΑΤΗΓΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΚΤΟΣ ΜΝΗΜΗΣ, ΟΠΟΥ Ο DATA LOADER ΤΡΟΦΟΔΟΤΕΙ ΤΜΗΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ ΣΤΟΝ ΥΠΟΛΟΓΙΣΤΙΚΟ ΠΥΡΗΝΑ, ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΟΥ ΟΠΟΙΟΥ ΟΠΤΙΚΟΠΟΙΟΥΝΤΑΙ ΜΕΣΩ ΤΟΥ ANALYSIS VISUALIZER. Η ΠΡΟ-ΥΠΟΛΟΓΙΣΜΕΝΗ ΒΙΒΛΙΟΘΗΚΗ XRF_LINES.JSON ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ ΓΙΑ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΣΤΟΙΧΕΙΩΝ	85
ΕΙΚΟΝΑ 17 : ΣΤΙΓΜΙΟΤΥΠΟ ΤΗΣ ΚΥΡΙΑΣ ΔΙΕΠΑΦΗΣ ΤΟΥ ΛΟΓΙΣΜΙΚΟΥ NICE (HDF5EXPLORERGUI) ΜΕΤΑ ΤΗ ΦΟΡΤΩΣΗ ΕΝΟΣ ΑΡΧΕΙΟΥ HDF5. ΑΠΕΙΚΟΝΙΖΕΤΑΙ Η ΔΕΝΔΡΙΚΗ ΔΟΜΗ ΤΟΥ ΑΡΧΕΙΟΥ, Ο ΠΙΝΑΚΑΣ ΜΕΤΑΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΟΝ ΕΠΙΛΕΓΜΕΝΟ ΚΟΜΒΟ ΚΑΙ ΤΑ ΚΥΡΙΑ ΚΟΥΜΠΙΑ ΕΛΕΓΧΟΥ ΤΗΣ ΑΝΑΛΥΣΗΣ. 97	
ΕΙΚΟΝΑ 18: ΛΕΠΤΟΜΕΡΗΣ ΠΡΟΒΟΛΗ ΤΗΣ ΙΕΡΑΡΧΙΚΗΣ ΔΟΜΗΣ ΤΟΥ ΑΡΧΕΙΟΥ ΔΕΔΟΜΕΝΩΝ ΕΝΤΟΣ ΤΟΥ HDF5EXPLORERGUI. ΔΙΑΚΡΙΝΟΝΤΑΙ ΟΙ ΟΜΑΔΕΣ (GROUP) ΚΑΙ ΤΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ (DATASET), ΚΑΘΩΣ ΚΑΙ ΟΙ ΠΛΗΡΟΦΟΡΙΕΣ ΓΙΑ ΤΟ ΜΕΓΕΘΟΣ ΚΑΙ ΤΟΝ ΤΥΠΟ ΤΟΥΣ.	99
ΕΙΚΟΝΑ 19: ΔΙΑΔΙΚΑΣΙΑ ΕΠΙΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΑΡΙΘΜΗΤΙΚΗΣ ΑΚΡΙΒΕΙΑΣ ΣΤΟ ΛΟΓΙΣΜΙΚΟ NICE.	105
ΕΙΚΟΝΑ 20: ΠΡΟΕΙΔΟΠΟΙΗΣΗ ΠΡΟΣ ΤΟΝ ΧΡΗΣΤΗ ΟΤΑΝ ΕΠΙΛΕΓΕΤΑΙ ΕΝΑΣ ΤΥΠΟΣ ΔΕΔΟΜΕΝΩΝ GROUP ΑΝΤΙ ΓΙΑ DATASET.	105
ΕΙΚΟΝΑ 21: Ο ΔΙΑΛΟΓΟΣ "DATA PRECISION" ΠΟΥ ΕΠΙΤΡΕΠΕΙ ΣΤΟΝ ΧΡΗΣΤΗ ΝΑ ΕΠΙΛΕΞΕΙ ΤΗΝ ΑΚΡΙΒΕΙΑ ΚΙΝΗΤΗΣ ΥΠΟΔΙΑΣΤΟΛΗΣ (FLOAT16, FLOAT32, FLOAT64) ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ.....	105
ΕΙΚΟΝΑ 22: Ο ΔΙΑΛΟΓΟΣ ΕΠΙΒΕΒΑΙΩΣΗΣ ΓΙΑ ΤΗ ΦΟΡΤΩΣΗ ΜΕΓΑΛΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ.	106
ΕΙΚΟΝΑ 23: ΤΟ ΣΥΣΤΗΜΑ ΚΑΡΤΕΛΩΝ ΠΟΥ ΠΑΡΕΧΕΙ ΠΡΟΣΒΑΣΗ ΣΤΙΣ ΔΙΑΦΟΡΕΣ ΑΝΑΛΥΤΙΚΕΣ ΛΕΙΤΟΥΡΓΙΕΣ.	106
ΕΙΚΟΝΑ 24: ΠΙΝΑΚΑΣ ΜΕ ΤΑ ΜΕΤΑΔΕΔΟΜΕΝΑ ΤΗΣ ΠΕΙΡΑΜΑΤΙΚΗΣ ΜΕΤΡΗΣΗΣ (MAPPING PARAMETERS). 107	
ΕΙΚΟΝΑ 25: Η ΚΑΡΤΕΛΑ "SPECTRUM PLOT" ΕΝΤΟΣ ΤΟΥ ANALYSIS WINDOW. ΑΠΕΙΚΟΝΙΖΕΤΑΙ ΤΟ ΑΘΡΟΙΣΤΙΚΟ ΦΑΣΜΑ (SUM SPECTRUM).....	108
ΕΙΚΟΝΑ 26: ΔΙΑΔΡΑΣΤΙΚΑ ΣΤΟΙΧΕΙΑ ΕΛΕΓΧΟΥ ΓΙΑ ΤΗΝ ΑΝΑΝΕΩΣΗ, ΤΗΝ ΑΛΛΑΓΗ ΚΛΙΜΑΚΑΣ (LOG SCALE), ΤΗΝ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ ΚΑΙ ΤΗΝ ΕΜΦΑΝΙΣΗ ΤΟΥ ΓΡΑΦΗΜΑΤΟΣ ΣΕ ΞΕΧΩΡΙΣΤΟ ΠΑΡΑΘΥΡΟ (POP OUT PLOT).	108
ΕΙΚΟΝΑ 27: ΧΑΡΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ ΦΘΟΡΙΣΜΟΥ (ENERGY DISTRIBUTION MAP) ΓΙΑ ΤΗΝ ΕΙΚΟΝΑ ΤΟΥ ΑΓΙΟΥ ΙΩΑΝΝΗ. ΟΙ ΦΩΤΕΙΝΟΤΕΡΕΣ ΠΕΡΙΟΧΕΣ (ΚΙΤΡΙΝΟ) ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΕ ΥΨΗΛΟΤΕΡΗ ΣΥΝΟΛΙΚΗ ΕΚΠΟΜΠΗ ΑΚΤΙΝΩΝ-Χ, ΥΠΟΔΕΙΚΝΥΟΝΤΑΣ ΤΗΝ ΠΑΡΟΥΣΙΑ ΣΤΟΙΧΕΙΩΝ ΜΕ ΥΨΗΛΟΤΕΡΟ ΑΤΟΜΙΚΟ ΑΡΙΘΜΟ Η/ΚΑΙ ΜΕΓΑΛΥΤΕΡΟ ΠΑΧΟΣ ΥΛΙΚΟΥ.	110

ΕΙΚΟΝΑ 28: ΤΟ ΜΕΣΟ ΦΑΣΜΑ ΦΘΟΡΙΣΜΟΥ (MEAN ENERGY SPECTRUM) ΠΟΥ ΠΡΟΚΥΠΤΕΙ ΑΠΟ ΤΟΝ ΥΠΟΛΟΓΙΣΜΟ ΤΟΥ ΜΕΣΟΥ ΟΡΟΥ ΟΛΩΝ ΤΩΝ ΕΠΙΜΕΡΟΥΣ ΦΑΣΜΑΤΩΝ ΤΗΣ ΕΙΚΟΝΑΣ. ΟΙ ΚΟΡΥΦΕΣ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΙΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΓΡΑΜΜΕΣ ΕΚΠΟΜΠΗΣ ΤΩΝ ΚΥΡΙΑΡΧΩΝ ΣΤΟΙΧΕΙΩΝ.....	111
ΕΙΚΟΝΑ 29: ΔΙΑΔΙΚΑΣΙΑ ΑΝΑΣΧΗΜΑΤΙΣΜΟΥ (RESHAPE) ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟ ANALYSIS WINDOW. (ΠΑΝΩ) Ο ΠΙΝΑΚΑΣ ΕΛΕΓΧΟΥ ΠΡΙΝ ΚΑΙ ΜΕΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ, ΟΠΟΥ ΟΙ ΤΡΙΣΔΙΑΣΤΑΤΕΣ ΔΙΑΣΤΑΣΕΙΣ (364, 274, 4096) ΜΕΤΑΤΡΕΠΟΝΤΑΙ ΣΕ ΔΙΣΔΙΑΣΤΑΤΟ ΠΙΝΑΚΑ (99736, 4096) ΚΑΤΑΛΛΗΛΟ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ.....	112
ΕΙΚΟΝΑ 30: ΤΟ ΜΗΝΥΜΑ ΕΠΙΒΕΒΑΙΩΣΗΣ ΤΗΣ ΕΠΙΤΥΧΟΥΣ ΟΛΟΚΛΗΡΩΣΗΣ.....	112
ΕΙΚΟΝΑ 31: ΕΝΗΜΕΡΩΜΕΝΟ ΠΑΡΑΘΥΡΟ ΔΕΔΟΜΕΝΩΝ.	113
ΕΙΚΟΝΑ 32: ΙΣΤΟΡΙΚΟ ΔΕΔΟΜΕΝΩΝ ΔΥΝΑΤΑΙ ΝΑ ΕΠΙΛΕΧΘΕΙ.....	113
ΕΙΚΟΝΑ 33: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA) ΣΤΑ ΛΟΓΑΡΙΘΜΙΚΑ ΜΕΤΑΣΧΗΜΑΤΙΣΜΕΝΑ ΔΕΔΟΜΕΝΑ.	114
ΕΙΚΟΝΑ 34: ΣΤΙΓΜΙΟΤΥΠΟ ΤΗΣ ΚΑΡΤΕΛΑΣ "ENERGY PLOTS" ΤΟΥ ΑΝΑΠΤΥΧΘΕΝΤΟΣ ΛΟΓΙΣΜΙΚΟΥ. ΑΠΕΙΚΟΝΙΖΟΝΤΑΙ Ο ΧΑΡΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION MAP), ΤΟ ΜΕΣΟ ΦΑΣΜΑ (MEAN ENERGY SPECTRUM) ΚΑΙ ΤΑ ΦΑΣΜΑΤΑ ΤΩΝ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ (PCA EIGENVECTORS) ΓΙΑ ΤΙΣ ΕΞΙ ΠΡΩΤΕΣ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ	116
ΕΙΚΟΝΑ 35: ΓΡΑΦΙΚΕΣ ΠΑΡΑΣΤΑΣΕΙΣ ΤΩΝ 30 ΕΝΕΡΓΕΙΑΚΩΝ ΚΑΝΑΛΙΩΝ (FEATURES) ΜΕ ΤΗ ΜΕΓΑΛΥΤΕΡΗ ΣΥΝΕΙΣΦΟΡΑ (LOADING VALUE) ΓΙΑ ΚΑΘΕΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΠΡΩΤΕΣ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ (PCA COMPONENTS 1-6). ΤΑ ΔΙΑΓΡΑΜΜΑΤΑ ΑΥΤΑ ΕΠΙΤΡΕΠΟΥΝ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΤΩΝ ΦΑΣΜΑΤΙΚΩΝ ΠΕΡΙΟΧΩΝ ΠΟΥ ΚΥΡΙΑΡΧΟΥΝ ΣΕ ΚΑΘΕ ΣΥΝΙΣΤΩΣΑ.....	118
ΕΙΚΟΝΑ 36: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (CORRELATION MATRIX) ΜΕΤΑΞΥ ΤΩΝ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ ΤΩΝ ΕΞΙ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ. Η ΔΙΑΓΩΝΙΟΣ ΜΕ ΤΙΜΕΣ 1.0 ΚΑΙ ΟΙ ΣΧΕΔΟΝ ΜΗΔΕΝΙΚΕΣ ΤΙΜΕΣ ΕΚΤΟΣ ΔΙΑΓΩΝΙΟΥ ΕΠΙΒΕΒΑΙΩΝΟΥΝ ΤΗΝ ΟΡΘΟΓΩΝΙΟΤΗΤΑ ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ ΤΗΣ PCA.	119
ΕΙΚΟΝΑ 37: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΠΟΥ ΑΠΕΙΚΟΝΙΖΕΙ ΤΗ ΣΧΕΣΗ ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ (SCORE) ΤΗΣ ΣΥΝΙΣΤΩΣΑΣ 4 ΕΝΑΝΤΙ ΤΗΣ ΣΥΝΙΣΤΩΣΑΣ 6 ΓΙΑ ΚΑΘΕ PIXEL.	120
ΕΙΚΟΝΑ 38: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΕΚΤΕΛΕΣΗ ΤΟΥ ΠΛΑΙΣΙΟΥ ΑΞΙΟΛΟΓΗΣΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΟΥ ΒΕΛΤΙΣΤΟΥ ΑΡΙΘΜΟΥ ΣΥΣΤΑΔΩΝ (Κ). ΤΑ ΔΙΑΓΡΑΜΜΑΤΑ ΑΠΕΙΚΟΝΙΖΟΥΝ ΤΙΣ ΜΕΤΡΙΚΕΣ ELBOW METHOD (INERTIA), SILHOUETTE SCORE, DAVIES-BOULDIN INDEX ΚΑΙ CALINSKI-HARABASZ INDEX ΓΙΑ Κ ΑΠΟ 2 ΕΩΣ 10. Η ΣΥΝΔΥΑΣΜΕΝΗ ΑΝΑΛΥΣΗ ΥΠΟΔΕΙΚΝΥΕΙ ΩΣ ΒΕΛΤΙΣΤΟ ΕΥΡΟΣ ΓΙΑ Κ ΤΙΣ ΤΙΜΕΣ 4 ΕΩΣ 6.....	121
ΕΙΚΟΝΑ 39: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ K-MEANS (K=6) ΣΤΑ ΔΕΔΟΜΕΝΑ PCA. Ο ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ (CLUSTER MAP) ΟΠΟΥ ΚΑΘΕ ΧΡΩΜΑ ΑΝΤΙΣΤΟΙΧΕΙ ΣΕ ΜΙΑ ΣΥΣΤΑΔΑ.	122
ΕΙΚΟΝΑ 40: ΤΟ ΜΕΣΟ ΦΑΣΜΑ ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ, ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΗΣ ΣΤΟΙΧΕΙΑΚΗΣ ΣΥΣΤΑΣΗΣ.	124
ΕΙΚΟΝΑ 41: ΤΟ ΔΙΑΓΡΑΜΜΑ SCREE PLOT ΠΟΥ ΔΕΙΧΝΕΙ ΤΗ ΣΩΡΕΥΤΙΚΗ ΚΑΙ ΑΤΟΜΙΚΗ ΕΞΗΓΟΥΜΕΝΗ ΔΙΑΚΥΜΑΝΣΗ ΓΙΑ ΚΑΘΕ ΣΥΝΙΣΤΩΣΑ.....	124
ΕΙΚΟΝΑ 42: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ GMM (K=6). Η ΜΕΘΟΔΟΣ ΠΑΡΕΧΕΙ ΣΑΦΕΣΤΕΡΟ ΔΙΑΧΩΡΙΣΜΟ ΣΕ ΛΕΠΤΟΜΕΡΕΙΕΣ ΟΠΩΣ Ο ΣΤΑΥΡΟΣ ΚΑΙ ΤΑ ΦΩΤΟΣΤΕΦΑΝΑ ΣΕ ΣΥΓΚΡΙΣΗ ΜΕ ΤΗΝ K-MEANS, ΑΝ ΚΑΙ ΜΕ ΜΕΓΑΛΥΤΕΡΟ ΥΠΟΛΟΓΙΣΤΙΚΟ ΚΟΣΤΟΣ.....	125
ΕΙΚΟΝΑ 43: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΠΟΥ ΑΠΕΙΚΟΝΙΖΕΙ ΤΗΝ ΚΑΤΑΝΟΜΗ ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION) ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ (GMM LABELS) ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΟΝ ΑΛΓΟΡΙΘΜΟ GMM. Η ΟΜΑΔΟΠΟΙΗΣΗ ΤΩΝ ΣΗΜΕΙΩΝ ΣΕ ΔΙΑΚΡΙΤΕΣ ΚΑΘΕΤΕΣ ΖΩΝΕΣ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΟΝ ΕΠΙΤΥΧΗ ΔΙΑΧΩΡΙΣΜΟ ΤΩΝ PIXELS ΣΕ ΣΥΣΤΑΔΕΣ.	126
ΕΙΚΟΝΑ 44: ΜΕΣΑ ΦΑΣΜΑΤΑ ΦΘΟΡΙΣΜΟΥ ΑΚΤΙΝΩΝ-X (AVERAGE SPECTRA) ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ. ΚΑΘΕ ΚΑΜΠΥΛΗ ΑΝΤΙΠΡΟΣΩΠΕΥΕΙ ΤΟ ΜΕΣΟ ΦΑΣΜΑ ΟΛΩΝ ΤΩΝ PIXELS ΠΟΥ ΑΝΗΚΟΥΝ ΣΕ ΜΙΑ ΣΥΓΚΕΚΡΙΜΕΝΗ ΣΥΣΤΑΔΑ, ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΗ ΧΗΜΙΚΗ ΕΡΜΗΝΕΙΑ ΚΑΙ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΤΩΝ ΚΥΡΙΑΡΧΩΝ ΥΛΙΚΩΝ ΣΕ ΚΑΘΕ ΟΜΑΔΑ.	127
ΕΙΚΟΝΑ 45: ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ (CLUSTER MAP) ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ DBSCAN. ΤΟ ΑΠΟΤΕΛΕΣΜΑ ΥΠΟΔΕΙΚΝΥΕΙ ΟΤΙ Ο ΑΛΓΟΡΙΘΜΟΣ, ΜΕ ΤΙΣ ΔΕΔΟΜΕΝΕΣ ΠΑΡΑΜΕΤΡΟΥΣ, ΑΠΕΤΥΧΕ ΝΑ ΔΙΑΧΩΡΙΣΕΙ ΟΥΣΙΑΣΤΙΚΕΣ ΔΟΜΕΣ, ΤΑΞΙΝΟΜΩΝΤΑΣ ΤΗ ΣΥΝΤΡΙΠΤΙΚΗ ΠΛΕΙΟΝΟΤΗΤΑ ΤΩΝ PIXELS ΣΕ ΜΙΑ ΜΟΝΟ ΣΥΣΤΑΔΑ (ΜΩΒ) ΚΑΙ ΧΑΡΑΚΤΗΡΙΖΟΝΤΑΣ ΤΑ ΥΠΟΛΟΙΠΑ ΩΣ ΘΟΡΥΒΟ (ΓΚΡΙ).	128
ΕΙΚΟΝΑ 46: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΠΟΥ ΣΥΣΧΕΤΙΖΕΙ ΤΗΝ ΚΑΤΑΝΟΜΗ ΤΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION) ΜΕ ΤΙΣ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΟΝ DBSCAN. Η ΕΝΤΟΝΗ ΣΥΓΚΕΝΤΡΩΣΗ ΣΗΜΕΙΩΝ ΣΤΗΝ ΕΤΙΚΕΤΑ "0" ΕΠΙΒΕΒΑΙΩΝΕΙ ΟΤΙ ΤΑ ΠΕΡΙΣΣΟΤΕΡΑ PIXELS ΟΜΑΔΟΠΟΙΗΘΗΚΑΝ ΣΕ ΜΙΑ ΚΥΡΙΑ ΣΥΣΤΑΔΑ, ΕΝΩ ΤΑ ΥΠΟΛΟΙΠΑ ΕΙΝΑΙ ΔΙΑΣΠΑΡΤΑ, ΑΝΤΙΣΤΟΙΧΩΝΤΑΣ ΣΕ ΘΟΡΥΒΟ.	129
ΕΙΚΟΝΑ 47: ΑΡΙΘΜΗΤΙΚΑ ΚΑΙ ΑΠΕΙΚΟΝΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ SPECTRAL CLUSTERING. Ο ΠΙΝΑΚΑΣ ΔΕΙΧΝΕΙ ΟΤΙ ΒΡΕΘΗΚΑΝ 6 ΣΥΣΤΑΔΕΣ ΚΑΙ ΠΑΡΑΘΕΤΕΙ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ PIXELS ΠΟΥ ΑΝΗΚΟΥΝ ΣΕ ΚΑΘΕΜΙΑ, ΠΑΡΕΧΟΝΤΑΣ ΜΙΑ ΠΟΣΟΤΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΤΟΥ ΜΕΓΕΘΟΥΣ ΤΩΝ ΟΜΑΔΩΝ.	129

ΕΙΚΟΝΑ 48: ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ (CLUSTER MAP) ΠΟΥ ΠΡΟΕΚΥΨΕ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ SPECTRAL CLUSTERING ($k=6$). ΚΑΘΕ ΧΡΩΜΑ ΑΝΤΙΠΡΟΣΩΠΕΥΕΙ ΜΙΑ ΔΙΑΚΡΙΤΗ ΣΥΣΤΑΔΑ, ΑΠΟΚΑΛΥΠΤΟΝΤΑΣ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΥΛΙΚΩΝ ΟΜΑΔΩΝ ΠΑΝΩ ΣΤΗΝ ΕΠΙΦΑΝΕΙΑ ΤΗΣ ΕΙΚΟΝΑΣ.	130
ΕΙΚΟΝΑ 49: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΠΟΥ ΑΠΕΙΚΟΝΙΖΕΙ ΤΗΝ ΚΑΤΑΝΟΜΗ ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION) ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ (SPECTRAL LABELS) ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΟΝ ΑΛΓΟΡΙΘΜΟ SPECTRAL CLUSTERING. Η ΟΜΑΔΟΠΟΙΗΣΗ ΤΩΝ ΣΗΜΕΙΩΝ ΣΕ ΔΙΑΚΡΙΤΕΣ ΚΑΘΕΤΕΣ ΖΩΝΕΣ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΟΝ ΕΠΙΤΥΧΗ ΔΙΑΧΩΡΙΣΜΟ ΤΩΝ PIXELS ΣΕ ΣΥΣΤΑΔΕΣ.	131
ΕΙΚΟΝΑ 50: ΜΕΣΑ ΦΑΣΜΑΤΑ ΦΘΟΡΙΣΜΟΥ ΑΚΤΙΝΩΝ-Χ (AVERAGE SPECTRA) ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΑΝΑΛΥΣΗ SPECTRAL CLUSTERING. ΚΑΘΕ ΚΑΜΠΥΛΗ ΑΝΤΙΠΡΟΣΩΠΕΥΕΙ ΤΟ ΜΕΣΟ ΦΑΣΜΑ ΟΛΩΝ ΤΩΝ PIXELS ΠΟΥ ΑΝΗΚΟΥΝ ΣΕ ΜΙΑ ΣΥΓΚΕΚΡΙΜΕΝΗ ΣΥΣΤΑΔΑ, ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΗ ΧΗΜΙΚΗ ΕΡΜΗΝΕΙΑ ΚΑΙ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΤΩΝ ΚΥΡΙΑΡΧΩΝ ΥΛΙΚΩΝ ΣΕ ΚΑΘΕ ΟΜΑΔΑ.	131
ΕΙΚΟΝΑ 51: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ NMF. ΟΙ ΧΑΡΤΕΣ ΤΩΝ ΕΞΙ ΣΥΝΙΣΤΩΣΩΝ NMF, ΟΙ ΟΠΟΙΟΙ ΛΟΓΩ ΤΗΣ ΜΗ-ΑΡΝΗΤΙΚΟΤΗΤΑΣ ΜΠΟΡΟΥΝ ΝΑ ΕΡΜΗΝΕΥΘΟΥΝ ΩΣ ΧΑΡΤΕΣ ΑΦΘΟΝΙΑΣ ΥΛΙΚΩΝ.	132
ΕΙΚΟΝΑ 52: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (HEATMAP) ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΑΝΑΛΥΣΗ NMF. ΣΕ ΑΝΤΙΘΕΣΗ ΜΕ ΤΗΝ PCA, ΟΙ ΣΥΝΙΣΤΩΣΕΣ ΤΗΣ NMF ΔΕΝ ΕΙΝΑΙ ΚΑΤ' ΑΝΑΓΚΗ ΟΡΘΟΓΩΝΙΕΣ. ΟΙ ΘΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ (ΓΑΛΑΖΙΕΣ ΑΠΟΧΡΩΣΕΙΣ) ΥΠΟΔΕΙΚΝΟΥΝ ΚΟΙΝΑ ΦΑΣΜΑΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ Η ΧΩΡΙΚΗ ΣΥΝΥΠΑΡΞΗ ΥΛΙΚΩΝ ΜΕΤΑΞΥ ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ.	134
ΕΙΚΟΝΑ 53: ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ (CLUSTER MAP) ΠΟΥ ΠΡΟΕΚΥΨΕ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ K-MEANS ($k=6$) ΣΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ NMF. ΚΑΘΕ ΧΡΩΜΑ ΑΝΤΙΠΡΟΣΩΠΕΥΕΙ ΜΙΑ ΔΙΑΚΡΙΤΗ ΥΛΙΚΗ ΟΜΑΔΑ, ΑΠΟΚΑΛΥΠΤΟΝΤΑΣ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΚΑΙ ΤΙΣ ΣΧΕΣΕΙΣ ΜΕΤΑΞΥ ΤΩΝ ΔΙΑΦΟΡΩΝ ΧΡΩΣΤΙΚΩΝ ΚΑΙ ΥΛΙΚΩΝ ΠΡΟΕΤΟΙΜΑΣΙΑΣ.	135
ΕΙΚΟΝΑ 54: ΜΕΣΑ ΦΑΣΜΑΤΑ ΦΘΟΡΙΣΜΟΥ ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΟΜΑΔΟΠΟΙΗΣΗ K-MEANS ΕΠΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ NMF. Η ΑΝΑΛΥΣΗ ΤΩΝ ΦΑΣΜΑΤΩΝ ΕΠΙΤΡΕΠΕΙ ΤΗΝ ΤΑΥΤΟΠΟΙΗΣΗ ΤΗΣ ΧΗΜΙΚΗΣ ΣΥΣΤΑΣΗΣ ΚΑΘΕ ΟΜΑΔΑΣ ΠΟΥ ΔΙΑΧΩΡΙΣΤΗΚΕ.	136
ΕΙΚΟΝΑ 55: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION) ΓΙΑ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΟΝ ΑΛΓΟΡΙΘΜΟ K-MEANS. Η ΣΑΦΗΣ ΟΜΑΔΟΠΟΙΗΣΗ ΤΩΝ ΣΗΜΕΙΩΝ ΣΕ ΚΑΘΕΤΕΣ ΖΩΝΕΣ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΗΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ ΤΟΥ ΔΙΑΧΩΡΙΣΜΟΥ.	137
ΕΙΚΟΝΑ 56: ΦΑΣΜΑ ΑΘΡΟΙΣΜΑΤΟΣ (SUM SPECTRUM) ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕΤΑ ΑΠΟ ΛΟΓΑΡΙΘΜΙΚΟ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟ $\log(\text{value}+1)$. Η ΕΦΑΡΜΟΓΗ ΤΗΣ ΛΟΓΑΡΙΘΜΙΚΗΣ ΚΛΙΜΑΚΑΣ ΕΝΙΣΧΥΕΙ ΤΙΣ ΚΟΡΥΦΕΣ ΧΑΜΗΛΗΣ ΕΝΤΑΣΗΣ, ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΗΝ ΚΑΛΥΤΕΡΗ ΟΠΤΙΚΗ ΕΠΙΘΕΩΡΗΣΗ ΤΩΝ ΙΧΝΟΣΤΟΙΧΕΙΩΝ ΠΟΥ ΕΝΔΕΧΕΤΑΙ ΝΑ ΧΑΝΟΝΤΑΙ ΣΤΗ ΓΡΑΜΜΙΚΗ ΚΛΙΜΑΚΑ.	138
ΕΙΚΟΝΑ 57: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΑΝΑΛΥΣΗΣ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA) ΣΤΑ ΛΟΓΑΡΙΘΜΙΚΑ ΜΕΤΑΣΧΗΜΑΤΙΣΜΕΝΑ ΔΕΔΟΜΕΝΑ. (ΠΑΝΩ) ΤΟ ΔΙΑΓΡΑΜΜΑ SCREE PLOT ΠΟΥ ΔΕΙΧΝΕΙ ΤΗ ΣΩΡΕΥΤΙΚΗ ΚΑΙ ΑΤΟΜΙΚΗ ΕΞΗΓΟΥΜΕΝΗ ΔΙΑΚΥΜΑΝΣΗ ΓΙΑ ΚΑΘΕ ΣΥΝΙΣΤΩΣΑ. (ΚΑΤΩ) ΟΙ ΧΩΡΙΚΟΙ ΧΑΡΤΕΣ (SCORE MAPS) ΓΙΑ ΤΙΣ ΕΞΙ ΠΡΩΤΕΣ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ.	139
ΕΙΚΟΝΑ 58: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (CORRELATION MATRIX) ΜΕΤΑΞΥ ΤΩΝ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ (EIGENVECTORS) ΤΩΝ ΕΞΙ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΑΝΑΛΥΣΗ ΤΩΝ ΛΟΓΑΡΙΘΜΙΚΩΝ ΔΕΔΟΜΕΝΩΝ. ΟΙ ΤΙΜΕΣ ΚΟΝΤΑ ΣΤΟ ΜΗΔΕΝ ΕΚΤΟΣ ΤΗΣ ΔΙΑΓΩΝΙΟΥ ΕΠΙΒΕΒΑΙΩΝΟΥΝ ΤΗΝ ΟΡΘΟΓΩΝΙΟΤΗΤΑ ΤΗΣ ΜΕΘΟΔΟΥ.	140
ΕΙΚΟΝΑ 59: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΕΚΤΕΛΕΣΗ ΤΟΥ ΠΛΑΙΣΙΟΥ ΑΞΙΟΛΟΓΗΣΗΣ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΟΥ ΒΕΛΤΙΣΤΟΥ ΑΡΙΘΜΟΥ ΣΥΣΤΑΔΩΝ (k) ΣΤΑ ΛΟΓΑΡΙΘΜΙΚΑ ΔΕΔΟΜΕΝΑ PCA. ΟΙ ΜΕΤΡΙΚΕΣ ΥΠΟΔΕΙΚΝΟΥΝ ΕΝΑ ΒΕΛΤΙΣΤΟ ΕΥΡΟΣ ΓΙΑ k ΜΕΤΑΞΥ 4 ΚΑΙ 7, ΕΠΙΒΕΒΑΙΩΝΟΝΤΑΣ ΤΗ ΔΙΑΤΗΡΗΣΗ ΤΟΥ ΑΡΙΘΜΟΥ ΤΩΝ 6 ΣΥΣΤΑΔΩΝ ΓΙΑ ΣΥΝΕΠΕΙΑ.	141
ΕΙΚΟΝΑ 60: ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ (CLUSTER MAP) ΠΟΥ ΠΡΟΚΥΠΤΕΙ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ K-MEANS ($k=6$) ΣΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ PCA ΕΠΙ ΤΩΝ ΛΟΓΑΡΙΘΜΙΚΩΝ ΔΕΔΟΜΕΝΩΝ. Ο ΧΑΡΤΗΣ ΑΥΤΟΣ ΑΠΟΚΑΛΥΠΤΕΙ ΚΑΛΥΤΕΡΟ ΔΙΑΧΩΡΙΣΜΟ ΣΕ ΠΕΡΙΟΧΕΣ ΜΕ ΥΛΙΚΑ ΥΨΗΛΟΥ Z (Π.Χ. ΧΡΥΣΟΣ, ΜΟΛΥΒΔΟΣ) ΣΕ ΣΥΓΚΡΙΣΗ ΜΕ ΤΗΝ ΑΝΑΛΥΣΗ ΣΤΗ ΓΡΑΜΜΙΚΗ ΚΛΙΜΑΚΑ.	142
ΕΙΚΟΝΑ 61: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ ΑΝΑ ΣΥΣΤΑΔΑ K-MEANS.	143
ΕΙΚΟΝΑ 62: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΗΣ ΜΕΘΟΔΟΥ NMF ΣΤΑ ΛΟΓΑΡΙΘΜΙΚΑ ΔΕΔΟΜΕΝΑ.	144
ΕΙΚΟΝΑ 63: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ NMF. ΣΕ ΑΝΤΙΘΕΣΗ ΜΕ ΤΗΝ PCA, ΟΙ ΣΥΝΙΣΤΩΣΕΣ ΤΗΣ NMF ΔΕΝ ΕΙΝΑΙ ΚΑΤ' ΑΝΑΓΚΗ ΟΡΘΟΓΩΝΙΕΣ, ΚΑΙ Ο ΠΙΝΑΚΑΣ ΑΠΟΚΑΛΥΠΤΕΙ ΜΕΤΡΙΕΣ ΕΩΣ ΙΣΧΥΡΕΣ ΘΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ, ΥΠΟΔΕΙΚΝΟΥΝΤΑΣ ΚΟΙΝΑ ΦΑΣΜΑΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ Η ΧΩΡΙΚΗ ΣΥΝΥΠΑΡΞΗ ΜΕΤΑΞΥ ΟΡΙΣΜΕΝΩΝ ΥΛΙΚΩΝ.	145
ΕΙΚΟΝΑ 64: ΟΠΤΙΚΟΠΟΙΗΣΗ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ K-MEANS ($k=6$) ΠΟΥ ΕΦΑΡΜΟΣΤΗΚΕ ΣΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ NMF. (ΠΑΝΩ) Ο ΤΕΛΙΚΟΣ ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ. (ΚΑΤΩ) ΤΟ ΔΙΑΓΡΑΜΜΑ ΚΑΤΑΝΟΜΗΣ ΤΗΣ ΕΝΕΡΓΕΙΑΣ ΑΝΑ ΣΥΣΤΑΔΑ.	146

ΕΙΚΟΝΑ 65: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΥΒΡΙΔΙΚΗ ΜΕΘΟΔΟ NMF-PCA. ΟΙ ΧΑΡΤΕΣ ΤΩΝ ΕΞΙ ΣΥΝΙΣΤΩΣΩΝ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΠΡΟΣΦΕΡΟΥΝ ΕΞΑΙΡΕΤΙΚΗ ΕΥΚΡΙΝΕΙΑ ΚΑΙ ΑΜΕΣΗ ΦΥΣΙΚΟΧΗΜΙΚΗ ΕΡΜΗΝΕΙΑ, ΔΙΑΧΩΡΙΖΟΝΤΑΣ ΜΕ ΣΑΦΗΝΕΙΑ ΤΟ ΥΠΟΣΤΡΩΜΑ, ΤΑ ΦΩΤΟΣΤΕΦΑΝΑ, ΤΑ ΕΝΔΥΜΑΤΑ ΚΑΙ ΤΑ ΣΑΡΚΩΜΑΤΑ.	147
ΕΙΚΟΝΑ 66: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ K-MEANS (K=6) ΣΤΑ ΔΕΔΟΜΕΝΑ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΥΒΡΙΔΙΚΗ ΑΝΑΛΥΣΗ NMF-PCA. (ΑΡΙΣΤΕΡΑ) Ο ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ ΑΠΕΙΚΟΝΙΖΕΙ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΥΛΙΚΩΝ ΟΜΑΔΩΝ. (ΔΕΞΙΑ) ΤΟ ΙΣΤΟΓΡΑΜΜΑ ΜΕΓΕΘΟΥΣ ΔΕΙΧΝΕΙ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ PIXELS ΣΕ ΚΑΘΕ ΣΥΣΤΑΔΑ, ΠΑΡΕΧΟΝΤΑΣ ΠΟΣΟΤΙΚΗ ΕΚΤΙΜΗΣΗ ΤΗΣ ΕΠΙΚΡΑΤΗΣΗΣ ΚΑΘΕ ΥΛΙΚΟΥ.	150
ΕΙΚΟΝΑ 67: ΜΕΣΑ ΦΑΣΜΑΤΑ ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ ΤΗΣ ΑΝΑΛΥΣΗΣ K-MEANS ΕΠΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ NMF-PCA. ΟΙ ΔΙΑΚΡΙΤΕΣ ΦΑΣΜΑΤΙΚΕΣ ΥΠΟΓΡΑΦΕΣ ΕΠΙΤΡΕΠΟΥΝ ΤΗΝ ΑΠΟΔΟΣΗ ΧΗΜΙΚΗΣ ΤΑΥΤΟΤΗΤΑΣ ΣΕ ΚΑΘΕ ΧΡΩΜΑΤΙΚΗ ΟΜΑΔΑ ΤΟΥ ΘΕΜΑΤΙΚΟΥ ΧΑΡΤΗ.	150
ΕΙΚΟΝΑ 68: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΤΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION) ΑΝΑ ΣΥΣΤΑΔΑ K-MEANS. Η ΧΡΩΜΑΤΙΚΗ ΚΑΙΜΑΚΑ (PIXEL COURT) ΑΠΕΙΚΟΝΙΖΕΙ ΤΗΝ ΠΥΚΝΟΤΗΤΑ ΤΩΝ ΣΗΜΕΙΩΝ, ΔΕΙΧΝΟΝΤΑΣ ΤΙΣ ΠΙΟ ΣΥΧΝΕΣ ΤΙΜΕΣ ΕΝΤΑΣΗΣ ΓΙΑ ΚΑΘΕ ΟΜΑΔΑ.	151
ΕΙΚΟΝΑ 69: ΟΠΤΙΚΗ ΑΠΟΜΟΝΩΣΗ ΤΗΣ "CLUSTER 1" ΑΠΟ ΤΟΝ ΘΕΜΑΤΙΚΟ ΧΑΡΤΗ ΤΗΣ ΥΒΡΙΔΙΚΗΣ ΜΕΘΟΔΟΥ NMF-PCA. Η ΠΕΡΙΟΧΗ ΕΝΔΙΑΦΕΡΟΝΤΟΣ (ΠΟΡΤΟΚΑΛΙ) ΑΝΑΔΕΙΚΝΥΕΤΑΙ, ΕΝΩ ΟΙ ΥΠΟΛΟΙΠΕΣ ΣΥΣΤΑΔΕΣ ΕΜΦΑΝΙΖΟΝΤΑΙ ΜΕ ΟΥΔΕΤΕΡΟ ΧΡΩΜΑ (ΑΝΟΙΧΤΟ ΓΚΡΙ) ΓΙΑ ΤΗΝ ΠΡΟΕΤΟΙΜΑΣΙΑ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΥΠΟ-ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ (SUB-CLUSTERING).	152
ΕΙΚΟΝΑ 70: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΙΕΡΑΡΧΙΚΗΣ ΑΝΑΛΥΣΗΣ (SUB-CLUSTERING) ΠΟΥ ΕΦΑΡΜΟΣΤΗΚΕ ΣΤΗΝ "CLUSTER 1". Ο ΤΕΛΙΚΟΣ ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ ΟΠΟΥ ΔΙΑΚΡΙΝΟΝΤΑΙ ΕΞΙ ΔΙΑΚΡΙΤΕΣ ΥΠΟ-ΟΜΑΔΕΣ (SUB-C 0-5), ΑΠΟΚΑΛΥΠΤΟΝΤΑΣ ΛΕΠΤΟΜΕΡΗ ΣΤΡΩΜΑΤΟΓΡΑΦΙΑ ΕΝΤΟΣ ΤΗΣ ΑΡΧΙΚΗΣ ΠΕΡΙΟΧΗΣ. .	153
ΕΙΚΟΝΑ 71: ΤΟ ΙΣΤΟΓΡΑΜΜΑ ΠΟΥ ΑΠΕΙΚΟΝΙΖΕΙ ΤΟ ΠΛΗΘΟΣ ΤΩΝ PIXELS ΓΙΑ ΚΑΘΕ ΥΠΟ-ΣΥΣΤΑΔΑ.	153
ΕΙΚΟΝΑ 72: ΟΠΤΙΚΗ ΑΠΟΜΟΝΩΣΗ ΤΗΣ "CLUSTER 3" (ΚΟΚΚΙΝΟ), Η ΟΠΟΙΑ ΑΝΤΙΣΤΟΙΧΕΙ ΚΥΡΙΩΣ ΣΤΑ ΕΝΔΥΜΑΤΑ, ΠΡΙΝ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ SUB-CLUSTERING ΓΙΑ ΤΗΝ ΠΕΡΑΙΤΕΡΩ ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΧΗΜΙΚΗΣ ΤΗΣ ΣΥΣΤΑΣΗΣ.	154
ΕΙΚΟΝΑ 73: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΙΕΡΑΡΧΙΚΗΣ ΑΝΑΛΥΣΗΣ (SUB-CLUSTERING) ΠΟΥ ΕΦΑΡΜΟΣΤΗΚΕ ΣΤΗΝ "CLUSTER 3". (ΠΑΝΩ) Ο ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ ΠΟΥ ΑΠΟΚΑΛΥΠΤΕΙ ΤΗΝ ΕΣΩΤΕΡΙΚΗ ΔΟΜΗ ΤΗΣ ΣΥΣΤΑΔΑΣ ΜΕ ΕΞΙ ΝΕΕΣ ΥΠΟ-ΟΜΑΔΕΣ. (ΚΑΤΩ) ΤΟ ΙΣΤΟΓΡΑΜΜΑ ΜΕΓΕΘΟΥΣ ΓΙΑ ΚΑΘΕ ΥΠΟ-ΣΥΣΤΑΔΑ.	155
ΕΙΚΟΝΑ 74: ΑΡΧΙΚΟ ΦΑΣΜΑ ΣΤΟ ΠΡΟΓΡΑΜΜΑ ΕΝΤΟΠΙΣΜΟΥ ΣΤΟΙΧΕΙΩΝ.	156
ΕΙΚΟΝΑ 75: ΔΙΑΔΙΚΑΣΙΑ ΒΑΘΜΟΝΟΜΗΣΗΣ ΚΑΙ ΤΑΥΤΟΠΟΙΗΣΗΣ ΣΤΟΙΧΕΙΩΝ ΣΤΟ ΜΕΣΟ ΦΑΣΜΑ. (ΠΑΝΩ) ΤΟ ΑΡΧΙΚΟ ΜΕΣΟ ΦΑΣΜΑ ΩΣ ΣΥΝΑΡΤΗΣΗ ΤΟΥ ΑΡΙΘΜΟΥ ΚΑΝΑΛΙΟΥ ΤΟΥ ΑΝΙΧΝΕΥΤΗ. (ΚΑΤΩ) ΤΟ ΒΑΘΜΟΝΟΜΗΜΕΝΟ ΦΑΣΜΑ ΣΕ ΚΛΙΜΑΚΑ ΕΝΕΡΓΕΙΑΣ (KEV), ΜΕ ΤΙΣ ΘΕΩΡΗΤΙΚΕΣ ΘΕΣΕΙΣ ΤΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΓΡΑΜΜΩΝ ΕΚΠΟΜΠΗΣ ΓΙΑ ΕΠΙΛΕΓΜΕΝΑ ΣΤΟΙΧΕΙΑ (CU, S, ZN, FE,...) ΝΑ ΕΠΙΣΗΜΑΙΝΟΝΤΑΙ ΜΕ ΔΙΑΚΕΚΟΜΜΕΝΕΣ ΓΡΑΜΜΕΣ.	156
ΕΙΚΟΝΑ 76: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΨΕΥΔΑΡΓΥΡΟ (ZN), ΠΟΥ ΑΠΕΙΚΟΝΙΖΟΥΝ ΤΗΝ ΕΝΤΑΣΗ ΤΩΝ ΓΡΑΜΜΩΝ ΕΚΠΟΜΠΗΣ ZN-K EDGE, KA1 ΚΑΙ KB1. Η ΣΠΟΡΑΔΙΚΗ ΚΑΙ ΤΟΠΙΚΑ ΠΕΡΙΟΡΙΣΜΕΝΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΣΤΟΙΧΕΙΟΥ ΑΠΟΤΕΛΕΙ ΙΣΧΥΡΗ ΕΝΔΕΙΞΗ ΜΕΤΑΓΕΝΕΣΤΕΡΩΝ ΕΠΕΜΒΑΣΕΩΝ ΣΥΝΤΗΡΗΣΗΣ ΜΕ ΧΡΗΣΗ ΛΕΥΚΟΥ ΤΟΥ ΨΕΥΔΑΡΓΥΡΟΥ (ZNO).	157
ΕΙΚΟΝΑ 77: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟ ΘΕΙΟ (S), ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ S-K EDGE, KA1, KA2 ΚΑΙ KB1. Η ΔΙΑΧΥΤΗ ΚΑΙ ΣΧΕΔΟΝ ΚΑΘΟΛΙΚΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΘΕΙΟΥ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΗ ΧΡΗΣΗ ΓΥΨΟΥ (CASO ₄ ·2H ₂ O) ΩΣ ΥΛΙΚΟ ΤΟΥ ΥΠΟΣΤΡΩΜΑΤΟΣ ΠΡΟΕΤΟΙΜΑΣΙΑΣ (GESSO).	158
ΕΙΚΟΝΑ 78: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΜΟΛΥΒΔΟ (PB), ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ LA1, LA2 ΚΑΙ LB1. Η ΥΨΗΛΗ ΣΥΓΚΕΝΤΡΩΣΗ ΤΟΥ ΜΟΛΥΒΔΟΥ ΣΤΑ ΣΑΡΚΩΜΑΤΑ ΚΑΙ ΣΤΑ "ΦΩΤΙΣΜΑΤΑ" ΤΩΝ ΕΝΔΥΜΑΤΩΝ ΤΑΥΤΟΠΟΙΕΙ ΤΗ ΧΡΗΣΗ ΤΟΥ ΛΕΥΚΟΥ ΤΟΥ ΜΟΛΥΒΔΟΥ (2PbCO ₃ ·Pb(OH) ₂), ΒΑΣΙΚΗΣ ΛΕΥΚΗΣ ΧΡΩΣΤΙΚΗΣ ΣΤΗΝ ΠΑΡΑΔΟΣΙΑΚΗ ΑΓΙΟΓΡΑΦΙΑ.	159
ΕΙΚΟΝΑ 79: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΣΙΔΗΡΟ (FE), ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ Fe-K EDGE, KA1, KA2 ΚΑΙ KB1. Η ΚΑΤΑΝΟΜΗ ΤΟΥ ΣΙΔΗΡΟΥ ΣΥΝΔΕΕΤΑΙ ΑΜΕΣΑ ΜΕ ΤΙΣ ΚΑΦΕ-ΚΟΚΚΙΝΕΣ ΚΑΙ ΩΧΡΕΣ ΠΕΡΙΟΧΕΣ, ΥΠΟΔΕΙΚΝΥΟΝΤΑΣ ΤΗ ΧΡΗΣΗ ΓΙΝΩΝ ΧΡΩΣΤΙΚΩΝ (ΩΧΡΕΣ) ΜΕ ΒΑΣΗ ΤΑ ΟΞΕΙΔΙΑ ΤΟΥ ΣΙΔΗΡΟΥ.	160
ΕΙΚΟΝΑ 80: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΧΑΛΚΟ (CU), ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ Cu-KA1, KA2 ΚΑΙ KB1. Η ΣΤΟΧΕΥΜΕΝΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΧΑΛΚΟΥ ΤΑΥΤΟΠΟΙΕΙ ΤΗ ΧΡΗΣΗ ΧΡΩΣΤΙΚΩΝ ΟΠΩΣ Ο ΜΑΛΑΧΙΤΗΣ Η Ο ΑΖΟΥΡΙΤΗΣ ΓΙΑ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΛΕΠΤΟΜΕΡΕΙΕΣ.	161
ΕΙΚΟΝΑ 81: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΧΡΥΣΟ (AU), ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ Au-LA1, LA2 ΚΑΙ LB1. Η ΑΠΟΚΛΕΙΣΤΙΚΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΧΡΥΣΟΥ ΣΤΑ ΦΩΤΟΣΤΕΦΑΝΑ ΚΑΙ ΣΤΙΣ ΔΙΑΚΟΣΜΗΤΙΚΕΣ ΓΡΑΜΜΕΣ (ΧΡΥΣΟΚΟΝΔΥΛΙΕΣ) ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΗΝ ΤΕΧΝΙΚΗ ΤΗΣ ΕΠΙΧΡΥΣΩΣΗΣ ΜΕ ΦΥΛΛΟ ΧΡΥΣΟΥ.	162

ΕΙΚΟΝΑ 82: : ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΧΑΛΚΟ (CA), ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ CA-KA1, KA2 ΚΑΙ KB1	163
ΕΙΚΟΝΑ 83: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (HEATMAP) ΜΕΤΑΞΥ ΤΩΝ ΚΥΡΙΟΤΕΡΩΝ ΣΤΟΙΧΕΙΑΚΩΝ ΓΡΑΜΜΩΝ ΠΟΥ ΑΝΙΧΝΕΥΘΗΚΑΝ. ΟΙ ΘΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ (ΚΟΚΚΙΝΟ) ΥΠΟΔΕΙΚΝΟΥΝ ΧΩΡΙΚΗ ΣΥΝΥΠΑΡΞΗ (Π.Χ., ΑΝΑΜΕΙΞΗ ΧΡΩΣΤΙΚΩΝ), ΕΝΩ ΟΙ ΑΡΝΗΤΙΚΕΣ (ΜΠΛΕ) ΑΜΟΙΒΑΙΟ ΑΠΟΚΛΕΙΣΜΟ.	164
ΕΙΚΟΝΑ 84: ΤΡΙΧΡΩΜΑΤΙΚΕΣ ΑΠΕΙΚΟΝΙΣΕΙΣ (RGB OVERLAYS) ΠΟΥ ΑΠΟΚΑΛΥΠΤΟΥΝ ΤΗ ΣΥΝΥΠΑΡΞΗ ΚΑΙ ΤΗΝ ΑΛΛΗΛΟΕΠΙΚΑΛΥΨΗ ΣΤΟΙΧΕΙΩΝ. (ΠΑΝΩ 84.Α) OVERLAY ΤΩΝ AU-LA1 (ΚΟΚΚΙΝΟ), PB-LA1 (ΠΡΑΣΙΝΟ) ΚΑΙ FE-KA1 (ΜΠΛΕ). (ΚΑΤΩ 84.Β) OVERLAY ΤΩΝ CU-KA1 (ΚΟΚΚΙΝΟ), FE-KA1 (ΠΡΑΣΙΝΟ) ΚΑΙ PB-LA1 (ΜΠΛΕ).	165
ΕΙΚΟΝΑ 85: ΤΡΙΧΡΩΜΑΤΙΚΗ ΑΠΕΙΚΟΝΙΣΗ (RGB OVERLAY) ΤΩΝ S-KA1 (ΚΟΚΚΙΝΟ), CU-KA1 (ΠΡΑΣΙΝΟ) ΚΑΙ CU-KB1 (ΜΠΛΕ), Η ΟΠΟΙΑ ΑΝΑΔΕΙΚΝΥΕΙ ΤΙΣ ΠΕΡΙΟΧΕΣ ΠΟΥ ΠΕΡΙΕΧΟΥΝ ΤΑΥΤΟΧΡΟΝΑ ΘΕΙΟ ΚΑΙ ΧΑΛΚΟ.	166
ΕΙΚΟΝΑ 86: Η ΕΙΚΟΝΑ ΤΟΥ ΑΓΙΟΥ ΙΩΑΝΝΗ ΤΟΥ ΠΡΟΔΡΟΜΟΥ, ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΜΕΛΕΤΗΣ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ. Η ΕΜΦΑΝΗΣ ΦΘΟΡΑ ΣΤΟ ΕΡΓΟ ΚΑΘΙΣΤΑ ΤΗΝ ΜΗ-ΚΑΤΑΣΤΡΟΦΙΚΗ ΑΝΑΛΥΣΗ ΜΑ-XRF ΑΠΑΡΑΙΤΗΤΗ ΓΙΑ ΤΗ ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΥΛΙΚΗΣ ΤΟΥ ΣΥΣΤΑΣΗΣ.	168
ΕΙΚΟΝΑ 87: ΛΕΙΠΤΟΜΕΡΗΣ ΚΑΤΑΓΡΑΦΗ ΤΩΝ ΠΕΙΡΑΜΑΤΙΚΩΝ ΠΑΡΑΜΕΤΡΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ ΓΙΑ ΤΗ ΣΑΡΩΣΗ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ (HIGH RESOLUTION) ΤΗΣ ΕΙΚΟΝΑΣ ΤΟΥ ΑΓΙΟΥ ΙΩΑΝΝΗ.	169
ΕΙΚΟΝΑ 88: ΦΑΣΜΑ ΑΘΡΟΙΣΜΑΤΟΣ (SUM SPECTRUM) ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ ΜΕΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΦΙΛΤΡΟΥ SAVITZKY-GOLAY. Η ΥΨΗΛΗ ΠΟΙΟΤΗΤΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ (ΑΥΞΗΜΕΝΟΣ ΛΟΓΟΣ ΣΗΜΑΤΟΣ ΠΡΟΣ ΘΟΡΥΒΟ) ΕΙΝΑΙ ΕΜΦΑΝΗΣ ΑΠΟ ΤΙΣ ΕΥΔΙΑΚΡΙΤΕΣ ΚΟΡΥΦΕΣ.	170
ΕΙΚΟΝΑ 89: ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΜΕΘΟΔΟΥ PCA ΓΙΑ ΤΗΝ ΕΠΙΛΟΓΗ ΤΟΥ ΚΑΤΑΛΛΗΛΟΥ ΑΡΙΘΜΟΥ ΣΥΝΙΣΤΩΣΩΝ ΣΤΗΝ ΑΝΑΛΥΣΗ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ. (ΠΑΝΩ) ΙΣΤΟΓΡΑΜΜΑ ΤΗΣ ΕΞΗΓΟΥΜΕΝΗΣ ΔΙΑΚΥΜΑΝΣΗΣ ΑΝΑ ΣΥΝΙΣΤΩΣΑ. (ΚΑΤΩ) ΤΟ ΔΙΑΓΡΑΜΜΑ SCREE PLOT ΠΟΥ ΔΕΙΧΝΕΙ ΤΗ ΣΩΡΕΥΤΙΚΗ ΔΙΑΚΥΜΑΝΣΗ, ΥΠΟΔΕΙΚΝΥΟΝΤΑΣ ΟΤΙ ΟΙ ΠΡΩΤΕΣ 6 ΣΥΝΙΣΤΩΣΕΣ ΔΙΑΤΗΡΟΥΝ ΠΑΝΩ ΑΠΟ ΤΟ 80% ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ.	171
ΕΙΚΟΝΑ 90: ΧΩΡΙΚΟΙ ΧΑΡΤΕΣ (SCORE MAPS) ΓΙΑ ΤΙΣ ΕΞΙ ΠΡΩΤΕΣ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ INCREMENTAL PCA ΣΤΑ ΔΕΔΟΜΕΝΑ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ. ΟΙ ΧΑΡΤΕΣ ΑΠΟΚΑΛΥΠΤΟΥΝ ΛΕΙΠΤΟΜΕΡΕΙΣ ΥΦΕΣ (TEXTURES), ΟΠΩΣ ΠΙΝΕΛΙΕΣ ΚΑΙ ΡΩΓΜΕΣ (CRAQUELURE).	173
ΕΙΚΟΝΑ 91: ΟΠΤΙΚΗ ΕΠΙΣΚΟΠΗΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ ΚΑΙ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΗΣ PCA. (ΠΑΝΩ ΑΡΙΣΤΕΡΑ) ΧΑΡΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ. (ΠΑΝΩ ΔΕΞΙΑ) ΜΕΣΟ ΦΑΣΜΑ. (ΚΑΤΩ) ΦΑΣΜΑΤΑ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ (PCA EIGENVECTORS/LOADINGS) ΓΙΑ ΤΙΣ ΕΞΙ ΚΥΡΙΕΣ ΣΥΝΙΣΤΩΣΕΣ.	174
ΕΙΚΟΝΑ 92: ΑΞΙΟΛΟΓΗΣΗ ΓΙΑ ΤΟΝ ΒΕΛΤΙΣΤΟ ΑΡΙΘΜΟ ΣΥΣΤΑΔΩΝ (K) ΣΤΑ ΔΕΔΟΜΕΝΑ PCA ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ.	175
ΕΙΚΟΝΑ 93: ΚΑΤΑΝΟΜΗ ΤΟΥ ΠΛΗΘΟΥΣ ΤΩΝ PIXELS ΓΙΑ ΤΙΣ 7 ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ K-MEANS.	176
ΕΙΚΟΝΑ 94: ΜΕΣΑ ΦΑΣΜΑΤΑ ΓΙΑ ΤΙΣ ΕΠΤΑ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΑΝΑΛΥΣΗ K-MEANS, ΑΠΕΙΚΟΝΙΖΟΜΕΝΑ ΩΣ ΣΥΝΑΡΤΗΣΗ ΤΟΥ ΚΑΝΑΛΙΟΥ ΤΟΥ ΑΝΙΧΝΕΥΤΗ (CHANNEL INDEX) ΠΡΙΝ ΑΠΟ ΤΗ ΒΑΘΜΟΝΟΜΗΣΗ ΣΕ ΕΝΕΡΓΕΙΑ (KEV).	177
ΕΙΚΟΝΑ 95: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ K-MEANS (K=7) ΕΠΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ PCA ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ. (ΑΡΙΣΤΕΡΑ) Ο ΤΕΛΙΚΟΣ ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ ΠΟΥ ΔΕΙΧΝΕΙ ΤΗ ΒΕΛΤΙΩΜΕΝΗ ΟΜΑΔΟΠΟΙΗΣΗ ΜΕ ΕΠΤΑ ΣΥΣΤΑΔΕΣ. (ΔΕΞΙΑ) ΤΟ ΙΣΤΟΓΡΑΜΜΑ ΜΕΓΕΘΟΥΣ ΠΟΥ ΠΟΣΟΤΙΚΟΠΟΙΕΙ ΤΟΝ ΑΡΙΘΜΟ ΤΩΝ PIXELS ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΠΤΑ ΣΥΣΤΑΔΕΣ.	177
ΕΙΚΟΝΑ 96: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΤΗΣ PCA (ΣΥΝΙΣΤΩΣΕΣ 4 ΕΝΑΝΤΙ 5), ΧΡΩΜΑΤΙΣΜΕΝΟ ΣΥΜΦΩΝΑ ΜΕ ΤΙΣ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ K-MEANS. ΑΥΤΗ Η ΑΠΕΙΚΟΝΙΣΗ ΔΕΙΧΝΕΙ ΠΩΣ ΟΙ ΟΜΑΔΕΣ ΠΟΥ ΒΡΕΘΗΚΑΝ ΑΠΟ ΤΗΝ K-MEANS ΚΑΤΑΝΕΜΟΝΤΑΙ ΣΤΟΝ ΧΩΡΟ ΤΩΝ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ.	178
ΕΙΚΟΝΑ 97: ΧΩΡΙΚΟΙ ΧΑΡΤΕΣ ΤΩΝ ΕΞΙ ΣΥΝΙΣΤΩΣΩΝ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΥΒΡΙΔΙΚΗ ΑΝΑΛΥΣΗ NMF-PCA. ΟΙ ΧΑΡΤΕΣ ΑΥΤΟΙ ΜΠΟΡΟΥΝ ΝΑ ΕΡΜΗΝΕΥΘΟΥΝ ΩΣ ΧΑΡΤΕΣ ΑΦΘΟΝΙΑΣ, ΑΝΑΔΕΙΚΝΥΟΝΤΑΣ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΔΙΑΚΡΙΤΩΝ ΧΗΜΙΚΩΝ ΣΥΝΙΣΤΩΣΩΝ ΜΕ ΕΞΑΙΡΕΤΙΚΗ ΕΥΚΡΙΝΕΙΑ.	181
ΕΙΚΟΝΑ 98: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (HEATMAP) ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ NMF-PCA. ΟΙ ΙΣΧΥΡΕΣ ΘΕΤΙΚΕΣ (ΚΟΚΚΙΝΟ) ΚΑΙ ΑΡΝΗΤΙΚΕΣ (ΜΠΛΕ) ΣΥΣΧΕΤΙΣΕΙΣ ΥΠΟΔΕΙΚΝΟΥΝ ΣΥΝΘΕΤΕΣ ΣΧΕΣΕΙΣ ΜΕΤΑΞΥ ΤΩΝ ΧΗΜΙΚΩΝ ΣΥΝΙΣΤΩΣΩΝ, ΟΠΩΣ ΣΥΝΥΠΑΡΞΗ Η ΑΜΟΙΒΑΙΟ ΑΠΟΚΛΕΙΣΜΟ.	182
ΕΙΚΟΝΑ 99: ΠΟΣΟΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΤΕΛΙΚΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗ K-MEANS (K=7) ΕΠΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ NMF-PCA. (ΑΡΙΣΤΕΡΑ) ΑΡΙΘΜΗΤΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ PIXELS ΑΝΑ ΣΥΣΤΑΔΑ. (ΔΕΞΙΑ) ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΚΑΤΑΝΟΜΗΣ ΜΕΓΕΘΟΥΣ ΤΩΝ ΣΥΣΤΑΔΩΝ.	183
ΕΙΚΟΝΑ 100: ΜΕΣΑ ΦΑΣΜΑΤΑ ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΠΤΑ ΣΥΣΤΑΔΕΣ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΟΜΑΔΟΠΟΙΗΣΗ K-MEANS ΕΠΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ NMF-PCA, ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΗΝ ΤΕΛΙΚΗ ΧΗΜΙΚΗ ΕΡΜΗΝΕΙΑ ΚΑΘΕ ΟΜΑΔΑΣ.	183

ΕΙΚΟΝΑ 101: Ο ΤΕΛΙΚΟΣ ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ K-MEANS (K=7) ΑΠΟ ΤΗΝ ΥΒΡΙΔΙΚΗ ΑΝΑΛΥΣΗ NMF-PCA ΣΤΑ ΔΕΔΟΜΕΝΑ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ. Ο ΧΑΡΤΗΣ ΑΥΤΟΣ ΔΕΝ ΕΙΝΑΙ ΑΠΛΩΣ ΕΝΑΣ ΧΑΡΤΗΣ ΥΛΙΚΩΝ, ΑΛΛΑ ΕΝΑΣ ΧΑΡΤΗΣ ΤΗΣ ΚΑΛΛΙΤΕΧΝΙΚΗΣ ΤΕΧΝΙΚΗΣ, ΔΙΑΚΡΙΝΟΝΤΑΣ ΠΑΧΙΕΣ ΑΠΟ ΛΕΠΤΕΣ ΣΤΡΩΣΕΙΣ ΚΑΙ ΦΩΤΙΣΜΑΤΑ ΑΠΟ ΣΚΙΕΣ.....	184
ΕΙΚΟΝΑ 102: ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION) ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΠΤΑ ΤΕΛΙΚΕΣ ΣΥΣΤΑΔΕΣ, ΕΠΙΒΕΒΑΙΩΝΟΝΤΑΣ ΤΟΝ ΕΠΙΤΥΧΗ ΚΑΙ ΔΙΑΚΡΙΤΟ ΔΙΑΧΩΡΙΣΜΟ ΤΟΥΣ.	185
ΕΙΚΟΝΑ 103: ΑΠΟΤΕΛΕΣΜΑ ΤΗΣ ΙΕΡΑΡΧΙΚΗΣ ΑΝΑΛΥΣΗΣ (SUB-CLUSTERING) ΣΤΗΝ ΠΕΡΙΟΧΗ ΤΩΝ ΣΑΡΚΩΜΑΤΩΝ (PARENT C6). ΟΙ ΥΠΟ-ΟΜΑΔΕΣ ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΕ ΔΙΑΚΡΙΤΑ ΣΤΑΔΙΑ ΤΗΣ ΖΩΓΡΑΦΙΚΗΣ ΔΙΑΔΙΚΑΣΙΑΣ, ΟΠΩΣ Η ΒΑΣΙΚΗ ΣΤΡΩΣΗ, ΤΑ ΦΩΤΙΣΜΑΤΑ ΚΑΙ ΟΙ ΣΚΙΕΣ.	186
ΕΙΚΟΝΑ 104: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΜΟΛΥΒΔΟ (Pb) , ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ ΕΚΠΟΜΠΗΣ Pb-La1, Pb-Lb1. Η ΥΨΗΛΗ ΣΥΓΚΕΝΤΡΩΣΗ ΤΟΥ ΜΟΛΥΒΔΟΥ ΣΤΑ ΣΑΡΚΩΜΑΤΑ ΚΑΙ ΣΤΑ "ΦΩΤΙΣΜΑΤΑ" ΤΑΥΤΟΠΟΙΕΙ ΤΗ ΧΡΗΣΗ ΤΟΥ ΛΕΥΚΟΥ ΤΟΥ ΜΟΛΥΒΔΟΥ, ΑΠΟΚΑΛΥΠΤΟΝΤΑΣ ΤΙΣ ΛΕΠΤΕΣ ΠΙΝΕΛΙΕΣ (ΨΙΜΥΘΙΕΣ) ΤΟΥ ΚΑΛΛΙΤΕΧΝΗ.	187
ΕΙΚΟΝΑ 105: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΥΔΑΡΓΥΡΟ (Hg) . Η ΣΤΟΧΕΥΜΕΝΗ ΠΑΡΟΥΣΙΑ ΤΟΥ (ΚΥΡΙΩΣ ΣΤΙΣ ΓΡΑΜΜΕΣ La1 ΚΑΙ Lb1) ΣΕ ΣΥΓΚΕΚΡΙΜΕΝΕΣ ΚΟΚΚΙΝΕΣ ΠΕΡΙΟΧΕΣ, ΟΠΩΣ Ο ΣΤΑΥΡΟΣ, ΤΑΥΤΟΠΟΙΕΙ ΤΗ ΧΡΗΣΗ ΤΗΣ ΠΟΛΥΤΙΜΗΣ ΧΡΩΣΤΙΚΗΣ ΚΙΝΝΑΒΑΡΙ (HgS).	188
ΕΙΚΟΝΑ 106: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΣΙΔΗΡΟ (Fe) . Η ΚΑΤΑΝΟΜΗ ΤΟΥ, ΚΥΡΙΩΣ ΜΕΣΩ ΤΩΝ ΓΡΑΜΜΩΝ Ka, ΣΥΝΔΕΕΤΑΙ ΑΜΕΣΑ ΜΕ ΤΙΣ ΚΑΦΕ-ΚΟΚΚΙΝΕΣ ΚΑΙ ΩΧΡΕΣ ΠΕΡΙΟΧΕΣ, ΕΠΙΒΕΒΑΙΩΝΟΝΤΑΣ ΤΗ ΧΡΗΣΗ ΓΗΙΝΩΝ ΧΡΩΣΤΙΚΩΝ (ΩΧΡΕΣ) ΜΕ ΒΑΣΗ ΤΑ ΘΕΙΔΙΑ ΤΟΥ ΣΙΔΗΡΟΥ	189
ΕΙΚΟΝΑ 107: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟ ΘΕΙΟ (S) (ΓΡΑΜΜΕΣ K-EDGE, Ka1,, Kb1). Η ΔΙΑΧΥΤΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΗ ΧΡΗΣΗ ΓΥΨΟΥ ΣΤΟ ΥΠΟΣΤΡΩΜΑ.	190
ΕΙΚΟΝΑ 108: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟ ΑΣΒΕΣΤΙΟ (Ca) (ΓΡΑΜΜΕΣ Ka1, Kb1). Η ΙΣΧΥΡΗ ΧΩΡΙΚΗ ΣΥΣΧΕΤΙΣΗ ΜΕ ΤΟ ΘΕΙΟ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΗΝ ΠΑΡΟΥΣΙΑ ΓΥΨΟΥ (CaSO ₄).	190
ΕΙΚΟΝΑ 109: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΧΡΥΣΟ (Au) (ΓΡΑΜΜΕΣ L). Η ΑΠΟΚΛΕΙΣΤΙΚΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΣΤΑ ΦΩΤΟΣΤΕΦΑΝΑ ΚΑΙ ΣΤΙΣ ΧΡΥΣΟΚΟΝΔΥΛΙΕΣ ΤΑΥΤΟΠΟΙΕΙ ΤΗΝ ΤΕΧΝΙΚΗ ΤΗΣ ΕΠΙΧΡΥΣΩΣΗΣ.	191
ΕΙΚΟΝΑ 110: ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΚΑΤΑΝΟΜΗΣ ΓΙΑ ΤΟΝ ΨΕΥΔΑΡΓΥΡΟ (Zn) . Η ΣΠΟΡΑΔΙΚΗ ΠΑΡΟΥΣΙΑ ΤΟΥ ΑΠΟΤΕΛΕΙ ΕΝΔΕΙΞΗ ΜΕΤΑΓΕΝΕΣΤΕΡΩΝ ΕΠΕΜΒΑΣΕΩΝ ΣΥΝΤΗΡΗΣΗΣ (RETOUCHING).	192
ΕΙΚΟΝΑ 111: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (HEATMAP) ΜΕΤΑΞΥ ΤΩΝ ΚΥΡΙΟΤΕΡΩΝ ΣΤΟΙΧΕΙΑΚΩΝ ΓΡΑΜΜΩΝ ΠΟΥ ΑΝΙΧΝΕΥΘΗΚΑΝ ΣΤΗΝ ΑΝΑΛΥΣΗ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ. ΟΙ ΘΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ (ΚΟΚΚΙΝΟ) ΥΠΟΔΕΙΚΝΟΥΝ ΧΩΡΙΚΗ ΣΥΝΥΠΑΡΞΗ ΥΛΙΚΩΝ, ΕΝΩ ΟΙ ΑΡΝΗΤΙΚΕΣ (ΜΠΛΕ) ΑΜΟΙΒΑΙΟ ΑΠΟΚΛΕΙΣΜΟ.	193
ΕΙΚΟΝΑ 112: ΣΥΓΚΡΙΣΗ ΤΟΥ ΣΤΟΙΧΕΙΑΚΟΥ ΧΑΡΤΗ ΓΙΑ ΤΟ ΑΣΒΕΣΤΙΟ (Ca-Ka) ΜΕΤΑΞΥ ΤΗΣ ΑΝΑΛΥΣΗΣ ΧΑΜΗΛΗΣ (ΑΡΙΣΤΕΡΑ) ΚΑΙ ΥΨΗΛΗΣ (ΔΕΞΙΑ) ΑΝΑΛΥΣΗΣ. Η ΥΨΗΛΟΤΕΡΗ ΑΝΑΛΥΣΗ ΑΠΟΚΑΛΥΠΤΕΙ ΜΕ ΜΕΓΑΛΥΤΕΡΗ ΛΕΠΤΟΜΕΡΕΙΑ ΤΙΣ ΦΘΟΡΕΣ ΚΑΙ ΤΗ ΔΟΜΗ ΤΟΥ ΥΠΟΣΤΡΩΜΑΤΟΣ.	197
ΕΙΚΟΝΑ 113: ΣΥΓΚΡΙΤΙΚΟΙ ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΓΙΑ ΤΟΝ ΣΙΔΗΡΟ (Fe). (ΠΑΝΩ) ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟ ΑΝΑΠΤΥΓΜΕΝΟ ΛΟΓΙΣΜΙΚΟ. (ΚΑΤΩ) ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟ ΕΡΓΑΣΤΗΡΙΑΚΟ ΛΟΓΙΣΜΙΚΟ.	198
ΕΙΚΟΝΑ 114: ΣΥΓΚΡΙΤΙΚΟΙ ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΓΙΑ ΤΟΝ ΥΔΑΡΓΥΡΟ (Hg) (ΠΑΝΩ) ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟ ΑΝΑΠΤΥΓΜΕΝΟ ΛΟΓΙΣΜΙΚΟ. ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΧΑΜΗΛΗΣ ΚΑΙ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ (ΚΑΤΩ) ΑΠΟ ΤΟ ΕΡΓΑΣΤΗΡΙΑΚΟ ΛΟΓΙΣΜΙΚΟ.	200
ΕΙΚΟΝΑ 115: ΣΥΓΚΡΙΤΙΚΟΙ ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΓΙΑ ΤΟΝ ΜΟΛΥΒΔΟ (Pb) (ΠΑΝΩ) ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟ ΑΝΑΠΤΥΓΜΕΝΟ ΛΟΓΙΣΜΙΚΟ. ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΧΑΜΗΛΗΣ ΚΑΙ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ (ΚΑΤΩ) ΑΠΟ ΤΟ ΕΡΓΑΣΤΗΡΙΑΚΟ ΛΟΓΙΣΜΙΚΟ.	201
ΕΙΚΟΝΑ 116: ΣΥΓΚΡΙΤΙΚΟΙ ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΓΙΑ ΤΟΝ ΧΡΥΣΟ (Au) (ΠΑΝΩ) ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟ ΑΝΑΠΤΥΓΜΕΝΟ ΛΟΓΙΣΜΙΚΟ. ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΧΑΜΗΛΗΣ ΚΑΙ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ (ΚΑΤΩ) ΑΠΟ ΤΟ ΕΡΓΑΣΤΗΡΙΑΚΟ ΛΟΓΙΣΜΙΚΟ.	203
ΕΙΚΟΝΑ 117: ΣΥΓΚΡΙΤΙΚΟΙ ΣΤΟΙΧΕΙΑΚΟΙ ΧΑΡΤΕΣ ΓΙΑ ΤΟΝ ΨΕΥΔΑΡΓΥΡΟ (Zn) (ΠΑΝΩ) ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΟ ΑΝΑΠΤΥΓΜΕΝΟ ΛΟΓΙΣΜΙΚΟ. ΣΥΓΚΡΙΣΗ ΜΕΤΑΞΥ ΧΑΜΗΛΗΣ ΚΑΙ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ (ΚΑΤΩ) ΑΠΟ ΤΟ ΕΡΓΑΣΤΗΡΙΑΚΟ ΛΟΓΙΣΜΙΚΟ.	205
ΕΙΚΟΝΑ 118: Η ΦΟΡΗΤΗ ΕΙΚΟΝΑ ΤΗΣ 'ΠΑΝΑΓΙΑΣ ΤΗΣ ΟΔΗΓΗΤΡΙΑΣ' ΠΟΥ ΑΠΟΤΕΛΕΣΕ ΤΟ ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΑΝΑΛΥΣΗΣ MA-XRF. Η ΕΙΚΟΝΑ ΠΑΡΟΥΣΙΑΖΕΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΦΘΟΡΕΣ ΚΑΙ ΑΠΟΤΕΛΕΙ ΤΥΠΙΚΟ ΔΕΙΓΜΑ ΜΕΤΑΒΥΖΑΝΤΙΝΗΣ ΑΓΙΟΓΡΑΦΙΑΣ.	207
ΕΙΚΟΝΑ 119: ΠΙΝΑΚΑΣ ΜΕ ΤΙΣ ΠΕΙΡΑΜΑΤΙΚΕΣ ΠΑΡΑΜΕΤΡΟΥΣ ΤΗΣ ΣΑΡΩΣΗΣ MA-XRF ΓΙΑ ΤΗΝ ΕΙΚΟΝΑ ΤΗΣ ΠΑΝΑΓΙΑΣ. ΠΕΡΙΛΑΜΒΑΝΕΙ ΠΛΗΡΟΦΟΡΙΕΣ ΓΙΑ ΤΙΣ ΔΙΑΣΤΑΣΕΙΣ ΤΗΣ ΣΑΡΩΣΗΣ (MAPPING PARAMETERS), ΤΙΣ ΣΥΝΘΗΚΕΣ ΑΠΟΚΤΗΣΗΣ ΔΕΔΟΜΕΝΩΝ (ACQUISITION PARAMETERS), ΤΙΣ ΡΥΘΜΙΣΕΙΣ ΤΗΣ ΛΥΧΝΙΑΣ ΑΚΤΙΝΩΝ-X (TUBE PARAMETER) ΚΑΙ ΤΙΣ ΠΑΡΑΜΕΤΡΟΥΣ ΤΟΥ ΑΝΙΧΝΕΥΤΗ (DETECTOR PARAMETERS).	208

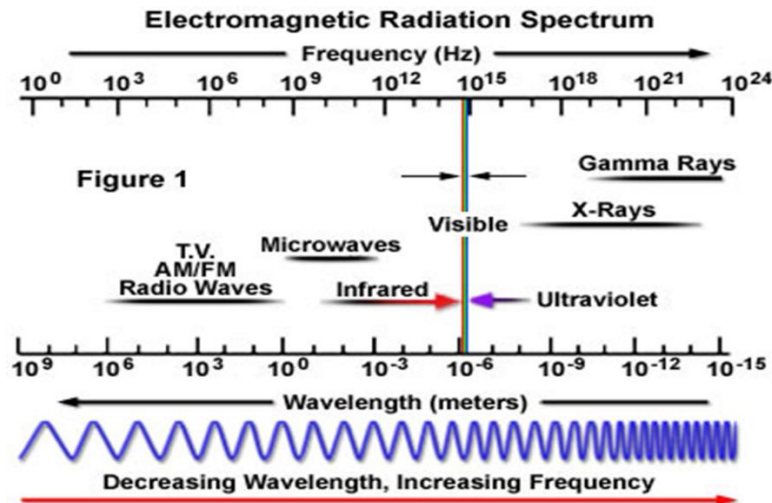
ΕΙΚΟΝΑ 120: ΤΟ ΦΑΣΜΑ ΑΘΡΟΙΣΜΑΤΟΣ (SUM SPECTRUM) ΤΩΝ ΔΕΔΟΜΕΝΩΝ MA-XRF ΠΡΙΝ ΑΠΟ ΤΗ ΒΑΘΜΟΝΟΜΗΣΗ. Ο ΟΡΙΖΟΝΤΙΟΣ ΑΞΟΝΑΣ ΑΝΤΙΣΤΟΙΧΕΙ ΣΤΑ ΚΑΝΑΛΙΑ (CHANNEL INDEX) ΤΟΥ ΑΝΙΧΝΕΥΤΗ. ΤΟ ΦΑΣΜΑ ΠΑΡΕΧΕΙ ΜΙΑ ΑΡΧΙΚΗ, ΣΥΝΟΛΙΚΗ ΕΙΚΟΝΑ ΤΗΣ ΣΤΟΙΧΕΙΑΚΗΣ ΣΥΣΤΑΣΗΣ ΤΗΣ ΕΙΚΟΝΑΣ.	208
ΕΙΚΟΝΑ 121: ΧΑΡΤΗΣ ΣΥΝΟΛΙΚΗΣ ΚΑΤΑΝΟΜΗΣ ΕΝΤΑΣΗΣ (ENERGY DISTRIBUTION MAP). ΟΙ ΦΩΤΕΙΝΟΤΕΡΕΣ ΠΕΡΙΟΧΕΣ (ΚΙΤΡΙΝΟ) ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΕ ΥΨΗΛΟΤΕΡΗ ΣΥΝΟΛΙΚΗ ΕΚΠΟΜΠΗ ΑΚΤΙΝΩΝ-Χ, ΥΠΟΔΕΙΚΝΥΟΝΤΑΣ ΤΗΝ ΠΑΡΟΥΣΙΑ ΣΤΟΙΧΕΙΩΝ ΜΕ ΥΨΗΛΟ ΑΤΟΜΙΚΟ ΑΡΙΘΜΟ (Π.Χ., Pb, Au) Η ΜΕΓΑΛΥΤΕΡΟ ΠΑΧΟΣ ΥΛΙΚΟΥ.....	209
ΕΙΚΟΝΑ 122: ΤΟ ΜΕΣΟ ΦΑΣΜΑ ΦΘΟΡΙΣΜΟΥ (MEAN ENERGY SPECTRUM) ΜΕΤΑ ΤΗ ΒΑΘΜΟΝΟΜΗΣΗ ΣΕ ΕΝΕΡΓΕΙΑ (ΚΕV). ΟΙ ΚΟΡΥΦΕΣ ΑΝΤΙΣΤΟΙΧΟΥΝ ΣΤΙΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΕΣ ΓΡΑΜΜΕΣ ΕΚΠΟΜΠΗΣ ΤΩΝ ΚΥΡΙΑΡΧΩΝ ΣΤΟΙΧΕΙΩΝ ΣΤΗΝ ΕΙΚΟΝΑ, ΟΠΩΣ Ca, S, Fe, Pb ΚΑΙ Au.	210
ΕΙΚΟΝΑ 123: SCREE PLOT	211
ΕΙΚΟΝΑ 124: ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΑΝΑΛΥΣΗΣ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA). (ΠΑΝΩ) ΤΟ ΔΙΑΓΡΑΜΜΑ SCREE PLOT ΔΕΙΧΝΕΙ ΤΗΝ ΑΤΟΜΙΚΗ (ΜΠΛΕ ΜΠΑΡΕΣ) ΚΑΙ ΣΩΡΕΥΤΙΚΗ (ΚΟΚΚΙΝΗ ΓΡΑΜΜΗ) ΕΞΗΓΟΥΜΕΝΗ ΔΙΑΚΥΜΑΝΣΗ. (ΚΑΤΩ) ΤΑ ΑΡΙΘΜΗΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ INCREMENTAL PCA (IPCA), ΠΟΥ ΕΠΙΒΕΒΑΙΩΝΟΥΝ ΟΤΙ ΟΙ ΠΡΩΤΕΣ 6 ΣΥΝΙΣΤΩΣΕΣ ΔΙΑΤΗΡΟΥΝ ΠΕΡΙΠΟΥ ΤΟ 90.2% ΤΗΣ ΣΥΝΟΛΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ.....	211
ΕΙΚΟΝΑ 125: ΧΩΡΙΚΟΙ ΧΑΡΤΕΣ (SCORE MAPS) ΤΩΝ ΕΞΙ ΠΡΩΤΩΝ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ (PCA COMPONENTS 1-6). ΚΑΘΕ ΧΑΡΤΗΣ ΑΠΕΙΚΟΝΙΖΕΙ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΜΙΑΣ ΟΡΘΟΓΩΝΙΑΣ ΠΗΓΗΣ ΦΑΣΜΑΤΙΚΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ, ΔΙΑΧΩΡΙΖΟΝΤΑΣ ΔΙΑΦΟΡΕΤΙΚΕΣ ΠΕΡΙΟΧΕΣ ΤΗΣ ΕΙΚΟΝΑΣ (Π.Χ., ΥΠΟΒΑΘΡΟ, ΦΩΤΟΣΤΕΦΑΝΑ, ΕΝΔΥΜΑΤΑ).	212
ΕΙΚΟΝΑ 126: ΦΑΣΜΑΤΙΚΗ ΕΡΜΗΝΕΙΑ ΤΩΝ ΚΥΡΙΩΝ ΣΥΝΙΣΤΩΣΩΝ. Ο ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (CORRELATION MATRIX) ΤΩΝ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ ΕΠΙΒΕΒΑΙΩΝΕΙ ΤΗΝ ΟΡΘΟΓΩΝΙΟΤΗΤΑ ΤΟΥΣ.	213
ΕΙΚΟΝΑ 127: ΤΑ ΦΑΣΜΑΤΑ ΤΩΝ ΙΔΙΟΔΙΑΝΥΣΜΑΤΩΝ (PCA EIGENVECTORS/LOADINGS), ΠΟΥ ΑΠΟΚΑΛΥΠΤΟΥΝ ΤΗ ΦΑΣΜΑΤΙΚΗ ΥΠΟΓΡΑΦΗ ΚΑΘΕ ΣΥΝΙΣΤΩΣΑΣ.....	214
ΕΙΚΟΝΑ 128: ΔΙΣΔΙΑΣΤΑΤΗ ΑΠΕΙΚΟΝΙΣΗ ΤΗΣ ΑΝΑΛΥΣΗΣ PCA. (ΠΑΝΩ) ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ ΤΩΝ ΣΥΝΤΕΛΕΣΤΩΝ (SCORES) ΤΗΣ ΣΥΝΙΣΤΩΣΑΣ 2 ΕΝΑΝΤΙ ΤΗΣ ΣΥΝΙΣΤΩΣΑΣ 1. (ΚΑΤΩ) ΔΙΑΓΡΑΜΜΑ ΤΩΝ ΑΝΤΙΣΤΟΙΧΩΝ LOADINGS, ΠΟΥ ΔΕΙΧΝΕΙ ΤΗ ΦΑΣΜΑΤΙΚΗ ΣΧΕΣΗ ΜΕΤΑΞΥ ΤΩΝ ΔΥΟ ΣΥΝΙΣΤΩΣΩΝ.	215
ΕΙΚΟΝΑ 129: ΑΠΟΤΕΛΕΣΜΑΤΑ ΑΠΟ ΤΗΝ ΕΚΤΕΛΕΣΗ ΤΟΥ ΠΛΑΙΣΙΟΥ ΑΞΙΟΛΟΓΗΣΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΓΙΑ ΤΟΝ ΠΡΟΣΔΙΟΡΙΣΜΟ ΤΟΥ ΒΕΛΤΙΣΤΟΥ ΑΡΙΘΜΟΥ ΣΥΣΤΑΔΩΝ (Κ). ΤΑ ΔΙΑΓΡΑΜΜΑΤΑ ΑΠΕΙΚΟΝΙΖΟΥΝ ΤΙΣ ΜΕΤΡΙΚΕΣ ELBOW METHOD, SILHOUETTE SCORE, DAVIES-BOULDIN INDEX ΚΑΙ CALINSKI-HARABASZ INDEX ΓΙΑ Κ ΑΠΟ 2 ΕΩΣ 10, ΥΠΟΔΕΙΚΝΥΟΝΤΑΣ ΩΣ ΒΕΛΤΙΣΤΟ ΕΥΡΟΣ ΤΙΣ 5-6 ΣΥΣΤΑΔΕΣ.	216
ΕΙΚΟΝΑ 130: (ΑΡΙΣΤΕΡΑ) ΑΡΙΘΜΗΤΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ PIXELS ΑΝΑ ΣΥΣΤΑΔΑ. (ΔΕΞΙΑ) ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΟΥ ΜΕΓΕΘΟΥΣ ΚΑΘΕ ΣΥΣΤΑΔΑΣ.	217
ΕΙΚΟΝΑ 131: ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ (CLUSTER MAP) ΠΟΥ ΠΡΟΕΚΥΨΕ ΑΠΟ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΑΛΓΟΡΙΘΜΟΥ K-MEANS (K=6) ΣΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ PCA. ΚΑΘΕ ΧΡΩΜΑ ΑΝΤΙΣΤΟΙΧΕΙ ΣΕ ΜΙΑ ΔΙΑΚΡΙΤΗ ΣΥΣΤΑΔΑ, ΑΠΟΚΑΛΥΠΤΟΝΤΑΣ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΤΩΝ ΚΥΡΙΑΡΧΩΝ ΥΛΙΚΩΝ ΟΜΑΔΩΝ.	217
ΕΙΚΟΝΑ 132: ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΩΝ ΣΥΣΤΑΔΩΝ K-MEANS. (ΠΑΝΩ) ΔΙΑΓΡΑΜΜΑ ΚΑΤΑΝΟΜΗΣ ΤΗΣ ΕΝΤΑΣΗΣ ΓΙΑ ΚΑΘΕ ΣΥΣΤΑΔΑ. (ΚΑΤΩ) ΤΑ ΜΕΣΑ ΦΑΣΜΑΤΑ ΦΘΟΡΙΣΜΟΥ ΓΙΑ ΚΑΘΕ ΜΙΑ ΑΠΟ ΤΙΣ ΕΞΙ ΣΥΣΤΑΔΕΣ, ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΗΝ ΑΚΡΙΒΗ ΧΗΜΙΚΗ ΤΟΥΣ ΕΡΜΗΝΕΙΑ.	218
ΕΙΚΟΝΑ 133: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΙΕΡΑΡΧΙΚΗΣ ΑΝΑΛΥΣΗΣ (SUB-CLUSTERING) ΠΟΥ ΕΦΑΡΜΟΣΤΗΚΕ ΣΤΗΝ 'CLUSTER 4' (ΦΩΤΟΣΤΕΦΑΝΑ). Ο ΧΑΡΤΗΣ ΑΠΟΚΑΛΥΠΤΕΙ ΤΗΝ ΕΣΩΤΕΡΙΚΗ ΔΟΜΗ ΤΗΣ ΣΥΣΤΑΔΑΣ ΜΕ ΕΞΙ ΝΕΕΣ ΥΠΟ-ΟΜΑΔΕΣ (SUB-C 0-5), ΕΠΙΤΡΕΠΟΝΤΑΣ ΤΗ ΔΙΕΡΕΥΝΗΣΗ ΛΕΠΤΟΜΕΡΩΝ ΥΛΙΚΩΝ ΔΙΑΦΟΡΟΠΟΙΗΣΕΩΝ (Π.Χ., ΔΙΑΦΟΡΕΤΙΚΑ ΠΑΧΗ ΦΥΛΛΟΥ ΧΡΥΣΟΥ, ΠΑΡΟΥΣΙΑ ΚΟΛΛΑΣ Η ΜΠΟΛ ΑΡΜΕΝΙΑΣ).	219
ΕΙΚΟΝΑ 134: ΧΩΡΙΚΟΙ ΧΑΡΤΕΣ ΤΩΝ ΕΞΙ ΣΥΝΙΣΤΩΣΩΝ ΠΟΥ ΠΡΟΕΚΥΨΑΝ ΑΠΟ ΤΗΝ ΑΝΑΛΥΣΗ NMF. ΛΟΓΩ ΤΗΣ ΜΗ-ΑΡΝΗΤΙΚΟΤΗΤΑΣ, ΟΙ ΧΑΡΤΕΣ ΑΥΤΟΙ ΜΠΟΡΟΥΝ ΝΑ ΕΡΜΗΝΕΥΘΟΥΝ ΩΣ ΧΑΡΤΕΣ ΑΦΘΟΝΙΑΣ, ΑΝΑΔΕΙΚΝΥΟΝΤΑΣ ΤΗ ΧΩΡΙΚΗ ΚΑΤΑΝΟΜΗ ΔΙΑΚΡΙΤΩΝ ΧΗΜΙΚΩΝ ΣΥΣΤΑΤΙΚΩΝ (Π.Χ., GESSO, ΧΡΥΣΟΣ, ΩΧΡΕΣ).....	221
ΕΙΚΟΝΑ 135: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ (HEATMAP) ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ NMF. ΣΕ ΑΝΤΙΘΕΣΗ ΜΕ ΤΗΝ PCA, ΟΙ ΣΥΝΙΣΤΩΣΕΣ ΔΕΝ ΕΙΝΑΙ ΚΑΤ' ΑΝΑΓΚΗ ΟΡΘΟΓΩΝΙΕΣ. ΟΙ ΘΕΤΙΚΕΣ ΣΥΣΧΕΤΙΣΕΙΣ ΥΠΟΔΕΙΚΝΥΟΥΝ ΚΟΙΝΑ ΦΑΣΜΑΤΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ Η ΧΩΡΙΚΗ ΣΥΝΥΠΑΡΞΗ ΥΛΙΚΩΝ ΜΕΤΑΞΥ ΤΩΝ ΣΥΝΙΣΤΩΣΩΝ.	223
ΕΙΚΟΝΑ 136: ΑΠΟΤΕΛΕΣΜΑΤΑ ΤΗΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ K-MEANS (K=6) ΕΠΙ ΤΩΝ ΔΕΔΟΜΕΝΩΝ NMF. (ΠΑΝΩ) Ο ΘΕΜΑΤΙΚΟΣ ΧΑΡΤΗΣ. (ΜΕΣΗ) ΤΑ ΜΕΣΑ ΦΑΣΜΑΤΑ ΓΙΑ ΚΑΘΕ ΣΥΣΤΑΔΑ. (ΚΑΤΩ) ΤΟ ΔΙΑΓΡΑΜΜΑ ΚΑΤΑΝΟΜΗΣ ΤΗΣ ΕΝΤΑΣΗΣ ΑΝΑ ΣΥΣΤΑΔΑ.....	225
ΕΙΚΟΝΑ 137: ΜΕΘΟΔΟΣ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ KMEANS ΕΠΙ ΑΝΩ ΣΤΑ ΔΕΔΟΜΕΝΑ ΤΗΣ NMF ΜΕ SUBCLUSTERING ΕΠΙ ΑΝΩ ΣΤΗΝ ΣΥΣΤΑΔΑ 5	226
ΕΙΚΟΝΑ 138: ΑΠΕΙΚΟΝΙΣΗ ΑΞΟΝΩΝ NMF ΜΕ ΕΠΙ ΑΝΩ ΣΤΗΝ ΕΙΚΟΝΑ, ΧΡΩΜΑΤΙΖΟΝΤΑΣ ΜΟΝΟ ΤΙΣ ΠΕΡΙΟΧΕΣ ΕΜΦΑΝΙΣΗΣ.....	227

ΕΙΚΟΝΑ 139: ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΗΣ ΑΞΟΝΩΝ ΜΕΘΟΔΟΥ NMF	228
ΕΙΚΟΝΑ 140: ΕΝΔΕΙΚΤΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΤΩΝ ΑΞΟΝΩΝ ΑΠΟ ΤΗΝ ΜΕΘΟΔΟ NMF ΜΕ ΒΑΣΗ ΚΑΘΕ ΣΥΣΤΑΔΑ ΔΙΑΦΟΡΕΤΙΚΟΥ ΧΡΩΜΑΤΟΣ	229
ΕΙΚΟΝΑ 141: ΓΡΑΦΙΚΗ ΑΠΕΙΚΟΝΙΣΗ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ, ΜΕ ΠΛΗΘΟΣ ΔΕΔΟΜΕΝΩΝ ΑΝΑ ΣΥΣΤΑΔΑ.....	230
ΕΙΚΟΝΑ 142: ΕΙΚΟΝΑ ΑΠΕΙΚΟΝΙΣΗΣ ΤΩΝ ΣΥΣΤΑΔΩΝ ΜΕ ΒΑΣΗ ΤΟ ΦΑΣΜΑ ΣΥΝΑΡΤΗΣΕΙ ΤΗΣ ΕΝΕΡΓΕΙΑΣ. .	230
ΕΙΚΟΝΑ 143: ΑΠΕΙΚΟΝΙΣΗ ΤΟΥ ΕΥΡΟΥΣ ΕΝΕΡΓΕΙΑΣ ΜΕ ΒΑΣΗ ΤΑ ΚΑΝΑΛΙΑ ΚΑΘΕ ΣΥΣΤΑΔΑΣ.....	231
ΕΙΚΟΝΑ 144: ΑΠΟΤΕΛΕΣΜΑ ΣΥΣΤΑΔΟΠΟΙΗΣΗΣ ΜΕΣΩ ΜΕΘΟΔΟΥ KMEANS.....	232
ΕΙΚΟΝΑ 145: ΣΥΣΤΑΔΑ 3 ΑΠΟΜΟΝΩΜΕΝΗ.....	233
ΕΙΚΟΝΑ 146: ΑΠΟΤΕΛΕΣΜΑ SUBCLUSTERING ΣΥΣΤΑΔΑΣ (3) ΜΕ ΕΞΙ ΣΥΣΤΑΔΕΣ	234
ΕΙΚΟΝΑ 147: ΑΠΟΤΕΛΕΣΜΑ SUBCLUSTERING ΣΥΣΤΑΔΑΣ (2) ΜΕ ΕΞΙ ΣΥΣΤΑΔΕΣ	235
ΕΙΚΟΝΑ 148: ΑΠΕΙΚΟΝΙΣΗ ΕΠΙΛΕΓΜΕΝΩΝ ΣΤΟΙΧΕΙΩΝ ΣΤΟ ΦΑΣΜΑ (ΕΠΑΝΩ), ΚΑΙ ΟΛΙΚΗ ΚΑΤΑΝΟΜΗ ΕΝΤΑΣΗΣ ΠΑΝΩ ΣΤΗΝ ΕΙΚΟΝΑ (ΚΑΤΩ ΑΡΙΣΤΕΡΑ) ΚΑΙ ΦΑΣΜΑ ΕΝΕΡΓΕΙΑΣ (ΚΑΤΩ ΔΕΞΙΑ).....	236
ΕΙΚΟΝΑ 149: ΑΠΟΤΕΛΕΣΜΑ ΚΑΤΑΝΟΜΗΣ ΚΟΡΥΦΩΝ ΜΕ ΒΑΣΗ ΤΟΝ ΑΛΓΟΡΙΘΜΟ ΑΠΟΛΥΤΗΣ ΤΑΥΤΙΣΗΣ. ΧΩΡΙΣ ΟΜΩΣ ΤΗΝ ΣΥΝΟΛΙΚΗ ΤΟ ΒΑΘΜΟΝΟΜΗΜΕΝΟ ΑΛΓΟΡΙΘΜΟ ΣΥΝΟΛΙΚΗΣ ΤΑΥΤΙΣΗΣ ΠΟΥ ΠΑΡΕΧΕΤΑΙ ΣΤΟ ΕΠΟΜΕΝΟ ΠΑΡΑΘΥΡΟ.....	237
ΕΙΚΟΝΑ 150: ΠΑΡΑΘΥΡΟ ΕΠΙΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ ΠΟΥ ΤΑΙΡΙΑΖΟΥΝ ΣΤΟ ΦΑΣΜΑ ΜΕ ΒΑΣΗ ΤΗΝ ΕΝΕΡΓΕΙΑ Η ΤΑ ΚΑΝΑΛΙΑ. ΟΠΩΣ ΕΠΙΣΗΣ ΚΑΙ ΕΠΙΛΟΓΗ ΚΟΡΥΦΩΝ ΓΙΑ ΑΝΑΓΝΩΣΗ ΠΙΘΑΝΩΝ ΣΤΟΙΧΕΙΩΝ.....	238
ΕΙΚΟΝΑ 151: ΠΑΡΑΘΥΡΟ ΥΠΟΛΟΓΙΣΜΟΥ ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΩΝ ΧΑΡΤΩΝ ΜΕΣΩ ΤΟΥ ΑΝΑΠΤΥΣΣΟΜΕΝΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ.....	239
ΕΙΚΟΝΑ 152: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΜΑΓΓΑΝΙΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K.....	239
ΕΙΚΟΝΑ 153: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΥΔΡΑΡΓΥΡΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L	240
ΕΙΚΟΝΑ 154: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΣΙΔΗΡΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K.....	240
ΕΙΚΟΝΑ 155: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΧΑΛΚΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K.....	241
ΕΙΚΟΝΑ 156: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΑΣΒΕΣΤΙΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K.....	241
ΕΙΚΟΝΑ 157: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΧΡΥΣΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L	242
ΕΙΚΟΝΑ 158: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΘΕΙΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K.....	243
ΕΙΚΟΝΑ 159: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΜΟΛΥΒΔΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L	243
ΕΙΚΟΝΑ 160: : ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΑΡΓΥΡΟΥ Ag ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L	244
ΕΙΚΟΝΑ 161: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΑΣΒΕΣΤΙΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K(ΕΠΑΝΩ) ΚΑΙ ΕΙΚΟΝΑ ΕΡΓΑΣΤΗΡΙΟΥ ΚΑΤΩ	247
ΕΙΚΟΝΑ 162: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΘΕΙΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K.....	247
ΕΙΚΟΝΑ 163: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΜΟΛΥΒΔΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L (ΕΠΑΝΩ) ΚΑΙ ΕΙΚΟΝΑ ΕΡΓΑΣΤΗΡΙΟΥ ΚΑΤΩ	248
ΕΙΚΟΝΑ 164: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΜΑΓΓΑΝΙΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K (ΕΠΑΝΩ) ΚΑΙ ΕΙΚΟΝΑ ΕΡΓΑΣΤΗΡΙΟΥ ΚΑΤΩ.....	249
ΕΙΚΟΝΑ 165: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΥΔΡΑΡΓΥΡΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L (ΕΠΑΝΩ) ΚΑΙ ΕΙΚΟΝΑ ΕΡΓΑΣΤΗΡΙΟΥ ΚΑΤΩ	250
ΕΙΚΟΝΑ 166: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΧΑΛΚΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ K (ΕΠΑΝΩ) ΚΑΙ ΕΙΚΟΝΑ ΕΡΓΑΣΤΗΡΙΟΥ ΚΑΤΩ.....	251
ΕΙΚΟΝΑ 167: ΣΤΟΙΧΕΙΟΜΕΤΡΙΚΟΣ ΧΑΡΤΗΣ ΧΡΥΣΟΥ ΑΠΟ ΤΟ ΠΡΟΓΡΑΜΜΑ ΑΝΑΛΥΣΗΣ. ΧΑΡΤΕΣ ΓΙΑ ΤΙΣ ΓΡΑΜΜΕΣ L (ΕΠΑΝΩ) ΚΑΙ ΕΙΚΟΝΑ ΕΡΓΑΣΤΗΡΙΟΥ ΚΑΤΩ	253

Α' ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ

1. ΗΛΕΚΤΡΟΜΑΓΝΗΤΙΚΗ ΑΚΤΙΝΟΒΟΛΙΑ X – Ray

Η ηλεκτρομαγνητική ακτινοβολία περιλαμβάνει ένα εύρος μηκών κύματος (Σχήμα 1), από τα ραδιοκύματα που εκτείνονται σε τάξεις μήκους χιλιομέτρων έως και τη σκληρή ακτινοβολία γάμμα με μήκη κύματος τόσο μικρά όσο τα πικομέτρα. Αυτά ακριβώς τα μήκη κύματος συναντώνται και στο ορατό φως, το οποίο αντιλαμβανόμαστε ως διαφορετικά χρώματα παντού γύρω μας, όπως για παράδειγμα σε ένα ουράνιο τόξο, σε μια φωτογραφία ακόμα και στο λευκό το χρώμα. Οι **ακτίνες X**, οι οποίες μοιράζονται τις ίδιες φυσικές ιδιότητες με το ορατό φως, είναι μια μορφή ηλεκτρομαγνητικής ακτινοβολίας με πολυάριθμες πρακτικές εφαρμογές σε διάφορες βιομηχανίες. Σε αυτό το εισαγωγικό κεφάλαιο, εμβαθύνουμε στις θεμελιώδεις έννοιες και διαδικασίες της φυσικής των **ακτίνων X** που σχετίζονται με τη φασματομετρία **ακτίνων X**. Δίνουμε ιδιαίτερη έμφαση στην εκπομπή συνεχών και χαρακτηριστικών **ακτίνων X**, καθώς και στις αλληλεπιδράσεις μεταξύ φωτονίων και ύλης. Ενώ παρέχουμε λεπτομερή κάλυψη των κυριότερων διεργασιών αλληλεπίδρασης, παρουσιάζουμε επίσης αναλυτικές εκφράσεις και δεδομένα σε πίνακες για διατομές και θεμελιώδεις παραμέτρους που σχετίζονται με την εκπομπή χαρακτηριστικών **ακτίνων X**. Παρουσιάζονται βασικές εξισώσεις για τον υπολογισμό της έντασης των χαρακτηριστικών **ακτίνων X** σε διάφορους τρόπους φασματομετρίας **ακτίνων X**, χωρίς να υπεισέλθουμε στην εξαγωγή τους. Λεπτομερέστερες εκφράσεις που συνδέουν την ένταση των εκπεμπόμενων χαρακτηριστικών **ακτίνων X** με τις συγκεντρώσεις στοιχείων στα δείγματα θα συζητηθούν σε επόμενα κεφάλαια αφιερωμένα σε συγκεκριμένους τύπους φασματομετρίας **ακτίνων X**.



Σχήμα 1: Ηλεκτρομαγνητικό Φάσμα με άξονες για τις συχνότητες και τα μήκη κύματος

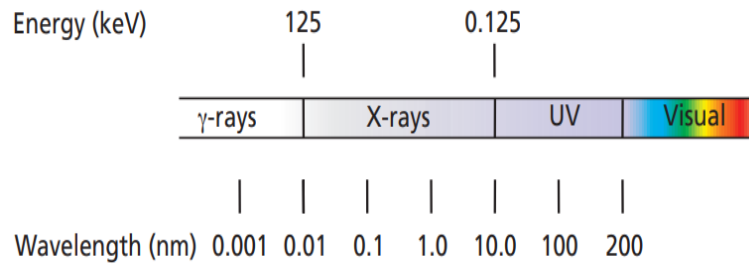
Ιστορία

Οι **ακτίνες X** ανακαλύφθηκαν για πρώτη φορά το **1895** από τον **Wilhelm Conrad Röntgen** στο Πανεπιστήμιο του **Würzburg** της Βαυαρίας. Ο **Röntgen** παρατήρησε ότι ορισμένοι κρύσταλλοι πλατινοκυανιούχου βαρίου που τοποθετήθηκαν κοντά σε έναν πλήρως κλειστό σωλήνα εκκένωσης φωσφορίζαν όταν γινόταν η εκκένωση. Εξετάζοντας τις σκιές που έριχναν οι ακτίνες, εντόπισε την προέλευσή τους στα τοιχώματα του σωλήνα εκκένωσης. Τα κρίσιμα πειράματα σχετικά με τις ιδιότητες των **ακτίνων X**, όπως η πόλωση, η περίθλαση, η ανάκλαση και η διάθλαση, δεν έδωσαν

καμία ένδειξη κυματοειδούς συμπεριφοράς και οδήγησαν στην ονομασία "X" για τις ακτίνες αυτές (άγνωστες). Αργότερα διαπιστώθηκε ότι οι **ακτίνες X** παράγονται όταν οι καθοδικές ακτίνες, που αναγνωρίστηκαν ως ηλεκτρόνια από τον **Thomson** το **1897**, προσκρούουν σε ανοδικό στόχο. Το πρωτοποριακό έργο του **Ρέντγκεν** του χάρισε το πρώτο βραβείο **Νόμπελ Φυσικής** το **1901**, σηματοδοτώντας την αρχή πολλών βραβείων Νόμπελ που απονεμήθηκαν για έρευνα σχετική με τις **ακτίνες X** μέχρι το **1927**. Ενώ οι **ακτίνες X** αρχικά παρουσίαζαν χαρακτηριστικά παρόμοια με το φως, προκάλεσαν μια επιστημονική συζήτηση σχετικά με την πραγματική τους φύση. Το **1906**, τα πειράματα σκέδασης του **Barkla** έδειξαν ότι οι **ακτίνες X** μπορούσαν να **πολωθούν**, υποδηλώνοντας μια **κυματική φύση**. Ωστόσο, οι μελέτες ιονισμού του **W. H. Bragg** υπέδειξαν μια **σωματιδιακή φύση**. Η **κυματική** φύση των **ακτίνων-X** τεκμηριώθηκε οριστικά το **1912**, όταν οι **Laue, Friedrich** και **Knipping** απέδειξαν την περίθλαση των **ακτίνων-X** από κρυστάλλους, αντιμετωπίζοντάς τους ως τρισδιάστατα πλέγματα περίθλασης. Το **1908**, οι **Barkla** και **Sadler** συμπέραναν την παρουσία συνιστωσών, των ακτινοβολιών **K** και **L**, στις **ακτίνες X** μέσω πειραμάτων σκέδασης. Περαιτέρω έρευνα από τον **W. H. Bragg** το **1913** επιβεβαίωσε την ύπαρξη ενός φάσματος γραμμών που επικάθεται σε ένα συνεχές φάσμα. Η εργασία του **Moseley** το **1913** αποκάλυψε ότι τα μήκη κύματος αυτών των γραμμών ήταν χαρακτηριστικά του στοιχείου του υλικού-στόχου και ακολουθούσαν την ακολουθία των ατομικών αριθμών, επιτρέποντας τον αδιαμφισβήτητο προσδιορισμό των ατομικών αριθμών. Τα χαρακτηριστικά των φασμάτων απορρόφησης και εκπομπής **ακτίνων X**, καθώς και η επίδραση της χημικής κατάστασης των απορροφητών και των εκπομπών, διερευνήθηκαν τα επόμενα χρόνια. Οι θεωρητικές εξελίξεις, όπως η θεωρία των φασμάτων **ακτίνων-X** και η παρατήρηση της **ανώμαλης διασποράς** και της εκτεταμένης λεπτής δομής στα φάσματα απορρόφησης **ακτίνων – X**, συνέβαλαν στην κατανόηση των **ακτίνων – X**. Επιπλέον, η φασματοσκοπία **ακτίνων – X φθορισμού** αναδείχθηκε ως πολύτιμο εργαλείο για τη στοιχειακή ανάλυση, η οποία διερευνήθηκε λεπτομερώς από το **1932** και μετά από ερευνητές όπως οι **Hevesy** και **Coster**.

Γενικά χαρακτηριστικά

Οι **ακτίνες – X**, γνωστές και ως ακτίνες **Ρέντγκεν**, αποτελούν ηλεκτρομαγνητική ακτινοβολία με μήκη κύματος που κυμαίνονται συνήθως από **0,005** έως **10 νανόμετρα (nm)**. Επικαλύπτονται με τις ακτίνες γάμμα στο άκρο του μικρού μήκους κύματος και πλησιάζουν την υπεριώδη ακτινοβολία στο άκρο του μεγάλου μήκους κύματος. Βασικές ιδιότητες των **ακτίνων – X** είναι το αόρατο, η διάδοση σε ευθείες γραμμές με ταχύτητα περίπου 3×10^8 μέτρα ανά δευτερόλεπτο (**m/s**), η ανεξαρτησία από ηλεκτρικά και μαγνητικά πεδία, η διαφορική **απορρόφηση** κατά τη **διέλευσή** τους από την ύλη, η **ανακλαστικότητα**, η **διάθλαση** και η **πόλωση**. Οι **ακτίνες – X** είναι επίσης ικανές να ionίζουν αέρια, να επηρεάζουν τις ηλεκτρικές ιδιότητες υγρών και στερεών, να μαυρίζουν φωτογραφικές πλάκες, να απελευθερώνουν φωτοηλεκτρόνια και ηλεκτρόνια ανάκρουσης, να προκαλούν βιολογικές αντιδράσεις και να εκπέμπουν τόσο συνεχή όσο και χαρακτηριστικά φάσματα. Η εκπομπή **χαρακτηριστικών ακτίνων – X** παρουσιάζει φάσματα γραμμών που είναι διακριτά για τα χημικά στοιχεία, καθιστώντας τα ανεκτίμητα για τη στοιχειακή ανάλυση.



Σχήμα 2: Ηλεκτρομαγνητικό Φάσμα με άξονα μηκών κύματος και ενεργειών

1.1 Εκπομπή συνεχούς ακτινοβολίας

Οι συνεχείς **ακτίνες X** παράγονται όταν φορτισμένα σωματίδια υψηλής ενέργειας, όπως τα ηλεκτρόνια, χάνουν ενέργεια κατά τη διέλευσή τους από το πεδίο **Coulomb** ενός πυρήνα. Αυτή η απώλεια ενέργειας έχει ως αποτέλεσμα την εκπομπή ενέργειας ακτινοβολίας, γνωστή ως ακτινοβολία πέδησης (**bremsstrahlung**). Η εκπομπή συνεχών **ακτίνων X** μπορεί να εξηγηθεί με τη χρήση της κλασικής ηλεκτρομαγνητικής θεωρίας, καθώς η επιτάχυνση των φορτισμένων σωματιδίων οδηγεί σε εκπομπή ακτινοβολίας. Τα ηλεκτρόνια υψηλής ενέργειας που προσκρούουν σε έναν στόχο υφίστανται ταχεία επιβράδυνση εντός του υλικού του στόχου, με αποτέλεσμα την αρνητική επιτάχυνση και την παραγωγή **bremsstrahlung**.

Το συνεχές φάσμα των **ακτίνων X** που παράγεται στους σωλήνες **ακτίνων X** χαρακτηρίζεται από ένα ελάχιστο μήκος κύματος (λ_{min}) που αντιστοιχεί στη μέγιστη ενέργεια των προσπίπτοντων ηλεκτρονίων:

$$\lambda_{min} = \frac{hc}{eV_0} \quad \{2.01\}$$

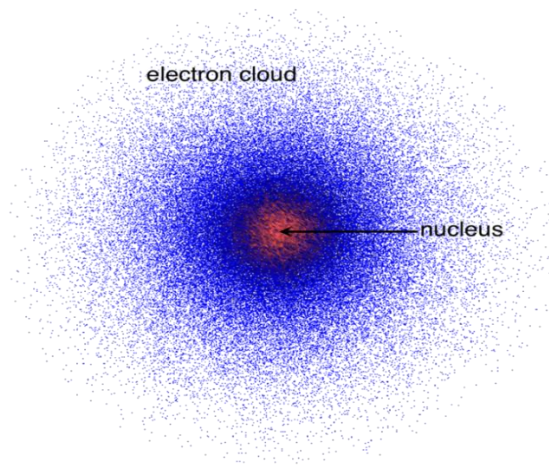
Εδώ, h αντιπροσωπεύει τη σταθερά του **Planck**, c είναι η ταχύτητα του φωτός, e δηλώνει το φορτίο ενός ηλεκτρονίου και (V_0) V_{max} είναι η μέγιστη τάση που εφαρμόζεται στο σωλήνα **ακτίνων-X**.

Ο λόγος μείωσης της ενέργειας εξαιτίας της ακτινοβολίας πέδησης (**bremsstrahlung**) προς αυτή της ενέργειας ιονισμού είναι υπολογίσιμη προσεγγιστικά μέσω του τύπου:

$$\left(\frac{m_0}{M_0}\right)^2 \frac{ZT}{1600m_0c^2} \quad \{2.02\}$$

1.2 Συνεχής φασματική κατανομή ακτίνων-X σε στόχους όπου προσκρούουν ηλεκτρόνια

Η παραγωγή **ακτίνων X** μέσω της αλληλεπίδρασης ηλεκτρονίων υψηλής ενέργειας με την ύλη παίζει καθοριστικό ρόλο σε διάφορες επιστημονικές και ιατρικές εφαρμογές. Η κατανόηση της φασματικής κατανομής αυτών των **ακτίνων X** είναι απαραίτητη για τη βελτιστοποίηση της χρήσης τους σε διάφορους τομείς, συμπεριλαμβανομένης της ιατρικής απεικόνισης, της ανάλυσης υλικών και της ταυτοποίησης στοιχείων. Η παρούσα συζήτηση εμβαθύνει στα θεμελιώδη χαρακτηριστικά των συνεχών φασμάτων **ακτίνων X** που εκπέμπονται από στόχους πάνω στους οποίους προσκρούουν ηλεκτρόνια και στη χωρική κατανομή τους, προσφέροντας πληροφορίες για την υποκείμενη φυσική και τις πρακτικές επιπτώσεις.



Σχήμα 3: Αναπαράσταση Ηλεκτρονικού νέφους γύρω από τον πυρήνα ατόμου

1.3 Φασματικά χαρακτηριστικά

1. Όριο μικρού μήκους κύματος (λ_{min})

Το συνεχές φάσμα **ακτίνων X** αρχίζει σε ένα συγκεκριμένο όριο μικρού μήκους κύματος, το οποίο συμβολίζεται ως λ_{min} . Το όριο αυτό είναι σημαντικό διότι οριοθετεί το όριο κάτω από το οποίο δεν παρατηρείται ακτινοβολία **ακτίνων X**. Η τιμή του λ_{min} είναι συνάρτηση των χαρακτηριστικών του σωλήνα **ακτίνων - X**, όπως η τάση, η κυματομορφή της τάσης και ο ατομικός αριθμός του **υλικού-στόχου**. Αυτές οι παράμετροι επηρεάζουν τα φωτόνια χαμηλότερης ενέργειας που μπορούν να παραχθούν μέσω της διαδικασίας **Bremsstrahlung**. Στις πρακτικές εφαρμογές, το λ_{min} καθορίζει το κατώτερο όριο για την ανίχνευση **ακτίνων X** και την ευαισθησία απεικόνισης.

2. Μήκος κύματος μέγιστης έντασης (λ_{max})

Μέσα στο συνεχές φάσμα των **ακτίνων - X**, υπάρχει ένα μήκος κύματος στο οποίο η εκπομπή **ακτίνων-X** φθάνει στη μέγιστη έντασή της, γνωστό ως λ_{max} . Ειδικότερα, το λ_{max} είναι περίπου 1.5 φορές το λ_{min} [Krane, K. S. (2012)], αλλά η σχέση αυτή δεν είναι σταθερή και μπορεί να μεταβάλλεται ανάλογα με τους προαναφερθέντες παράγοντες. Η κατανόηση της μεταβολής του λ_{max} με αυτές τις παραμέτρους είναι απαραίτητη για τη βελτιστοποίηση των πηγών **ακτίνων-X** για συγκεκριμένες εφαρμογές. Η ακριβής γνώση της λ_{max} βοηθά στην επιλογή του κατάλληλου εξοπλισμού ανίχνευσης ακτίνων-X και στο σχεδιασμό αποδοτικών συστημάτων απεικόνισης **ακτίνων-X**.

3. Εξάρτηση από την ένταση

Η ένταση του συνεχούς φάσματος **ακτίνων-X** συνδέεται στενά με το τετράγωνο της τάσης που εφαρμόζεται στη λυχνία **ακτίνων-X** και την πρώτη δύναμη του ατομικού αριθμού (Z) του **υλικού-στόχου**. Εμπειρικές μελέτες, ιδίως το έργο του **Kulenkampff**, παρείχαν μαθηματικές εκφράσεις για την περιγραφή αυτής της κατανομής ενέργειας. Η διατύπωση έχει την ακόλουθη μορφή:

$$I(v)dv = i[aZ(v_0 - v) + bZ^2]dv \quad \{2.03\}$$

Όπου το $I(\nu) d\nu$ αντιπροσωπεύει την ένταση εντός ενός συγκεκριμένου εύρους συχνοτήτων ($\nu, \nu + d\nu$), i είναι το ρεύμα ηλεκτρονίων που προσπίπτει στο στόχο, Z είναι ο ατομικός αριθμός του υλικού του στόχου και a, b είναι σταθερές. Αξίζει να σημειωθεί ότι ο δεύτερος όρος της εξίσωσης είναι συνήθως μικρός σε σύγκριση με τον πρώτο και συχνά αγνοείται. Κατά συνέπεια, η συνολική ολοκληρωμένη ένταση εκφράζεται ως εξής:

$$I = i(a'ZV_0^2 + b'Z^2V_0) \quad \{2.04\}$$

Όπου $a' = a(e^2/h^2)/2$ και $b' = b(e/h)$. Για τις σταθερές αυτές έχουν καθοριστεί εμπειρικές τιμές, διευκολύνοντας τους πρακτικούς υπολογισμούς.

Η απόδοση (Eff) της μετατροπής της εισερχόμενης ηλεκτρικής ισχύος σε ακτίνες X δίνεται από τη σχέση: $Eff = \frac{I}{V_0 i} = a'Z(V_0 + 16.3Z)$ {2.05}

όπου V_0 είναι σε **Volt (V)**. Πειραματικοί προσδιορισμοί έχουν δώσει μια κατά προσέγγιση τιμή για το a_0 , μια κρίσιμη παράμετρο για τους υπολογισμούς της απόδοσης παραγωγής **ακτίνων X**.

Κβαντική θεωρία και φασματική κατανομή

Η ολοκληρωμένη κατανόηση του συνεχούς φάσματος των **ακτίνων-X** οφείλεται σε μεγάλο βαθμό στην κβαντική θεωρία, ιδίως στις διατυπώσεις των **Kramers (1923)** και **Wentzel (1924)**. Ο **Kramers** χρησιμοποίησε την αρχή της αντιστοιχίας για να εξάγει εκφράσεις για την ενεργειακή κατανομή των συνεχών **ακτίνων-X** σε **λεπτούς στόχους**. Οι εξισώσεις του είναι καθοριστικές για την κατανόηση της ενεργειακής κατανομής των εκπεμπόμενων ακτίνων X.

$$I(\nu)d\nu = \frac{16\pi^2 A Z^2 e^5}{3\sqrt{3}m_0 V_0 c^3} d\nu, \quad \nu < \nu_0 \quad \{2.06\}$$

$$I(\nu)d\nu = 0, \quad \nu > \nu_0$$

Όπου το A αντιπροσωπεύει την ατομική μάζα του υλικού-στόχου.

$$I(\nu)d\nu = \frac{8\pi e^2 h}{3\sqrt{3}lm_0 c^3} Z(\nu_0 - \nu)d\nu \quad \{2.07\}$$

Η απόδοση της παραγωγής ακτίνων X που υπολογίζεται μέσω του νόμου του **Kramers** δίνεται από τη σχέση

$$Eff = 9.2 \times 10^{-10} Z V_0 \quad \{2.08\}$$

Ειδικότερα, οι εξισώσεις του **Kramers** παρέχουν ένα πολύτιμο πλαίσιο για την εξήγηση των φασματικών χαρακτηριστικών των **ακτίνων X** που εκπέμπονται από λεπτούς στόχους.

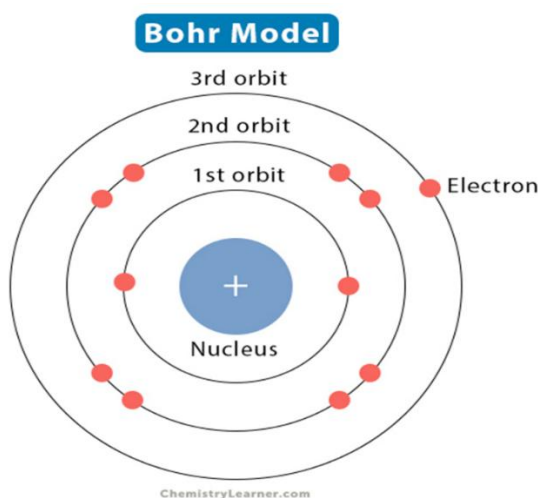
2. ΕΚΠΟΜΠΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΚΤΙΝΩΝ - X

Η εκπομπή χαρακτηριστικών ακτίνων-X περιλαμβάνει μεταβάσεις ατομικών ηλεκτρονίων εντός των υλικών-στόχων, οι οποίες προκαλούνται συνήθως είτε από βομβαρδισμό ηλεκτρονίων είτε από απορρόφηση φωτονίων. Τα ηλεκτρόνια ενός ατόμου καταλαμβάνουν διακριτούς φλοιούς (**Shells**) ή, ισοδύναμα, ενεργειακά επίπεδα

(**Energy Levels**). Οι φλοιοί αυτοί, που περιγράφονται και ως στοιβάδες, περιβάλλουν τον πυρήνα και η ενέργεια των ηλεκτρονίων που τους καταλαμβάνουν ποικίλλει ανάλογα με την απόστασή τους από αυτόν (βλ. Σχήμα 4).

Για την απομάκρυνση ενός ηλεκτρονίου από έναν εσωτερικό φλοιό απαιτείται η προσφορά ενός συγκεκριμένου, ελάχιστου ποσού ενέργειας. Είναι σημαντικό να σημειωθεί ότι όσο πιο μακριά από τον πυρήνα βρίσκεται ένας ηλεκτρονικός φλοιός, τόσο μικρότερη είναι η ενέργεια που δεσμεύει τα ηλεκτρόνια του στο άτομο. Οι φλοιοί αυτοί συμβολίζονται με γράμματα της λατινικής αλφαβήτου: **K**, **L**, **M**, **N**, κ.ο.κ., με τον φλοιό **K** να είναι ο εσωτερικότερος και να μπορεί να φιλοξενήσει έως **2** (δύο) ηλεκτρόνια. Ο φλοιός **L** αποτελείται από τρία υποεπίπεδα (**LI**, **LII**, **LIII**) και μπορεί να περιέχει έως **8** (οκτώ) ηλεκτρόνια, ενώ ο φλοιός **M** διαθέτει πέντε υποεπίπεδα (**MI** έως **MV**) και μπορεί να φιλοξενήσει έως **18** (δεκαοκτώ) ηλεκτρόνια.

Για παράδειγμα, για την απομάκρυνση ενός ηλεκτρονίου από τον φλοιό **K**, πρέπει να προσδοθεί στο άτομο ενέργεια τουλάχιστον ίση με την ενέργεια σύνδεσης (**Binding Energy**) του ηλεκτρονίου αυτού. Αντίστοιχα, για την απομάκρυνση ενός ηλεκτρονίου από τον φλοιό **L**, η απαιτούμενη ενέργεια είναι μικρότερη από αυτή του φλοιού **K**, και ούτω καθεξής. Η ελάχιστη αυτή ενέργεια που απαιτείται για την απομάκρυνση ενός ηλεκτρονίου από έναν συγκεκριμένο φλοιό ονομάζεται ενέργεια σύνδεσης (**Binding Energy**) ή, ισοδύναμα, ακμή απορρόφησης (**Absorption Edge**) του συγκεκριμένου φλοιού. Οι τιμές αυτές είναι χαρακτηριστικές για κάθε στοιχείο και κάθε φλοιό. Επομένως, η ενέργεια ενός ηλεκτρονίου εξαρτάται άμεσα από τον φλοιό στον οποίο βρίσκεται και από το στοιχείο στο οποίο ανήκει.



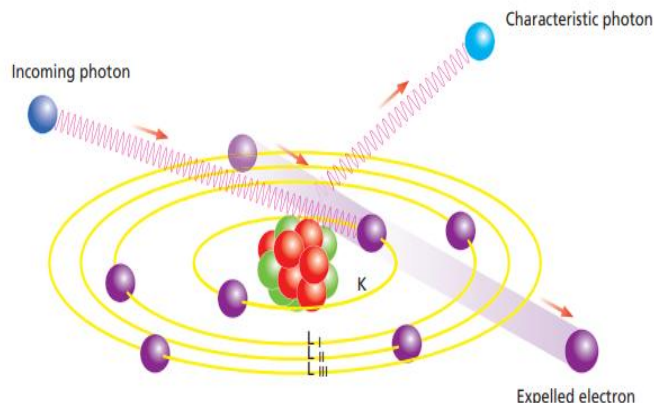
Σχήμα 4: Το ατομικό πρότυπο του Bohr

$$\text{Energy Level} = \text{Binding Energy} = \text{Absorption Edge}$$

Μια πιο διαισθητική κατανόηση αυτών των μηχανισμών και της κβαντωμένης φύσης των ενεργειακών επιπέδων των ηλεκτρονίων σε ένα άτομο παρέχεται από το ατομικό μοντέλο του **Bohr** (Σχήμα 4). Αν και απλοϊκό, το μοντέλο του Bohr εισήγαγε την ιδέα των διακριτών ενεργειακών επιπέδων και αποτέλεσε σημαντικό βήμα για την κατανόηση της ατομικής δομής.

Όταν δημιουργηθεί μια κενή θέση σε έναν εσωτερικό φλοιό, όπως ο **K**, αυτή μπορεί να πληρωθεί από ένα ηλεκτρόνιο προερχόμενο από έναν εξωτερικότερο φλοιό, για παράδειγμα τον **L** ή τον **M**. Εάν η πλήρωση γίνει από ηλεκτρόνιο του φλοιού **L**,

εκπέμπεται χαρακτηριστική ακτινοβολία X της σειράς $K\alpha$. Εάν η πλήρωση γίνει από ηλεκτρόνιο του φλοιού M , εκπέμπεται ακτινοβολία της σειράς $K\beta$. Αντίστοιχα, κενές θέσεις που δημιουργούνται στους φλοιούς L ή M (δευτερογενής ιονισμός) μπορούν επίσης να πληρωθούν από ηλεκτρόνια εξωτερικότερων φλοιών, οδηγώντας στην εκπομπή χαρακτηριστικών ακτίνων X των σειρών L , M κ.ο.κ (Σχήμα 5).



Σχήμα 5: Διέγερση ατόμου από προσκρουόμενο φωτόνιο με εκπομπή χαρακτηριστικού φωτονίου και ηλεκτρονίου

2.1 Φασματική σειρά στις ακτίνες X

Οι γραμμές εκπομπής στα φάσματα των ακτίνων X αντιστοιχούν σε μεταβάσεις μεταξύ συγκεκριμένων ενεργειακών επιπέδων των ατόμων. Η συχνότητα (ν) μιας γραμμής ακτίνων X καθορίζεται από τη διαφορά ενέργειας ($E_1 - E_2$) μεταξύ αυτών των επιπέδων, ενώ η h (σταθερά του **Planck**) χρησιμοποιείται για τη μετατροπή της σε μήκος κύματος.

$$\nu = \frac{E_1 - E_2}{h} \quad \{2.07\}$$

Οι ενεργειακές στάθμες περιγράφονται από κβαντικούς αριθμούς (n , l , s και j) και οι ενέργειές τους (E) για δεδομένο ατομικό αριθμό (Z) υπολογίζονται χρησιμοποιώντας μια έκφραση που περιλαμβάνει τη σταθερά **Rydberg** (R), τις ατομικές παραμέτρους και τη σταθερά λεπτής δομής (a). Οι μεταβάσεις **ακτίνων X** πρέπει να τηρούν κανόνες επιλογής. Για τις διπολικές μεταβάσεις, η μεταβολή της στροφορμής (DI) πρέπει να είναι 1 και η μεταβολή της συνολικής στροφορμής (Dj) μπορεί να είναι 0 ή 1 .

$$\frac{E}{Rh} = \frac{(Z - S_{n,l})^2}{n^2} + \alpha^2 \frac{(Z - d_{n,l,j})^2}{n^3} \left(\frac{1}{l + \frac{1}{2}} - \frac{3}{4n} \right) - \alpha^2 \frac{(Z - d_{n,l,j})^4 j(j+1) - l(l+1) - s(s+1)}{2l(l + \frac{1}{2})(l+1)} \quad \{2.08\}$$

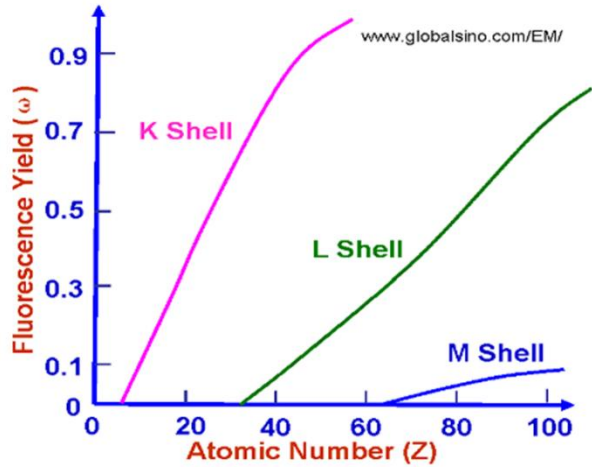
Οι μεταβάσεις που απαγορεύονται για διπολική ακτινοβολία μπορούν να εμφανιστούν ως **πολυπολικές μεταβάσεις**, όπως οι μεταβάσεις ηλεκτρικού τετραπόλου και μαγνητικού διπόλου, που υπόκεινται σε διαφορετικούς κανόνες επιλογής. Ο νόμος του **Moseley** συσχετίζει το μήκος κύματος μιας γραμμής **ακτίνων- X** με τον ατομικό αριθμό (Z), βοηθώντας στη συστηματοποίηση των φασμάτων **ακτίνων- X** .

$$\frac{1}{\lambda} = k(Z - \sigma)^2 \quad \{2.09\}$$

$$\alpha = \frac{2\pi e^2}{hc} \quad \{2.10\}$$

2.2 Απόδοση φθορισμού

Η απόδοση φθορισμού (**fluorescence yield**, ω) ενός συγκεκριμένου ατομικού φλοιού εκφράζει την πιθανότητα η αποδιέγερση του ατόμου μετά τον ιονισμό αυτού του φλοιού να γίνει μέσω της εκπομπής ενός χαρακτηριστικού φωτονίου ακτίνων-Χ (ακτινοβόλος μετάβαση), αντί μέσω μιας μη ακτινοβόλου διαδικασίας (όπως η εκπομπή ηλεκτρονίου **Auger**) Σχήμα 6. Η απόδοση φθορισμού εξαρτάται από τον συγκεκριμένο φλοιό (π.χ., **K**, **L**, **M**), τον υποφλοιό (π.χ., **LI**, **LII**, **LIII**) και τον ατομικό αριθμό του στοιχείου. Για τον φλοιό **K**, η



Σχήμα 6: Απόδοση φθορισμού εξαρτωμένη του ατομικού αριθμού

απόδοση φθορισμού (ω_K) ορίζεται ως ο λόγος του αριθμού των εκπεμπόμενων φωτονίων της σειράς **K** (I_K) προς τον συνολικό αριθμό των πρωτογενών κενών θέσεων που δημιουργήθηκαν στον φλοιό **K** (n_K), δηλαδή: $\omega_K = \frac{I_K}{n_K}$ {2.12}, όπου I_K είναι ο συνολικός αριθμός των φωτονίων **ακτίνων X K** που εκπέμπονται και n_K είναι ο αριθμός των κενών θέσεων του πρωτογενούς κελύφους **K**. Οι αποδόσεις φθορισμού για τους εξωτερικότερους φλοιούς (**L**, **M**, κ.λπ.) είναι πιο περίπλοκες, καθώς επηρεάζονται από την ύπαρξη πολλαπλών υποφλοιών και την πιθανότητα μη ακτινοβόλων μεταβάσεων μεταξύ αυτών των υποφλοιών, γνωστών ως μεταβάσεις Coster-Kronig.

$$\omega_i^X = \frac{I_i^X}{n_i^X} \quad \{2.13\}$$

Οι μέσες αποδόσεις φθορισμού για ένα κέλυφος (ω^X) υπολογίζονται με βάση τους σχετικούς αριθμούς των πρωτογενών κενών θέσεων (n_i^X) και των αποδόσεων φθορισμού των υποκελυφών (ω_i^X).

$$\overline{\omega_X} = \sum_{i=1}^k N_i^X \omega_i^X \quad \{2.14\}$$

2.3 Αλληλεπίδραση των φωτονίων με την ύλη

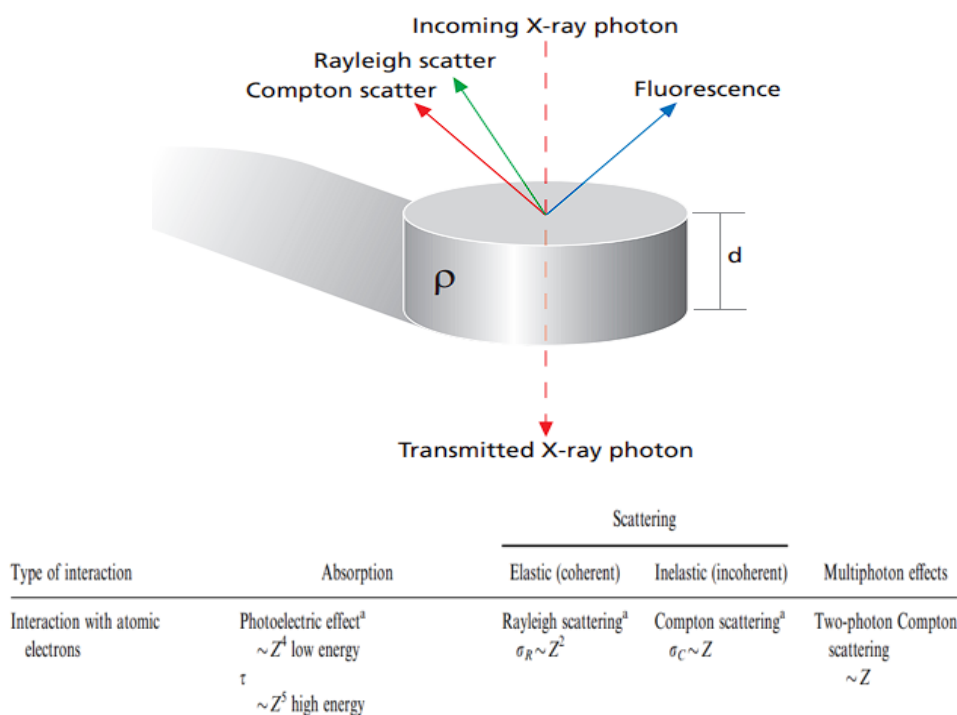
Όταν τα φωτόνια αλληλεπιδρούν με την ύλη, μπορούν να σκεδαστούν ή να απορροφηθούν, οδηγώντας σε διαφορετικά φαινόμενα με βάση τον στόχο (π.χ. ηλεκτρόνια, άτομα ή πυρήνες) και τον τύπο του συμβάντος (απορρόφηση, σκέδαση ή παραγωγή ζεύγους). Είναι σημαντικό να σημειωθεί ότι αυτές οι αλληλεπιδράσεις είναι ανεξάρτητες από την προέλευση του φωτονίου, είτε πρόκειται για πυρηνική ή ηλεκτρονική μετάβαση. Για λόγους απλότητας, θα αναφερόμαστε εδώ τόσο στις ακτίνες γάμμα (γ) όσο και στις **ακτίνες X** ως "**φωτόνια**". Οι πιθανές αλληλεπιδράσεις συνοψίζονται στο (**Hubbell, 1969**), όπου το " **τ** " αντιπροσωπεύει τη συνολική διατομή φωτοηλεκτρικής απορρόφησης ανά άτομο και τα " **σ_R** " και " **σ_C** " υποδηλώνουν τις διατομές σύγκρουσης **Rayleigh** και **Compton**, ωστόσο η παραγωγή ζεύγους προϋποθέτει ενέργειες μεγαλύτερες του 1.022 MeV αντίστοιχα. Η πιθανότητα κάθε

αλληλεπίδρασης μπορεί να εκφραστεί ως διατομή σύγκρουσης ανά άτομο, ηλεκτρόνιο ή πυρήνα εντός του απορροφητή. Το άθροισμα όλων αυτών των διατομών, κανονικοποιημένο ανά άτομο, δίνει την πιθανότητα " σ_{tot} " ότι ένα προσπίπτον φωτόνιο θα αλληλεπιδράσει κατά τη διέλευσή του από έναν πολύ λεπτό απορροφητή που περιέχει ένα άτομο ανά τετραγωνικό εκατοστό της επιφάνειας που είναι κάθετη στη διαδρομή του:

$$\sigma_{tot} = \tau + \sigma_R + \sigma_C \quad \{2.21\}$$

Η συνολική διατομή σύγκρουσης ανά άτομο " σ_{tot} ", όταν πολλαπλασιαστεί με τον αριθμό των ατόμων ανά κυβικό εκατοστό του απορροφητή, δίνει τον γραμμικό συντελεστή εξασθένησης " μ " ανά εκατοστό διαδρομής στον απορροφητή:

$$\mu^* \left(\frac{1}{\text{cm}} \right) = \sigma_{tot} \left(\frac{\text{cm}^2}{\text{atom}} \right) \rho \left(\frac{\text{g}}{\text{cm}^3} \right) \frac{N_0}{A} \left(\frac{\text{atoms}}{\text{g}} \right) \quad \{2.22\}$$



Σχήμα 7: Πίνακας αλληλεπιδράσεων των υλικών με τις Ακτίνες - X (κάτω) και αναπαράσταση των αλληλεπιδράσεων (επάνω)

Εδώ, το " ρ " είναι η πυκνότητα του μέσου, το " N_0 " είναι ο αριθμός *Avogadro* (6.0225×10^{23} άτομα/mol) και το " A " αντιπροσωπεύει την ατομική μάζα. Αξίζει να σημειωθεί ότι ο συντελεστής απορρόφησης είναι πιο συγκεκριμένος από τον συντελεστή εξασθένησης. Η εξασθένηση περιλαμβάνει τις καθαρά ελαστικές διεργασίες όπου το φωτόνιο εκτρέπεται χωρίς να χάνει ενέργεια, περιλαμβάνοντας μόνο έναν συντελεστή σκέδασης. Σε μια φωτοηλεκτρική αλληλεπίδραση, ολόκληρη η ενέργεια του φωτονίου απορροφάται από ένα άτομο του μέσου. Στο φαινόμενο **Compton**, μέρος της ενέργειας απορροφάται και εμφανίζεται ως κινητική ενέργεια ενός ηλεκτρονίου ανάκρουσης **Compton**, ενώ το υπόλοιπο παραμένει ως σκεδαζόμενο φωτόνιο **Compton**. Έτσι, η απορρόφηση μετατρέπει την ενέργεια του φωτονίου σε κινητική ενέργεια φορτισμένου σωματιδίου, ενώ η σκέδαση έχει ως αποτέλεσμα την εκτροπή του προσπίπτοντος φωτονίου. Για στενές, παράλληλες και μονοχρωματικές δέσμες, η εξασθένηση φωτονίων στην ομογενή ύλη ακολουθεί τον εκθετικό νόμο:

$$I = I_0 e^{-\mu^* t} \quad \{2.23\}$$

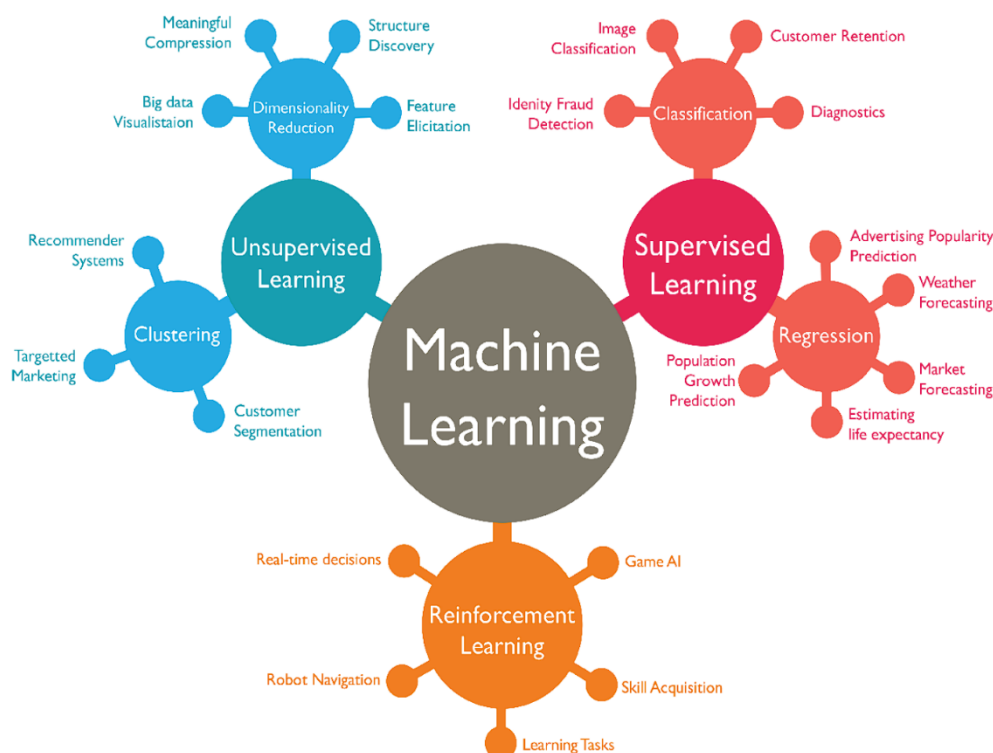
Όπου " I " είναι η μεταδιδόμενη ένταση, " I_0 " είναι η προσπίπτουσα ένταση, " t " είναι το πάχος του απορροφητή σε εκατοστά. Σε περιπτώσεις που αφορούν χημικές ενώσεις ή μείγματα, ο συντελεστής εξασθένησης μάζας " μ " μπορεί να προσεγγιστεί χρησιμοποιώντας σταθμισμένους μέσους όρους των συντελεστών των συστατικών στοιχείων: $\mu = \sum_{i=1}^n W_i \mu_i$ {2.24}

Εδώ, " W_i " είναι το κλάσμα βάρους του *i-οστού* στοιχείου και " n " είναι ο συνολικός αριθμός των στοιχείων στον απορροφητή.

Για πιο πολύπλοκες γεωμετρίες και καταστάσεις, η εξασθένηση εξακολουθεί να είναι ουσιαστικά εκθετική, αλλά τροποποιείται από δύο πρόσθετους παράγοντες: τον παράγοντα γεωμετρίας και τον παράγοντα συσσώρευσης. Αυτοί οι παράγοντες λαμβάνουν υπόψη τα δευτερογενή φωτόνια που παράγονται στον απορροφητή, κυρίως λόγω σκέδασης *Compton*.

Β' ΘΕΩΡΗΤΙΚΟ ΜΕΡΟΣ

ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ



Εικόνα 1: Γραφική απεικόνιση του τομέα της Μηχανικής Μαθησης

Κεφάλαιο: Εισαγωγή στη Μηχανική Μάθηση για την Ανάλυση Έργων Τέχνης

Προχωρώντας βηματικά προς το την πειραματική διαδικασία έχουμε μια ακόμα «κορυφή» να κατακτήσουμε. Πλέον στις μέρες της ανάπτυξης της τεχνητής νοημοσύνης με αλματώδη βήματα γίνεται όλο και πιο νοητό να την χρησιμοποιήσουμε σ' όλες τις πτυχές τις ζωής μας. Έχει κατακλύσει από κινητά τηλέφωνα, κάμερες, υπολογιστές, προγράμματα, βιομηχανίες ακόμα και τηλεφωνικά κέντρα και υπηρεσίες. Είναι πλέον αναπόσπαστο κομμάτι, διότι προσφέρει γρήγορες απαντήσεις και διευκόλυνση στον άνθρωπο, πως όμως καταλήξαμε να έχουμε αυτή την «βοήθεια»; Η σύντομη απάντηση είναι μέσω της μηχανικής μάθησης και των πτυχών της, όπως και φαίνεται και στο παραπάνω σχήμα.

Ορισμός και Σημασία της Μηχανικής Μάθησης & Εξόρυξης Δεδομένων

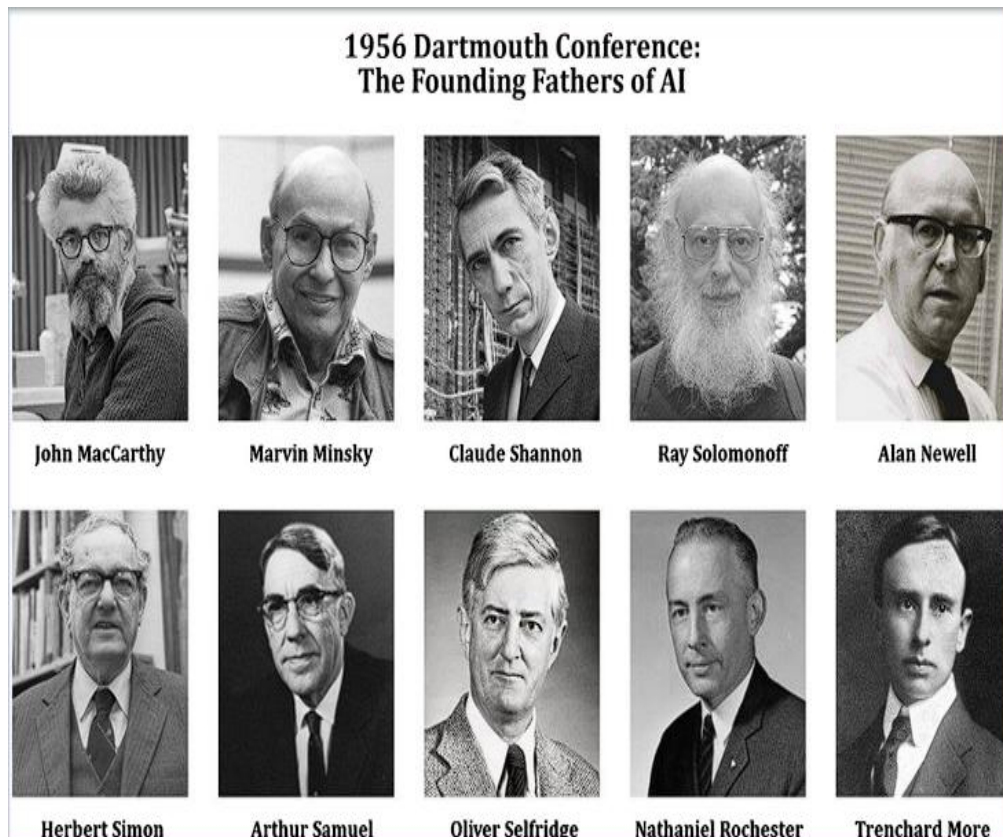
Η μηχανική μάθηση είναι ένας ταχέως εξελισσόμενος τομέας της επιστήμης των υπολογιστών, που διαθέτει ένα ευρύ φάσμα εφαρμογών [ML-11, ML-14]. Ο βασικός σκοπός της περιστρέφεται γύρω από την αυτόματη ανίχνευση σημαντικών μοτίβων σε δεδομένα, γεγονός που καθιστά αναγκαία τη χρήση ακριβούς γλώσσας, γραμματικής ακρίβειας και επίσημου ύφους. Η συμμόρφωση με τις συμβατικές δομές ενοτήτων, η χρήση σαφών και τεκμηριωμένων επικεφαλίδων και η διατήρηση της συνέπειας στο στυλ παραπομπών και υποσημειώσεων αποτελούν απαραίτητα στοιχεία της αποτελεσματικής επικοινωνίας. Κατά την αξιοποίηση των εργαλείων μηχανικής μάθησης, είναι υψίστης σημασίας η διατήρηση της αντικειμενικότητας, ενώ εναρμονίζονται τα αποδεικτικά στοιχεία με τις αιτιώδεις συνδέσεις στις δηλώσεις. Για

την ενίσχυση της σαφήνειας, οι συντομογραφίες τεχνικών όρων θα πρέπει να διευκρινίζονται κατά την αρχική τους χρήση.

Η μηχανική μάθηση (**Machine Learning**) έχει εξελιχθεί σε ακρογωνιαίό λίθο της τεχνολογίας των πληροφοριών, με απρόσκοπτη ενσωμάτωση στην καθημερινή μας ζωή. Με την εκθετική αύξηση της παραγωγής δεδομένων, έχει γίνει ένα απαραίτητο εργαλείο τόσο για τις βιομηχανίες όσο και για τον ακαδημαϊκό χώρο [ML-11]. Η έλευση της εξελιγμένης ανάλυσης δεδομένων εδραιώνει περαιτέρω τον ρόλο της ως βασικού συστατικού που οδηγεί στην τεχνολογική πρόοδο.

Η εξόρυξη δεδομένων (**Data Mining**) αποτελεί μία από τις σημαντικότερες εφαρμογές της μηχανικής μάθησης. Οι αναλυτικές διαδικασίες πρέπει να είναι απαλλαγμένες από υποκειμενικότητα, με τη δέουσα προσοχή στα πιθανά σφάλματα. Η καθιέρωση περίπλοκων συνδέσεων μεταξύ διαφόρων χαρακτηριστικών είναι υψίστης σημασίας. Η εξόρυξη δεδομένων και η μηχανική μάθηση είναι άρρηκτα συνδεδεμένες, προσφέροντας τη δυνατότητα για βαθιές γνώσεις μέσω σχολαστικά επεξεργασμένων αλγορίθμων μάθησης.

Ο τομέας της εξόρυξης δεδομένων έχει κάνει αξιοσημείωτα βήματα προόδου, με συνεχείς εξελίξεις στον ορίζοντα. Η εμφάνιση της μηχανικής μάθησης είναι στενά συνδεδεμένη με την ανάπτυξη της πληροφορικής και της νανοτεχνολογίας, πυροδοτώντας μια θέρμη για την αποκάλυψη κρυμμένων μοτίβων δεδομένων για αξιοποιήσιμες γνώσεις. Η συγχώνευση της στατιστικής, της επιστήμης των δεδομένων και της μηχανικής μάθησης έχει καταλύσει σημαντική πρόοδο σε διάφορους τομείς [ML-15]. Υποστηριζόμενος από μια ισχυρή μαθηματική βάση και ενισχυμένος με ισχυρά εργαλεία που περιλαμβάνουν την πληροφορική και τη θεωρία πληροφοριών, αυτός ο επιστημονικός κλάδος ευδοκimeί. Οι αλγόριθμοι μηχανικής μάθησης κατηγοριοποιούνται με βάση τα επιδιωκόμενα αποτελέσματά τους, με ιδιαίτερα υψηλή ζήτηση για την επιβλεπόμενη μάθηση (**Supervised Learning**). Η αδιάκοπη εισροή δεδομένων έχει ωθήσει τη διαρκή εξέλιξη των τεχνικών μηχανικής μάθησης, οδηγώντας σε ένα πλούσιο φάσμα αλγορίθμων που ανταποκρίνονται τόσο σε εποπτευόμενα όσο και σε μη εποπτευόμενα παραδείγματα [ML-10, ML-12].



Εικόνα 2: Προτεργάτες της Τεχνίτης Νοημοσύνης

Ιστορία Μηχανικής Μάθησης

Η μηχανική μάθηση (**ML**) είναι ιδανική για την αξιοποίηση του ανεκμετάλλετου δυναμικού που κρύβεται στο αχανές βασίλειο των μεγάλων δεδομένων (**Big Data**) [KMeans-1, ML-11]. Η εγγενής δύναμή της έγκειται στην ικανότητα απόσταξης πολύτιμων πληροφοριών από περίπλοκες και ποικίλες πηγές δεδομένων με ελάχιστη ανθρώπινη παρέμβαση, η οποία υποστηρίζεται από τον επικεντρωμένο στα δεδομένα και κλιμακούμενο χαρακτήρα της. Η **ML** αποδεικνύεται ιδιαίτερα ικανή στο χειρισμό των περιπλοκών των ετερογενών πηγών δεδομένων και της εκτεταμένης σειράς μεταβλητών και όγκων δεδομένων που χαρακτηρίζουν τα διευρυνόμενα σύνολα δεδομένων. Καθώς περισσότερα δεδομένα διοχετεύονται σε ένα πλαίσιο **ML**, η ικανότητά του για εκπαίδευση και παραγωγή ανώτερων γνώσεων αυξάνεται αναλογικά. Ανεμπόδιστη από τους περιορισμούς της ανάλυσης σε ατομικό επίπεδο, η **ML** υπερέχει στην αποκάλυψη και τη διαλεύκανση μοτίβων που κρύβονται μέσα στα δεδομένα. Ο *Nilsson (2010)* οριοθετεί το **ML** ως ένα υποσύνολο της τεχνητής νοημοσύνης που κατασκευάζει δυναμικούς αλγορίθμους προικισμένους με την ικανότητα να λαμβάνουν αποφάσεις με βάση τα δεδομένα, διακρίνοντάς τους από τις στατικές εντολές προγραμματισμού. Η αλληλεπίδραση μεταξύ αυτών των δύο πεδίων είναι αναπόσπαστη και σχεδόν αδιαχώριστη. Στη συνέχεια παρατίθενται συνοπτικές πληροφορίες για την ιστορική εξέλιξη και των δύο τομέων.

- I. **1950:** Ο *Alan Turing* εισήγαγε το εμβληματικό **Turing Test** (*Turing, 1950*), μια κομβική έννοια κατά την οποία μια μηχανή ή ένας υπολογιστής πρέπει να υποδυθεί πειστικά έναν άνθρωπο για να θεωρηθεί ευφυής.

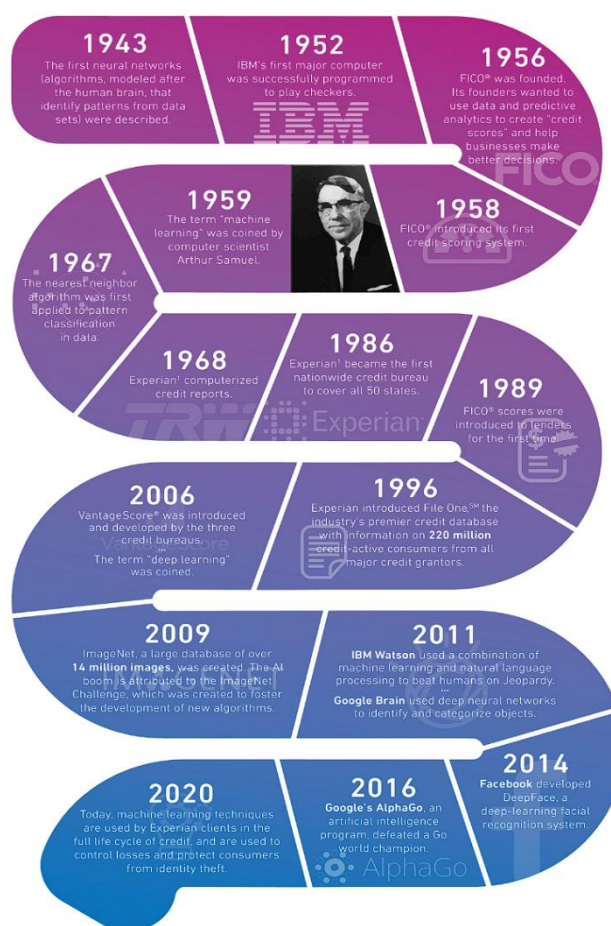
- II. **1951:** Ο *Minsky (1951)* και ο *Dean Edmonds* πρωτοπορούν με το πρώτο τεχνητό νευρωνικό δίκτυο (*ANN*) που ονομάστηκε *Stochastic Neural Analog Reinforcement Computer (SNARC)* και αποτελείται από **40 νευρώνες**.
- III. **1952:** Ο *Arthur Samuel*, πρωτοπόρος στα ηλεκτρονικά παιχνίδια και την τεχνητή νοημοσύνη, δημιούργησε το πρώτο πρόγραμμα υπολογιστή ικανό να μαθαίνει δυναμικά, με παράδειγμα την ικανότητά του να παίζει ντάμα και να ξεπερνά γρήγορα τις ανθρώπινες επιδόσεις.
- IV. **1956:** Ο όρος "τεχνητή νοημοσύνη" επινοήθηκε στο συνέδριο του **Dartmouth**, μια στιγμή καμπίς όπου ο *John McCarthy*, *Claude Shannon*, και *Nathan Rochester* συγκεντρώθηκαν για να διερευνήσουν τα νευρωνικά δίκτυα, την αυτοματοποιημένη θεωρία και τις μελέτες νοημοσύνης. Η πρόταση του συνεδρίου διαβεβαίωνε τη δυνατότητα ακριβούς περιγραφής όλων των πτυχών της μάθησης και της νοημοσύνης για τη δημιουργία μηχανών που θα μπορούσαν να τις προσομοιώσουν. Στους συμμετέχοντες περιλαμβάνονταν αξιόλογες προσωπικότητες όπως οι *Arthur Samuel (IBM)*, *Trenchard More (Princeton)*, *Oliver Selfridge* και *Ray Solomonoff (MIT)*. Οι *Herbert Simon* και *Allen Newell*, και οι δύο από το *Carnegie Tech*, είχαν ένα εν εξελίξει έργο, το *Logic Theorist*, το οποίο αναγνωρίστηκε μετά το συνέδριο ως το πρωτοποριακό σύστημα τεχνητής νοημοσύνης.
- V. **1958:** Ο *Frank Rosenblatt* επινόησε το *perceptron*, το πρώτο *ANN* για υπολογιστές, με κύριο στόχο την αναγνώριση μοτίβων και σχημάτων (*Rosenblatt, 1958*).
- VI. Το **1959**, οι *Widrow* και *Hoff (1960)* παρουσίασαν πρωτοποριακά μοντέλα γνωστά ως "*ADALINE*", το οποίο διέπρεψε στην αναγνώριση δυαδικών προτύπων, και "*MADALINE*", το πρωτοποριακό νευρωνικό δίκτυο που σχεδιάστηκε για την εξάλειψη της ηχούς στις τηλεφωνικές γραμμές. Και τα δύο μοντέλα αξιοποίησαν την έννοια των πολλαπλών *ADaptive LINear Elements*. Κατά τη διάρκεια αυτής της εποχής, ο *Arthur Samuel* συνέβαλε καθοριστικά με την αντίληψη της μηχανικής μάθησης (*ML*) ως ενός πεδίου όπου οι υπολογιστές θα μπορούσαν να αποκτήσουν γνώση χωρίς ρητό προγραμματισμό. Παρά την αρχική επιτυχία του *MADALINE*, η πρόοδος παρέμεινε περιορισμένη μέχρι τα τέλη της δεκαετίας του **1980**, γεγονός που αποδίδεται σε διάφορους παράγοντες, συμπεριλαμβανομένης της διαρκούς δημοτικότητας και απλότητας της αρχιτεκτονικής **Von Neumann**.
- VII. Το **1967**, οι *Cover et al. (1967)* εισήγαγαν τον κανόνα του πλησιέστερου γείτονα (*NN*), ο οποίος είχε προηγουμένως αναφερθεί ως "ταξινομητής ελάχιστης απόστασης" από τον *Nilsson (1965)* και ως "αλγόριθμος εγγύτητας" από τον *Sebestyen (1962)*.
- VIII. Το κομβικό έτος **1982** είδε την παρουσίαση του *Hopfield (1982)* στην Εθνική Ακαδημία Επιστημών, προτείνοντας την ανάπτυξη πιο πρακτικών μηχανών που χρησιμοποιούν αμφίδρομες γραμμές αντί για συμβατικούς νευρώνες. Επιπλέον, το **1982** πραγματοποιήθηκε ένα σημαντικό κοινό συνέδριο ΗΠΑ – Ιαπωνίας για τα συνεργατικά/ανταγωνιστικά νευρωνικά δίκτυα, όπου η Ιαπωνία παρουσίασε το φιλόδοξο σχέδιο υπολογιστών πέμπτης γενιάς που αφορούσε την τεχνητή νοημοσύνη.

IX. Το 1985, το *Αμερικανικό Ινστιτούτο Φυσικής* εγκαινίασε την ετήσια συνάντηση των νευρωνικών δικτύων στην πληροφορική, θέτοντας τις βάσεις για την πρόοδο στον τομέα. Αυτή η δυναμική κορυφώθηκε με το πρώτο Διεθνές Συνέδριο για τα Νευρωνικά Δίκτυα από το *Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE)* το 1987.

Η θεμελιώδης παραδοχή της ML περιστρέφεται γύρω από την ικανότητα των μηχανών (Εικόνα 3), με τη χρήση αλγορίθμων και μοντέλων, να δημιουργούν προβλέψεις με βάση δεδομένα [ML-5]. Οι αλγόριθμοι ML ακολουθούν διάφορες

προσεγγίσεις, συμπεριλαμβανομένων των μεθόδων με επίβλεψη, χωρίς επίβλεψη και με ημиеπίβλεψη [ML-12, ML-13]. Κοινό χαρακτηριστικό όλων των μεθόδων ML είναι η επαναληπτική συγχώνευση της στατιστικής ανάλυσης και της ελαχιστοποίησης του σφάλματος ή της μεγιστοποίησης της ανταμοιβής, που εφαρμόζεται σε διάφορους βαθμούς. Πολλοί αλγόριθμοι ML επιδίδονται σε εξαντλητικές αναζητήσεις πολυάριθμων πιθανών αποτελεσμάτων για τον εντοπισμό του βέλτιστου αποτελέσματος, όπως ορίζεται από τον χρήστη για ένα δεδομένο πρόβλημα. Η υπολογιστική πολυπλοκότητα

αυτών των επαναλήψεων κατέστησε ανέφικτους τους χειροκίνητους υπολογισμούς, συμβάλλοντας στον πρόσφατο εκδημοκρατισμό αυτών των μεθόδων μεταξύ μεμονωμένων ερευνητών. Η γενική αρχή είναι η χρήση προσαρμοσίμων, αυτοματοποιημένων προσεγγίσεων για την ανίχνευση προτύπων εντός των δεδομένων, με κύριο στόχο την πρόβλεψη μελλοντικών δεδομένων. Η ML προσφέρει ένα σαφές πλεονέκτημα έναντι των παραδοσιακών στατιστικών τεχνικών, ιδίως στο πλαίσιο των γεωεπιστημών, καθώς υπερέχει στη μοντελοποίηση δεδομένων υψηλής διάστασης και μη γραμμικών δεδομένων που χαρακτηρίζονται από πολύπλοκες αλληλεπιδράσεις και ελλείπουσες τιμές. Τα δεδομένα της αστρονομίας και των γεωεπιστημών, συγκεκριμένα, παρουσιάζουν μη γραμμικά και υψηλών διαστάσεων χαρακτηριστικά με περίπλοκα φαινόμενα αλληλεπίδρασης [ML-11]. Οι παραδοσιακές μέθοδοι συχνά υποθέτουν γραμμικότητα και δυσκολεύονται να αντιμετωπίσουν τα φαινόμενα αλληλεπίδρασης. Οι ερευνητές έχουν καταφύγει σε στρατηγικές όπως οι μετασχηματισμοί δεδομένων, οι οποίοι μπορεί να θέσουν σε κίνδυνο την ερμηνευσιμότητα, ή σε μεθόδους αποσύνθεσης/ανασύνθεσης για την απλοποίηση

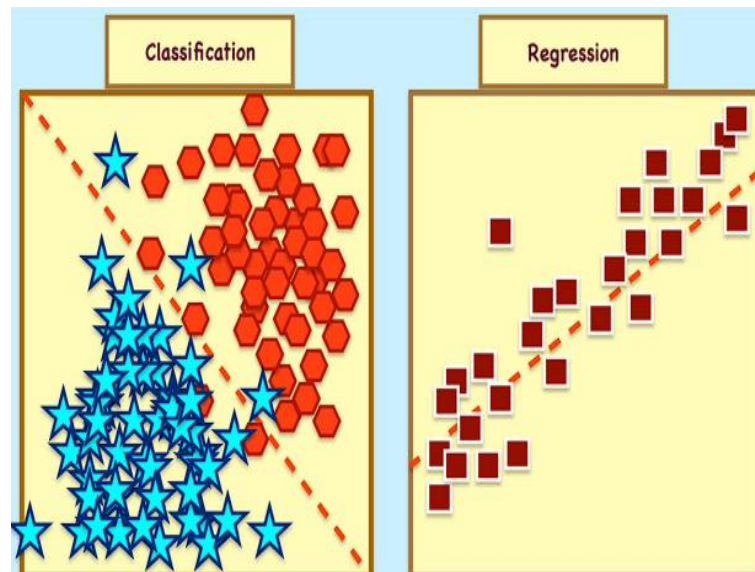


Εικόνα 3: Ιστορική πορεία της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης

πολύπλοκων δυναμικών. Εναλλακτικά, ορισμένοι επέλεξαν να επιβάλουν γραμμικότητα χωρίς να τροποποιήσουν τα δεδομένα. Αρκετές συγκριτικές μελέτες έχουν καταδείξει ότι οι τεχνικές *ML* μπορούν να υπερτερούν των παραδοσιακών στατιστικών προσεγγίσεων σε ένα ευρύ φάσμα προβλημάτων (*Levine et al., 1996-Prasad et al., 2006- Zhao et al., 2011*). Ωστόσο, είναι σημαντικό να προσεγγίζεται η σύγκριση των τεχνικών με περίσκεψη (*Fielding, 2006*). Η οριοθέτηση μεταξύ των μεθόδων *ML* και των παραδοσιακών στατιστικών τεχνικών δεν είναι πάντα σαφής και οι μέθοδοι *ML* δεν υπερτερούν καθολικά έναντι της παραδοσιακής στατιστικής. Η επιλογή της/των μεθόδου/ων πρέπει να ευθυγραμμίζεται με το συγκεκριμένο πρόβλημα που τίθεται, και μια μετα-προσέγγιση, λαμβάνοντας υπόψη τα αποτελέσματα πολλαπλών αλγορίθμων, αποδεικνύεται συχνά βέλτιστη.

Διαδικασία εκμάθησης δεδομένων

Στη μηχανική μάθηση, το ταξίδι ξεκινά με ένα σύνολο δεδομένων που περιλαμβάνει τόσο ανεξάρτητες όσο και εξαρτημένες μεταβλητές. Αυτό το σύνολο δεδομένων χρησιμεύει ως βάση για να κατανοήσει η μηχανή τη σύνδεση μεταξύ των ανεξάρτητων μεταβλητών (είσοδος) και της εξαρτημένης μεταβλητής (έξοδος). Στη συνέχεια, όταν ο αλγόριθμος συναντά νέα δεδομένα εισόδου, μπορεί να εφαρμόσει



Εικόνα 4: Απεικόνιση Μεθόδων, Ταξινόμησης αριστερά και Παλινδρόμησης δεξιά

αυτή τη σχέση που έμαθε για να δημιουργήσει προβλέψεις. Μετά τη φάση της εκπαίδευσης, ο αλγόριθμος πρέπει να υποβληθεί σε δοκιμές για να αξιολογηθεί η ικανότητά του να παράγει ακριβείς προβλέψεις με νέα, μη εμφανή δεδομένα. Αυτή η αξιολόγηση απαιτεί ένα ξεχωριστό σύνολο δεδομένων που περιλαμβάνει τόσο ανεξάρτητες όσο και εξαρτημένες μεταβλητές, με τις εξαρτημένες μεταβλητές να αποκρύπτονται σκόπιμα από τον αλγόριθμο. Στη συνέχεια, οι προβλέψεις του αλγορίθμου αντιπαραβάλλονται με τα κρυμμένα δεδομένα για να εκτιμηθεί η ποιότητα των προβλέψεών του. Είναι επιτακτική ανάγκη να χρησιμοποιηθεί ένα σύνολο δεδομένων επαρκούς μεγέθους που να μπορεί να χωριστεί σε δύο τμήματα τόσο για την εκπαίδευση όσο και για τη δοκιμή. Η επιλογή της μεθόδου μηχανικής μάθησης, το μέγεθος και τα χαρακτηριστικά των συνόλων δεδομένων εκπαίδευσης και δοκιμής, καθώς και η μεθοδολογία αξιολόγησης πρέπει να γίνονται με σύνεση ώστε να επιτυγχάνεται ισορροπία μεταξύ μεροληψίας και ακρίβειας, αποδίδοντας ουσιαστικά αποτελέσματα προσαρμοσμένα στο συγκεκριμένο πρόβλημα.

Παραδείγματα μάθησης

Η μηχανική μάθηση περιλαμβάνει μια πληθώρα παραδειγμάτων μάθησης, καθένα από τα οποία είναι προσαρμοσμένο για την επίτευξη συγκεκριμένων στόχων με βάση το

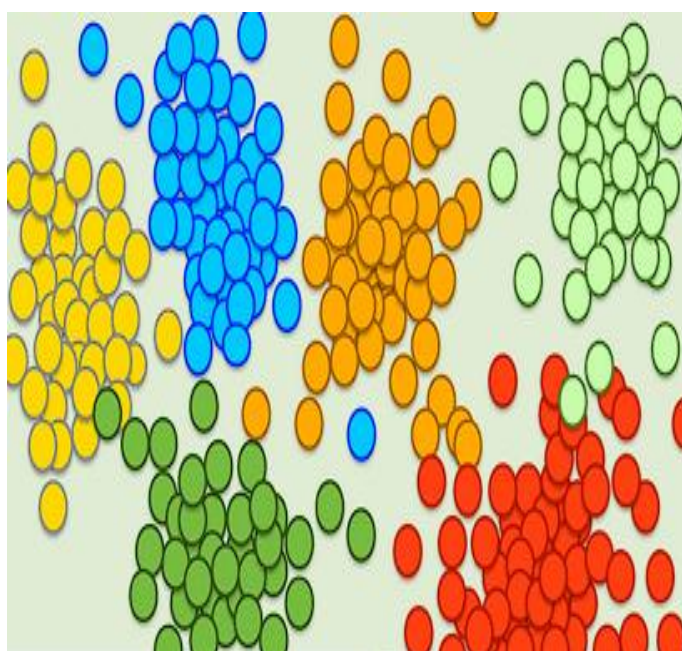
επιθυμητό αποτέλεσμα του αλγορίθμου. Εδώ, παρέχουμε συνοπτικές περιγραφές τριών σημαντικών παραδειγμάτων μάθησης:

Μάθηση με επίβλεψη (Supervised Learning)

Η μάθηση με επίβλεψη ξεχωρίζει ως μία από τις πιο εκτεταμένα χρησιμοποιούμενες τεχνικές μηχανικής μάθησης. Σε αυτό το παράδειγμα, το σύνολο δεδομένων εκπαίδευσης έρχεται προ-επισημασμένο, παρέχοντας στον αλγόριθμο σωστές απαντήσεις ή επιθυμητές εξόδους (ετικέτες) [ML-5]. Αυτά τα επισημειωμένα ζεύγη εισόδου-εξόδου χρησιμοποιούνται για την εξαγωγή ενός αλγορίθμου μάθησης ή μιας συνάρτησης αντιστοίχισης από τη μεταβλητή εισόδου (X) στη μεταβλητή εξόδου (Y). Οι σωστές απαντήσεις, που συμβολίζονται ως ετικέτες, είναι ήδη γνωστές, αποτελώντας ένα επισημασμένο σύνολο ζευγών εισόδου-εξόδου [ML-10, ML-15].

Μάθηση χωρίς επίβλεψη (Unsupervised Learning)

Η μάθηση χωρίς επίβλεψη, γνωστή και ως ανακάλυψη γνώσης, λειτουργεί με μη επισημασμένα, μη ταξινομημένα και μη δομημένα δεδομένα εκπαίδευσης [ML-12, ML-13]. Ο κύριος στόχος της είναι να αποκαλύψει κρυμμένα και ενδιαφέροντα μοτίβα μέσα σε μη επισημειωμένα δεδομένα. Σε αντίθεση με την επιβλεπόμενη μάθηση, οι μη επιβλεπόμενες μέθοδοι δεν μπορούν να εφαρμοστούν άμεσα σε προβλήματα παλινδρόμησης ή ταξινόμησης, καθώς δεν υπάρχει εκ των προτέρων γνώση των τιμών εξόδου. Μια εξέχουσα τεχνική μάθησης χωρίς επίβλεψη είναι η ομαδοποίηση. Εικόνα 5, η



Εικόνα 5: Γραφική απεικόνιση συσταδοποίησης (ομαδοποίησης) δεδομένων

οποία χρησιμοποιείται για τη διερεύνηση δεδομένων, τον εντοπισμό λανθανόντων προτύπων ή την ομαδοποίηση σημείων δεδομένων [ML-4, ML-16]. Οι εφαρμογές της ανάλυσης συστάδων περιλαμβάνουν ποικίλους τομείς, όπως η ανάλυση γονιδιακής ακολουθίας, η έρευνα αγοράς και η αναγνώριση αντικειμένων.

Οι αλγόριθμοι μάθησης χωρίς επίβλεψη παίζουν σημαντικό ρόλο στην ανάλυση δεδομένων, στην αποκάλυψη κρυφών μοτίβων και στον εντοπισμό ομαδοποιήσεων εντός συνόλων δεδομένων. Οι εφαρμογές τους καλύπτουν ποικίλους τομείς, όπως η ανάλυση γονιδιακών ακολουθιών, η έρευνα αγοράς και η αναγνώριση αντικειμένων. Αξιοσημείωτες τεχνικές μάθησης χωρίς επίβλεψη περιλαμβάνουν την ομαδοποίηση, την ανίχνευση ανωμαλιών, τα νευρωνικά δίκτυα και τις μεθόδους μάθησης λανθάνουσας μεταβλητής [ML-17].

Μάθηση με ημιεπίβλεψη (Semi – Supervised Learning)

Η ημιεπιτηρούμενη μάθηση σηματοδοτεί τη συγχώνευση τεχνικών μηχανικής μάθησης με επίβλεψη και χωρίς επίβλεψη. Αυτοί οι αλγόριθμοι αξιοποιούν μερικώς

επισημασμένα δεδομένα εκπαίδευσης, τα οποία συνήθως χαρακτηρίζονται από ένα μικρό σύνολο επισημασμένων δεδομένων που συνυπάρχουν με ένα σημαντικό όγκο μη επισημασμένων δεδομένων. Οι αλγόριθμοι με ημιεπίβλεψη υποβάλλονται σε εκπαίδευση χρησιμοποιώντας αυτό το συνονθύλευμα επισημασμένων και μη επισημασμένων δεδομένων, οδηγώντας σε αυξημένη ακρίβεια μάθησης.

Επιλογή του κατάλληλου αλγορίθμου μάθησης

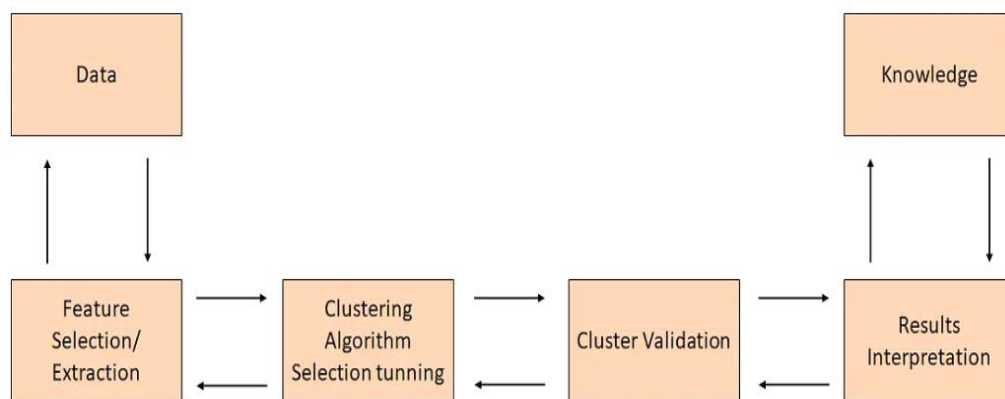
Η απόφαση σχετικά με την επιλογή ενός συγκεκριμένου αλγορίθμου μάθησης αποτελεί ένα κομβικό και διαφοροποιημένο βήμα στο πεδίο της μηχανικής μάθησης. Μόλις οι προκαταρκτικές δοκιμές αποδώσουν ικανοποιητικά αποτελέσματα, ο ταξινομητής είναι έτοιμος για τακτική χρήση. Το πρωταρχικό κριτήριο για την αξιολόγηση του ταξινομητή συνήθως περιστρέφεται γύρω από την ακρίβεια πρόβλεψής του. Για τη σύγκριση διαφόρων αλγορίθμων μηχανικής μάθησης (*ML*) με επίβλεψη, μια συνήθης πρακτική είναι η διεξαγωγή στατιστικών αξιολογήσεων της ακρίβειας που επιτυγχάνουν οι εκπαιδευμένοι ταξινομητές σε διαφορετικά σύνολα δεδομένων. Σε περιπτώσεις όπου είναι διαθέσιμη άφθονη παροχή δεδομένων, καθίσταται εφικτή η δημιουργία πολλαπλών συνόλων εκπαίδευσης μεγέθους N . Στη συνέχεια, εκτελούνται και οι δύο αλγόριθμοι μάθησης σε καθένα από αυτά τα σύνολα εκπαίδευσης και εκτιμάται η διαφορά στην ακρίβεια για κάθε ζεύγος ταξινομητών με τη χρήση ενός σημαντικού συνόλου δοκιμών. Ο μέσος όρος αυτών των διαφορών ακρίβειας παρέχει μια εκτίμηση της αναμενόμενης διακύμανσης του σφάλματος γενίκευσης σε όλα τα πιθανά σύνολα εκπαίδευσης μεγέθους N . Εν τω μεταξύ, η διακύμανση της απόδοσης των ταξινομητών θεωρείται μια εκτίμηση της μεταβλητότητας του συνολικού συνόλου. Το βήμα που ακολουθεί περιλαμβάνει την εκτέλεση ενός ζευγαρωτού $t - test$ για την αξιολόγηση της μηδενικής υπόθεσης, η οποία υποθέτει ότι η μέση διαφορά μεταξύ των ταξινομητών ισούται με μηδέν. Είναι σημαντικό να αναγνωρίσουμε ότι αυτή η στατιστική δοκιμή μπορεί δυνητικά να αποδώσει δύο τύπους σφαλμάτων. Το σφάλμα τύπου I , το οποίο χαρακτηρίζεται από την εσφαλμένη απόρριψη της μηδενικής υπόθεσης από το τεστ, ουσιαστικά υποδεικνύει μια "σημαντική" διαφορά ενώ δεν υπάρχει. Αντίθετα, το σφάλμα τύπου II δηλώνει την πιθανότητα να μην απορριφθεί η μηδενική υπόθεση, ενώ πράγματι υπάρχει διαφορά.

Στην πράξη, οι ερευνητές αντιμετωπίζουν συχνά ένα σενάριο όπου είναι προσβάσιμο μόνο ένα σύνολο δεδομένων μεγέθους N . Για να αντιμετωπιστεί αυτός ο περιορισμός, δημιουργούνται ξεχωριστά σύνολα εκπαίδευσης μέσω υποδειγματοληψίας, με τις περιπτώσεις που δεν περιλαμβάνονται στο σύνολο εκπαίδευσης να προορίζονται για δοκιμή. Δυστυχώς, η προσέγγιση αυτή παραβιάζει την υπόθεση ανεξαρτησίας που είναι ζωτικής σημασίας για τον ορθό έλεγχο σημαντικότητας. Κατά συνέπεια, τα σφάλματα τύπου I τείνουν να υπερβαίνουν το επιλεγμένο επίπεδο σημαντικότητας, καθιστώντας την κατάσταση λιγότερο από ιδανική. Για τον μετριασμό αυτού του ζητήματος, έχουν επινοηθεί με την πάροδο των ετών διάφορες ευρετικές προσαρμογές του $t - test$. Ιδανικά, το αποτέλεσμα του τεστ θα πρέπει να παρουσιάζει ανεξαρτησία από τη συγκεκριμένη διαμέριση των δεδομένων που προκύπτει από τη διαδικασία τυχαιοποίησης. Μια τέτοια ανεξαρτησία θα διευκόλυνε σημαντικά την αναπαραγωγή των πειραματικών ευρημάτων που δημοσιεύονται στη βιβλιογραφία. Στην πράξη, ωστόσο, υπάρχει συχνά κάποιος βαθμός ευαισθησίας στην επιλεγμένη διαμέριση. Για να εκτιμήσουν την αναπαραγωγικότητα των αποτελεσμάτων, οι ερευνητές αναλαμβάνουν πολλαπλές επαναλήψεις της ίδιας δοκιμής στο ίδιο σύνολο δεδομένων, χρησιμοποιώντας διαφορετικές τυχαίες καταταμίσεις. Συνήθως, αυτό περιλαμβάνει τη

διεξαγωγή της δοκιμής δέκα ή περισσότερες φορές και την αξιολόγηση του πόσο συχνά τα αποτελέσματα ευθυγραμμίζονται. Η επιβλεπόμενη ταξινόμηση αποτελεί μια από τις πιο συχνά εκτελούμενες εργασίες στον τομέα των ευφύων συστημάτων. Ως αποτέλεσμα, έχει προκύψει ένα ευρύ φάσμα τεχνικών, που προέρχονται από την **Τεχνητή Νοημοσύνη** (συμπεριλαμβανομένων των λογικών/συμβολικών προσεγγίσεων), τις μεθόδους που βασίζονται σε Perceptron και τη Στατιστική (συμπεριλαμβανομένων των **Bayesian Networks** και των τεχνικών που βασίζονται σε **Instance**).

Unsupervised Learning

Φανταστείτε μια μηχανή ή έναν οργανισμό που λαμβάνει τακτικά μια ροή εισόδων, οι οποίες συμβολίζονται ως x_1, x_2, x_3 κ.ο.κ. Κάθε x_i σηματοδοτεί μια αισθητηριακή είσοδο σε μια συγκεκριμένη χρονική στιγμή. Οι εισοδοί μπορεί να έχουν διάφορες μορφές, όπως εικόνες που συλλαμβάνονται από τον αμφιβληστροειδή χιτώνα ή εικονοστοιχεία από μια φωτογραφική μηχανή, ή ηχογραφήσεις ηχητικών κυμάτων. Μπορεί επίσης να περιλαμβάνουν πιο λεπτές μορφές αισθητηριακών δεδομένων, όπως η διατύπωση σε ένα άρθρο ειδήσεων ή ένας κατάλογος προϊόντων σε ένα καρότσι για ψώνια. Στη σφαίρα της μηχανικής μάθησης, υπάρχουν δύο διακριτές κατηγορίες, με την πρώτη να είναι Μάθηση με επίβλεψη.



Στη σφαίρα της μάθησης με επίβλεψη, η μηχανή λαμβάνει μια ακολουθία εισόδων ($x_1, x_2, \dots$) και μια συνοδευτική ακολουθία επιθυμητών εξόδων ($y_1, y_2, \dots$). Ο πρωταρχικός στόχος είναι η μηχανή να μάθει μια αντιστοίχιση που θα της επιτρέψει να παράγει τη σωστή έξοδο για νέες, αθέατες εισόδους. Η έξοδος μπορεί να έχει τη μορφή μιας ετικέτας κλάσης για εργασίες ταξινόμησης ή ενός πραγματικού αριθμού για εργασίες παλινδρόμησης.

Η μάθηση χωρίς επίβλεψη συνεπάγεται ότι η μηχανή λαμβάνει δεδομένα εισόδου ($x_1, x_2, \dots$) χωρίς πρόσβαση σε αντίστοιχους εποπτευόμενους στόχους ή ανταμοιβές. Αν και αρχικά μπορεί να φαίνεται αινιγματικό το πώς η μηχανή μπορεί να μάθει χωρίς ρητή ανατροφοδότηση, μπορεί να δημιουργηθεί ένα επίσημο πλαίσιο για τη μάθηση χωρίς επίβλεψη. Σε αυτό το πλαίσιο, ο σκοπός της μηχανής είναι να παράγει αναπαραστάσεις των δεδομένων εισόδου που εξυπηρετούν διάφορες λειτουργίες, όπως η διευκόλυνση της λήψης αποφάσεων, η πρόβλεψη μελλοντικών εισόδων και η αποτελεσματική επικοινωνία με άλλες μηχανές. Οι τεχνικές συντομογραφίες θα εξηγούνται κατά την πρώτη χρήση τους. Ο θεμελιώδης στόχος της μάθησης χωρίς επίβλεψη είναι η αποκάλυψη μοτίβων δεδομένων πέρα από τον τυχαίο θόρυβο. Αυτό περιλαμβάνει την ομαδοποίηση, η οποία ομαδοποιεί παρόμοια σημεία δεδομένων, και τη μείωση της

διάστασης. Η σαφής, συνοπτική γλώσσα και οι λογικές εξηγήσεις θα εξασφαλίσουν τη βέλτιστη κατανόηση. Θα διατηρηθούν καθ' όλη τη διάρκεια κοινά ακαδημαϊκά στοιχεία, συνεπής παραπομπή και επίσημο μητρώο. Η περιττή ορολογία, η μεροληπτική γλώσσα και οι διακοσμητικές φράσεις θα αποφεύγονται, διατηρώντας την αντικειμενικότητα. Επιπλέον, θα δοθεί προσεκτική προσοχή στη γραμματική, την ορθογραφία και τη στίξη, εξασφαλίζοντας ακρίβεια και ευχέρεια.

Μηχανική μάθηση, στατιστική και θεωρία πληροφοριών (1.2): Μεγάλο μέρος της έρευνας στη μάθηση χωρίς επίβλεψη επικεντρώνεται στη δημιουργία ενός πιθανοτικού μοντέλου δεδομένων. Η εκτίμηση ενός μοντέλου που αντικατοπτρίζει την κατανομή πιθανοτήτων των μελλοντικών εισόδων (x_I) από προηγούμενες εισόδους ($x_1, \dots, x_{I-1}$) μπορεί να είναι επωφελής, ακόμη και όταν η μηχανή δεν έχει εποπτεία ή κίνητρα. Αυτή η μοντελοποίηση είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου η σειρά των εισόδων είναι άγνωστη ή άσχετη.

Για παράδειγμα, μια μηχανή μπορεί να δημιουργήσει ένα μοντέλο με την υπόθεση ότι τα σημεία δεδομένων $x_1, x_2, \dots$ αντλούνται ανεξάρτητα και πανομοιότυπα από μια κατανομή γνωστή ως $P(x)$. Αυτά τα μοντέλα είναι χρήσιμα για την παρακολούθηση και την ανίχνευση ακραίων τιμών. Για παράδειγμα, εάν το x αντιπροσωπεύει τα μοτίβα των ενδείξεων αισθητήρων από έναν πυρηνικό σταθμό παραγωγής ενέργειας και η $P(x)$ μαθαίνεται από τα δεδομένα που συλλέγονται κατά τη διάρκεια της κανονικής λειτουργίας, το μοντέλο αυτό μπορεί να υπολογίσει την πιθανότητα μιας νέας μέτρησης αισθητήρα. Εάν αυτή η πιθανότητα είναι σημαντικά χαμηλή, θα μπορούσε να αποτελεί ένδειξη ενός μοντέλου κάτω του μέσου όρου ή ασυνήθιστης συμπεριφοράς του εργοστασίου, η οποία μπορεί να απαιτεί περαιτέρω διερεύνηση ή κλείσιμο. Τα πιθανοτικά μοντέλα είναι ιδιαίτερα πολύτιμα στην ταξινόμηση. Τα μοντέλα, $P_1(x)$ και $P_2(x)$, επιτρέπουν την ανάλυση των χαρακτηριστικών των κατόχων πιστωτικών καρτών που έχουν τηρήσει τις προθεσμίες τους και εκείνων που δεν τις έχουν τηρήσει. Χρησιμοποιώντας αυτά τα μοντέλα, η μηχανή μπορεί να κατηγοριοποιήσει μια νέα αίτηση, x' , προσδιορίζοντας τις συγκριτικές πιθανότητες των $P_1(x')$ και $P_2(x')$. Οι συντομογραφίες των τεχνικών όρων εξηγούνται όταν χρησιμοποιούνται για πρώτη φορά και η γλώσσα διατηρείται επίσημη και ισορροπημένη, αποφεύγοντας τη χρήση μεροληπτικών ή συναισθηματικών διατυπώσεων. Σύμφωνα με τα βρετανικά αγγλικά, ακολουθούνται οι ορθογραφικές συμβάσεις και τηρείται η γραμματική ορθότητα σε όλο το κείμενο.

Επιπλέον, η χρήση πιθανολογικών μοντέλων αποτελεί κρίσιμη πτυχή της διευκόλυνσης της επιτυχούς επικοινωνίας και συμπίεσης δεδομένων. Κατά τη μετάδοση τυχαία επιλεγμένων συμβόλων x από μια κατανομή $P(x)$ για ψηφιακή επικοινωνία, τα δεδομένα συνήθως κωδικοποιούνται ώστε να διασφαλίζεται ότι τα σύμβολα με υψηλότερες συχνότητες λαμβάνουν μικρότερες κωδικές λέξεις, εξοικονομώντας εύρος ζώνης. Το θεώρημα κωδικοποίησης πηγής του *Shannon* καθορίζει τον βέλτιστο αριθμό bits που απαιτούνται για την κωδικοποίηση ενός συμβόλου με πιθανότητα $P(x)$, ορίζοντας την πιθανότητα ως $-\log_2(P(x))$. Το κόστος που σχετίζεται με την κωδικοποίηση μπορεί να ελαχιστοποιηθεί με τη χρήση αυτής της μεθόδου, η οποία συσχετίζεται με την εντροπία της κατανομής P . Είναι δύσκολο να γνωρίζουμε την πραγματική κατανομή των δεδομένων, αλλά είναι δυνατόν να αναπτύξουμε μια ικανότητα χρήσης του μοντέλου $Q(x)$ για την εκτίμησή της. Ο βέλτιστος κώδικας του μοντέλου χρησιμοποιεί $-\log_2(Q(x))$ bits για κάθε σύμβολο x . Το αναμενόμενο κόστος κωδικοποίησης, με βάση την πραγματική κατανομή, μετρά την απόκλιση KL

(**Kullback-Leibler**) και αποκαλύπτει την αναποτελεσματικότητα της κωδικοποίησης στη συμπίεση δεδομένων που λαμβάνονται από την πραγματική κατανομή P χρησιμοποιώντας το μοντέλο Q . Η απόκλιση KL είναι θετική και μηδενίζεται μόνο όταν το P ισούται με το Q . Συνεπώς, η ποιότητα του μοντέλου δεδομένων επηρεάζει σημαντικά την ικανότητά μας να συμπιέζουμε και να μεταδίδουμε αποτελεσματικά φρέσκα δεδομένα. Αναδεικνύεται η θεμελιώδης σύνδεση μεταξύ μηχανικής μάθησης, στατιστικής και θεωρίας της πληροφορίας. Η μηχανική μάθηση περιλαμβάνει ποικίλες μεθόδους, συμπεριλαμβανομένης της επιβλεπόμενης και της μη επιβλεπόμενης μάθησης. Η τελευταία περιλαμβάνει την κατασκευή πιθανοτικών μοντέλων για την αναγνώριση μοτίβων και την απεικόνιση δεδομένων, ακόμη και χωρίς συγκεκριμένη καθοδήγηση. Τα μοντέλα έχουν ευρύ φάσμα εφαρμογών, όπως ο εντοπισμός ανωμαλιών και η επίτευξη αποτελεσματικής συμπίεσης δεδομένων. Αυτό επιτυγχάνεται μέσω ενός συνδυασμού μηχανικής μάθησης, στατιστικής και θεωρίας της πληροφορίας.

$$H(P) = - \sum_x P(x) \log_2 P(x) \quad \{6.03\}$$

$$- \sum_x P(x) \log_2 Q(x) \quad \{6.04\}$$

$$KL(P | Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad \{6.05\}$$

	Decision Trees	Neural Networks	Naïve Bayes	kNN	SVM	Rule-learners
Accuracy in general	**	***	*	**	****	**
Speed of learning with respect to number of attributes and the number of instances	***	*	****	****	*	**
Speed of classification	****	****	****	*	****	****
Tolerance to missing values	***	*	****	*	**	**
Tolerance to irrelevant attributes	***	*	**	**	****	**
Tolerance to redundant attributes	**	**	*	**	***	**
Tolerance to highly interdependent attributes (e.g. parity problems)	**	***	*	*	***	**
Dealing with discrete/binary/continuous attributes	****	*** (not discrete)	*** (not continuous)	*** (not directly discrete)	** (not discrete)	*** (not directly continuous)
Tolerance to noise	**	**	***	*	**	*
Dealing with danger of overfitting	**	*	***	***	**	**
Attempts for incremental learning	**	***	****	****	**	*
Explanation ability/transparency of knowledge/classifications	****	*	****	**	*	****
Model parameter handling	***	*	****	***	*	***

Table 4. Comparing learning algorithms (**** stars represent the best and * star the worst performance)

Εικόνα 6: Σύγκριση μεταξύ μεθόδων ως προς την ακρίβεια, την ταχύτητα και την ανεκτικότητα τους σε θόρυβο, λάθος δεδομένα

Μέθοδοι Μείωσης Διαστατικότητας και Συσταδοποίησης για Υπερφασματικά Δεδομένα

Τα υπερφασματικά δεδομένα, όπως αυτά που προκύπτουν από την τεχνική MA-XRF, χαρακτηρίζονται από εξαιρετικά υψηλή διαστατικότητα, καθώς κάθε pixel

περιγράφεται από χιλιάδες ενεργειακά κανάλια. Αυτή η πολυπλοκότητα, γνωστή ως "κατάρτα της διάστασης", καθιστά την άμεση ανάλυση και οπτικοποίηση των δεδομένων δύσκολη και υπολογιστικά απαιτητική. Για την αντιμετώπιση αυτής της πρόκλησης, εφαρμόζονται τεχνικές **μείωσης διαστατικότητας (dimensionality reduction)**. Στόχος αυτών των μεθόδων είναι ο μετασχηματισμός των δεδομένων από τον αρχικό, υψηλής διάστασης χώρο, σε έναν νέο χώρο πολύ χαμηλότερων διαστάσεων, διατηρώντας παράλληλα το μέγιστο δυνατό ποσοστό της χρήσιμης πληροφορίας και απομακρύνοντας τον θόρυβο. Οι μέθοδοι Ανάλυσης Κύριων Συνιστωσών (**PCA**) και Μη Αρνητικής Παραγοντοποίησης Πινάκων (**NMF**), που αποτελούν τον πυρήνα της παρούσας ανάλυσης, είναι πρωτίστως τεχνικές μείωσης διαστατικότητας.

Ανάλυση Κύριων Συνιστωσών PCA

Η Ανάλυση Κύριων Συνιστωσών (**PCA**) είναι μια ευρέως χρησιμοποιούμενη πολυμεταβλητή στατιστική τεχνική με μακρά ιστορία που χρονολογείται από σημαντικές προσωπικότητες όπως οι **Pearson, Cauchy, Jordan, Cayley, Silvester** και **Hamilton**. Ενώ οι ρίζες της μπορούν να εντοπιστούν σε αυτές τις πρώιμες συνεισφορές, η σύγχρονη διατύπωσή της και ο όρος "κύρια συνιστώσα" επισημοποιήθηκαν από τον **Hotelling**. Η **PCA** χρησιμοποιείται σε διάφορους επιστημονικούς κλάδους για την ανάλυση πινάκων δεδομένων που αντιπροσωπεύουν παρατηρήσεις που περιγράφονται από πολλαπλές συσχετιζόμενες εξαρτημένες μεταβλητές. Ο πρωταρχικός της στόχος είναι να εξάγει βασικές πληροφορίες από αυτούς τους πίνακες δεδομένων και να τις εκφράζει ως ένα σύνολο νέων ορθογώνιων μεταβλητών που αναφέρονται ως κύριες συνιστώσες. Επιπλέον, η **PCA** βοηθά στην αναπαράσταση των μοτίβων ομοιότητας μεταξύ παρατηρήσεων και μεταβλητών με την απεικόνισή τους ως σημεία σε χάρτες.

Η **PCA** είναι μια μέθοδος εξαγωγής χαρακτηριστικών που μετασχηματίζει γραμμικά διανύσματα εισόδου υψηλής διάστασης σε διανύσματα χαμηλότερης διάστασης με ασυσχέτιστες συνιστώσες. Μη γραμμικές παραλλαγές της **PCA** έχουν αναπτυχθεί με τη χρήση διαφορετικών αλγορίθμων, με την ανάλυση κύριων συνιστωσών **Kernel (Kernel Principal Component Analysis, KPCA)** να είναι μία από αυτές. Η **KPCA** επεκτείνει τη μέθοδο πυρήνα, η οποία αρχικά σχεδιάστηκε για τις Μηχανές Διανυσμάτων Υποστήριξης (**SVM**), στην **PCA**. Η **KPCA** περιλαμβάνει την αντιστοίχιση των αρχικών εισόδων σε έναν χώρο χαρακτηριστικών υψηλών διαστάσεων με τη χρήση της μεθόδου πυρήνα και στη συνέχεια την εκτέλεση **PCA** σε αυτόν τον μετασχηματισμένο χώρο, με αποτέλεσμα τη μη γραμμική **PCA** στον αρχικό χώρο εισόδου. Μια άλλη μέθοδος γραμμικού μετασχηματισμού, η Ανάλυση ανεξάρτητων συνιστωσών (**ICA**), έχει επίσης αναπτυχθεί. Σε αντίθεση με την **PCA**, η οποία επιδιώκει ασυσχέτιστες συνιστώσες, η **ICA** στοχεύει στην επίτευξη στατιστικά ανεξάρτητων συνιστωσών στα μετασχηματισμένα διανύσματα. Η **ICA** αναπτύχθηκε αρχικά για τον τυφλό διαχωρισμό πηγών και έκτοτε έχει γενικευτεί για την εξαγωγή χαρακτηριστικών.

Η **PCA**, με την παραδοχή της γραμμικότητας, στοχεύει στην εύρεση μιας νέας βάσης που μπορεί να επανεκφράσει καλύτερα το σύνολο δεδομένων.

Η υπόθεση αυτή απλοποιεί σημαντικά το πρόβλημα περιορίζοντας τις πιθανές βάσεις σε γραμμικούς συνδυασμούς των αρχικών διανυσμάτων βάσης. Με αυτή την υπόθεση, η **PCA** μπορεί να επανεκφράσει τα δεδομένα ως γραμμικό συνδυασμό των διανυσμάτων βάσης. Ας υποθέσουμε ότι το X αναπαριστά το αρχικό σύνολο

δεδομένων, όπου κάθε στήλη είναι ένα δείγμα, και το Y είναι μια νέα αναπαράσταση αυτού του συνόλου δεδομένων.

$$PX = Y \quad \{6.38\}$$

Ας ορίσουμε τις ακόλουθες ποσότητες:

- p_i είναι οι γραμμές του P .
- x_i είναι οι στήλες του X (μεμονωμένα δείγματα)
- y_i είναι οι στήλες του Y

Η εξίσωση αντιπροσωπεύει μια αλλαγή βάσης, η οποία μπορεί να ερμηνευτεί με διάφορους τρόπους:

Ο P είναι ένας πίνακας που μετασχηματίζει το X σε Y .

Γεωμετρικά, ο P είναι μια περιστροφή και μια έκταση που μετασχηματίζει το X σε Y . Οι γραμμές του P , $\{p_1, \dots, p_m\}$, είναι ένα νέο σύνολο διανυσμάτων βάσης για την έκφραση των στηλών του X . Κάθε συντελεστής του y_i είναι ένα τετραγωνικό γινόμενο του x_i με την αντίστοιχη γραμμή του P , που σημαίνει ότι οι γραμμές του P χρησιμεύουν ως ένα νέο σύνολο διανυσμάτων βάσης για την αναπαράσταση των στηλών του X . Για τον προσδιορισμό της κατάλληλης βάσης P , απαιτούνται πρόσθετες παραδοχές πέραν της γραμμικότητας. Η επιλογή αυτών των υποθέσεων είναι κρίσιμη για την επίτευξη ουσιαστικών αποτελεσμάτων. Προκύπτουν τα ακόλουθα ερωτήματα: Ποιος είναι ο καλύτερος τρόπος επανεκφράσεως του X ; Ποια είναι μια καλή επιλογή της βάσης P ; Η απάντηση σε αυτά τα ερωτήματα περιλαμβάνει τον καθορισμό των χαρακτηριστικών που πρέπει να παρουσιάζουν τα μετασχηματισμένα δεδομένα Y . Πέρα από τη γραμμικότητα, απαιτούνται πρόσθετες παραδοχές για να καταλήξουμε σε μια λογική λύση. Οι υποθέσεις αυτές εξετάζονται στις ακόλουθες ενότητες.

$$PX = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$$

$$Y = \begin{bmatrix} p_1 \cdot x_1 & \cdots & p_1 \cdot x_n \\ \vdots & \ddots & \vdots \\ p_m \cdot x_1 & \cdots & p_m \cdot x_n \end{bmatrix} \quad \{6.39\}$$

Η διακύμανση και ο στόχος: Για να καθορίσει κανείς τι σημαίνει "εκφράζει καλύτερα τα δεδομένα", πρέπει να εξετάσει την ελαχιστοποίηση του πλεονασμού, που μετράται από το μέγεθος της συνδιακύμανσης, και τη μεγιστοποίηση του σήματος, που μετράται από τη διακύμανση. Ο στόχος είναι η βελτιστοποίηση του πίνακα συμμεταβλητότητας (*Covariance Matrix*) C_Y . Όλοι οι εκτός διαγωνίου όροι του C_Y θα πρέπει να είναι μηδέν, υποδεικνύοντας ότι το Y είναι αποδιαρθρωμένο (*decorrelated*). Οι διαδοχικές διαστάσεις του Y θα πρέπει να είναι ταξινομημένες με βάση τη διακύμανση. Η **PCA** στοχεύει στη διαγωνιοποίηση του πίνακα συνδιακύμανσης C_Y για την επίτευξη αυτών των στόχων. Η υπόθεση ορθοκανονικότητας στην **PCA** απλοποιεί αυτή τη διαδικασία και επιτρέπει την απλή επιλογή των κύριων συνιστωσών. Η **PCA** υποθέτει ότι όλα τα διανύσματα βάσης $\{p_1, \dots, p_m\}$ είναι ορθοκανονικά, καθιστώντας εφικτή τη διαδικασία

ευθυγράμμισης μιας βάσης με τον άξονα μέγιστης διακύμανσης. Ο αλγόριθμος για την επιλογή των κύριων συνιστωσών ακολουθεί τα εξής βήματα:

$$y_i = \begin{bmatrix} \mathbf{p}_1 \cdot \mathbf{x}_i \\ \vdots \\ \mathbf{p}_m \cdot \mathbf{x}_i \end{bmatrix} \quad \{6.40\}$$

Επιλέξτε μια κανονικοποιημένη κατεύθυνση στον m-διάστατο χώρο που μεγιστοποιεί τη διακύμανση στο X. Αποθηκεύστε αυτό το διάνυσμα ως $\mathbf{p}_1$.

Βρείτε μια άλλη κατεύθυνση που μεγιστοποιεί τη διακύμανση, αλλά περιορίστε την αναζήτηση σε κατευθύνσεις ορθογώνιες προς όλες τις προηγούμενες επιλεγμένες κατευθύνσεις. Αποθηκεύστε αυτό το διάνυσμα ως $\mathbf{p}_i$.

Επαναλάβετε αυτή τη διαδικασία μέχρι να επιλεγούν m διανύσματα. Το διατεταγμένο σύνολο των p που προκύπτει γίνεται οι κύριες συνιστώσες.

$$a = [a_1 a_2 \cdots a_n] \quad \{6.41\}$$

$$b = [b_1 b_2 \cdots b_n] \quad \{6.42\}$$

$$\text{covariance of A and B} \equiv \sigma_{AB}^2 = \frac{1}{n} \sum_i a_i b_i \quad \{6.43\}$$

$$\sigma_{ab}^2 \equiv \frac{1}{n} ab^T \quad \{6.44\}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} \quad \{6.45\}$$

$$C_x \equiv \frac{1}{n} \mathbf{X} \mathbf{X}^T \quad \{6.46\}$$

Τα ιδιοδιανύσματα $\{\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^r\}$ και τα διανύσματα $\{\mathbf{u}^1, \mathbf{u}^2, \dots, \mathbf{u}^r\}$ σχηματίζουν ορθοκανονικές βάσεις σε έναν r-διάστατο χώρο. Για να εδραιώσουμε αυτές τις πληροφορίες, μπορούμε να τις αναπαραστήσουμε ως πολλαπλασιασμό πινάκων. Αρχικά, δημιουργούμε έναν νέο διαγώνιο **πίνακα** Σ ως εξής:

$$\Sigma \equiv \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ 0 & & & 0 & \ddots \\ & & & & & 0 \end{bmatrix} \quad \{6.47\}$$

Εδώ, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$ αντιπροσωπεύει τις ταξινομημένες κατά σειρά μοναδικές τιμές. Κατασκευάζουμε επίσης συνοδευτικούς ορθογώνιους πίνακες, V και U , ως εξής:

$$V = [\widehat{v_1} \widehat{v_2} \cdots \widehat{v_m}] \quad U = [\widehat{u_1} \widehat{u_2} \cdots \widehat{u_n}] \quad \{6.48\}$$

Για να αντιμετωπίσουμε πιθανά ζητήματα εκφυλισμού, προσθέτουμε $(m-r)$ και $(n-r)$ πρόσθετα ορθοκανονικά διανύσματα σε αυτούς τους πίνακες, εξασφαλίζοντας την πληρότητά τους.

Αυτή η βασισμένη στον πίνακα **Singular Value Decomposition (SVD)** μπορεί να εκφραστεί ως $XV = U\Sigma$, όπου κάθε στήλη των V και U ακολουθεί την κλιμακωτή εκδοχή της αποσύνθεσης. Δεδομένου ότι το V είναι ορθογώνιο, μπορούμε να πολλαπλασιάσουμε και τις δύο πλευρές με

$$XV = U\Sigma \quad \{6.49\}$$

$$\text{για να καταλήξουμε στην τελική μορφή της διάσπασης: } X = U\Sigma V^T \quad \{6.50\}$$

Παρόλο που η Εξίσωση μπορεί να εμφανίζεται χωρίς σαφές κίνητρο, αντιπροσωπεύει μια ισχυρή αποσύνθεση. Υποδηλώνει ότι κάθε αυθαίρετος πίνακας X μπορεί να μετασχηματιστεί σε έναν ορθογώνιο πίνακα (μια περιστροφή), έναν διαγώνιο πίνακα (μια έκταση) και έναν άλλο ορθογώνιο πίνακα (μια δεύτερη περιστροφή). Ερμηνεύοντας την **SVD** σε αυτό το πλαίσιο, μπορούμε να επανερμηνεύσουμε την παραπάνω εξίσωση ως εξής: $Xa=kb$, όπου a και b είναι διανύσματα στήλης και k είναι μια κλιμακωτή σταθερά. Τα σύνολα $\{v^1, v^2, \dots, v^m\}$ και $\{u^1, u^2, \dots, u^n\}$ χρησιμεύουν ως ορθοκανονικά σύνολα διανυσμάτων που καλύπτουν m ή n -διάστατους χώρους, αντίστοιχα, αντιπροσωπεύοντας όλες τις πιθανές "εισόδους" (a) και "εξόδους" (b).

Για να επισημοποιήσουμε την έννοια ότι τα $\{v^1, v^2, \dots, v^m\}$ και $\{u^1, u^2, \dots, u^n\}$ καλύπτουν όλες τις πιθανές "εισόδους" και "εξόδους", μπορούμε να χειριστούμε την Εξίσωση για να κάνουμε αυτή την ιδέα πιο ακριβή:

$$\begin{aligned} X &= U\Sigma V^T \\ U^T X &= \Sigma V^T \\ U^T X &= Z \end{aligned} \quad \{6.51\}$$

$Z \equiv \Sigma V^T$, και οι προηγούμενες στήλες $\{u^1, u^2, \dots, u^n\}$ είναι τώρα γραμμές στο U^T . Κατά τη σύγκριση αυτής της εξίσωσης με την εξίσωση 1, οι $\{u^1, u^2, \dots, u^n\}$ επιτελούν παρόμοιο ρόλο με τους $\{p^1, p^2, \dots, p^m\}$. Επομένως, η U^T αντιπροσωπεύει μια αλλαγή βάσης από το X στο Z , υποδεικνύοντας ότι μετασχηματίζουμε διανύσματα στήλης, όπως ακριβώς κάναμε και προηγουμένως. Εφόσον η ορθοκανονική βάση U^T (ή P) μετασχηματίζει διανύσματα στήλης, είναι μια βάση που καλύπτει τις στήλες του X . Αυτές οι βάσεις που καλύπτουν τις στήλες αναφέρονται ως χώρος στηλών του X , επισημοποιώντας αυτό που αποτελεί τις πιθανές "εξόδους" οποιουδήποτε πίνακα. Είναι ενδιαφέρον ότι υπάρχει μια συμμετρική έννοια στην **SVD** που ονομάζεται χώρος γραμμών. Μπορούμε να εκτελέσουμε μια παρόμοια ανάλυση για να δείξουμε ότι το V εκτείνεται στον χώρο γραμμών, ο οποίος τυποποιεί την έννοια των πιθανών "εισόδων" σε έναν αυθαίρετο πίνακα. Αν και οι περιπλοκές της **SVD** προσφέρουν σημαντικό βάθος, για τους σκοπούς αυτού της εργασίας, έχουμε αποκτήσει αρκετές γνώσεις για να κατανοήσουμε πώς η **PCA** εντάσσεται σε αυτό το πλαίσιο.

SVD και PCA

Η **PCA** και η **SVD** είναι στενά συνυφασμένες μεταξύ τους. Ας επιστρέψουμε στον αρχικό πίνακα δεδομένων $m \times n$ X . Μπορούμε να ορίσουμε έναν νέο πίνακα Y ως πίνακα $n \times m$ ως εξής: $Y \equiv \frac{1}{\sqrt{n}} X^T$ {6.52}

Κάθε στήλη του Y έχει μηδενικό μέσο όρο. Αναλύοντας το $Y^T Y$, διαπιστώνουμε ότι ισούται με τον πίνακα συνδιακύμανσης του X (C_X). Όπως συζητήθηκε προηγουμένως, οι κύριες συνιστώσες του X είναι τα ιδιοδιανύσματα του C_X .

$$\begin{aligned} Y^T Y &= \left(\frac{1}{\sqrt{n}} X^T \right)^T \left(\frac{1}{\sqrt{n}} X^T \right) \\ &= \frac{1}{n} X X^T \\ Y^T Y &= C_X \end{aligned} \quad \{6.53\}$$

Όταν υπολογίζουμε το **SVD** του Y , οι στήλες του πίνακα V περιέχουν τα ιδιοδιανύσματα του $Y^T Y$, το οποίο είναι ισοδύναμο με το C_X . Επομένως, οι στήλες του V αντιπροσωπεύουν τις κύριες συνιστώσες του X . Αυτή η σχέση δείχνει ότι η εύρεση των κύριων συνιστωσών είναι ισοδύναμη με τον προσδιορισμό μιας ορθοκανονικής βάσης που καλύπτει το χώρο των στηλών του X .

Quick Summary of PCA

1. Organize data as an $m \times n$ matrix, where m is the number of measurement types and n is the number of samples.
2. Subtract off the mean for each measurement type.
3. Calculate the SVD or the eigenvectors of the covariance.

Εικόνα 7: Πίνακας ανακεφαλαίωσης για την μέθοδο της PCA

Μη Αρνητική Παραγοντοποίηση Πινάκων NMF

Η Μη Αρνητική Παραγοντοποίηση Πινάκων (**Non-negative Matrix Factorization, NMF**) αποτελεί μια ισχυρή μέθοδο μείωσης διαστατικότητας και εξαγωγής χαρακτηριστικών, η οποία έχει βρει ευρεία εφαρμογή σε πολυεπιστημονικά πεδία, συμπεριλαμβανομένης της ανάλυσης υπερφασματικών εικόνων [**NMF-2, NMF-10**]. Η θεμελιώδης αρχή της **NMF** βασίζεται στην παραγοντοποίηση ενός μη-αρνητικού πίνακα δεδομένων X (όπου οι στήλες είναι τα φάσματα και οι γραμμές τα ενεργειακά κανάλια) σε ένα γινόμενο δύο, επίσης μη-αρνητικών, πινάκων: ενός πίνακα βάσης U (που περιέχει τις "φασματικές υπογραφές" των συνιστωσών) και ενός πίνακα συντελεστών V (που περιέχει τις "ποσότητες" κάθε συνιστώσας σε κάθε *pixel*). Η προσέγγιση αυτή βασίζεται στην αρχή ότι πολύπλοκα συστήματα μπορούν να αναπαρασταθούν ως προσθετικοί συνδυασμοί απλούστερων, θεμελιωδών μερών. [**NMF-5, NMF-11, NMF-12**].

Σε αντίθεση με άλλες μεθόδους παραγοντοποίησης πινάκων, όπως η **PCA**, η **NMF** επιβάλλει τον αυστηρό περιορισμό της μη-αρνητικότητας. Αυτή η ιδιότητα είναι καθοριστικής σημασίας σε εφαρμογές όπως η φασματοσκοπία, όπου οι αρνητικές τιμές

σε φάσματα ή σε συγκεντρώσεις υλικών στερούνται φυσικού νοήματος [*NMF-11, Lee & Seung, 2001*]. Ο περιορισμός αυτός οδηγεί σε μια προσθετική αναπαράσταση (*parts-based representation*), όπου τα δεδομένα αναλύονται στα συστατικά τους μέρη, καθιστώντας τα αποτελέσματα εξαιρετικά ερμηνεύσιμα. Η επίλυση του προβλήματος *NMF* είναι ένα μη-κυρτό πρόβλημα βελτιστοποίησης, για το οποίο έχουν αναπτυχθεί διάφοροι αλγόριθμοι, με τους πιο γνωστούς να είναι οι πολλαπλασιαστικοί κανόνες ενημέρωσης (*multiplicative updates*) και οι μέθοδοι εναλλασσόμενων μη-αρνητικών ελαχίστων τετραγώνων (*ANLS*).

Οι εφαρμογές της *NMF* καλύπτουν ένα ευρύ φάσμα επιστημονικών και τεχνολογικών πεδίων. Στην επεξεργασία εικόνας, η *NMF* έχει χρησιμοποιηθεί για την εξαγωγή χαρακτηριστικών, την αποθορυβοποίηση και την αναγνώριση αντικειμένων. Στη βιοπληροφορική, η μέθοδος έχει αποδειχθεί ιδιαίτερα χρήσιμη για την ανάλυση δεδομένων γονιδιακής έκφρασης και την ταξινόμηση βιολογικών δειγμάτων. Στον τομέα της επεξεργασίας φυσικής γλώσσας, η *NMF* έχει εφαρμοστεί με επιτυχία σε προβλήματα ανάλυσης θεμάτων και συστάσεων. Κάθε μία από αυτές τις εφαρμογές επωφελείται από τις μοναδικές ιδιότητες της μεθόδου, ιδιαίτερα από την ικανότητά της να παράγει ερμηνεύσιμες και φυσιολογικά πιστές αναπαραστάσεις.

Οι θεωρητικές προκλήσεις που σχετίζονται με την *NMF* παραμένουν πολυάριθμες και ενδιαφέρουσες. Μια από τις κύριες δυσκολίες αφορά την κατανόηση των συνθηκών σύγκλισης των διαφόρων αλγορίθμων. Επιπλέον, η ανάπτυξη σταθερών στατιστικών θεμελίων για τη μέθοδο αποτελεί αντικείμενο συνεχούς έρευνας. Αυτά τα θεωρητικά ερωτήματα δεν έχουν μόνο ακαδημαϊκό ενδιαφέρον, αλλά επηρεάζουν άμεσα την πρακτική εφαρμογή της μεθόδου σε πραγματικά προβλήματα.

Η σύγκριση της *NMF* με άλλες μεθόδους μείωσης διαστάσεων αποκαλύπτει τόσο τα πλεονεκτήματα όσο και τους περιορισμούς της. Σε αντίθεση με την Ανάλυση Κύριων Συνιστωσών (*PCA*), η *NMF* παράγει αναπαραστάσεις που διατηρούν τη φυσική σημασία των αρχικών δεδομένων. Ωστόσο, αυτό το πλεονέκτημα συχνά έρχεται με το κόστος της αυξημένης υπολογιστικής πολυπλοκότητας και της δυσκολίας στην ερμηνεία των αποτελεσμάτων σε πολύ μεγάλες διαστάσεις. Αυτές οι διαφορές υπογραμμίζουν την ανάγκη για προσεκτική επιλογή της μεθόδου ανάλυσης ανάλογα με τα χαρακτηριστικά του προβλήματος και τη φύση των δεδομένων.

Οι μελλοντικές κατευθύνσεις έρευνας για την *NMF* εστιάζονται σε πολλά ενδιαφέροντα θέματα. Η ενσωμάτωση της μεθόδου με τεχνικές βαθιάς μάθησης αποτελεί μια πολλά υποσχόμενη κατεύθυνση, η οποία θα μπορούσε να οδηγήσει σε υβριδικά μοντέλα με βελτιωμένη απόδοση. Παράλληλα, η ανάπτυξη νέων αλγορίθμων για την αντιμετώπιση μεγάλων και δυναμικών συνόλων δεδομένων παραμένει μια σημαντική πρόκληση. Τέλος, η διερεύνηση νέων εφαρμογών της *NMF* σε τομείς όπως η κβαντική μηχανική και η κλιματική μοντελοποίηση ανοίγει νέους ορίζοντες για την επιστημονική έρευνα. Η *NMF* έχει εφαρμοστεί με επιτυχία στην ανάλυση υπερφασματικών εικόνων [*NMF-2, Feng et al., 2022; NMF-4, Dong et al., 2021*] και συγκεκριμένα σε δεδομένα XRF [*NMF-3, Alfeld et al., 2014; NMF-10, Lach et al., 2021*].

Συμπερασματικά, η θεωρία της *NMF* έχει προχωρήσει σημαντικά, αλλά παραμένει ένα έργο σε εξέλιξη. Οι ιδιότητες του *NMF* έχουν διερευνηθεί βαθύτερα, αν και μια σταθερή στατιστική βάση, συγκρίσιμη με τις παραδοσιακές μεθόδους παραγοντοποίησης όπως η *PCA* ή η *LDA*, βρίσκεται ακόμη υπό ανάπτυξη λόγω της

πολυπλοκότητάς του. Ορισμένα προβλήματα έχουν επιλυθεί, ιδίως εκείνα που αφορούν πρόσθετους περιορισμούς, αλλά πολλά ερωτήματα παραμένουν ανοιχτά. Οι αλγόριθμοι *NMF* κατηγοριοποιούνται σε τέσσερις κύριες κατηγορίες: **Basic NMF**, **Constrained NMF**, **Structured NMF** και **Generalized NMF**. Αυτή η κατηγοριοποίηση παρέχει ένα ευρύ πλαίσιο για την κατανόηση των διαφόρων μοντέλων *NMF* και των εφαρμογών τους. Ο βασικός *NMF* αποτελεί το θεμέλιο πάνω στο οποίο βασίζονται όλα τα άλλα μοντέλα *NMF*, και θα εμβαθύνουμε στη βελτιστοποίηση και τις υπολογιστικές μεθόδους για την αποτελεσματική και εύρωστη επίλυσή του. Επιπλέον, θα ασχοληθούμε με πρακτικά ζητήματα που σχετίζονται με το *NMF* για σύνολα δεδομένων μεγάλης κλίμακας και **online** επεξεργασία. Η περιορισμένη *NMF* περιλαμβάνει τέσσερις υποκατηγορίες: Η καθεμία από αυτές επιβάλλει συγκεκριμένους περιορισμούς για να βελτιώσει την ερμηνευσιμότητα και την αποτελεσματικότητα των λύσεων *NMF*.

Αντίστοιχα, η δομημένη *NMF* κατηγοριοποιείται σε τρεις υποκατηγορίες:

- **Weighted NMF (WNMF)**, η οποία αποδίδει διαφορετικά βάρη στα στοιχεία με βάση τη σχετική τους σημασία.
- **Convolutional NMF (CVNMF)**, η οποία λειτουργεί στο πεδίο χρόνου-συχνότητας.
- Μη αρνητική τριπαγοντική μήτρα (**Nonnegative Matrix Trifactorization - NMTF**), η οποία αναλύει τον πίνακα δεδομένων σε τρεις παραγοντικούς πίνακες.

Επιπλέον, η γενικευμένη *NMF* ταξινομείται σε τέσσερις υποκατηγορίες:

- **Semi-NMF**, η οποία χαλαρώνει τον περιορισμό μη αρνητικότητας σε συγκεκριμένους πίνακες παραγόντων.
- Μη αρνητική παραγοντοποίηση τανυστών (**Nonnegative Tensor Factorization - NTF**), η οποία επεκτείνει τα δεδομένα σε μορφή πίνακα σε τανυστές υψηλότερων διαστάσεων.
- Μη αρνητική παραγοντοποίηση πινάκων-συνόλων (**NMSF**), η οποία επεκτείνει τα σύνολα δεδομένων από πίνακες σε πίνακες-σύνολα.
- **Kernel NMF (KNMF)**, μια μη γραμμική παραλλαγή του *NMF*.

Ορισμός και Μαθηματική Διατύπωση

Η Μη Αρνητική Παραγοντοποίηση Πινάκων (**Non-negative Matrix Factorization, NMF**) αποτελεί μια σημαντική μέθοδο ανάλυσης δεδομένων που εφαρμόζεται σε μη αρνητικούς πίνακες. Δεδομένου ενός πίνακα δεδομένων $X \in \mathbb{R}^{M \times N} \geq 0$, η *NMF* επιδιώκει την ανάλυσή του σε δύο μη αρνητικούς πίνακες: τον πίνακα βάσης $U \in \mathbb{R}^{M \times L} \geq 0$ και τον πίνακα συντελεστών $V \in \mathbb{R}^{L \times N} \geq 0$, έτσι ώστε $X \approx UV$. Αυτή η προσέγγιση μπορεί να εκφραστεί ισοδύναμα σε διανυσματική μορφή ως $x_j \approx \sum (u_i v_{ij})$, όπου κάθε παρατήρηση x_j αναλύεται σε γραμμικό συνδυασμό των διανυσμάτων βάσης u_i με βάρη v_{ij} .

Θεωρητικές Ιδιότητες και Προκλήσεις

Ο βαθμός παραγοντοποίησης (factorization rank) L αντιπροσωπεύει τον αριθμό των συνιστωσών που θα εξαχθούν και συνήθως επιλέγεται $L \ll \min(M, N)$ για την επίτευξη μείωσης διαστατικότητας. Ωστόσο, η λύση του προβλήματος *NMF* είναι μη-μοναδική. Αυτό οφείλεται κυρίως στην περιστροφική ασάφεια, όπου για οποιονδήποτε κατάλληλο πίνακα D , η λύση $(UD)(D^{-1}V)$ είναι επίσης μια έγκυρη παραγοντοποίηση,

και στη γεωμετρική ασάφεια, που σχετίζεται με την επιλογή των διανυσμάτων βάσης. Για την αντιμετώπιση αυτών των ασαφειών και την επίτευξη πιο σταθερών και ερμηνεύσιμων λύσεων, μπορούν να εισαχθούν πρόσθετοι περιορισμοί, όπως η αραιότητα (sparsity) στους πίνακες U ή V .

Αλγοριθμικές Προσεγγίσεις και Εφαρμογές

Η εύρεση των πινάκων U και V είναι ένα πρόβλημα βελτιστοποίησης, όπου ο στόχος είναι η ελαχιστοποίηση μιας συνάρτησης κόστους που μετρά την απόσταση ανακατασκευής μεταξύ του X και του UV . Οι δύο πιο συχνά χρησιμοποιούμενες συναρτήσεις κόστους είναι το Τετραγωνικό Σφάλμα (Frobenius Norm) και η Απόκλιση Kullback-Leibler. Για την ελαχιστοποίηση, έχουν αναπτυχθεί διάφοροι αλγόριθμοι, με τους πιο κλασικούς να είναι οι πολλαπλασιαστικοί κανόνες ενημέρωσης (Multiplicative Updates) των Lee και Seung, οι οποίοι, αν και απλοί, μπορεί να έχουν αργή σύγκλιση. Η ποιότητα της τελικής λύσης εξαρτάται σε μεγάλο βαθμό από την αρχικοποίηση των πινάκων.

Λόγω της ικανότητάς της να εξάγει ερμηνεύσιμες, προσθετικές συνιστώσες, η NMF έχει βρει ευρεία εφαρμογή σε πεδία όπως η επεξεργασία εικόνων για την εξαγωγή χαρακτηριστικών, η βιοπληροφορική για την ανάλυση γονιδιακής έκφρασης και, όπως στην παρούσα εργασία, στην ανάλυση υπερφασματικών δεδομένων για τον διαχωρισμό χημικών ανόργανων στοιχείων.

Η Υβριδική Προσέγγιση NMF-PCA

Η παρούσα εργασία υιοθετεί μια υβριδική, πολυσταδιακή προσέγγιση που συνδυάζει τις παραπάνω μεθόδους. Η ροή εργασίας είναι η εξής: πρώτον, τα δεδομένα υποβάλλονται σε **PCA** για τη μείωση του θορύβου και τη συμπίκνωση της πληροφορίας. Δεύτερον, τα αποθορυβοποιημένα δεδομένα της PCA χρησιμοποιούνται ως είσοδος στον αλγόριθμο **NMF** για να εξαχθούν μη-αρνητικοί, φυσικά ερμηνεύσιμοι χάρτες αφθονίας. Τέλος, αυτοί οι καθαροί χάρτες τροφοδοτούν τον αλγόριθμο **K-Means** για να δημιουργηθεί ο τελικός, διακριτός θεματικός χάρτης υλικών. Αυτή η συνδυαστική προσέγγιση αποσκοπεί στο να παρέχει ένα τελικό αποτέλεσμα που είναι ταυτόχρονα στατιστικά εύρωστο, υπολογιστικά αποδοτικό και, κυρίως, άμεσα ερμηνεύσιμο από χημική και καλλιτεχνική σκοπιά.

Μέθοδοι Συσταδοποίησης

Μετά τη μείωση της διαστατικότητας, το επόμενο βήμα είναι η ανακάλυψη της εγγενούς δομής στα δεδομένα, κάτι που επιτυγχάνεται μέσω της **συσταδοποίησης (clustering)**. Η συσταδοποίηση αποτελεί μια θεμελιώδη μέθοδο της Μάθησης χωρίς Επίβλεψη, με στόχο την ομαδοποίηση μιας συλλογής αντικειμένων, όπως τα pixels, έτσι ώστε τα αντικείμενα εντός της ίδιας ομάδας (συστάδας) να είναι πιο "παρόμοια" μεταξύ τους από ό,τι με τα αντικείμενα άλλων ομάδων. Η διαδικασία αυτή, ωστόσο, παρουσιάζει εγγενείς προκλήσεις. Η έννοια της "ομοιότητας" δεν είναι πάντα μεταβατική, γεγονός που μπορεί να οδηγήσει σε αντιφατικά αποτελέσματα ανάλογα με τον αλγόριθμο. Η σημαντικότερη πρόκληση, όμως, πηγάζει από την ίδια τη φύση της μάθησης χωρίς επίβλεψη: η απουσία μιας προκαθορισμένης "βασικής αλήθειας" (ground truth), γεγονός που καθιστά την αξιολόγηση της απόδοσης ένα μη-τετριμμένο πρόβλημα.

Η διαδικασία της συσταδοποίησης περιλαμβάνει διάφορες θεμελιώδεις επιλογές, όπως η αναπαράσταση των δεδομένων, συνήθως ως ένας πίνακας χαρακτηριστικών (feature matrix), και η επιλογή ενός κατάλληλου μέτρου εγγύτητας, όπως η Ευκλείδεια απόσταση, για την ποσοτικοποίηση της ομοιότητας μεταξύ των δειγμάτων. Κάθε αλγόριθμος χρησιμοποιεί ένα διαφορετικό κριτήριο ομαδοποίησης, το οποίο συχνά εκφράζεται ως μια συνάρτηση κόστους που προσπαθεί να βελτιστοποιήσει. Οι αλγόριθμοι διακρίνονται κυρίως σε ιεραρχικούς, που δημιουργούν μια δενδρική δομή, και σε διαμεριστικούς, όπως ο K-Means, που χωρίζουν τα δεδομένα σε έναν προκαθορισμένο αριθμό μη-επικαλυπτόμενων συστάδων.

Η διαδικασία δημιουργίας μιας συνάρτησης συσταδοποίησης περιλαμβάνει διάφορα βασικά βήματα:

- **Επιλογή χαρακτηριστικών:** Η προσεκτική επιλογή χαρακτηριστικών είναι απαραίτητη για την κωδικοποίηση της μέγιστης πληροφορίας σχετικά με τη συνάρτηση αξίας, ελαχιστοποιώντας παράλληλα τον πλεονασμό μεταξύ των χαρακτηριστικών. Μπορεί επίσης να απαιτείται προεπεξεργασία.
- **Μέτρο εγγύτητας:** Το μέτρο αυτό καθορίζει την ομοιότητα ή τη διαφορετικότητα μεταξύ διανυσμάτων χαρακτηριστικών. Είναι ζωτικής σημασίας να διασφαλιστεί ότι όλα τα επιλεγμένα χαρακτηριστικά συμβάλλουν εξίσου στο μέτρο εγγύτητας και να αποφευχθεί η κυριαρχία των χαρακτηριστικών, η οποία συνήθως αντιμετωπίζεται κατά την προεπεξεργασία.
- **Κριτήριο ομαδοποίησης:** Το κριτήριο για την ομαδοποίηση ορίζεται από έναν εμπειρογνώμονα με βάση τη δομή των συστάδων δεδομένων. Μπορεί να εκφραστεί μέσω μιας συνάρτησης κόστους ή άλλων κανόνων.
- **Αλγόριθμοι συσταδοποίησης:** Η επιλογή ενός κατάλληλου αλγοριθμικού σχήματος για την αποκάλυψη της δομής συσταδοποίησης του συνόλου δεδομένων, λαμβάνοντας υπόψη το επιλεγμένο μέτρο εγγύτητας και το κριτήριο συσταδοποίησης.
- **Επικύρωση των αποτελεσμάτων:** Αφού ληφθούν αποτελέσματα από τον αλγόριθμο συσταδοποίησης, η ορθότητά τους επαληθεύεται με τη χρήση κατάλληλων μέτρων.

Η αξιολόγηση της ποιότητας μιας συσταδοποίησης, ακριβώς επειδή δεν υπάρχει αντικειμενική "σωστή" απάντηση, είναι μια εγγενώς δύσκολη διαδικασία. Η επικύρωση των αποτελεσμάτων μπορεί να γίνει μέσω εσωτερικών μετρικών, οι οποίες αξιολογούν τη συνοχή και τον διαχωρισμό των συστάδων με βάση τα ίδια τα δεδομένα. Μέτρα όπως ο δείκτης Silhouette, ο δείκτης Davies-Bouldin και η μέθοδος του "αγκώνα" (Elbow method), που χρησιμοποιήθηκαν και στην παρούσα εργασία, εμπίπτουν σε αυτή την κατηγορία και είναι κρίσιμα για τον προσδιορισμό του βέλτιστου αριθμού συστάδων (k). Τελικά, η ερμηνεία των αποτελεσμάτων απαιτεί τον συνδυασμό των ποσοτικών μέτρων με την επιστημονική γνώση του πεδίου εφαρμογής για την εξαγωγή ουσιαστικών συμπερασμάτων.

Η ανάλυση συστάδων βρίσκει εφαρμογές σε διάφορους τομείς:

- [1]. **Μείωση δεδομένων:** Σε σενάρια με μεγάλο όγκο δεδομένων (N), η επεξεργασία δεδομένων γίνεται πρόκληση. Η ανάλυση συστάδων μπορεί να βοηθήσει στην οργάνωση των δεδομένων σε "σημαντικές" συστάδες, αντιμετωπίζοντας κάθε συστάδα ως ενιαία οντότητα. Για παράδειγμα, κατά τη

μετάδοση δεδομένων, ορίζεται ένας εκπρόσωπος για κάθε συστάδα και αντί να μεταδίδονται μεμονωμένα δείγματα δεδομένων, αποστέλλεται ένας κωδικός αριθμός που αντιστοιχεί στον εκπρόσωπο της συστάδας, επιτυγχάνοντας συμπίεση δεδομένων.

- [2]. **Δημιουργία υποθέσεων:** Η ανάλυση συστάδων χρησιμοποιείται για την εξαγωγή υποθέσεων σχετικά με τα δεδομένα. Χρησιμεύει ως εργαλείο για την πρόταση υποθέσεων με βάση πρότυπα που αποκαλύπτονται μέσω της ομαδοποίησης. Η δοκιμή αυτών των υποθέσεων με άλλα σύνολα δεδομένων είναι ζωτικής σημασίας.
- [3]. **Έλεγχος υποθέσεων:** Η ανάλυση συστάδων μπορεί επίσης να χρησιμοποιηθεί για τον έλεγχο της εγκυρότητας μιας δεδομένης υπόθεσης.
- [4]. **Πρόβλεψη με βάση τις ομάδες:** Η ανάλυση συστάδων προσδιορίζει ομάδες με βάση τα χαρακτηριστικά των προτύπων. Στη συνέχεια, όταν αντιμετωπίζεται ένα διφορούμενο μοτίβο, είναι δυνατόν να αξιολογηθεί η συστάδα στην οποία πιθανότατα ανήκει και να κατηγοριοποιηθεί αναλόγως με βάση τα χαρακτηριστικά της συστάδας (*Theodoridis et al., 2010*).

Ο αριθμός των πιθανών ομαδοποιήσεων επηρεάζεται από διάφορες μετρικές εγγύτητας, οι οποίες περιγράφουν τις σχέσεις ομοιότητας και ανομοιότητας εντός των ομαδοποιήσεων. Ο συνδυασμός ενός μέτρου εγγύτητας και ενός σχήματος συσταδοποίησης μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα, περιπλέκοντας την ερμηνεία. Σε ένα ιδανικό σενάριο, κάποιος θα εξέταζε όλες τις πιθανές καταταμήσεις και θα επέλεγε την πιο λογική χρησιμοποιώντας ένα προκαθορισμένο κριτήριο. Ωστόσο, για μέτριες τιμές του N , αυτή η εξαντλητική προσέγγιση καθίσταται μη πρακτική (*Theodoridis et al., 2010*).

Η ανάλυση συστάδων χρησιμοποιεί διάφορους αλγόριθμους που εξετάζουν μόνο ένα υποσύνολο όλων των πιθανών διαμερισμάτων, οδηγώντας σε διαφορετικά αποτελέσματα ανάλογα με τον αλγόριθμο και τα κριτήρια που χρησιμοποιούνται. Αυτοί οι αλγόριθμοι αποσκοπούν στον εντοπισμό των συγκεκριμένων χαρακτηριστικών συστάδων εντός του συνόλου δεδομένων. Οι αλγόριθμοι συσταδοποίησης μπορούν να κατηγοριοποιηθούν ως εξής:

- 1. **Διαδοχικοί αλγόριθμοι:** Αυτοί οι αλγόριθμοι δημιουργούν μια ενιαία συστάδα και είναι σχετικά απλοί και γρήγοροι. Το τελικό αποτέλεσμα εξαρτάται συχνά από τη σειρά με την οποία επεξεργάζονται τα διανύσματα χαρακτηριστικών και τείνουν να δημιουργούν συμπαγείς, υπερσφαιρικές ή υπερελλειψοειδείς συστάδες με βάση τη μετρική απόστασης που χρησιμοποιείται.
- 2. **Αλγόριθμοι ιεραρχικής συσταδοποίησης:** Αυτές οι μέθοδοι χωρίζονται περαιτέρω σε δύο ομάδες:
 - I. **Συσσωρευτικοί αλγόριθμοι:** Δημιουργούν μια ακολουθία φθινουσών συστάδων σε κάθε στάδιο συγχωνεύοντας δύο συστάδες σε μία. Οι αλγόριθμοι μονής και πλήρους σύνδεσης είναι εξέχοντες εκπρόσωποι, κατάλληλοι για την ανάκτηση μεγάλων και συμπαγών συστάδων.
 - II. **Διαχωριστικοί αλγόριθμοι:** Αυτοί οι αλγόριθμοι λειτουργούν προς την αντίθετη κατεύθυνση, διαιρώντας μια ενιαία συστάδα σε δύο σε κάθε στάδιο.

3. **Αλγόριθμοι συσταδοποίησης με βάση τη βελτιστοποίηση της συνάρτησης κόστους:** Αυτή η κατηγορία περιλαμβάνει αλγορίθμους όπου μια συνάρτηση κόστους (J) ποσοτικοποιεί την ευαισθησία στον προσδιορισμό της συσταδοποίησης. Ο αριθμός των συστάδων (m) συνήθως διατηρείται σταθερός. Αυτοί οι αλγόριθμοι χρησιμοποιούν συχνά αρχές διαφορικού λογισμού για τη βελτιστοποίηση του J και τερματίζουν όταν επιτυγχάνεται τοπικό βέλτιστο. Οι υποκατηγορίες αυτής της κατηγορίας περιλαμβάνουν:
 - I. **Σκληροί ή τραχείς αλγόριθμοι συσταδοποίησης:** Αναθέτουν κάθε διάνυσμα αποκλειστικά σε μία μόνο συστάδα με βάση το επιλεγμένο κριτήριο βελτιστοποίησης, με τον αλγόριθμο Lloyd να αποτελεί ένα δημοφιλές παράδειγμα.
 - II. **Πιθανοτικοί αλγόριθμοι ομαδοποίησης:** Ένας υποτύπος των σκληρών αλγορίθμων συσταδοποίησης που χρησιμοποιούν την ταξινόμηση κατά *Bayes* για την ανάθεση διανυσμάτων σε συστάδες με βάση τις εκ των υστέρων πιθανότητες.
 - III. **Ασαφείς αλγόριθμοι συσταδοποίησης:** Επιτρέπουν στα διανύσματα να ανήκουν σε συστάδες σε ορισμένο βαθμό.
 - IV. **Πιθανολογικοί αλγόριθμοι συσταδοποίησης:** Αξιολογούν την πιθανότητα ένα διάνυσμα να ανήκει σε μια συστάδα.

Αλγόριθμοι ανίχνευσης ορίων: Αυτοί οι αλγόριθμοι ενημερώνουν επαναληπτικά τα όρια των συστάδων αντί να εντοπίζουν απευθείας συστάδες με βάση τα διανύσματα χαρακτηριστικών. Εκτός από αυτούς, υπάρχουν διάφοροι άλλοι αλγόριθμοι ομαδοποίησης, όπως η ομαδοποίηση με διακλαδώσεις και δεσμεύσεις, η γενετική ομαδοποίηση, οι μέθοδοι στοχαστικής χαλάρωσης, η ομαδοποίηση που αναζητά κοιλάδες, η ανταγωνιστική μάθηση, οι αλγόριθμοι που βασίζονται σε μορφολογικούς μετασχηματισμούς, οι αλγόριθμοι που βασίζονται στην πυκνότητα, η ομαδοποίηση υποδιαστημάτων και οι μέθοδοι που βασίζονται σε πυρήνες (*Theodoridis et al., 2010*).

Συνοπτικά, η ανάλυση συστάδων δημιουργεί προκλήσεις στην αξιολόγηση των δομών ομαδοποίησης, καθώς εμπίπτει στον τομέα της μάθησης χωρίς επίβλεψη. Παρ' όλα αυτά, βρίσκει εφαρμογές στη μείωση δεδομένων, στη δημιουργία και τον έλεγχο υποθέσεων και στην πρόβλεψη βάσει ομάδων. Ο αριθμός των πιθανών ομαδοποιήσεων είναι τεράστιος, καθιστώντας την εξαντλητική διερεύνηση μη πρακτική. Οι αλγόριθμοι ομαδοποίησης ποικίλλουν ως προς την προσέγγιση και το αποτέλεσμα, περιλαμβάνοντας διαδοχικές, ιεραρχικές και βασισμένες στη βελτιστοποίηση συναρτήσεων κόστους μεθόδους, καθεμία από τις οποίες είναι κατάλληλη για διαφορετικά σενάρια.

Αλγόριθμος Συσταδοποίησης K – MEANS

Η ομαδοποίηση K-means Σχήμα 7 είναι μια διαδικασία που ξεκινά με τον ορισμό μιας συνάρτησης κόστους σε ένα σύνολο πιθανών λύσεων ομαδοποίησης, με στόχο την εύρεση της βέλτιστης ομαδοποίησης μέσω ενός προβλήματος βελτιστοποίησης. Σε αυτό το μοντέλο, ο στόχος της ομαδοποίησης καθορίζεται από μια αντικειμενική συνάρτηση που συμβολίζεται ως G , η οποία απεικονίζει μια είσοδο (X, d) και μια προτεινόμενη λύση ομαδοποίησης $C = (C_1, \dots, C_k)$ σε θετικούς πραγματικούς αριθμούς, με στόχο την ελαχιστοποίηση της $G((X, d), C)$. Η επίτευξη αυτού του στόχου απαιτεί τη χρήση ενός κατάλληλου αλγορίθμου αναζήτησης. Είναι σημαντικό να σημειωθεί ότι η ομαδοποίηση **K-means** είναι συγκεκριμένα ένας αλγόριθμος προσέγγισης και όχι η

ακριβής λύση του προβλήματος ελαχιστοποίησης της συνάρτησης κόστους **SSEWC** (Sum of Squared Errors Within Clusters).

$$SSEWC = \argmin_c \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2 = \argmin_c \sum_{i=1}^k |C_i| \text{var} C_i \quad \{6.34\}$$

$$d(x_i, \mu_i) = \sqrt{(x_i - \mu_i)^2}$$

Συνήθως, οι αντικειμενικές συναρτήσεις που χρησιμοποιούνται στην ομαδοποίηση περιλαμβάνουν μια παράμετρο για τον αριθμό των συστάδων, που συμβολίζεται ως k . Στην πράξη, η επιλογή της παραμέτρου k εναπόκειται συχνά στον χρήστη του αλγορίθμου ομαδοποίησης και μπορούν να εφαρμοστούν διαφορετικές αντικειμενικές συναρτήσεις. Μια ευρέως χρησιμοποιούμενη αντικειμενική συνάρτηση είναι η αντικειμενική συνάρτηση ***k-means***, η οποία μετρά την τετραγωνική απόσταση κάθε σημείου δεδομένων στο X από το κεντροειδές (μέσος όρος) της συστάδας του. Αυτή η αντικειμενική συνάρτηση είναι ιδιαίτερα σημαντική σε εφαρμογές όπως η ψηφιακή επικοινωνία, όπου το X αντιπροσωπεύει ένα σύνολο σημάτων προς μετάδοση. Ωστόσο, η εύρεση της βέλτιστης λύσης για το ***k-means*** είναι υπολογιστικά ανέφικτη, οδηγώντας στη χρήση επαναληπτικών αλγορίθμων για την προσέγγιση της βέλτιστης λύσης. Ως εκ τούτου, ο όρος "ομαδοποίηση ***k-means***" αναφέρεται συνήθως στο αποτέλεσμα αυτού του επαναληπτικού αλγορίθμου και όχι στην ακριβή ελαχιστοποίηση του στόχου ***k-means*** (Shalev-Shwartz & Ben-David, 2014).

$$SSEWC = \argmin_c \sum_{i=1}^k \frac{1}{2|C_i|} \sum_{x,y \in C_i} |x - y|^2 \quad \{6.35\}$$

$$\sum_{x \in C_i} |x - \mu_i|^2 = \sum_{x \neq y \in C_i} (x - \mu_i)^T (\mu_i - y) \quad \{6.36\}$$

$$SSEBC = \argmax_c \sum_{i=1}^k |C_i| |\mu - \mu_i|^2 \quad \{6.37\}$$

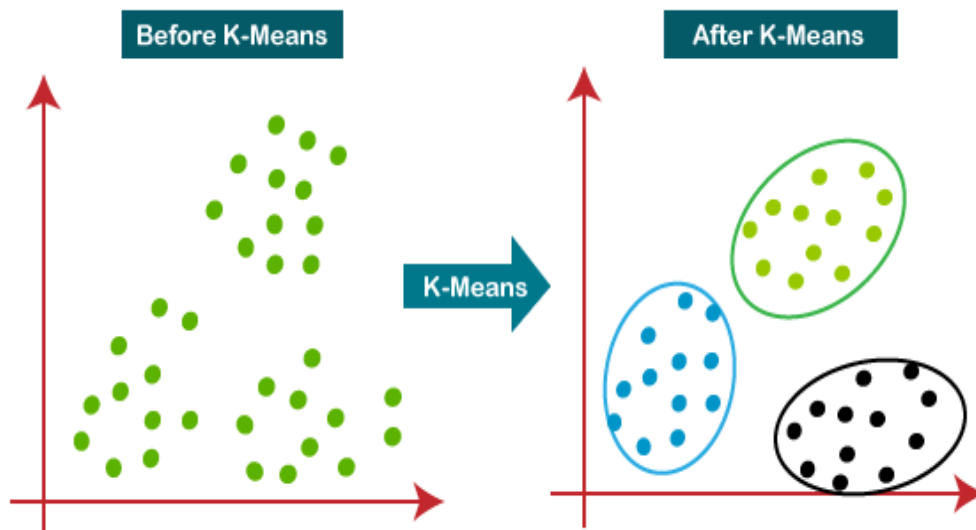
Οι παραλλαγές του αλγορίθμου ***k-means*** περιστρέφονται γύρω από τη θεμελιώδη έννοια της ελαχιστοποίησης του αθροίσματος των τετραγωνικών σφαλμάτων εντός των συστάδων (***SSEWC***). Η ελαχιστοποίηση αυτή επιτυγχάνεται λαμβάνοντας υπόψη τις κατά ζεύγη τετραγωνικές αποκλίσεις μεταξύ σημείων δεδομένων εντός της ίδιας συστάδας. Τα βήματα αυτά περιλαμβάνουν:

- Δημιουργία K κεντροειδών και ανάθεση κάθε σημείου δεδομένων στο πλησιέστερο κεντροειδές.
- Επαναυπολογισμός των κεντροειδών επαναληπτικά έως ότου δεν υπάρχει περαιτέρω αλλαγή στην κατανομή των σημείων δεδομένων σε συστάδες.

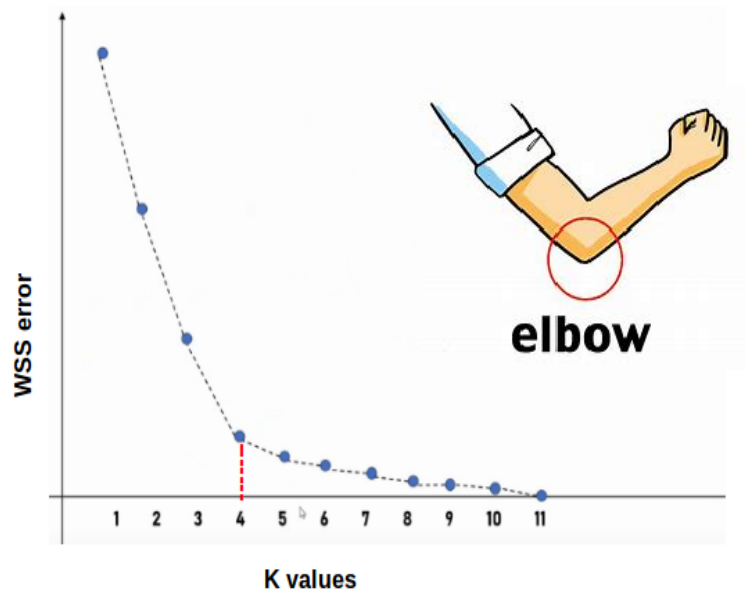
Ο αλγόριθμος ***k-means*** προσφέρει αρκετά πλεονεκτήματα, όπως το ότι είναι μια μέθοδος μάθησης χωρίς επίβλεψη, η απλότητα στην αρχή και την εφαρμογή του και τα ερμηνεύσιμα αποτελέσματα. Ωστόσο, έχει τα μειονεκτήματά του, όπως η πρόκληση της επιλογής μιας κατάλληλης τιμής για το K , η πιθανότητα σύγκλισης σε τοπικά ελάχιστα, η βραδύτερη σύγκλιση σε μεγάλα σύνολα δεδομένων και η ευαισθησία σε ακραίες τιμές. Επομένως, ο καθορισμός του αριθμού των συστάδων και η κατανόηση του πεδίου εφαρμογής είναι κρίσιμα ζητήματα κατά τη χρήση του αλγορίθμου ***k-means***.

Θα πρέπει να σημειωθεί ότι η συνάρτηση αθροίσματος τετραγωνικών σφαλμάτων (*SSE*) υποθέτει ότι οι συστάδες διαθέτουν χαρακτηριστικά που μοιάζουν με μίγμα *Gaussians*. Υπό αυτή την υπόθεση, κάθε συστάδα θεωρείται ότι έχει σφαιρικό σχήμα και τα αντικείμενα δεδομένων συγκεντρώνονται κυρίως γύρω από τα αντίστοιχα κεντροειδή τους. Κατά συνέπεια, ο αλγόριθμος *k-means* μπορεί να δυσκολευτεί να εντοπίσει μια σημαντική κατάτμηση όταν έχει να κάνει με μη σφαιρικές συστάδες ή συστάδες που διαφέρουν σημαντικά σε μέγεθος. Επιπλέον, η τυπική αντικειμενική συνάρτηση *k-means* είναι ευάλωτη στην επίδραση των ανωμαλιών. Τέτοιες ανωμαλίες μπορούν να αλλοιώσουν την κατασκευή του κεντροειδούς, με αποτέλεσμα διανύσματα που δεν απεικονίζουν επαρκώς τις συστάδες δεδομένων, τα οποία μπορούν να παράγουν ιδιαίτερα παραμορφωμένες συστάδες.

Οι συγγραφείς έχουν προτείνει την ενσωμάτωση μιας "συστάδας ακραίων τιμών" για να ξεπεραστούν αυτές οι δυσκολίες. Αυτή η συστάδα έχει σχεδιαστεί για να φιλοξενεί αντικείμενα που δεν εντάσσονται σε καμία άλλη ομάδα. Ορισμένοι έχουν προτείνει μια μέθοδο που συνεπάγεται την επανειλημμένη εφαρμογή του αλγορίθμου συσταδοποίησης και στη συνέχεια την εξάλειψη των ανεπαρκώς συσταδοποιημένων δεδομένων στο τέλος κάθε εκτέλεσης. Παρ' όλα αυτά, και οι δύο προσεγγίσεις απαιτούν τον καθορισμό ενός αυθαίρετου κατωφλίου για να καθοριστεί αν ένα αντικείμενο απέχει αρκετά από το υπάρχον κεντροειδές ώστε να αναγνωριστεί ως ακραίο. Μια περαιτέρω πρόκληση προκύπτει όταν προκύπτουν κενές συστάδες κατά τη διάρκεια της επαναληπτικής διαδικασίας βελτίωσης. Σε τέτοιες περιπτώσεις, μια συνήθης τακτική είναι να κατανέμεται το πιο απομακρυσμένο αντικείμενο (δηλαδή το αντικείμενο που βρίσκεται πιο μακριά από το υπάρχον κεντροειδές του) στην κενή συστάδα. Ωστόσο, εάν το πρόβλημα επιμένει, υποδηλώνει ότι το σφάλμα έγκειται κυρίως στην επιλογή του μοντέλου ομαδοποίησης, συμπεριλαμβανομένης της λανθασμένης επιλογής της τιμής για το k [*KMeans-3*, *KMeans-5*, *KMeans-9*].



Elbow method

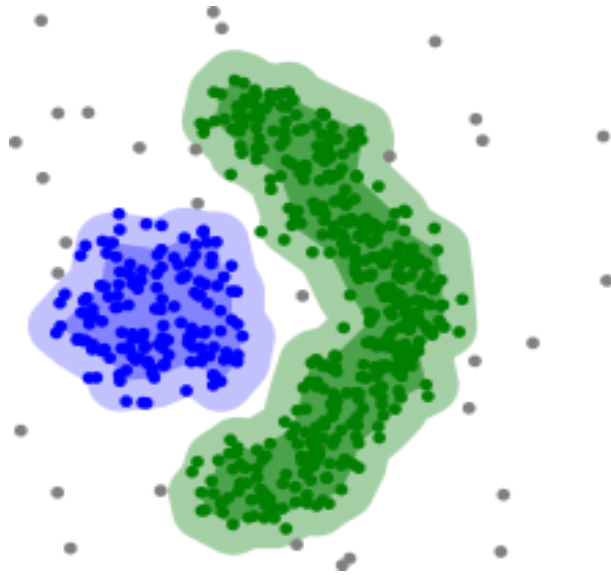


Σχήμα 7: Εικόνες Αναπαράστασης δεδομένων πριν και μετά την Συσταδοποίηση (ορθή). Επιπλέον η κάτω εικόνα αποτελεί την μέθοδο εύρεσης σωστού αριθμού συστάδων με την τεχνική του αγκώνα

Αλγόριθμος Συσταδοποίησης DBScan

Οι αλγόριθμοι ομαδοποίησης με βάση την πυκνότητα, όπως ο **DBSCAN** (**Density-Based Spatial Clustering of Applications with Noise**), έχουν σχεδιαστεί για να ανακαλύπτουν ομάδες στα δεδομένα εντοπίζοντας περιοχές με υψηλή πυκνότητα σημείων δεδομένων [DBScan-1] **Σχήμα 8**. Σε αντίθεση με άλλες τεχνικές συσταδοποίησης, οι αλγόριθμοι αυτοί δεν επιβάλλουν αυστηρές υποθέσεις σχετικά με το σχήμα των συστάδων, καθιστώντας τους ικανούς να ανιχνεύουν συστάδες αυθαίρετου σχήματος, ενώ παράλληλα χειρίζονται αποτελεσματικά τις ακραίες τιμές. Επιπλέον, οι αλγόριθμοι που βασίζονται στην πυκνότητα τείνουν να έχουν χαμηλότερη χρονική πολυπλοκότητα, καθιστώντας τους κατάλληλους για την επεξεργασία μεγάλων συνόλων δεδομένων. Ορισμένοι γνωστοί αλγόριθμοι συσταδοποίησης με βάση την πυκνότητα περιλαμβάνουν τους **DBSCAN**, **DBCLASD**, **DENCLUE** και **OPTICS**. Ενώ αυτοί οι αλγόριθμοι μοιράζονται μια κοινή βασική φιλοσοφία, διαφέρουν στον τρόπο με τον οποίο ποσοτικοποιούν την πυκνότητα (*Theodoridis et al., 2010*). Ο **DBSCAN** (**Density-Based Spatial Clustering of Applications with Noise**) υπολογίζει την πυκνότητα εξετάζοντας τον αριθμό των σημείων δεδομένων εντός μιας καθορισμένης περιοχής γύρω από κάθε σημείο σε έναν μονοδιάστατο χώρο. Τα αποτελέσματα αυτού του αλγορίθμου εξαρτώνται σε μεγάλο βαθμό από τις επιλογές των παραμέτρων ϵ και q . Διαφορετικές τιμές των παραμέτρων μπορούν να οδηγήσουν σε εντελώς διαφορετικά αποτελέσματα. Ως εκ τούτου, η επιλογή των κατάλληλων τιμών για τα ϵ και q είναι ουσιαστική και συχνά απαιτείται πειραματισμός με διαφορετικούς συνδυασμούς για τον εντοπισμό της καλύτερης προσαρμογής για ένα δεδομένο σύνολο δεδομένων. Η **DBSCAN** ενδέχεται να μην αποδίδει καλά σε σενάρια με σημαντικές διακυμάνσεις της πυκνότητας μεταξύ συστάδων ή σε δεδομένα υψηλής διάστασης. Μια επέκταση της **DBSCAN**, που ονομάζεται **OPTICS** (**Ordering Points to Identify**

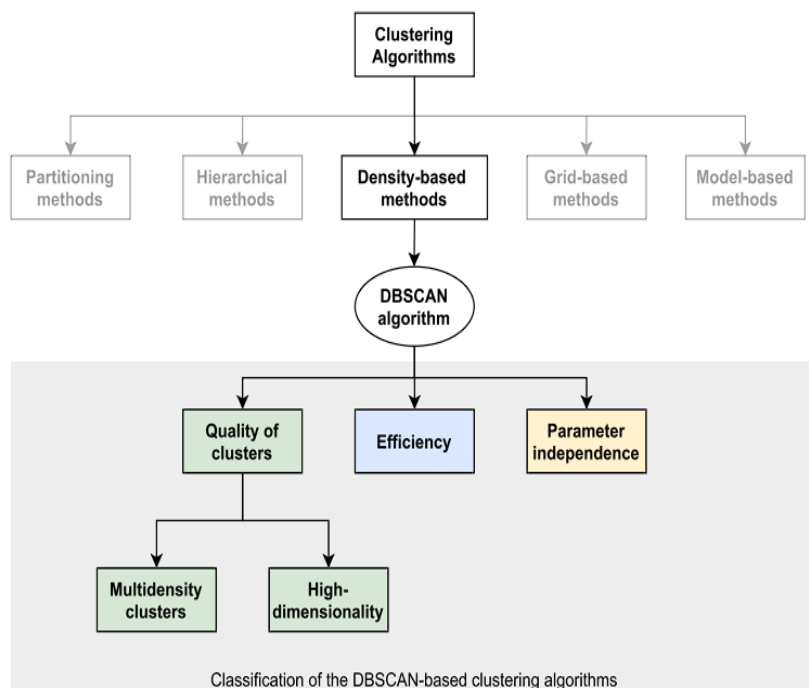
the Clustering Structure), αντιμετωπίζει την πρόκληση της επιλογής παραμέτρων και παρέχει μια ιεραρχική άποψη της δομής των συστάδων. Παρουσιάζει επίσης χαμηλότερη υπολογιστική πολυπλοκότητα σε σύγκριση με το **DBSCAN** (*Theodoridis et al., 2010*). Το **DBSCAN** βασίζεται στην έννοια ότι οι συστάδες είναι περιοχές υψηλής πυκνότητας σημείων δεδομένων που διαχωρίζονται από αραιότερες περιοχές στο χώρο των δεδομένων. Ο αλγόριθμος χρησιμοποιεί δύο παραμέτρους εισόδου: ϵ (έψιλον) και *minpts* (ελάχιστα σημεία) για την εκτίμηση της τοπικής πυκνότητας γύρω από ένα σημείο δεδομένων. Συγκεκριμένα, το ϵ ορίζει την ακτίνα της τοπικής γειτονιάς και το *minpts* είναι ο ελάχιστος αριθμός σημείων δεδομένων που απαιτούνται εντός αυτής της ακτίνας για να σχηματιστεί μια συστάδα. Δεδομένου ενός συνόλου δεδομένων X , το **DBSCAN** υπολογίζει την τοπική πυκνότητα κάθε αντικειμένου x_i στο X ως τον αριθμό των αντικειμένων εντός της ϵ -γειτονιάς του (συμβολίζεται ως $Ne(x_i)$). Αυτή η προσέγγιση επιτρέπει την ανακάλυψη συστάδων με αυθαίρετα σχήματα.



Σχήμα 8: Αναπαράσταση Συσταδοποίησης δεδομένων σε δυο (2) ομάδες

Η **DBSCAN** ταξινομεί τα αντικείμενα σε τρεις κατηγορίες: αντικείμενα πυρήνα, αντικείμενα ορίων και αντικείμενα θορύβου. Τα αντικείμενα πυρήνα είναι εκείνα με τουλάχιστον *minPts* σημεία δεδομένων στην ϵ -γειτονιά τους, που σημαίνει ότι ανήκουν σε μια συστάδα. Τα συνοριακά αντικείμενα έχουν λιγότερους από *minPts* γείτονες, αλλά μπορούν να συνδεθούν με τη γειτονιά ενός αντικειμένου πυρήνα. Τα αντικείμενα θορύβου δεν πληρούν τα κριτήρια ούτε για τα αντικείμενα πυρήνα ούτε για τα συνοριακά αντικείμενα και αντιμετωπίζονται ως ακραίες τιμές. Ο αλγόριθμος **DBSCAN** προχωρά επιλέγοντας ένα τυχαίο μη ταξινομημένο αντικείμενο πυρήνα ως "σπόρο" και συλλέγοντας αναδρομικά αντικείμενα που είναι προσβάσιμα σε πυκνότητα για να σχηματίσουν μια συστάδα. Η διαδικασία αυτή συνεχίζεται έως ότου δεν απομένουν μη ταξινομημένα αντικείμενα. Ενώ ο αρχικός αλγόριθμος **DBSCAN** παρουσιάζει χρονική πολυπλοκότητα $O(n^2)$, βελτιστοποιήσεις, όπως η χρήση δομών χωρικού ευρετηρίου όπως τα *R*-δέντρα*, μπορούν να μειώσουν την πολυπλοκότητα σε $O(n \log n)$. Ο **DBSCAN** εισήγαγε το πλεονέκτημα του εντοπισμού συστάδων με ακανόνιστα σχήματα, της ανίχνευσης ακραίων τιμών και του προσδιορισμού συστάδων χωρίς προηγούμενη γνώση των ετικετών κλάσεων. Η ευελιξία της οδήγησε σε εφαρμογές σε διάφορους τομείς, όπως η πολιτική μηχανική, η χημεία, οι κοινωνικές επιστήμες, η ιατρική διάγνωση, η τηλεπισκόπηση, η όραση υπολογιστών και η

ανίχνευση ανωμαλιών. Η επιτυχία του αλγορίθμου συνέβαλε στην αναγνώρισή του με το βραβείο δοκιμής του χρόνου της ομάδας ειδικού ενδιαφέροντος για την ανακάλυψη γνώσης σε βάσεις δεδομένων (**SIGKDD**). Επιπλέον, ο **DBSCAN** ενέπνευσε την ανάπτυξη πολυάριθμων αλγορίθμων ομαδοποίησης με στόχο την υπέρβαση των περιορισμών του και τη βελτίωση των δυνατοτήτων του.



Εικόνα 6: Διάγραμμα UML αλγορίθμου συσταδοποίησης DBScan με τα πλεονεκτήματά του

Συνοπτικά, ο **DBSCAN** Εικόνα 6 είναι ένας πρωτοποριακός αλγόριθμος ομαδοποίησης με βάση την πυκνότητα, γνωστός για την ικανότητά του να εντοπίζει ομάδες αυθαίρετου σχήματος και να χειρίζεται αποτελεσματικά τις ακραίες τιμές. Παρέχει μια πρακτική λύση για την ομαδοποίηση σε ποικίλες εφαρμογές και άνοιξε το δρόμο για περαιτέρω εξελίξεις στον τομέα της ομαδοποίησης με βάση την πυκνότητα.

Ο αλγόριθμος **DBSCAN**, όπως παρουσιάζεται, λειτουργεί εντοπίζοντας συστάδες με βάση την έννοια των περιοχών με υψηλή πυκνότητα αντικειμένων δεδομένων που χωρίζονται από περιοχές με χαμηλότερη πυκνότητα σημείων δεδομένων σε έναν δεδομένο χώρο δεδομένων. Για να το επιτύχει αυτό, ο **DBSCAN** χρησιμοποιεί δύο κρίσιμες παραμέτρους εισόδου: ϵ (έψιλον) και **minPts** (ελάχιστα σημεία), οι οποίες διαδραματίζουν ζωτικό ρόλο στην εκτίμηση της πυκνότητας εντός μιας συγκεκριμένης περιοχής που περιβάλλει ένα σημείο δεδομένων. Αναλυτικότερα, το ϵ αντιπροσωπεύει την ακτίνα της τοπικής γειτονιάς, ενώ το **minPts** υποδηλώνει τον ελάχιστο αριθμό αντικειμένων δεδομένων που απαιτούνται εντός αυτής της ακτίνας για τον ορισμό μιας συστάδας. Για ένα δεδομένο σύνολο δεδομένων X , η **DBSCAN** υπολογίζει την τοπική πυκνότητα κάθε σημείου δεδομένων xi , όπου $xi \in X$, προσδιορίζοντας τον συνολικό αριθμό των αντικειμένων εντός της ϵ -γειτονιάς του. Αυτό μπορεί να εκφραστεί ως η καρτελικότητα του $Ne(xi)$, που συμβολίζεται ως $|Ne(xi)|$, όπου το $Ne(xi)$ ορίζεται ως ένα σύνολο σημείων δεδομένων $xj \in X$ που ικανοποιούν τη συνθήκη $\forall j, dist(xi, xj) < \epsilon$. Μέσω της χρήσης της εξίσωσης, ο αλγόριθμος **DBSCAN** καθίσταται ικανός να εντοπίζει συστάδες αυθαίρετου σχήματος. Επιπλέον, λαμβάνοντας υπόψη τον αριθμό των αντικειμένων δεδομένων εντός μιας γειτονιάς, προκύπτουν τρεις διαφορετικοί

τύποι αντικειμένων. Πρώτον, τα αντικείμενα πυρήνα είναι εκείνα για τα οποία η ε-γειτονιά περιέχει τουλάχιστον $\minPts$ αντικείμενα δεδομένων, που μαθηματικά αναπαρίσταται ως $|Ne(xi)| \geq \minPts$, όπου $xi \in X$. Τα αντικείμενα πυρήνα υποδεικνύουν την παρουσία μιας συστάδας όταν λαμβάνονται υπόψη οι καθορισμένες τιμές ϵ και $\minPts$, με τα γειτονικά τους αντικείμενα να σχηματίζουν μια συστάδα. Δεύτερον, τα συνοριακά αντικείμενα έχουν λιγότερα αντικείμενα δεδομένων στη γειτονιά τους ϵ από την απαιτούμενη τιμή $\minPts$. Ωστόσο, για ένα συνοριακό αντικείμενο xj , υπάρχει ένα αντικείμενο πυρήνα xi τέτοιο ώστε $xj \in Ne(xi)$. Τέλος, τα αντικείμενα θορύβου είναι εκείνα τα σημεία δεδομένων που αποτυγχάνουν να ικανοποιήσουν την ελάχιστη απαίτηση $\minPts$ στην *ε-γειτονιά* τους και δεν ανήκουν στη γειτονιά κανενός αντικειμένου πυρήνα.

Επιπλέον, ο αλγόριθμος **DBSCAN** εισάγει πρόσθετους ορισμούς για την καθιέρωση της έννοιας της προσπελασιμότητας μεταξύ ζευγών αντικειμένων δεδομένων. Οι ορισμοί αυτοί ορίζονται ως εξής:

A. ΑΜΕΣΑ ΠΡΟΣΒΑΣΙΜΟΤΗΤΑ ΠΥΚΝΟΤΗΤΑΣ:

Ένα αντικείμενο $xi \in X$ θεωρείται άμεσα προσβάσιμο σε πυκνότητα από ένα αντικείμενο $xj \in X$ εντός των δεδομένων παραμέτρων ϵ και $\minPts$ εάν πληρούνται δύο προϋποθέσεις: (α) το xi είναι μέλος της ε-γειτονιάς του xj και (β) το xj ταξινομείται ως αντικείμενο πυρήνα. Αυτό εκφράζεται ως $DirReach(xi, xj) \Leftrightarrow xi \in Ne(xj) \wedge |Ne(xj)| \geq \minPts$.

B. ΠΡΟΣΒΑΣΙΜΟΤΗΤΑ ΠΥΚΝΟΤΗΤΑΣ:

Ένα αντικείμενο $xi \in X$ θεωρείται προσβάσιμο σε πυκνότητα από ένα άλλο αντικείμενο $xj \in X$ όσον αφορά τα καθορισμένα ϵ και $\minPts$ εάν υπάρχει μια ακολουθία αντικειμένων $o1, \dots, on$, όπου $o1 = xj$ και $on = xi$. Η προϋπόθεση για να ισχύει αυτό είναι ότι κάθε επόμενο αντικείμενο $oi+1$ είναι άμεσα προσβάσιμο κατά πυκνότητα από το προηγούμενο αντικείμενο oi . Μαθηματικά, $Reach(xi, xj) \Leftrightarrow \exists o1, \dots, on \in X : o1 = xj \wedge on = xi \wedge \forall i \in \{1 \dots n - 1\} : DirReach(oi, oi+1)$. Η εισαγωγή αυτών των ορισμών εξασφαλίζει ότι δύο συνοριακά αντικείμενα, ακόμη και αν ανήκουν στην ίδια συστάδα, μπορεί να μην είναι προσβάσιμα σε πυκνότητα το ένα από το άλλο λόγω των περιορισμών πυρήνα-αντικειμένου που επιβάλλονται στην ακολουθία των αντικειμένων $o1, \dots, on$. Ωστόσο, είναι άμεσα προσβάσιμα σε πυκνότητα από ένα κοινό αντικείμενο πυρήνα, εισάγοντας έτσι την έννοια της συνδεσιμότητας πυκνότητας.

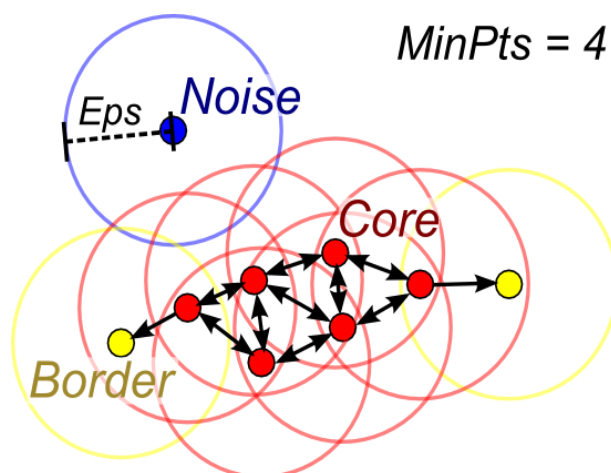
C. ΣΥΝΔΕΣΜΟΤΗΤΑ ΠΥΚΝΟΤΗΤΑΣ:

Ένα αντικείμενο $xi \in X$ θεωρείται συνδεδεμένο με πυκνότητα με ένα άλλο αντικείμενο $xj \in X$ εντός των καθορισμένων ϵ και $\minPts$ εάν υπάρχει ένα άλλο αντικείμενο o έτσι ώστε τόσο το xi όσο και το xj να είναι προσβάσιμα με πυκνότητα από το o όσον αφορά τα ϵ και $\minPts$. Αυτό εκφράζεται ως $Connect(xi, xj) \Leftrightarrow \exists o \in X : Reach(o, xi) \wedge Reach(o, xj)$. Με βάση αυτές τις καθορισμένες έννοιες, η **DBSCAN** ορίζει μια συστάδα ως ένα μέγιστο σύνολο αντικειμένων που συνδέονται με πυκνότητα σε σχέση με την προσπελασιμότητα πυκνότητας. Με άλλα λόγια, μια συστάδα C , που καθορίζεται από το ϵ και το $\minPts$, είναι ένα μη κενό υποσύνολο του X που ικανοποιεί τις ακόλουθες συνθήκες: α . Για όλα τα xi, xj : αν $xi \in C$ και το xj είναι προσβάσιμο από το xi με βάση

την πυκνότητα σε σχέση με το ϵ και το *minPts*, τότε το $x_j \in C$, και β . Για όλα τα $x_i, x_j \in C$: το x_i είναι συνδεδεμένο με πυκνότητα με το x_j σε σχέση με το ϵ και το *minPts*.

Τα αντικείμενα θορύβου, από την άλλη πλευρά, αναφέρονται σε ένα σύνολο σημείων δεδομένων εντός του συνόλου δεδομένων που δεν ανήκουν σε καμία από τις αναγνωρισμένες συστάδες. Ο αλγόριθμος **DBSCAN** χρησιμοποιεί έναν μηχανισμό δύο βημάτων για την ταξινόμηση των αντικειμένων στο σύνολο δεδομένων με βάση τους παρεχόμενους ορισμούς. Αυτά τα μέγιστα σύνολα αντικειμένων που είναι προσβάσιμα σε πυκνότητα αντιστοιχίζονται στη συνέχεια στις αντίστοιχες συστάδες τους. Ο αλγόριθμος ξεκινά επιλέγοντας τυχαία ένα μη ταξινομημένο βασικό αντικείμενο από το σύνολο δεδομένων, ορίζοντάς το ως "σπόρο" και στη συνέχεια συγκεντρώνοντας όλα τα αντικείμενα που είναι προσβάσιμα από την πυκνότητα από αυτόν τον σπόρο για να κατασκευάσει μια συστάδα που περιλαμβάνει τον ίδιο τον σπόρο. Αυτή η διαδικασία δύο βημάτων εφαρμόζεται επαναληπτικά μέχρι να μην απομείνουν μη ταξινομημένα αντικείμενα.

Η **DBSCAN** είναι γνωστή για την ικανότητά της να δημιουργεί συστάδες χωρίς να



απαιτείται προηγούμενη γνώση του αριθμού των συστάδων. Αυτό το χαρακτηριστικό τον διαφοροποιεί από αλγόριθμους όπως ο *k-means*, οι οποίοι απαιτούν τον αριθμό των συστάδων ως παράμετρο εισόδου. Ωστόσο, ο **DBSCAN** έχει ορισμένους αξιοσημείωτους περιορισμούς. Ο αλγόριθμος βασίζεται σε μεγάλο βαθμό στην παράμετρο ϵ που ορίζει ο χρήστης. Οι συστάδες που προσδιορίζονται από τον αλγόριθμο μπορεί να

διαφέρουν σημαντικά με βάση την επιλεγμένη τιμή ϵ . Για παράδειγμα, μια υψηλότερη τιμή ϵ μπορεί να συγχωνεύσει κατά λάθος αντικείμενα δεδομένων που θα έπρεπε να ανήκουν σε ξεχωριστές συστάδες, ενώ μια χαμηλότερη τιμή ϵ μπορεί να ταξινομήσει αντικείμενα δεδομένων ως θόρυβο ενώ θα έπρεπε να ανήκουν σε μια συστάδα. Επιπλέον, η αντιμετώπιση του ϵ ως μιας παγκόσμιας παραμέτρου πυκνότητας περιορίζει την ικανότητα του αλγορίθμου να ανιχνεύει συστάδες με διαφορετικές τοπικές πυκνότητες. Πολλά σύνολα δεδομένων του πραγματικού κόσμου παρουσιάζουν αντικείμενα δεδομένων ομαδοποιημένα σε περιοχές με διαφορετικά επίπεδα πυκνότητας. Η χρήση μιας παγκόσμιας τιμής ϵ από τον **DBSCAN** οδηγεί στο πρόβλημα είτε της συγχώνευσης πολλαπλών συστάδων σε μία (με μεγάλο ϵ) είτε της εσφαλμένης ταξινόμησης αντικειμένων δεδομένων σε περιοχές με χαμηλότερη πυκνότητα (με μικρό ϵ).

Η αποτελεσματικότητα της διαδικασίας **EXPANDCLUSTER** στο πλαίσιο του αλγορίθμου **DBSCAN** είναι μια άλλη πτυχή που αξίζει να εξεταστεί. Αυτή η διαδικασία συλλέγει αντικείμενα με δυνατότητα προσέγγισης της πυκνότητας από ένα αντικείμενο πυρήνα για να σχηματίσει μια συστάδα. Ωστόσο, περιλαμβάνει την εκτέλεση της λειτουργίας **REGIONQUERY** σε κάθε αντικείμενο εντός της *ε-γειτονιάς* του αντικειμένου πυρήνα. Η λειτουργία αυτή συμβάλλει σημαντικά στην υπολογιστική πολυπλοκότητα του αλγορίθμου, καθώς περιλαμβάνει την επανάληψη όλων των

αντικειμένων δεδομένων στο σύνολο δεδομένων. Κατά συνέπεια, η συνολική πολυπλοκότητα του αλγορίθμου γίνεται $O(n^2)$ όταν λαμβάνονται υπόψη όλες οι εκτελέσεις **REGIONQUERY** για κάθε σπόρο κατά τη διάρκεια της εκτέλεσης του αλγορίθμου. Για να μετριαστεί αυτή η πολυπλοκότητα, μπορούν να χρησιμοποιηθούν δομές χωρικών ευρετηρίων όπως το **R*-tree**, οι οποίες απαιτούν $O(\log n)$ για ερωτήματα περιοχής. Αυτή η προσέγγιση επιτρέπει στον **DBSCAN** να βρίσκει συστάδες σε ένα σύνολο δεδομένων με πολυπλοκότητα $O(n \log n)$. Ο αρχικός αλγόριθμος **DBSCAN** υπερέχει στον εντοπισμό συστάδων με αυθαίρετα σχήματα, ένα κρίσιμο χαρακτηριστικό, ιδίως σε χωρικές βάσεις δεδομένων όπου οι συστάδες μπορεί να έχουν ακανόνιστα σχήματα. Σε αντίθεση με τις μεθόδους μερικής συσταδοποίησης όπως ο **k-means**, οι οποίες υποθέτουν σφαιρικά σχήματα συστάδων, ο **DBSCAN** προσαρμόζεται στα εγγενή σχήματα συστάδων του συνόλου δεδομένων. Επιπλέον, η **DBSCAN** αναγνωρίζει αποτελεσματικά απομονωμένα αντικείμενα δεδομένων και τα χαρακτηρίζει ως θόρυβο. Αντίθετα, οι μέθοδοι συσταδοποίησης με μερική κατανομή δεν διαθέτουν μηχανισμούς που να λαμβάνουν υπόψη τους την παρουσία αντικειμένων θορύβου, καθώς απλώς διαιρούν το χώρο δεδομένων σε έναν προκαθορισμένο αριθμό συστάδων. Παρά τα πλεονεκτήματα αυτά, η **DBSCAN** έχει περιορισμούς όσον αφορά το χειρισμό συνόλων δεδομένων με δομές πολλαπλής πυκνότητας και υψηλής διαστατικότητας. Οι βάσεις δεδομένων του πραγματικού κόσμου συχνά παρουσιάζουν ποικίλες πυκνότητες σημείων δεδομένων και χαρακτηριστικά υψηλής διάστασης, θέτοντας προκλήσεις για τον αρχικό αλγόριθμο. Για την αντιμετώπιση αυτών των ζητημάτων, έχουν προταθεί διάφοροι αλγόριθμοι ομαδοποίησης με βάση τον **DBSCAN**.

1. **Σύνολα δεδομένων πολλαπλής πυκνότητας:** Μια πρόκληση που αντιμετωπίζει ο **DBSCAN** είναι ο εντοπισμός συστάδων με διαφορετικές πυκνότητες σε ένα σύνολο δεδομένων. Δεδομένου ότι το ϵ χρησιμεύει ως συνολική παράμετρος, μπορεί να ταξινομήσει μια ομάδα αντικειμένων σε μια περιοχή υψηλής πυκνότητας ως θόρυβο. Για να αντιμετωπιστεί αυτό, εισήχθησαν αλγόριθμοι όπως οι **OPTICS** και **HDBSCAN**. Ο **OPTICS** παράγει μια διάταξη συστάδων που αντιστοιχεί σε διαφορετικές τιμές παραμέτρων, επιτρέποντας μεγαλύτερη ευελιξία. Εισάγει τις έννοιες της απόστασης πυρήνα και της απόστασης προσπελασιμότητας, οι οποίες χρησιμοποιούνται για τη δημιουργία ενός διαγράμματος προσπελασιμότητας. Το **HDBSCAN** βασίζεται στο **OPTICS**, χρησιμοποιώντας το γράφημα αμοιβαίας προσπελασιμότητας για τον αποτελεσματικότερο εντοπισμό συστάδων. Η **VDBSCAN**, μια άλλη μέθοδος, χρησιμοποιεί το γράφημα **k-dist** για τον δυναμικό προσδιορισμό των τιμών ϵ και την εύρεση συστάδων με διαφορετικές πυκνότητες.

2. **Ποιότητα των συστάδων:** Η **DBSCAN** υπερέχει στο χειρισμό συστάδων ακανόνιστου σχήματος, αλλά αντιμετωπίζει προκλήσεις σε σύνολα δεδομένων με ποικίλες πυκνότητες σημείων δεδομένων και υψηλές διαστάσεις. Για να ξεπεραστούν αυτές οι προκλήσεις, έχουν προταθεί αλγόριθμοι όπως ο αναθεωρημένος **DBSCAN**. Αυτή η αναθεωρημένη έκδοση επικεντρώνεται σε αλυσίδες αντικειμένων με δυνατότητα προσέγγισης της πυκνότητας για τον αποτελεσματικό εντοπισμό οριακών αντικειμένων. Με την εισαγωγή αντικειμένων προσπελάσιμων από την πυκνότητα πυρήνα, αντιμετωπίζει το ζήτημα της παραγωγής διαφορετικών δομών συστάδων με πολλαπλές επαναλήψεις του **DBSCAN**.

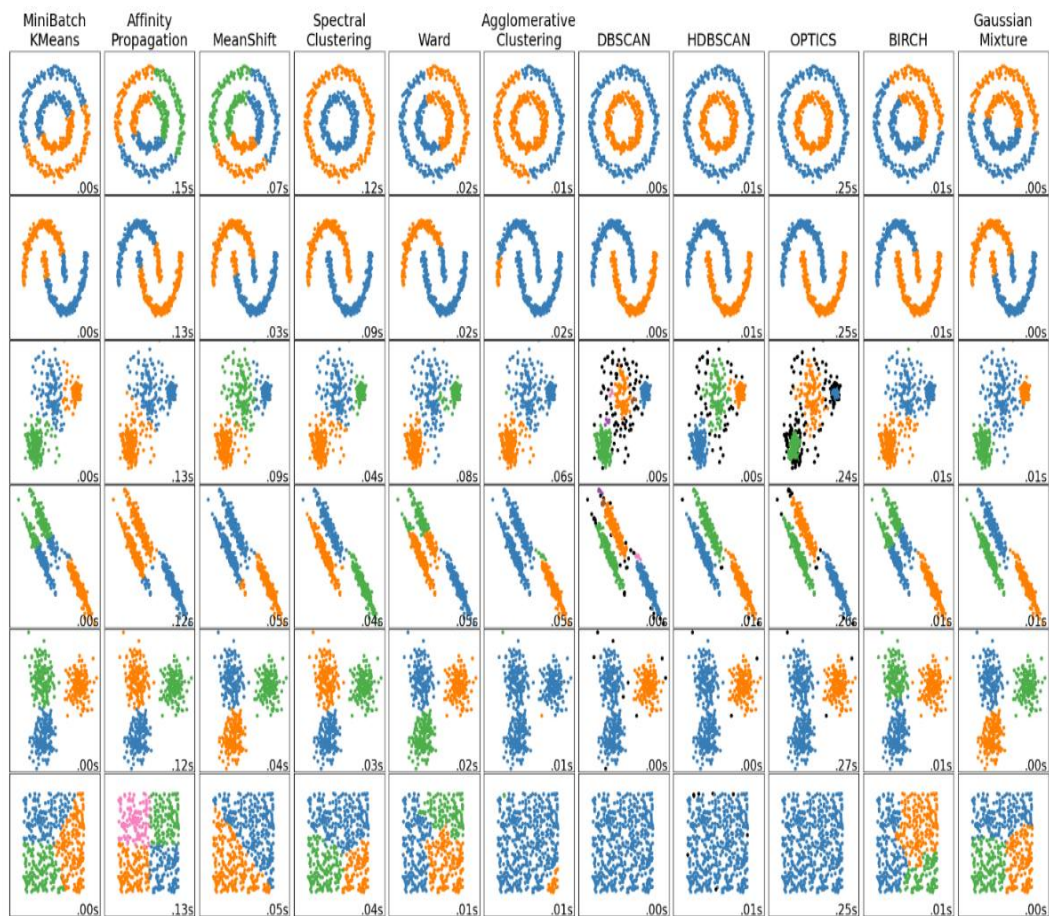
Αυτοί οι αλγόριθμοι συσταδοποίησης με βάση το **DBSCAN** προσφέρουν βελτιώσεις σε σχέση με τον αρχικό **DBSCAN** αντιμετωπίζοντας τους περιορισμούς του, καθιστώντας τους πολύτιμα εργαλεία για διάφορες εργασίες ανάλυσης δεδομένων.

Υψηλής διάστασης σύνολα δεδομένων: Τα σύνολα δεδομένων υψηλής διάστασης αποτελούν σημαντική πρόκληση για τους παραδοσιακούς αλγορίθμους ομαδοποίησης, συμπεριλαμβανομένου του αρχικού **DBSCAN**. Αυτή η πρόκληση, γνωστή ως "κατάρρα της διάστασης", προκύπτει όταν τα σύνολα δεδομένων έχουν σημαντικές συστάδες ενσωματωμένες βαθιά μέσα στον χώρο των χαρακτηριστικών τους. Δυστυχώς, αλγόριθμοι όπως ο **DBSCAN**, οι οποίοι εξετάζουν ολόκληρο τον πολυδιάστατο χώρο, συχνά δυσκολεύονται να εντοπίσουν αυτές τις συστάδες. Το πρόβλημα αυτό επιδεινώνεται από τη χρήση μέτρων απόστασης μεταξύ αντικειμένων δεδομένων για τη μέτρηση της ομοιότητας, καθώς άσχετες διαστάσεις μπορούν να αποκρύψουν σημαντικές συστάδες, καθιστώντας δύσκολο για τον **DBSCAN** να τις εντοπίσει. Για την αντιμετώπιση αυτού του προβλήματος, διάφορες μέθοδοι στη βιβλιογραφία έχουν προτείνει τεχνικές μείωσης της διαστατικότητας. Οι τεχνικές αυτές αποσκοπούν στο φιλτράρισμα των άσχετων ή περιττών χαρακτηριστικών για τη μείωση της διαστατικότητας του χώρου δεδομένων και, κατά συνέπεια, τη βελτίωση της απόδοσης της συσταδοποίησης. Ορισμένες μέθοδοι περιλαμβάνουν μετασχηματισμό χαρακτηριστικών, όπως η ανάλυση κύριων συνιστωσών (**PCA**), η οποία παράγει νέες συναρτήσεις χαρακτηριστικών. Ωστόσο, αυτές οι μέθοδοι έχουν τις δικές τους προκλήσεις, όπως δυσκολίες στην ερμηνεία των παραγόμενων χαρακτηριστικών και περιορισμένη επιτυχία σε ορισμένες περιπτώσεις.

Οι τεχνικές ομαδοποίησης υποδιαστημάτων έχουν αναδειχθεί ως μια πολλά υποσχόμενη λύση σε αυτές τις προκλήσεις. Οι μέθοδοι αυτές ανιχνεύουν αυτόματα συστάδες σε πολλαπλούς υποχώρους του αρχικού χώρου δεδομένων, επιτρέποντας μια πιο εστιασμένη και αποτελεσματική διαδικασία συσταδοποίησης. Μια τέτοια προσέγγιση, γνωστή ως **Density-Connected Subspace Clustering (SUBCLU)**, αξιοποιεί τεχνικές συσταδοποίησης υποχωρών για συσταδοποίηση με βάση την πυκνότητα. Η **SUBCLU** προσδιορίζει υποσύνολα διαστάσεων (υποχώρους) στα οποία υπάρχουν συστάδες και εφαρμόζει την έννοια της μονοτονίας των συνόλων που συνδέονται με την πυκνότητα για τον εντοπισμό αυτών των υποχωρών. Η ιδέα πίσω από τη μονοτονία είναι ότι εάν δύο αντικείμενα συνδέονται με πυκνότητα σε έναν υποχώρο, θα συνδέονται επίσης με πυκνότητα σε οποιονδήποτε υποχώρο που περιέχει τον αρχικό υποχώρο.

Αποδοτικότητα: Η αποδοτικότητα του αρχικού αλγορίθμου **DBSCAN** μπορεί να βελτιωθεί με την ελαχιστοποίηση του αριθμού των κλήσεων της λειτουργίας **REGIONQUERY** ή με τη βελτιστοποίηση της ίδιας της λειτουργίας. Μια προσέγγιση είναι ο γρήγορος αλγόριθμος ομαδοποίησης με βάση την πυκνότητα για μεγάλες βάσεις δεδομένων (**Fast Density-Based Clustering Algorithm for Large Databases - FDBSCAN**), ο οποίος μειώνει τον πλεονασμό στις κλήσεις **REGIONQUERY**. Κατά τη διάρκεια της φάσης **EXPANDCLUSTER** του **DBSCAN**, ο **FDBSCAN** επιλέγει αντιπροσωπευτικά αντικείμενα εντός μιας περιοχής για να αποφύγει επαναλαμβανόμενες κλήσεις **REGIONQUERY** σε αντικείμενα γειτονικών περιοχών που τέμνονται στις ε-γειτονίες των αντικειμένων του πυρήνα. Η προσέγγιση αυτή εξασφαλίζει ότι το **REGIONQUERY** δεν εκτελείται πολλές φορές στο ίδιο αντικείμενο, βελτιώνοντας σημαντικά την αποδοτικότητα.

Ένας άλλος αλγόριθμος, ο **Rough-DBSCAN**, χρησιμοποιεί μια υβριδική προσέγγιση για τη μείωση της χρονικής πολυπλοκότητας του **DBSCAN**. Αρχικά εντοπίζει αντικείμενα μοντέλου που ονομάζονται **leaders** χρησιμοποιώντας τη μέθοδο ομαδοποίησης **leaders**, η οποία απαιτεί χρόνο **O(n)**. Αυτά τα αντικείμενα-ηγέτες περιέχουν πληροφορίες πυκνότητας και στη συνέχεια χρησιμοποιούνται για τη δημιουργία συστάδων μέσω της συσταδοποίησης με βάση την πυκνότητα, μειώνοντας το συνολικό χρόνο υπολογισμού σε **O(n)** σε σύγκριση με τον αρχικό **DBSCAN**. Αυτές οι προσεγγίσεις αντιμετωπίζουν ορισμένες από τις προκλήσεις αποδοτικότητας και επεκτασιμότητας που αντιμετωπίζει ο αλγόριθμος **DBSCAN** όταν ασχολείται με σύνολα δεδομένων υψηλών διαστάσεων, καθιστώντας τις πολύτιμα εργαλεία για την ομαδοποίηση σε τέτοια σενάρια.



Σχήμα 9: Αναπαράσταση Τεχνικών Συσταδοποίησης σε διάφορους τύπους δεδομένων

Μαθηματικός Ορισμός της Πυκνότητας

Ο DBSCAN ορίζει την πυκνότητα μέσω της ε-γειτονιάς ενός σημείου. Για ένα σύνολο δεδομένων $X = \{x_1, x_2, \dots, x_n\}$, η ε-γειτονιά ενός σημείου x_i δίνεται από:

$$N_{\epsilon}(x_i) = \{x_j \in X \mid \text{dist}(x_i, x_j) \leq \epsilon\}$$

Ένα σημείο χαρακτηρίζεται ως **πυρήνας** (core point) αν η ε-γειτονιά του περιέχει τουλάχιστον **minPts** σημεία, δηλαδή $|N_{\epsilon}(x_i)| \geq \text{minPts}$.

Ταξινόμηση Σημείων

Τα σημεία του συνόλου δεδομένων ταξινομούνται σε τρεις κατηγορίες:

- **Σημεία πυρήνα:** Αποτελούν το κέντρο μιας συστάδας.
- **Σημεία ορίου:** Βρίσκονται στα άκρα μιας συστάδας και έχουν λιγότερους από **minPts** γείτονες, αλλά ανήκουν στην ε-γειτονιά ενός πυρήνα.
- **Θόρυβος:** Σημεία που δεν ανήκουν σε καμία συστάδα.

Σχηματισμός Συστάδων

Μια συστάδα ορίζεται ως ένα μέγιστο σύνολο σημείων που είναι **πυκνά συνδεδεμένα**. Δύο σημεία x_i και x_j θεωρούνται συνδεδεμένα αν υπάρχει μια ακολουθία πυρήνων που τα συνδέει. Μαθηματικά:

$$\text{Connect}(x_i, x_j) \Leftrightarrow \exists o \in X: \text{Reach}(o, x_i) \wedge \text{Reach}(o, x_j)$$

Χαρακτηριστικό	DBSCAN	k-means	Ιεραρχική
Σχήμα συστάδων	Αυθαίρετο	Σφαιρικό	Εξαρτάται
Θόρυβος	Ανθεκτικός	Ευαίσθητος	Ευαίσθητος
Παράμετροι	ϵ , minPts	k	Μέτρο ομοιότητας
Πολυπλοκότητα	$O(n \log n)$	$O(nkI)$	$O(n^3)$

Clustering?	MethodFamily	Sub. Abbr	SubFamily Approaches
Clustering	k-Partition (KPart) : Based on the distance of data points, assign them into k exclusive partitions or groups, where each partition represents a cluster.	KM	K-means: Each data point is assigned into the cluster according to the distance with the centroid of a cluster, which is the mean of all points within the cluster. The initial number of clusters is assigned manually.
		XM	X-means: An extended K-Means which tries to automatically determine the number of clusters.
		KMD	K-medoids: A clustering algorithm related to the k-means, while using an actual point in the cluster as centroid to clustering.
	Hierarchical (HC) : Grouping data into a hierarchy of clusters.	HC	Hierarchical clustering: Can be either agglomerative (bottom-up) or divisive (top-down).
	Density (Density) : Discover non-spherical clusters by finding core objects with dense neighborhoods.	DBC	Density based clustering: Connects core objects and their neighborhoods to form dense regions as clusters.
	Neural Network (NN) : Clustering based on Neural Network.	NGas	Neural-Gas clustering: A competitive learning technique with SoftMax learning rule.
		SOMs	Self-Organizing Maps: Produces a low-dimensional representation of the input space.
	Fuzzy Logic (Fuzzy) : Each data point can belong to two or more clusters in fuzzy clustering.	FCM	Fuzzy C-Means: Assigns membership to each data point corresponding to each cluster center on the basis of distance between the center and itself.
		FSubC	Fuzzy subtractive clustering: Uses fuzzy inferences based on rules generated by subtractive clustering.
		FSOMs	Fuzzy Self-Organizing Maps: Combines SOMs with the concept of fuzziness in fuzzy clustering.
	Spectral Clustering (SC) : Make use of the eigenvalues of the similarity matrix to reduce dimensionality before clustering.	SC	Partitions a dataset based on the connectivity between its nodes in a graph with spectral clustering.
	Expectation maximization clustering (EM) : Get probabilities of cluster memberships based on probability distributions.	EM	EM clustering: Finds the maximum likelihood by iteratively alternating between expectation step and maximization step.
	Optimization (Opt) : Taking clustering as an optimization problem, such as apply Particle Swarm Optimization to solve.	PSO	Particle Swarm Optimization clustering: Find the optimal clusters by iteratively improving a candidate solution with regard to a given measure.
	Affinity Propagation (AP) : Finds exemplars, members of the input set that are representative of clusters.	AP	Affinity Propagation clustering: Obtains the data similarities, and iteratively exchanges the real-valued messages between data points to clustering.
	Self Learning (SL) : Clustering by comparing each metric value with the corresponding median or average value.	CLA	Clustering and labelling: Compute instances violation degree based on the relation between metric values and corresponding median values, then clustering by the violation or consistency information.
		CLAMI	Based on CLA, add two extra steps: Metric selection and instance selection based on violation degree.
		ACL	Average Clustering and labelling: Clustering by metrics of instances violation score (MIVS), which is computed based on the relation between each metric value and their average metric value.
	Clustering Ensemble (CLEM) : Ensemble multiple clustering solutions.	CLEM	Clustering Ensemble: Uses multiple algorithms to clustering, then combines the different clustering solutions and produces a single cluster.
No Clustering	Threshold (THD) ¹ : Threshold for selected metric. It might be determined by experiences, rules, calculations, or machine learning, etc.	THD	Threshold: Use some metric thresholds to classify instances directly. e.g., given a metric threshold vector, if any metric of an instance exceeds the corresponding threshold, it will be defect-prone.
	Expert (EXP) : Determine by experience.	EXP	Expert determine the labels of each instance directly by their experience.
	Ranking (MR) : Metric Ranking	MR	Metric Ranking: For some change metrics in just-in-time prediction, using 1/Metric to rank instances in ascending order.

Σχήμα 10: Τεχνικές Συσταδοποίηση

Γ' Πειραματική διάταξη

M6 JETSTREAM

Το **M6-JETSTREAM** της **Bruker** είναι ένα σαρωτής φασματοσκοπίας φθορισμού ακτίνων-X (**XRF scanner**) που επιτρέπει τη ποιοτική και ποσοτική χωρική στοιχειακή ανάλυση δειγμάτων, με χωρική διακριτική ικανότητα της τάξης των **100 μm ($\mu\text{-XRF}$, *micro-XRF*)**). Επιπλέον έχει τη δυνατότητα σάρωσης μεγάλων επιφανειών (**80×60 cm^2**) επιτρέποντας την παραγωγή στοιχειακών χαρτών των εξεταζόμενων αντικειμένων. Έχει κατασκευαστεί για τη μη καταστροφική, μη επεμβατική στοιχειακή ανάλυση της κατανομής των στοιχείων σε αντικείμενα πολιτιστικής κληρονομιάς, για την ανάλυση γεωλογικών δειγμάτων, για τον ποιοτικό έλεγχο βιομηχανικών προϊόντων και γενικότερα για τη μελέτη αντικειμένων τα οποία παρουσιάζουν χωρική ανομοιογένεια, στην κλίμακα των **100 μm** . Η απεικονιστική φασματοσκοπία φθορισμού ακτίνων-X είναι ιδανική για μη ομογενή δείγματα, απαιτείται ελάχιστη ή καθόλου προετοιμασία του δείγματος, επιτρέπει την εξαγωγή στοιχειακών χαρτών και τον προσδιορισμό συσχετίσεων μεταξύ των στοιχείων.

4.1.1 Χαρακτηριστικά του φασματόμετρου M6 JETSTREAM

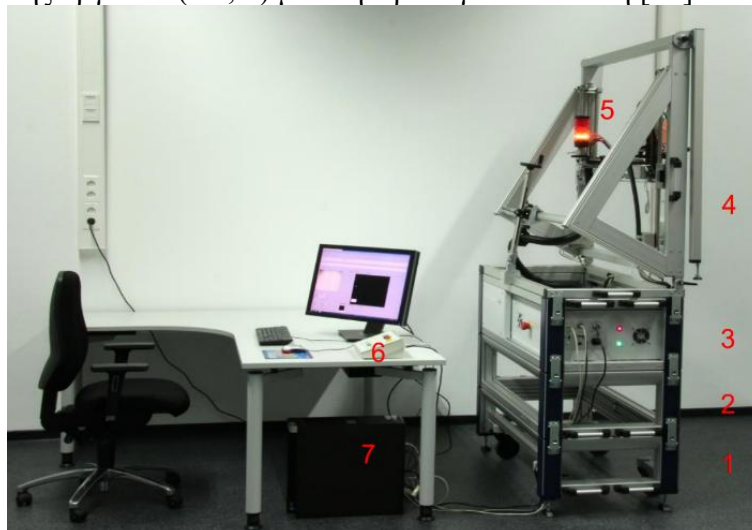
Το **M6 JETSTREAM** έχει τα ακόλουθα χαρακτηριστικά:

- Ανοιχτός σχεδιασμός με κινούμενη κεφαλή μέτρησης για να ξεπεραστεί η ανάγκη μετακίνησης του δείγματος
- Αρθρωτή σχεδίαση για εύκολη μεταφορά και επιτόπια εγκατάσταση εντός **30 λεπτών**
- Σάρωση σε όρθια ή κατακόρυφη θέση, ανάλογα με το υπό εξέταση αντικείμενο
- Περιοχή σάρωσης **800 mm x 600 mm ($W \times H$ σε όρθια θέση, $W \times D$ σε κεκλιμένη θέση)**
- Διέγερση μικρών περιοχών δειγμάτων έως και **100 μm** λόγω της πολυκατευθυντικής οπτικής ακτίνων X
- Ακριβής τοποθέτηση του δείγματος με τη βοήθεια του γρήγορου και ακριβούς σταδίου **X-Y-Z** που ελέγχεται από λογισμικό και ποντίκι
- Έλεγχος της θέσης του δείγματος από δύο συστήματα βίντεο με διαφορετική μεγέθυνση (**10x** και **100x**)
- Υψηλή ικανότητα ρυθμού καταμέτρησης και φασματοσκοπική ανάλυση λόγω της τελευταίας τεχνολογίας ανιχνευτών **SDD (XFlash®)**
- Έλεγχος του οργάνου μέσω γραφικής διεπαφής χρήστη (**GUI**) για εύκολο χειρισμό με λειτουργίες ελέγχου για του οργάνου, όπως:
 - Συνθήκες διέγερσης (παράμετροι λυχνίας, φίλτρα, χρόνος μέτρησης)
 - Ανιχνευτής ενεργειακού διαχωρισμού
 - Στάδιο ελέγχου θέσης
 - Κάμερα (μεγέθυνση, φωτισμός του θαλάμου δείγματος)
 - Μέτρηση των φασμάτων κατανομής των στοιχείων
 - Εμφάνιση και αξιολόγηση δεδομένων (φάσματα, κατανομές στοιχείων)
- Κατάλληλο για ποικίλους τύπους δειγμάτων και εφαρμογές (με έμφαση σε επίπεδα δείγματα)
- Χαμηλό λειτουργικό κόστος [12]

4.1.2 Δομή του συστήματος M6 JETSTREAM

Το M6 JETSTREAM αποτελείται από τα ακόλουθα εξαρτήματα του συστήματος (Εικόνα 18):

- Κάτω μέρος οργάνου (1) με 4 τροχούς με φρένα και 4 ρυθμιζόμενα βάθρα για τη στήριξη των του οργάνου
- Διαχωριστικό μπλοκ (2) απαραίτητο για οριζόντια λειτουργία εργασίας, μπορεί να χρησιμοποιηθεί για ρύθμιση ύψους σε κατακόρυφη λειτουργία
- Μέσο τμήμα (3) του οργάνου με ηλεκτρονική μονάδα και διεπαφές για έλεγχο μέσω υπολογιστή, με σερβομηχανισμούς για την κίνηση όλης της κεφαλής, προειδοποιητική λυχνία, κεφαλή μέτρησης και τροφοδοσία ρεύματος
- Ανώτερο τμήμα οργάνου (4) με κινηματική, διαδρομή για την κεφαλή μέτρησης (5) και μονάδα κλίσης. Η κεφαλή μέτρησης περιέχει:
 - Πηγή ακτινοβολίας ακτίνων X
 - Πολυτριχοειδής φακούς ακτίνων X
 - Ανιχνευτή ολίσθησης πυριτίου (SDD)
 - 2 κάμερες CCD, η μία με ρυθμιζόμενη εστιακή απόσταση
 - Φωτισμός δείγματος
 - Αισθητήρες σύγκρουσης υπερήχων
- Ακροδέκτης (6) για τον τηλεχειρισμό των λειτουργιών ασφαλείας
- Υπολογιστής οργάνου (PC, 7) με πληκτρολόγιο και οθόνη [12]



Εικόνα 7 - Εξαρτήματα συστήματος M6 JETSTREAM της εταιρείας Bruker [12].

4.1.3 Μέρη του φασματομέτρου

4.1.3.1 Κινήσεις στους άξονες $-x$, $-y$, $-z$

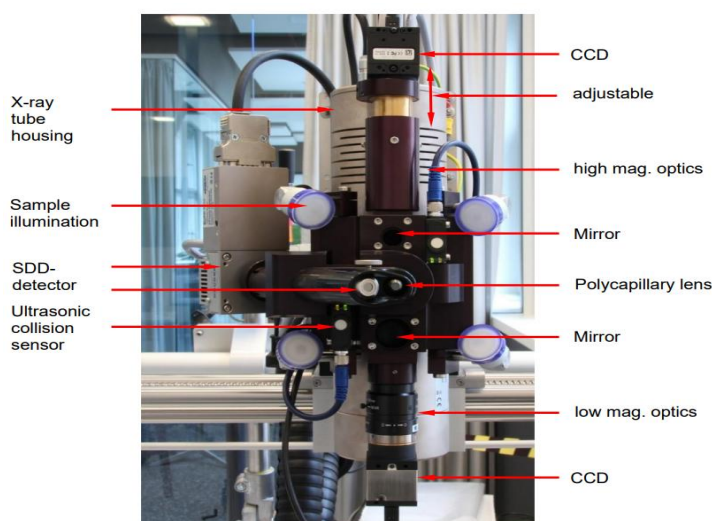
Η μονάδα σάρωσης x - y - z του **M6 JETSTREAM** είναι μια σερβοκινηματική 4 αξόνων, που χρησιμοποιεί δύο ηλεκτρονικά συγχρονισμένους άξονες- y , και έχει τη δυνατότητα για την ανύψωση του βάρους της κεφαλής μέτρησης κατά τον άξονα $-z$. Όλες οι κινήσεις συντονίζονται από τον σερβοελεγκτή 4 αξόνων [12].

Πίνακας 1 - Παράμετροι της κινηματικής-xyz [12].

Παράμετροι	Αξία
Κατευθύνσεις σάρωσης	xyz
Αριθμός μηχανοκίνητων αξόνων	4
Αρχή	Σφαιρικοί κοχλίες που κινούνται κωδικοποιητή που ελέγχεται από κινητήρες συνεχούς ρεύματος με ιμάντα 2:1
Βήμα σφαιρικού κοχλία	x/y: 2 mm, z: 1 mm
Εύρος μετακίνησης	x: 800 mm, y: 600 mm, z: 95 mm
Μέγιστη ταχύτητα	x/y: 100 mm/s, z: 20 mm/s
Διακόπτες αναφοράς	Δύο οπτικοί διακόπτες για κάθε άξονα
Φρένα	Μαγνητικά φρένα για τους άξονες y και z

4.1.3.2 Κεφαλή μέτρησης

Η κεφαλή μέτρησης (Εικόνα 9) είναι το κεντρικό λειτουργικό στοιχείο του M6 JETSTREAM. Τοποθετείται στη διάταξη κίνησης και περιέχει την πηγή ακτίνων-X, τον ανιχνευτή, το σύστημα κάμερας, το φωτισμό του δείγματος και τους αισθητήρες σύγκρουσης υπερήχων.



Στον

Πίνακα 3,

Εικόνα 8: Λεπτομερής απεικόνιση της κεφαλής μέτρησης του συστήματος M6 JETSTREAM. Διακρίνονται ο σωλήνας ακτίνων-X (X-ray tube housing), ο ανιχνευτής SDD, το σύστημα οπτικής απεικόνισης (Sample illumination) και η πολυτριχοειδής οπτική (Polycapillary lens)

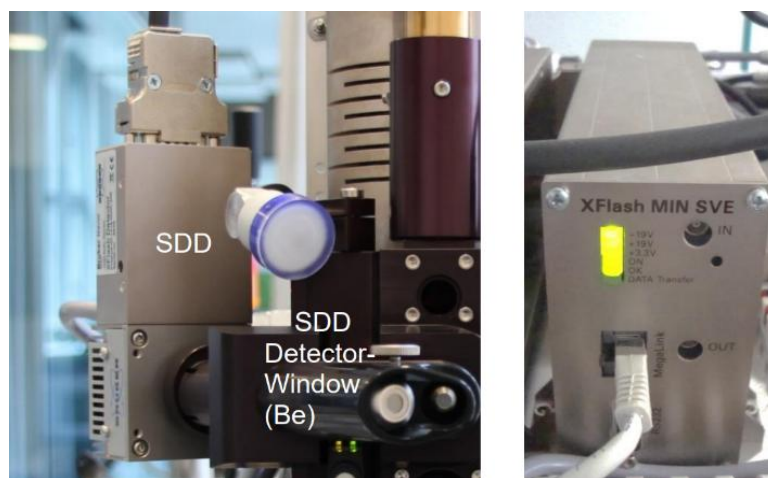
παρουσιάζονται όλες οι βασικές λειτουργίες των διάφορων τμημάτων της κεφαλής μέτρησης του φασματομέτρου [12].

Πίνακας 2 - Τμήματα πηγής ακτίνων-X [12].

Μονάδα	Λειτουργία
Περίβλημα σωλήνα	Μηχανική ενσωμάτωση όλων των εξαρτημάτων, δρομολόγηση του αέρα ψύξης, θωράκιση των ακτίνων-X
Σωλήνας ακτίνων-X	Πηγή ακτινοβολίας ακτίνων-X
Σωλήνας και στοιχείο ψύξης	Πρωταρχική θωράκιση, απαγωγή θερμότητας, τοποθέτηση του σωλήνα
Μονάδα κλείστρου	Απελευθέρωση της δέσμης διέγερσης κατά τη διάρκεια της μέτρησης
Τροχός φίλτρου	Τροχός με 5 φίλτρα και μία κενή θέση για τροποποίηση του φάσματος διέγερσης
Ελεγκτής περιβλήματος σωλήνα	Ρύθμιση θερμοκρασίας, εξοικονόμηση λειτουργικών παραμέτρων
Ανεμιστήρες	Ψύξη του σωλήνα ακτίνων-X
Μηχανική διεπαφή	Προσάρτηση και τοποθέτηση της πηγής ακτίνων-X
Προειδοποιητικές λάμπες	Ένδειξη της κατάστασης λειτουργίας
Πολύκατευθυντικά οπτικά συστήματα ακτίνων-X	Καθορισμός της γεωμετρίας δέσμης

4.1.3.3 Ανιχνευτής ακτίνων-X

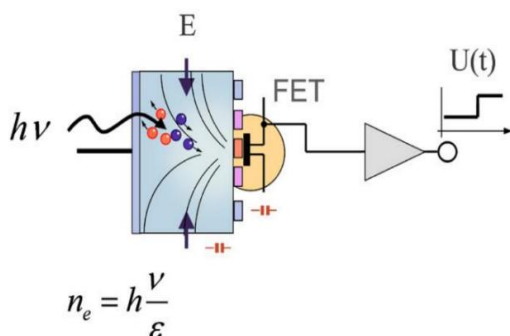
Το **M6 JETSTREAM** χρησιμοποιεί έναν ανιχνευτή πυριτίου με ενεργειακή διασπορά (**SDD detector**) και ενεργό επιφάνεια **30 mm²** για την ανίχνευση της ακτινοβολίας φθορισμού. Αυτός ο ανιχνευτής προσφέρει ικανοποιητική ικανότητα ρυθμού καταμέτρησης σε συνδυασμό με πολύ σταθερή ενεργειακή ανάλυση. Η ηλεκτρονική μονάδα **MIN SVE** είναι η τροφοδοσία ρεύματος και ο φασματοσκοπικός ενισχυτής για τον **SDD**. Δεδομένα φάσματος μεταφέρονται μέσω σύνδεσης **Ethernet** από το **MIN SVE** στον ελεγκτή επικοινωνίας και από το από εκεί στον **H/Y**. Η χαρακτηριστική ακτινοβολία ακτίνων X που εκπέμπεται από το δείγμα περνάει σε έναν ανιχνευτή πυριτίου (Εικόνα 20).



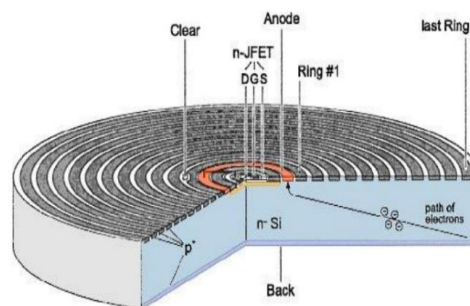
Εικόνα 9 - Ανιχνευτής ολίσθησης πυριτίου και MIN SVE (μέρος του ηλεκτρονικού κουτιού).

Κάθε κβάντο ακτίνων-X στον ενεργό όγκο του ανιχνευτή παράγει ένα νέφος φορέων φορτίου το οποίο τροφοδοτείται σε ένα ηλεκτρόδιο μέσω ενός εσωτερικού ηλεκτρικού πεδίου τάσης ολίσθησης, όπου δημιουργεί έναν παλμό (Εικόνες 4,5). Αυτός ο παλμός ενισχύεται από έναν προενισχυτή και διαβιβάζεται στον επεξεργαστή ψηφιακού σήματος ο οποίος πραγματοποιεί φιλτράρισμα του σήματος για τη μείωση του ηλεκτρονικού θορύβου και το διαβιβάζει στον *ADC* που κάνει ανάλυση του ύψους του παλμού που παράγει το φάσμα. Το *SDD* ψύχεται ηλεκτρικά για τη μείωση του ηλεκτρονικού θορύβου [12].

4.1.3.4 Σύστημα κάμερας



Εικόνα 11 - Τρόπος λειτουργίας ενός ανιχνευτή ακτίνων-X SDD.



Εικόνα 10 - Διατομή ενός ανιχνευτή ακτίνων-X SDD.

Το σύστημα κάμερας αποτελείται από δύο κάμερες οι οποίες είναι τοποθετημένες "πάνω" και "κάτω" από την κεφαλή της λυχνίας ακτίνων-X. Το οπτικό τους πεδίο καταγράφει διαφορετικές περιοχές του δείγματος. Η κάτω κάμερα δίνει σε ένα δείγμα επισκόπηση με μέγεθος περίπου $32 \times 21 \text{ mm}^2$. Για την κάμερα υψηλής μεγέθυνσης, η εστιακή απόσταση μπορεί να αλλάζει σε πέντε διαφορετικές θέσεις. Αυτό επιτρέπει την αλλαγή της απόστασης εργασίας (η απόσταση μεταξύ κεφαλής μέτρησης και του δείγματος) και κατ' επέκταση (λόγω της αποκλίνουσας δέσμης ακτίνων-X) το μέγεθος του ίχνους της δέσμης στην επιφάνεια του δείγματος. Η απόσταση εργασίας μπορεί να τροποποιηθεί σε πέντε βήματα με μέγεθος βήματος 3.2 mm αλλάζοντας το μέγεθος της περιοχής μελέτης μεταξύ $100 \mu\text{m}$ και $700 \mu\text{m}$ [12].

4.1.3.5 Ηλεκτρονική μονάδα με τερματικό

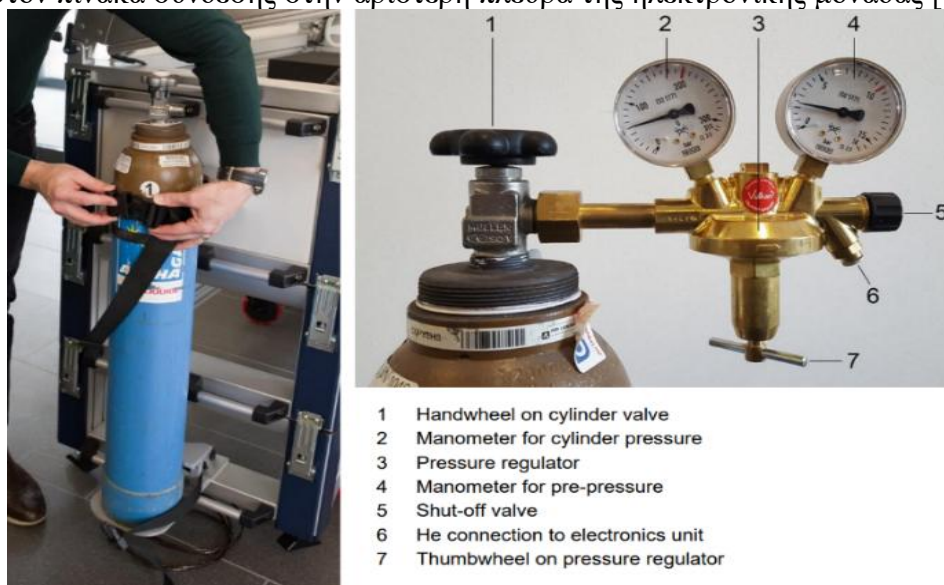
Το τερματικό είναι μια κονσόλα χειριστή με τις σημαντικότερες λειτουργίες ελέγχου και απεικόνισης. Η ενεργοποίηση της συσκευής και το άνοιγμα του κλείστρου είναι δυνατά μόνο από το τερματικό [12].

4.1.3.6 Μονάδα ελέγχου

Το σύνολο της συσκευής ελέγχεται από μια μονάδα ελέγχου (= H/Y οργάνου). Η μονάδα ελέγχου απαιτεί τουλάχιστον **4 GB (8 GB για λειτουργικό σύστημα 64-bit) RAM** και μεγάλο σκληρό δίσκο (**1TB**). Ως λειτουργικό σύστημα χρησιμοποιούνται τα **Windows 10**. Η μονάδα ελέγχου συνδέεται με το **M6 JETSTREAM** μέσω **USB** και Ethernet και με τις περιφερειακές μονάδες οθόνη, πληκτρολόγιο και ποντίκι. Η μονάδα ελέγχου ελέγχει τις παραμέτρους του οργάνου, τη μέτρηση και την αξιολόγηση των δεδομένων με τη χρήση του πακέτου λογισμικού **M6 JETSTREAM** [12].

4.1.3.7 Ροή ηλίου (He)

Το **M6 JETSTREAM** διαθέτει σύστημα ροής ηλίου. Εάν τα δείγματα μετρούνται υπό αέρα, οι χαρακτηριστικές ακτίνες X από ελαφρά στοιχεία θα απορροφηθούν σημαντικά από το κύριο συστατικά του αέρα (N_2 και O_2). Ο χρήστης είναι υπεύθυνος για τη διασφάλιση της παροχής αερίου ηλίου. Να σημειωθεί ότι το ήλιο με καθαρότητα **98%** είναι επαρκές για την ανίχνευση στοιχείων με $Z \geq 11$ για το **M6 JETSTREAM**. Η Εικόνα 23, δείχνει μια τυπική διαμόρφωση με χρήση φιάλης συμπιεσμένου αερίου **He**. Είναι ωφέλιμο να χρησιμοποιούνται δύο ξεχωριστά μανόμετρα για την ένδειξη της πίεσης της φιάλης και της προπίεσης στη σύνδεση της γραμμής τροφοδοσίας. Η προπίεση πρέπει να κυμαίνεται μεταξύ **2.5** και **3.0 bar**. Ο ρυθμιστής ροής για τον έλεγχο και τη ρύθμιση της ροής αερίου βρίσκεται στο εσωτερικό του οργάνου. Ένας εύκαμπτος σωλήνας χρησιμοποιείται για τη σύνδεση της παροχής αερίου στην "**είσοδο He**" στον πίνακα σύνδεσης στην αριστερή πλευρά της ηλεκτρονικής μονάδας [12].



Εικόνα 12 - Φιάλη αερίου ηλίου με τα εξαρτήματα σύνδεσης για μετρήσεις σε ροή He.

Δ' ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

Υλικά, Μέθοδοι και Αναλυτική Ροή Εργασίας

Αντικείμενα Μελέτης: Ιστορικό και Τεχνοτροπικό Πλαίσιο

Οι ελληνικές εικόνες που δημιουργήθηκαν μεταξύ του **18ου** και **19ου** αιώνα αποτελούν συνειδητή διατήρηση της βυζαντινής και μεταβυζαντινής καλλιτεχνικής παράδοσης, παρά την ευρύτερη ευρωπαϊκή τάση προς την ελαιογραφία και τον νατουραλισμό. Τα εξεταζόμενα έργα, τα οποία περιλαμβάνουν την εικόνα της Παναγίας Οδηγήτριας, την εικόνα του Αγίου Ιωάννη του Προδρόμου με Ιεράρχη και μια εκκλησιαστική επιγραφή, φιλοτεχνήθηκαν με παραδοσιακές μεθόδους: αβγότεμπερα σε ξύλινες σανίδες με γύψο και επίστρωση με φύλλα χρυσού. Αυτή η προσήλωση στην τεχνική *maniera greca* υπογραμμίζει μια πολιτιστική αντίσταση, δίνοντας έμφαση στον θεολογικό συμβολισμό και τη λατρευτική λειτουργία έναντι της αισθητικής καινοτομίας. Η εικονογραφία τους, διατηρώντας συνθέσεις που ανάγονται στον βυζαντινό χώρο, αποκαλύπτει μια συνειδητή προσήλωση στα βυζαντινά πρότυπα, ακόμη και όταν καλλιτεχνικά κινήματα όπως ο Ρομαντισμός και ο Νεοκλασικισμός κυριαρχούσαν αλλού στην Ευρώπη.

Η τεχνολογική συνέχεια στις μεσαιωνικές πρακτικές, αν και διασφάλιζε την πνευματική αυθεντικότητα, σήμερα παρουσιάζει σημαντικές προκλήσεις στη συντήρηση, από την αποκόλληση των χρωμάτων έως τη μικροβιακή φθορά. Η αντιμετώπιση αυτών των φθορών απαιτεί προηγμένες μη-καταστροφικές τεχνικές ανάλυσης. Η φασματομετρία φθορισμού ακτίνων-X (XRF), και ειδικότερα η μακρο-σάρωση (**MA-XRF**), έχει καθιερωθεί ως θεμελιώδης αναλυτική τεχνική για τον χαρακτηρισμό πολιτιστικών αγαθών, καθώς επιτρέπει τη δημιουργία λεπτομερών χαρτών κατανομής στοιχείων. Ωστόσο, η τεχνική αυτή παράγει τεράστια σύνολα δεδομένων, τα οποία θέτουν σημαντικές προκλήσεις στην επεξεργασία και ερμηνεία.

Για την αντιμετώπιση αυτών των προκλήσεων, η παρούσα μελέτη εφάρμοσε μια πολυεπίπεδη χημειομετρική προσέγγιση. Αξιοποιήθηκαν τεχνικές **μείωσης διαστάσεων**, όπως η Ανάλυση Κύριων Συνιστωσών (**PCA**), η Μη Αρνητική Παραγοντοποίηση Πινάκων (**NMF**) και η **t-Stochastic Neighbor Embedding (t-SNE)**, σε συνδυασμό με αλγόριθμους **ομαδοποίησης (clustering)**, όπως οι k-means, **DBScan** και **Spectral Clustering**. Η επιλογή του βέλτιστου αριθμού συστάδων έγινε με τη χρήση του κριτηρίου silhouette, επιτρέποντας τον υπολογισμό αντιπροσωπευτικών φασμάτων για κάθε υλική ομάδα. Αυτές οι μέθοδοι διευκόλυναν την εξαγωγή σημαντικών μοτίβων από τα πολύπλοκα σύνολα δεδομένων, αποκαλύπτοντας πληροφορίες για τη σύσταση των υλικών και τη συντηρητική φύση των μεταβυζαντινών καλλιτεχνικών πρακτικών. Η συνδυασμένη αυτή ανάλυση παρέχει ένα ισχυρό πλαίσιο για τη βαθύτερη κατανόηση της υλικής κουλτούρας ιστορικών περιόδων.

Εικονογραφική και Θεολογική Σημασία

Οι επιλεγμένες εικόνες (Εικόνα 13) αντικατοπτρίζουν βασικά θέματα της Ορθόδοξης παράδοσης, αποδεικνύοντας πώς η παραδοσιακή εικονογραφία επιβίωσε ακόμη και κατά την οθωμανική περίοδο και τα πρώτα χρόνια του ελληνικού κράτους. Η *Παναγία Οδηγήτρια*, για παράδειγμα, διατηρεί μια σύνθεση που ανάγεται στον βυζαντινό χώρο, τονίζοντας τον ρόλο της Θεοτόκου ως μεσιτεύτριας. Η σκηνή με τον *Άγιο Γεώργιο* συνθέτει εσχολογικές και διδακτικές εικόνες, ενισχύοντας τη δογματική διδασκαλία της Εκκλησίας. Αυτές οι επιλογές αποκαλύπτουν μια συνειδητή προσήλωση στα

βυζαντινά πρότυπα, ακόμη και όταν καλλιτεχνικά κινήματα όπως ο Ρομαντισμός και ο Νεοκλασικισμός κυριαρχούσαν αλλού στην Ευρώπη.



Εικόνα 13: Οι τρεις θρησκευτικές φορητές εικόνες που αποτέλεσαν αντικείμενο της πειραματικής μελέτης. (a) Η Παναγία η Οδηγήτρια, (b) Ο Άγιος Ιωάννης ο Πρόδρομος (σε κατάσταση φθοράς), και (c) η πίσω μεριά της εικόνας (b)

Υλικά και Τεχνικές: Μια Μεσαιωνική Κληρονομιά

Τεχνολογικά, αυτές οι εικόνες ακολουθούν πιστά μεσαιωνικές πρακτικές. Η χρήση της *αβγότεμπερας*—ενός μέσου που εκτιμάται για την ανθεκτικότητα και τη ματ λάμψη του— τις διαχωρίζει από τις δυτικές ελαιογραφίες, οι οποίες προτιμούσαν τον εκλεπτυσμό και τον φωτοσκιασμό. Οι ξύλινες σανίδες, επιστρωμένες με γύψο, παρείχαν σταθερό υπόστρωμα, αν και η ευαισθησία τους σε περιβαλλοντικές ζημιές είναι εμφανής στην κατεστραμμένη εικόνα του *Αγίου Ιωάννη του Προδρόμου*. Τα φύλλα χρυσού, εφαρμοσμένα με την τεχνική της υδατοχρύσωσης, όχι μόνο μετέδιδαν το θεϊκό φως αλλά επιβεβαίωναν και την ιερότητα των εικόνων. Τέτοιες μέθοδοι, που μεταβιβάζονταν μέσω εργαστηρίων (συχνά σε μοναστηριακό πλαίσιο), διασφάλιζαν τεχνική συνέχεια αλλά περιόριζαν την καινοτομία. Αυτός ο συντηρητισμός στα υλικά υπογραμμίζει ένα ευρύτερο πολιτιστικό σκοπό: τη διατήρηση της πνευματικής αυθεντικότητας έναντι της καλλιτεχνικής πρωτοτυπίας.

Ιστορικό Πλαίσιο και Προκλήσεις Συντήρησης

Ο **18ος** και **19ος** αιώνας ήταν περίοδοι αναστάτωσης για την ελληνική τέχνη, μεταξύ οθωμανικής κυριαρχίας και του νεοσύστατου ελληνικού κράτους. Ενώ δυτικές επιρροές άρχισαν να διεισδύουν σε αστικά κέντρα, τα επαρχιακά και μοναστηριακά εργαστήρια αντιστάθηκαν, διατηρώντας τη *maniera greca*. Η επιμονή των βυζαντινών συμβάσεων σε αυτές τις εικόνες μπορεί να αντικατοπτρίζει ένα ευρύτερο αφήγημα ταυτότητας, ιδιαίτερα μετά την Επανάσταση του **1821**, όταν η βυζαντινή κληρονομιά χρησιμοποιήθηκε ως σύμβολο ελληνικής διαφορετικότητας. Σήμερα, αυτά τα έργα παρουσιάζουν σημαντικές προκλήσεις στη συντήρηση, από την αποκόλληση των

χρωμάτων έως τη μικροβιακή φθορά, απαιτώντας τεχνικές όπως η πολυφασματική απεικόνιση και η φασματομετρία ακτίνων X για την αποκατάστασή τους.

Η φασματομετρία φθορισμού ακτίνων X (*XRF*) έχει καθιερωθεί ως θεμελιώδης αναλυτική τεχνική στην επιστήμη της πολιτιστικής κληρονομιάς [*XRF-12, Van Grieken and Markowicz, 2001; XRF-27, Shackley, 2018*], χάρη στην ικανότητά της να παρέχει γρήγορη και μη καταστρεπτική στοιχειακή χαρακτηρισμό αντικειμένων. Η *μακρο-XRF (MA-XRF)* σάρωση, ειδικότερα, επιτρέπει τη δημιουργία λεπτομερών χαρτών κατανομής στοιχείων που αποκαλύπτουν κρίσιμες πληροφορίες σχετικά με τα υλικά, τις τεχνικές και το ιστορικό συντήρησης ενός έργου τέχνης [*XRF-10, Mazzinghi et al., 2021; XRF-15, Gerodimos et al., 2022*]. Ωστόσο, η τεχνική παράγει εξαιρετικά μεγάλα σύνολα δεδομένων, συχνά εκατομμυρίων φασμάτων, τα οποία θέτουν σημαντικές προκλήσεις στην επεξεργασία και ερμηνεία των δεδομένων.

Για την αντιμετώπιση αυτών των προκλήσεων, απαιτούνται εξελιγμένες μέθοδοι ανάλυσης δεδομένων. Η παρούσα μελέτη εφάρμοσε δύο κύριες προσεγγίσεις: ανάλυση συσταδοποίησης και μείωση διαστάσεων. Για τη συσταδοποίηση, χρησιμοποιήθηκε πληθώρα αλγορίθμων όπως *k-means, DBScan, Hierarchical Clustering, Spectral Clustering* [*KMeans-5, Kanungo et al., 2002*]. Βεβαία λόγω της υπολογιστικής του αποδοτικότητας και της αποτελεσματικότητάς του στην ομαδοποίηση παρόμοιων φασμάτων βάσει Ευκλείδειας απόστασης προτιμήθηκε κυρίως ο *k-means*. Κρίσιμη πτυχή αυτής της διαδικασίας ήταν ο προσδιορισμός του βέλτιστου αριθμού συστάδων, ο οποίος επιτεύχθηκε χρησιμοποιώντας το κριτήριο *silhouette* για την αξιολόγηση της ποιότητας της συσταδοποίησης. Αυτή η προσέγγιση επέτρεψε τον υπολογισμό αντιπροσωπευτικών φασμάτων για κάθε συστάδα, παρέχοντας ουσιαστικές συνοψίσεις των φασματικών δεδομένων.

Οι τεχνικές μείωσης διαστάσεων ήταν εξίσου κρίσιμες για τη διαχείριση της υψηλής διάστασης των δεδομένων *MA-XRF* [*PCA-1, Huang et al., 2019*]. Η ανάλυση εφάρμοσε αρχικά την Ανάλυση Κύριων Συνιστωσών (*PCA*), μια γραμμική μέθοδο προβολής που μειώνει αποτελεσματικά τις διαστάσεις των δεδομένων διατηρώντας ταυτόχρονα τις βασικές πληροφορίες [*PCA-2, Shlens, 2014; PCA-3, Abdi and Williams, 2010*]. Στη συνέχεια, χρησιμοποιήθηκε η μη γραμμική μέθοδος *t-Stochastic Neighbor Embedding (t-SNE)* για περαιτέρω μείωση των δεδομένων σε δύο διαστάσεις, επιτρέποντας μια πιο ενστικτώδη απεικόνιση των φασματικών σχέσεων. Επιπρόσθετα μέθοδοι όπως *NMF* φάνηκαν αρκετά χρήσιμες και μπορούν να αναγνωστούν στο κατάλληλο κεφάλαιο. Μαζί, αυτές οι μέθοδοι διευκόλυναν την εξαγωγή σημαντικών μοτίβων από πολύπλοκα σύνολα δεδομένων.

Η πρακτική εφαρμογή αυτών των τεχνικών επιδείχθηκε μέσω της ανάλυσης μιας ελληνικής εικόνας του 18ου αιώνα που απεικονίζει την Παναγία Οδηγήτρια, αποδιδόμενη σε Καπεσοβίτες ζωγράφους. Αυτή η μελέτη περίπτωσης τόνισε τη συνέχεια των βυζαντινών καλλιτεχνικών παραδόσεων κατά τη μεταβυζαντινή περίοδο, ιδίως τη συνεχή χρήση αβγότεμπερας σε γυψοστρωμένες σανίδες παρά την αυξανόμενη επικράτηση της ελαιογραφίας στη Δυτική Ευρώπη. Η τεχνική ανάλυση αποκάλυψε όχι μόνο τη σύσταση των υλικών της εικόνας, αλλά και σημαντικές πληροφορίες σχετικά με τη συντηρητική φύση των μεταβυζαντινών καλλιτεχνικών πρακτικών, οι οποίες επηρεάζονταν έντονα από τα εγχειρίδια τεχνιτών και τις μεσαιωνικές παραδόσεις [*XRF-15, Gerodimos et al., 2022; NMF-10, Łach et al., 2021*].

Η σύζευξη της **MA-XRF** με προηγμένες μεθόδους ανάλυσης δεδομένων προσφέρει δυνατότητες για την έρευνα της πολιτιστικής κληρονομιάς. Συνδυάζοντας τη χαρτογράφηση στοιχείων με τεχνικές μηχανικής μάθησης, οι ερευνητές δεν περιορίζονται πλέον στην απλή χαρτογράφηση στοιχείων αλλά μπορούν να αποκαλύψουν λεπτομερείς πληροφορίες σχετικά με καλλιτεχνικά υλικά, τεχνικές και ιστορία συντήρησης. Επιπλέον, αυτή η προσέγγιση παρέχει ένα εργαλείο για την κατανόηση της υλικής κουλτούρας ιστορικών περιόδων, ιδίως σε πλαίσια όπου οι παραδοσιακές πρακτικές επιβίωσαν παρά την ευρύτερη καλλιτεχνική εξέλιξη. Η μελέτη μεταβυζαντινών εικόνων, για παράδειγμα, ωφελείται σημαντικά από τέτοιες αναλύσεις, καθώς αποκαλύπτουν τόσο τις τεχνικές πτυχές της παραγωγής εικόνων όσο και την πολιτιστική ανθεκτικότητα των καλλιτεχνικών παραδόσεων υπό την οθωμανική κυριαρχία.

Αυτή η μεθοδολογία αντιπροσωπεύει ένα ισχυρό πλαίσιο για τη μη επεμβατική έρευνα αντικειμένων πολιτιστικής κληρονομιάς, συνδυάζοντας προηγμένες αναλυτικές τεχνικές με καλλιτεχνική και ιστορική έρευνα. Μελλοντικές μελέτες θα μπορούσαν να επεκτείνουν αυτές τις μεθόδους ενσωματώνοντας πρόσθετες τεχνικές απεικόνισης ή βελτιώνοντας τους αλγορίθμους μηχανικής μάθησης για να ενισχύσουν την ερμηνεία των δεδομένων. Η συνεχής ανάπτυξη τέτοιων προσεγγίσεων θα συμβάλει αναμφίβολα σε μια βαθύτερη κατανόηση των καλλιτεχνικών πρακτικών και της υλικής κουλτούρας σε διαφορετικές ιστορικές περιόδους και γεωγραφικές περιοχές.

Έχοντας ολοκληρώσει μια εισαγωγή στις βασικές έννοιες και απαραίτητες γνώσεις θα προχωρήσουμε στο βασικό κομμάτι της εργασίας που είναι η πραγματοποίηση του πειράματος υπό ορισμένες συνθήκες.

Στο τρέχον πείραμα η εικόνα που θα χρησιμοποιηθεί είναι αυτή που απεικονίζεται παραπάνω (**a**), με σκοπό την ανάλυση των χρωμάτων και των υλικών της εικόνας μέσω της μεθόδου Μακρο Φασματομετρίας Ακτίνων X (**mar**).

Η φασματομετρική ανάλυση πραγματοποιήθηκε με το σύστημα **μ -XRF M1-Mistral (Bruker)**, εξοπλισμένο με σωλήνα **ακτίνων X ροδίου Rh υψηλής απόδοσης**. Το σύστημα διαθέτει παχύ γυάλινο παράθυρο (**2 mm**) και παράγει συνεχές φάσμα διέγερσης με ενεργειακό όριο **25 keV**. Η χωρική ανάλυση ρυθμίζεται μέσω της αλλαγής της εστιακής απόστασης που καθορίζει το ίχνος της δέσμης των ακτίνων X πάνω στο δείγμα (συγκεντρωτών δέσμης), οι οποίοι προσδιορίζουν την εστιακή κηλίδα στο δείγμα, με δυνατότητα επιλογής διαφόρων διαμέτρων ανάλογα με τις απαιτήσεις ανάλυσης.

Το σύστημα ενσωματώνει μηχανοκίνητο τριγωνικό τραπέζι θέσης (**X-Y-Z**) με εύρος κίνησης **180 mm** στους οριζόντιους άξονες και ελάχιστο βήμα **10 μ m**. Ο **Z-άξονας** επιτρέπει την αυτόματη προσαρμογή της απόστασης δείγματος-ανιχνευτή, διασφαλίζοντας βέλτιστες συνθήκες μέτρησης ακόμη και για ανομοιομορφες επιφάνειες. Αυτή η λειτουργία υποστηρίζεται από οπτικό μικροσκόπιο **30x** με σύστημα τεσσάρων **LED** φωτισμού, το οποίο επιτρέπει την ακριβή εστίαση και την οπτική παρακολούθηση του δείγματος κατά τη διάρκεια των μετρήσεων.

Για την ανίχνευση των χαρακτηριστικών ακτίνων X χρησιμοποιήθηκε ανιχνευτής πυριτίου (**SDD**) με ενεργή επιφάνεια **30 mm²** και πάχος **450 μ m**. Η ενεργειακή ανάλυση του συστήματος περιγράφεται από τη σχέση **$fwhm(eV) = 2.47 \cdot E(eV) + 4400$** , όπου το **fwhm** (πλήρες εύρος στο μισό μέγιστο) εκφράζει την ικανότητα του συστήματος να

διαχωρίζει φασματικές γραμμές με κοντινές ενέργειες. Η γραμμική αυτή σχέση δείχνει ότι η ανάλυση χαλαρεί γιατί οι γκαουσιανές διευρύνονται και αυξάνει το FWHM με την αύξηση της ενέργειας, με τυπική τιμή **139 eV** για την γραμμή **Mn Ka (5.9 keV)**.

Η διαδικασία σκανάριατος εκτελέστηκε σε λειτουργία βήμα-προς-βήμα, καλύπτοντας επιφάνεια **64 × 40 mm²** με βήμα 1-3 mm και χρόνο μέτρησης **2 sec ανά σημείο** ανάλογα την πειραματική διαδικασία και εικόνα υπό εξέταση. Η δέσμη διέγερσης χαρακτηρίζονταν από **Gaussian** κατανομή με χωρικό εύρος **800 μm (fwhm)**. Οι συνθήκες λειτουργίας του σωλήνα ορίστηκαν σε **50 kV και 800 μA**, με συνολικό χρόνο συλλογής δεδομένων περίπου **1.5 ώρες**. Κατά τη διάρκεια της μέτρησης καταγράφηκαν **2.665 φάσματα** (ένα ανά σημείο μέτρησης), καθένα από τα οποία περιέχει πληροφορία για την κατανομή φωτονίων **σε 2.048 ενεργειακά κανάλια**.

Τα δεδομένα οργανώθηκαν και αποθηκεύτηκαν σε μορφή **X-ray cube (HDF)**, η οποία επιτρέπει την αποδοτική διαχείριση και ανάλυση του μεγάλου όγκου πληροφορίας. Η δομή αυτή περιλαμβάνει τρεις χωρικές διαστάσεις (**x, y θέση**) και μία ενεργειακή διάσταση, καθιστώντας δυνατή την ταυτόχρονη ανάλυση χωρικής κατανομής και φασματικών χαρακτηριστικών. Ως βασικό εργαλείο ανάλυσης χρησιμοποιήθηκε το φάσμα αθροιστικό φάσμα (**sum spectrum**), που προκύπτει από το άθροισμα όλων των μεμονωμένων φασμάτων και παρέχει μια γενική εικόνα της στοιχειακής σύστασης του δείγματος.

Η μεθοδολογία αυτή προσφέρει πολλαπλά πλεονεκτήματα για τη μελέτη πολιτιστικών αντικειμένων:

1. Μη καταστρεπτικός χαρακτήρας μέτρησης
2. Υψηλή χωρική ανάλυση (έως **30 μm**)
3. Ποσοτική και ποιοτική ανάλυση πολλαπλών στοιχείων ταυτόχρονα
4. Δυνατότητα ανίχνευσης τόσο κύριων όσο και ιχνοστοιχείων
5. Ολοκληρωμένη ανάλυση χωρικής κατανομής στοιχείων

Η εφαρμογή της τεχνικής αυτής σε έργα τέχνης και αρχαιολογικά ευρήματα επιτρέπει την αποκάλυψη κρίσιμων πληροφοριών σχετικά με:

- Τα υλικά και τις τεχνικές κατασκευής
- Την προέλευση των πρώτων υλών
- Ιστορικές επεμβάσεις και επισκευές
- Διαδικασίες γήρανσης και φθοράς

Η συνδυασμένη ανάλυση των χωρικών και ενεργειακών δεδομένων παρέχει μοναδικές δυνατότητες για την κατανόηση της τεχνολογίας κατασκευής και της ιστορικής εξέλιξης πολιτιστικών αντικειμένων, καθιστώντας τη φασματομετρία **μ-XRF** ένα ισχυρό εργαλείο στη σύγχρονη έρευνα της πολιτιστικής κληρονομιάς.

Η χαρτογράφηση της στοιχειακής κατανομής πραγματοποιήθηκε με το σύστημα σάρωσης **Bruker M6-Jetstream**. Το σύστημα είναι εξοπλισμένο με σωλήνα ακτίνων-X χαμηλής ισχύος (30W) με **άνοδο Ροδίου (Rh)** και πολυτριχοειδείς φακούς (polycapillary optics) για την εστίαση της δέσμης. Η ανίχνευση της φθορίζουσας

ακτινοβολίας γίνεται με ανιχνευτή ολίσθησης πυριτίου (SDD) με ενεργό επιφάνεια **30 mm²**. Δυνατότητα σάρωσης έως **80 × 60 cm²**.

- **Σωλήνα ακτίνων X Ροδίου 30 W**
- Τάση λειτουργίας **50 kV** και ένταση **600 μΑ**
- Πολυτριχοειδή οπτική για εστίαση της δέσμης
- Ανιχνευτή πυριτίου (**SDD**) **30 mm²** με ανάλυση **139 eV (Mn Kα)**
- Καταγραφή φασμάτων σε **4096 κανάλια**
- Πραγματοποιήθηκαν διακριτές σαρώσεις (σύνολα δεδομένων "Α" έως "C") με παραμέτρους που καταγράφονται στον Πίνακα παρακάτω. Κάθε σύνολο περιλαμβάνει χιλιάδες μεμονωμένα φάσματα XRF, επιτρέποντας τόσο χωρική ανάλυση όσο και μελέτη της μέσης σύστασης.
- Σύνθεση του υποβάθρου
- Όλες οι εικόνες παρουσίασαν έντονες εκπομπές **Ca Kα/Kβ** και **S Kα**, χαρακτηριστικές του γύψου (**CaSO₄·2H₂O**) στο **υποστρώμα**. Οι εκπομπές **Sr K** υποδεικνύουν την παρουσία σελεστίνης (**SrSO₄**) ως φυσικής προσμείξεως.

Επομένως ας παραθέσουμε μια σύνοψη για το πειραματικό πρωτόκολλο:

Πειραματικό Πρωτόκολλο

Πραγματοποιήθηκαν σαρώσεις υψηλής χωρικής ανάλυσης στις υπό μελέτη εικόνες για την απόκτηση των υπερφασματικών δεδομένων που αποτέλεσαν τη βάση της χημειομετρικής ανάλυσης. Οι συνθήκες σάρωσης ήταν οι εξής:

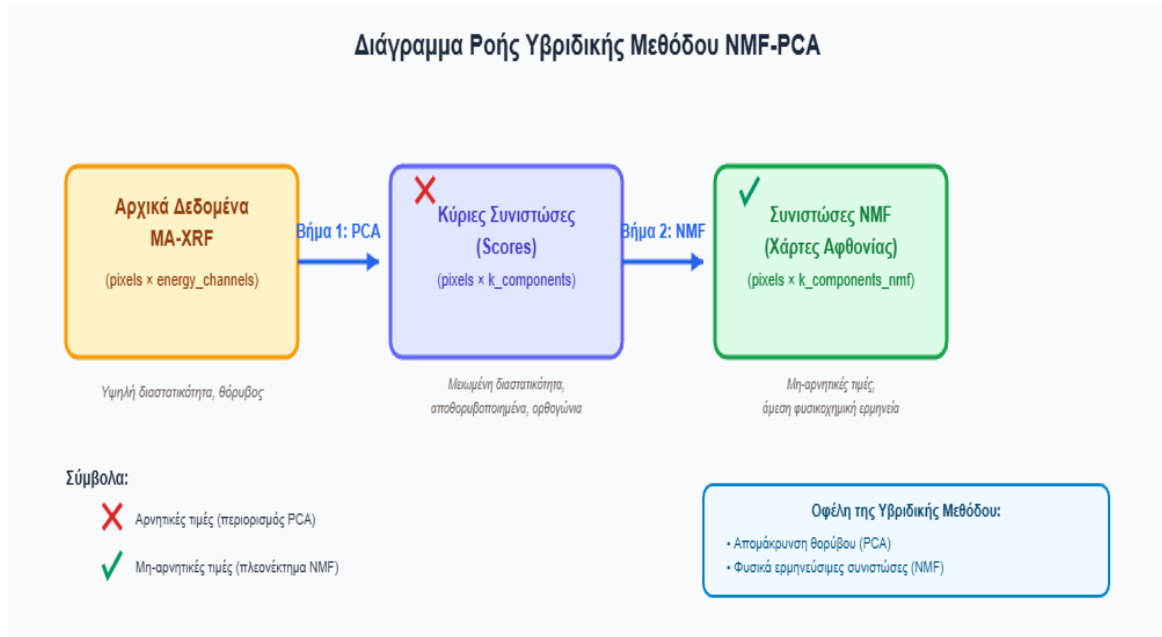
- Συνθήκες Λειτουργίας Σωλήνα: Τάση **50 kV**, Ένταση **600 μΑ**.
- Μέγεθος Δέσμης (**Spot Size**): Διάμετρος **100 μm**.
- Βήμα Σάρωσης (**Pixel Size**):
 - **200 μm** για την ανάλυση της ζωγραφικής επιφάνειας του Αγίου Ιωάννη.
 - **300 μm** για την ανάλυση της Παναγίας της Οδηγήτριας και των ξύλινων υποστηριγμάτων.
- Χρόνος Μέτρησης ανά **Pixel**:
 - **20 ms** για την ανάλυση στα **200 μm**.
 - **10 ms** για την ανάλυση στα **300 μm**.
- Ατμόσφαιρα: Αέρας (**Air, 1000 mbar**).
- Φίλτρο: Δεν χρησιμοποιήθηκε.

Οι παραπάνω ρυθμίσεις οδήγησαν στη συλλογή εκατοντάδων χιλιάδων (π.χ., **612,000** για τον Άγιο Ιωάννη) μεμονωμένων φασμάτων, τα οποία αποθηκεύτηκαν σε μορφή **HDF5** για περαιτέρω επεξεργασία.

Χημειομετρική Ανάλυση με το Πλαίσιο NICE

Η τεράστια ποσότητα δεδομένων (τάξης *gigabytes*) που παρήχθη από τις σαρώσεις *MA-XRF* έθεσε σημαντικές υπολογιστικές προκλήσεις. Για την αντιμετώπισή τους, η ανάλυση πραγματοποιήθηκε εξ ολοκλήρου με το εξειδικευμένο λογισμικό *NICE* που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας. Η αναλυτική ροή εργασίας Εικόνα 14 που ακολουθήθηκε για κάθε εικόνα ήταν η εξής:

1. Εισαγωγή και Επισκόπηση Δεδομένων: Τα αρχεία *HDF5* φορτώθηκαν μέσω του *NICE_Explorer_GUI.py*. Σε αυτό το στάδιο, επιλέχθηκε η επιθυμητή αριθμητική ακρίβεια (float16 για λόγους απόδοσης) και έγινε η αρχική επισκόπηση των μεταδεδομένων.
2. Προεπεξεργασία: Εντός του *AnalysisWindow*, υπολογίστηκε το φάσμα συνολικής έντασης (*sum spectrum*) και εφαρμόστηκε φίλτρο *Savitzky-Golay* για την απομάκρυνση του θορύβου (σε περιοχές που κρίθηκε απαραίτητο).
3. Μείωση Διαστάσεων: Εφαρμόστηκε η υβριδική μέθοδος *NMF-PCA*. Αρχικά, η Incremental *PCA* μείωσε τη διαστατικότητα των δεδομένων διατηρώντας πάνω από το 80% της πληροφορίας. Στη συνέχεια, η *NMF* εφαρμόστηκε στα αποτελέσματα της *PCA* για την εξαγωγή μη-αρνητικών συνιστωσών με άμεση φυσικοχημική ερμηνεία.
4. Ομαδοποίηση (*Clustering*): Τα αποτελέσματα της *NMF-PCA* χρησιμοποιήθηκαν ως είσοδος στον αλγόριθμο K-Means. Ο βέλτιστος αριθμός συστάδων ($k=6$ ή 7) προσδιορίστηκε με τη βοήθεια των διαγραμμάτων αξιολόγησης (*Elbow*, *Silhouette*, etc.).
5. Ιεραρχική Ανάλυση (*Sub-clustering*): Για περιοχές ιδιαίτερου ενδιαφέροντος (π.χ., φωτοστέφανα, ενδύματα), τα *pixels* που ανήκαν σε μια συγκεκριμένη αρχική συστάδα απομονώθηκαν και ο αλγόριθμος *K-Means* εφαρμόστηκε εκ νέου σε αυτό το υποσύνολο, αποκαλύπτοντας λεπτομερέστερη στρωματογραφία.
6. Στοιχειακή Χαρτογράφηση και Επικύρωση: Τέλος, δημιουργήθηκαν παραδοσιακοί χάρτες κατανομής στοιχείων (*Pb*, *Hg*, *Fe*, *Au*, κ.λπ.) για την επικύρωση των ταυτοποιήσεων που προέκυψαν από την ανάλυση ομαδοποίησης.



Εικόνα 14: Διάγραμμα ροής της υβριδικής μεθόδου NMF-PCA. Η PCA χρησιμοποιείται αρχικά για την απομάκρυνση του θορύβου και τη μείωση της διαστατικότητας. Στη συνέχεια, η NMF εφαρμόζεται στα αποθρομβωποιημένα δεδομένα της PCA, μετατρέποντας τις αφηρημένες, ορθογώνιες συνιστώσες σε μη-αρνητικούς, φυσικά ερμηνεύσιμους χάρτες αφθονίας υλικών.

Πρόγραμμα Ανάλυσης/ Αλγόριθμος

Η αποτελεσματική διερεύνηση πολυδιάστατων δεδομένων από τεχνικές απεικόνισης απαιτεί τη χρήση προηγμένων υπολογιστικών μεθόδων. Ωστόσο, η ad-hoc εφαρμογή τους εγκυμονεί κινδύνους για την εγκυρότητα και την επαναληψιμότητα των αποτελεσμάτων. Για την αντιμετώπιση αυτής της πρόκλησης, αναπτύχθηκε το **NICE** (*N-dimensional Interactive Computing Environment*), ένα ολοκληρωμένο, αρθρωτό (*modular*) πλαίσιο λογισμικού γραμμένο σε **Python**, το οποίο αυτοματοποιεί και τυποποιεί πολύπλοκες αναλυτικές ροές εργασίας.

Πρωταρχικός στόχος είναι η δυνατότητα επιλογής του εκάστοτε αρχείου δεδομένων μορφής **HDF5**. Τα αρχεία **HDF5** (*Hierarchical Data Format version 5*) χρησιμοποιούνται ευρέως για την αποτελεσματική οργάνωση και αποθήκευση των πολυδιάστατων δεδομένων που παράγονται από μετρήσεις **MA-XRF**. Η ιεραρχική δομή τους επιτρέπει την ταυτόχρονη αποθήκευση των πρωτογενών φασματικών δεδομένων, των χωρικών συντεταγμένων και των μεταδεδομένων σε ένα ενιαίο, αυτοσυννοδευόμενο αρχείο. Κάθε σάρωση **MA-XRF** αποθηκεύεται ως ξεχωριστή ομάδα (*group*) που περιέχει: **(1)** έναν τριδιάστατο πίνακα ($x_pixels \times y_pixels \times energy_channels$) με τα φασματικά δεδομένα, **(2)** έναν δισδιάστατο πίνακα με τις ακριβείς χωρικές συντεταγμένες κάθε σημείου μέτρησης, και **(3)** ένα σύνολο μεταδεδομένων (*attributes*) που περιγράφουν τις πειραματικές συνθήκες (τάση/ρεύμα σωλήνα, χρόνο μέτρησης ανά σημείο, γεωμετρία μέτρησης). Η μορφή υποστηρίζει αλγορίθμους συμπίεσης (π.χ. **gzip**) που μειώνουν σημαντικά το μέγεθος των αρχείων χωρίς απώλεια πληροφορίας, με τυπικές αναλογίες συμπίεσης **3:1** για φασματικά δεδομένα. Η δυνατότητα **chunking** επιτρέπει την αποδοτική ανάγνωση επιμέρους τμημάτων των δεδομένων χωρίς φόρτωση ολόκληρου του αρχείου στη μνήμη, ενώ η ενσωμάτωση μεταδεδομένων σε επίπεδο ομάδων και datasets διασφαλίζει την πλήρη αναπαραγωγιμότητα των αναλύσεων. Για την επεξεργασία των αρχείων αυτών, η

βιβλιοθήκη *h5py* σε συνδυασμό με το *NumPy* προσφέρουν ένα ισχυρό εργαλείο, επιτρέποντας λ.χ. την ταυτόχρονη ανάγνωση πολλαπλών σαρώσεων από το ίδιο αρχείο, ή την επιλογή συγκεκριμένων ενεργειακών παραθύρων για γρήγορη χαρτογράφηση στοιχείων.

Ένα τυπικό *MA-XRF* αρχείο $64 \times 40 \text{ mm}^2$ (1 mm βήμα, 4096 κανάλια) καταλαμβάνει $\sim 250 \text{ MB}$ με συμπίεση. Η διαχείριση αυτής της πολυπλοκότητας υλοποιείται μέσω του *NICE_hdf5_core.py*, το οποίο παρέχει μια αφαίρεση υψηλού επιπέδου για τις λειτουργίες αρχείων, όπως θα αναλυθεί παρακάτω.

Έπειτα η ανάγνωση των αρχείων γίνεται μέσω των καταλλήλων βιβλιοθηκών. Από την ανανέωση και έπειτα ο χρήστης θα πρέπει να βρίσκεται σε θέση να μπορεί να απεικονίσει τα δεδομένα αυτά τόσο σε έναν γράφο όπου θα απεικονίζεται το φάσμα της έντασης όσο και το φάσμα των ενεργειών και της ιδίας της εικόνας υπό εξέταση.

Στη συνέχεια, ο χρήστης θα μπορεί να επιλέξει να αναλύσει ολόκληρο το φάσμα ή ένα συγκεκριμένο μέρος αυτού, με οποιαδήποτε μέθοδο επιλέξει από τις εισαγόμενες στον αλγόριθμο. Για μείωση διαστάσεων *PCA*, *NMF*, *t-SNE* ενώ για ομαδοποίηση (*clustering*) έχουμε *k-means*, *DBScan*, *Hierarchical Clustering*. Με την δυνατότητα βέβαια προσθήκης περισσότερων μεθόδων εάν κριθεί απαραίτητο. Εφόσον επιλέξει από ένα μενού μορφής λίστας τις επιθυμητές μεθόδους θα πρέπει να συμπληρώσει τον αριθμό τόσο των διαστάσεων που επιθυμεί όσο και τον αριθμό των συστάδων.

Η ορθή επιλογή του αριθμού συστάδων μπορεί να βοηθηθεί μέσω της εισαγωγής εύρεσης καλύτερου αριθμού από το *silhouette score*, τον δείκτη *Calinski-Harabasz* και *Davies-Bouldin* με τα αντίστοιχα διαγράμματα τους και αποτελέσματα.

Δείκτης	Μαθηματικός Ορισμός	Πλεονεκτήματα	Όρια Εφαρμογής
Silhouette Score	Μετρά τη συνοχή (a) και τον διαχωρισμό (b) για κάθε δείγμα: $\frac{(b-a)}{\max(a,b)}$.	Ευαίσθητος στη διαχωριστικότητα και πυκνότητα συστάδων.	Υπολογιστικά απαιτητικός. Λιγότερο αποτελεσματικός σε μη-κυρτές συστάδες.
Calinski-Harabasz	Αναλογία της διασποράς μεταξύ συστάδων προς τη διασπορά εντός των συστάδων.	Υπολογιστικά αποδοτικός.	Τείνει να ευνοεί σφαιρικές συστάδες και μεγαλύτερο k.
Davies-Bouldin	Μέση "ομοιότητα" κάθε συστάδας με την πιο όμοιά της. Χαμηλότερες τιμές είναι καλύτερες.	Απαιτεί λιγότερους υπολογισμούς.	Προϋποθέτει κυρτές, ισοτροπικές συστάδες.

Επομένως έχοντας εξοπλίσει τον χρήστη με τα απαραίτητα αποτελέσματα και τις κατάλληλες παραμέτρους αφήνοντας όμως ελευθέρια στην επιλογή ο χρήστης μπορεί να προχωρήσει στην ανάλυση. Από την ανάλυση θα λάβει τα επιθυμητά αποτελέσματα

και Γράφους, τόσο για τις ενέργειες και τις συστάδες, όσο και για την εξάρτηση των κυρίων διαστάσεων. Τα αποτελέσματα αποθηκεύονται τόσο σε μορφή .csv όσο και στην ταχεία μνήμη **RAM** του υπολογιστή δίνοντας έτσι την δυνατότητα στον χρήστη να προσεγγίσει μια ακόμα ανάλυση στα δεδομένα αυτά. Με νέες αναλύσεις στις παραμέτρους και στις τιμές. Για μια ολοκληρωμένη και ευχάριστη εμπειρία του χρήστη ο αλγόριθμος κάνει χρήση γραφικών απεικονίσεων **GUI (Graphic User Interface)** όπου όλες οι αλληλεπιδράσεις του χρήστη θα γίνονται μέσω κουμπιών και επίλογων τιμών από λίστες και μέσω του ιδίου του συστήματος.

Ο τρόπος υλοποίησης του αλγορίθμου αλλά και ο ίδιος ο αλγόριθμος απεικονίζονται παρακάτω μέσω βημάτων επεξήγησης. Η γλώσσα γραφής του αλγορίθμου είναι η **Python**.

Ανάλυση Μεθοδολογίας και Αλγοριθμικής Υλοποίησης του Πλαισίου NICE

Το παρόν σύστημα λογισμικού, με την ονομασία **NICE (N-dimensional Interactive Computing Environment)**, αποτελεί ένα ολοκληρωμένο πλαίσιο για την επιστημονική ανάλυση δεδομένων, γραμμένο στη γλώσσα **Python**. Ο σχεδιασμός του είναι αρθρωτός (**modular**), κατανέμοντας τη λειτουργικότητα σε διακριτά σενάρια (**modules**) για τη βελτίωση της συντηρησιμότητας και της επεκτασιμότητας (Εικόνα 15). Η αρχιτεκτονική του επιτρέπει την αλληλεπίδραση του χρήστη μέσω ενός εξελεγχμένου γραφικού περιβάλλοντος (**GUI**), το οποίο ενοποιεί την πρόσβαση στα δεδομένα, την εκτέλεση πολύπλοκων αναλύσεων και την οπτικοποίηση των αποτελεσμάτων. Η μεθοδολογία αυτή εξοπλίζει τον ερευνητή με ισχυρά εργαλεία, παρέχοντας ταυτόχρονα την ευελιξία για την παραμετροποίηση της αναλυτικής διαδικασίας.

Το σύστημα έχει σχεδιαστεί για να διαχειρίζεται μεγάλα και πολύπλοκα σύνολα δεδομένων, όπως αυτά που προκύπτουν από φασματοσκοπικές τεχνικές απεικόνισης, και υποστηρίζει ολόκληρη τη ροή εργασίας: από την αρχική εξερεύνηση των δεδομένων έως την εξαγωγή τελικών συμπερασμάτων. Τα αποτελέσματα, συμπεριλαμβανομένων των αριθμητικών δεδομένων και των γραφικών παραστάσεων, μπορούν να αποθηκευτούν σε διάφορες μορφές (π.χ., .csv), ενώ τα ενδιαμέσως αποτελέσματα παραμένουν διαθέσιμα στην ταχεία μνήμη (RAM) για επαναληπτικές αναλύσεις. Αυτή η δυνατότητα επιτρέπει στον χρήστη να εφαρμόζει διαδοχικές αναλύσεις στα παραγόμενα δεδομένα (π.χ., ομαδοποίηση στα αποτελέσματα της μείωσης διαστάσεων), διευκολύνοντας μια βαθύτερη και πολυεπίπεδη διερεύνηση.

Τα βασικά σενάρια που συγκροτούν το πλαίσιο είναι:

- **GENERATE_XRF_LIBRARY.py**: Δημιουργία βιβλιοθήκης φασματικών αναφορών.
- **NICE_HDF5_CORE.py**: Θεμελιώδης διαχείριση και πρόσβαση σε αρχεία HDF5.
- **ANALYSIS_METHODS_4_NEW18.py**: Ο υπολογιστικός πυρήνας με τους αλγόριθμους ανάλυσης.
- **ANALYSIS_VISUALIZER_3_NEW18.py**: Εξειδικευμένη μονάδα για τη δημιουργία οπτικοποιήσεων.
- **NICE_Explorer_GUI_5_NEW18.py**: Η κύρια διεπαφή για την πλοήγηση στα αρχεία.

- **ANALYSIS_WINDOW_6_NEW18.py:** Το κεντρικό, ολοκληρωμένο περιβάλλον ανάλυσης.

Προκλήσεις και Αρχιτεκτονικές Λύσεις

Ο σχεδιασμός του πλαισίου NICE καθοδηγήθηκε από την ανάγκη αντιμετώπισης τριών θεμελιωδών προκλήσεων στην ανάλυση επιστημονικών δεδομένων απεικόνισης:

- **Διαχείριση Μεγάλου Όγκου Δεδομένων (Big Data):** Τα αρχεία MA-XRF μπορούν εύκολα να υπερβούν τη διαθέσιμη μνήμη RAM ενός τυπικού υπολογιστή. Η απλοϊκή φόρτωση ολόκληρου του αρχείου είναι ανέφικτη.

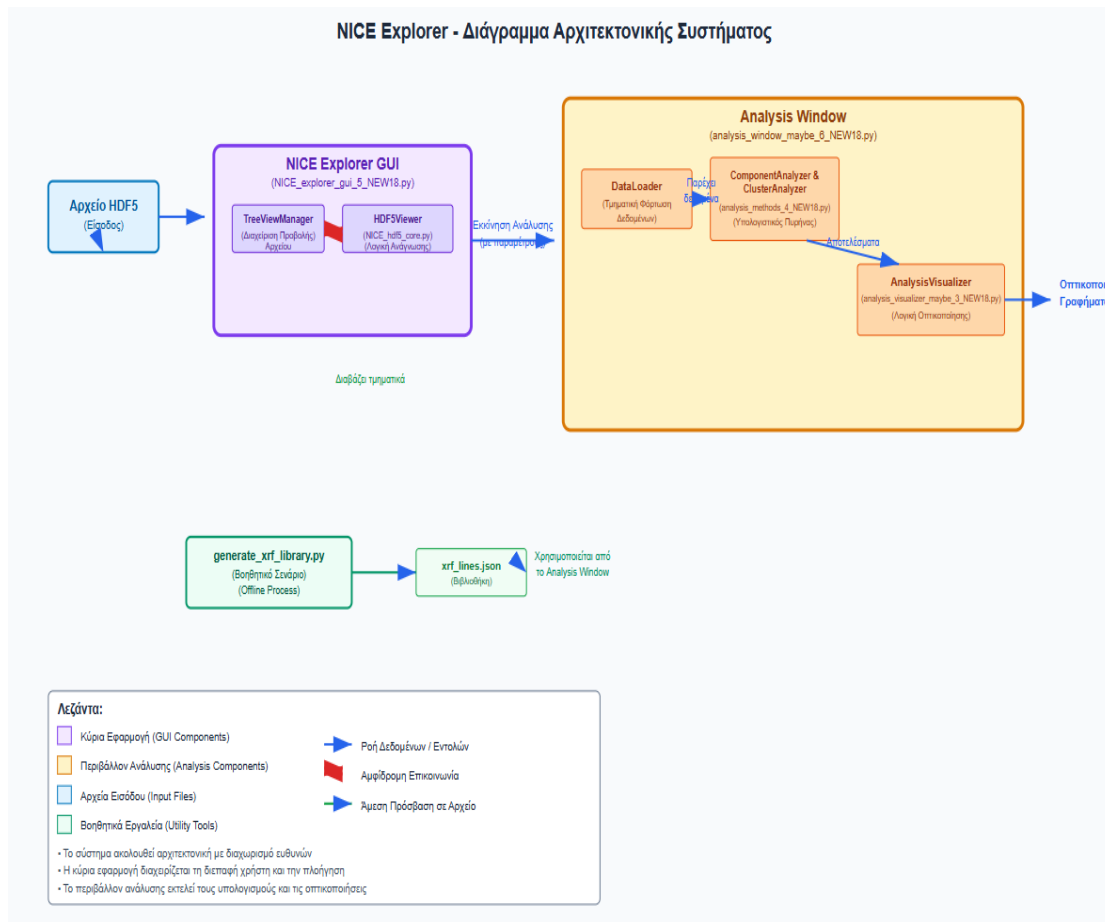
Λύση: Υιοθετήθηκε μια στρατηγική επεξεργασίας εκτός μνήμης (out-of-core processing). Η κλάση DataLoader στο ANALYSIS_WINDOW_6_NEW18.py δεν φορτώνει ποτέ ολόκληρο το dataset, αλλά το διαβάζει τμηματικά (chunk by chunk) απευθείας από τον δίσκο. Αλγόριθμοι όπως η IncrementalPCA στο analysis_methods_4_NEW18.py σχεδιάστηκαν ειδικά για να λειτουργούν με αυτή την τμηματική ροή δεδομένων.

- **Εξασφάλιση Επαναληψιμότητας:** Μια σύνθετη ανάλυση περιλαμβάνει πολλά βήματα, με διαφορετικές παραμέτρους. Η μη αυτόματη καταγραφή αυτής της διαδικασίας είναι επιρρεπής σε σφάλματα και καθιστά δύσκολη την επαλήθευση των αποτελεσμάτων από άλλους ερευνητές.

Λύση: Το σύστημα διατηρεί ένα πλήρες ιστορικό ανάλυσης (data_history). Κάθε ενέργεια του χρήστη, μαζί με τις παραμέτρους της και τα παραγόμενα αποτελέσματα, καταγράφεται σε μια δομή δεδομένων. Η λειτουργία "Save/Load Project" σειριοποιεί ολόκληρη αυτή την κατάσταση, συμπεριλαμβανομένων των εκπαιδευμένων μοντέλων, σε ένα ενιαίο αρχείο, εγγυώμενο την απόλυτη αναπαραγωγικότητα της ανάλυσης.

- **Διαδραστικότητα και Διερεύνηση σε Πραγματικό Χρόνο:** Οι αλγόριθμοι μηχανικής μάθησης συχνά λειτουργούν ως "μαύρα κουτιά". Ο ερευνητής χρειάζεται εργαλεία για να συνδέσει τα αφηρημένα αποτελέσματα (π.χ., συνιστώσες PCA) με τον πραγματικό, χωρικό τομέα των δεδομένων.

Λύση: Το GUI είναι πλούσιο σε διαδραστικά στοιχεία. Η επιλογή φασματικής Περιοχής Ενδιαφέροντος (ROI) με το SpanSelector και η λειτουργία "Plot Linking" (σύνδεση γραφημάτων), όπου η επιλογή σημείων σε ένα διάγραμμα διασποράς επισημαίνει τα αντίστοιχα pixels στην εικόνα, αποτελούν παραδείγματα εργαλείων που γεφυρώνουν το χάσμα μεταξύ ανάλυσης και ερμηνείας, μετατρέποντας το λογισμικό σε ένα πραγματικό περιβάλλον επιστημονικής ανακάλυψης.



Εικόνα 15: Διάγραμμα αρχιτεκτονικής και ροής δεδομένων του πλαισίου NICE. Η κύρια εφαρμογή (NICE Explorer GUI) διαχειρίζεται την πλοήγηση στα αρχεία HDF5 και εκκινεί το εξειδικευμένο AnalysisWindow. Το παράθυρο ανάλυσης υλοποιεί μια στρατηγική επεξεργασίας εκτός μνήμης, όπου ο DataLoader τροφοδοτεί τμηματικά δεδομένα στον υπολογιστικό πυρήνα, τα αποτελέσματα του οποίου οπτικοποιούνται μέσω του AnalysisVisualizer. Η προ-υπολογισμένη βιβλιοθήκη xrf_lines.json χρησιμοποιείται για την ταυτοποίηση στοιχείων

GENERATE_XRF_LIBRARY.py

Το σενάριο *generate_xrf_library.py* αποτελεί μια θεμελιώδη, προπαρασκευαστική συνιστώσα του αναλυτικού πλαισίου *NICE*, η οποία έχει σχεδιαστεί για να επιλύσει ένα κρίσιμο πρόβλημα αποδοτικότητας και διαχείρισης εξαρτήσεων. Ο πρωταρχικός του σκοπός είναι η δημιουργία μιας στατικής, προ-υπολογισμένης βιβλιοθήκης αναφοράς, η οποία περιέχει τις ενεργειακές τιμές (σε μονάδες *keV*) των χαρακτηριστικών γραμμών φθορισμού ακτίνων X (*XRF*) και των ακμών απορρόφησης για ένα εκτενές σύνολο χημικών στοιχείων. Η υιοθέτηση αυτής της μεθοδολογίας της προ-δημιουργίας αντί για δυναμικούς υπολογισμούς κατά το χρόνο εκτέλεσης βασίζεται σε μια στρατηγική απόφαση για την αποσύζευξη (decoupling) της κύριας εφαρμογής ανάλυσης από την εξειδικευμένη επιστημονική βιβλιοθήκη xraylib. Αυτή η αποσύζευξη παρέχει δύο σημαντικά πλεονεκτήματα: αφενός, απλοποιεί τη διαδικασία διανομής και εγκατάστασης του τελικού λογισμικού, καθώς η xraylib δεν αποτελεί πλέον άμεση εξάρτηση της διαδραστικής εφαρμογής για αυτή τη συγκεκριμένη λειτουργία, και αφετέρου, βελτιστοποιεί δραματικά την απόδοση, καθώς η ανάγνωση ενός τοπικού αρχείου *JSON* είναι υπολογιστικά ασήμαντη σε σύγκριση με τις επαναλαμβανόμενες κλήσεις σε μια εξωτερική βιβλιοθήκη κατά τη διάρκεια μιας διαδραστικής αναλυτικής συνεδρίας. Η αλγοριθμική υλοποίηση είναι συστηματική και ευέλικτη.

Το σενάριο ορίζει αρχικά δύο δομές δεδομένων τύπου λεξικού, τις *emission_lines_to_include* και *absorption_edges_to_include*, οι οποίες ως ένας διαμορφώσιμος κατάλογος.

```
1. # Απόσπασμα από generate_xrf_library.py:
2. # Παράδειγμα διαμορφώσιμου καταλόγου για τις γραμμές εκπομπής.
3.
4. emission_lines_to_include = {
5.     "Kα1": xraylib.KA1_LINE,
6.     "Kα2": xraylib.KA2_LINE,
7.     "Kβ1": xraylib.KB1_LINE,
8.     "Lα1": xraylib.LA1_LINE,
9.     # ... και άλλες επιθυμητές γραμμές ...
10. }
```

Αυτές οι δομές αντιστοιχίζουν συμβολοσειρές αναγνωρίσιμες από τον άνθρωπο, όπως *"Kα1"*, με τις αντίστοιχες αριθμητικές σταθερές που χρησιμοποιεί η *xraylib* για την αναγνώριση των ατομικών μεταπτώσεων. Αυτή η αρχιτεκτονική επιτρέπει στους ερευνητές να προσαρμόζουν με ευκολία το περιεχόμενο της παραγόμενης βιβλιοθήκης, προσθέτοντας ή αφαιρώντας γραμμές ανάλογα με τις ειδικές απαιτήσεις του πειράματός τους, χωρίς να χρειάζεται να τροποποιήσουν τον κύριο λογικό πυρήνα του σεναρίου.

Η διαδικασία εκτελείται μέσω ενός ένθετου βρόχου, ο οποίος διατρέχει συστηματικά όλους τους συνδυασμούς των προκαθορισμένων ατομικών αριθμών και των επιλεγμένων γραμμών, υποβάλλοντας ερωτήματα στις συναρτήσεις *xraylib.LineEnergy* και *xraylib.EdgeEnergy*.

Ένα κρίσιμο στοιχείο της μεθοδολογίας είναι η ενσωμάτωση ενός ισχυρού μηχανισμού χειρισμού σφαλμάτων, ο οποίος αντιμετωπίζει το αναμενόμενο πρόβλημα της ανυπαρξίας ορισμένων φασματικών γραμμών για συγκεκριμένα στοιχεία, ιδίως για εκείνα με χαμηλό ατομικό αριθμό. Η *xraylib* σηματοδοτεί αυτές τις περιπτώσεις είτε επιστρέφοντας μηδενική ενέργεια είτε προκαλώντας μια εξαίρεση τύπου *ValueError*. Η λύση που εφαρμόζεται είναι η περικλείουσα κάθε κλήσης σε ένα μπλοκ *try-except*. Αυτός ο μηχανισμός αγνοεί με χάρη τις μη έγκυρες περιπτώσεις, επιτρέποντας στο σενάριο να συνεχίσει την εκτέλεσή του απρόσκοπτα και διασφαλίζοντας ότι η τελική βιβλιοθήκη περιέχει αποκλειστικά έγκυρες, θετικές ενεργειακές τιμές, στρογγυλοποιημένες σε έξι δεκαδικά ψηφία για βέλτιστη ακρίβεια. Το τελικό προϊόν, ένα αρχείο *JSON*, αποτελεί μια σταθερή, αναπαραγώγιμη και εξαιρετικά αποδοτική πηγή αναφοράς για τις ανώτερες βαθμίδες του αναλυτικού πλαισίου.

```
1. for z in atomic_numbers:
2.     element_data = {}
3.     try:
4.         symbol = xraylib.AtomicNumberToSymbol(z)
5.         print(f"Processing Z={z} ({symbol})...")
6.
7.         # Επεξεργασία γραμμών εκπομπής
8.         for line_name, line_code in emission_lines_to_include.items():
9.             try:
10.                 energy_kev = xraylib.LineEnergy(z, line_code)
11.                 # Η xraylib επιστρέφει 0 αν η γραμμή δεν υπάρχει για το στοιχείο
12.                 if energy_kev > 0:
13.                     element_data[line_name] = round(energy_kev, 6)
14.             except ValueError:
15.                 # Ο μηχανισμός αγνοεί με χάρη (gracefully ignores) τις μη έγκυρες
16.                 # περιπτώσεις, επιτρέποντας στο σενάριο να συνεχίσει απρόσκοπτα.
17.                 pass
18.
```

```

19. # ... (αντίστοιχη λογική για τις ακμές απορρόφησης) ...
20.
21. if element_data: # Προσθήκη μόνο εάν βρέθηκαν δεδομένα
22.     element_library[symbol] = element_data
23.
24. except Exception as element_err:
25.     print(f"Error processing Z={z}: {element_err}", file=sys.stderr)

```

NICE_HDF5_CORE.py: Θεμελίωση της Πρόσβασης και Επικύρωσης Δεδομένων

Αυτό το σενάριο υλοποιεί μια αφαίρεση υψηλού επιπέδου για τον χειρισμό αρχείων HDF5. Η κλάση **HDF5Viewer** ενθυλακώνει τις λειτουργίες που αφορούν το σύστημα αρχείων, όπως το άνοιγμα, η ανάκτηση μεταδεδομένων και η εξαγωγή δεδομένων, αντιμετωπίζοντας το πρόβλημα της διαλειτουργικότητας. Η κλάση **DataReceiver** λειτουργεί ως σημείο ελέγχου για τα δεδομένα που φορτώνονται στη μνήμη. Ο αλγόριθμός της περιλαμβάνει την άμεση επικύρωση του τύπου των δεδομένων (*np.ndarray* ή *pd.DataFrame*) και τον υπολογισμό περιγραφικών στατιστικών (*min*, *max*, *mean*, *std*, κλπ.). Αυτή η ποσοτική επισκόπηση είναι θεμελιώδης για την Εξερευνητική Ανάλυση Δεδομένων (**EDA**), παρέχοντας στον χρήστη μια άμεση κατανόηση της φύσης των δεδομένων.

DataReceiver Class: Διαχείριση και Επικύρωση Δεδομένων

Η κλάση **DataReceiver** είναι υπεύθυνη για τη λήψη δεδομένων, την επικύρωσή τους και την εκτέλεση προκαταρκτικής ανάλυσης. Τα δεδομένα μπορεί να είναι είτε *numpy.ndarray* είτε *pandas.DataFrame*. Η κλάση χρησιμοποιεί *type hints* για να επιβάλει την ακεραιότητα των δεδομένων, διασφαλίζοντας ότι επεξεργάζονται μόνο έγκυροι τύποι δεδομένων. Κατά τη λήψη δεδομένων μέσω της μεθόδου *receive_data*, η κλάση πραγματοποιεί άμεση επικύρωση για να επιβεβαιώσει τον τύπο των δεδομένων, αποτρέποντας έτσι σφάλματα κατά την εκτέλεση που σχετίζονται με μη συμβατές δομές δεδομένων. Η μέθοδος αποθηκεύει επίσης μεταδεδομένα που σχετίζονται με τα δεδομένα, επιτρέποντας την ιχνηλάτηση και την επεξεργασία με γνώση του περιβάλλοντος.

Η μέθοδος *log_reception* είναι υπεύθυνη για την καταγραφή βασικών λεπτομερειών, όπως η διαδρομή προέλευσης των δεδομένων, το σχήμα και ο τύπος δεδομένων, παρέχοντας έτσι μια διαφανή εγγραφή της διαδικασίας εισαγωγής δεδομένων. Η μέθοδος *_basic_stats* είναι μια στατιστική διαδικασία που υπολογίζει ένα ολοκληρωμένο σύνολο αριθμητικών μετρήσεων δεδομένων. Αυτές περιλαμβάνουν, αλλά δεν περιορίζονται σε, ελάχιστη, μέγιστη, μέση, διάμεσο, τυπική απόκλιση και τεταρτημόρια. Οι στατιστικές εκτυπώνονται για άμεση επιθεώρηση και επιστρέφονται ως λεξικό για περαιτέρω προγραμματιστική χρήση. Αυτή η προσέγγιση εγγυάται την προσβασιμότητα της κατανομής και των κεντρικών τάσεων των δεδομένων, επιτρέποντας γρήγορους διαγνωστικούς ελέγχους και εξερευνητική ανάλυση δεδομένων.

```

1. # Απόσπασμα από DataReceiver.receive_data και _basic_stats:
2.
3. def receive_data(self, data: np.ndarray, path: str, metadata: Dict) -> None:
4.     """Store and validate received data with metadata"""
5.     if not isinstance(data, (np.ndarray, pd.DataFrame)):
6.         raise ValueError("Data must be numpy array or pandas DataFrame")
7.     # ... αποθήκευση δεδομένων ...

```

```

8. self._basic_stats()
9.
10. def _basic_stats(self) -> Dict[str, float]:
11.     """Calculate enhanced statistics for numerical data"""
12.     if not hasattr(self.received_data, 'dtype') or not np.issubdtype(self.received_data.dtype, np.number):
13.         return {}
14.     stats = {
15.         "min": np.min(self.received_data),
16.         "max": np.max(self.received_data),
17.         "mean": np.mean(self.received_data),
18.         # ... και άλλες στατιστικές ...
19.     }
20.     return stats

```

HDF5Viewer Class: Λειτουργίες Αρχείων και Εξαγωγή Δεδομένων

Η κλάση **HDF5Viewer** ενσωματώνει λειτουργίες για αλληλεπίδραση με αρχεία **HDF5**. Αυτές περιλαμβάνουν το άνοιγμα αρχείων, την ανάκτηση πληροφοριών κόμβων και την εξαγωγή δεδομένων. Η μέθοδος **open_file** είναι υπεύθυνη για την επικύρωση της επέκτασης του αρχείου (**.h5** ή **.hdf5**) πριν από το άνοιγμά του, διασφαλίζοντας έτσι τη συμβατότητα με τη μορφή **HDF5**. Το λογισμικό διατηρεί επίσης μια εγγραφή πρόσφατα προσπελασμένων αρχείων, βελτιώνοντας τη χρηστικότητα με την επιτρέποντας γρήγορη πρόσβαση σε συχνά χρησιμοποιούμενα σύνολα δεδομένων. Η μέθοδος χρησιμοποιεί χειρισμό εξαιρέσεων για τη διαχείριση σφαλμάτων πρόσβασης σε αρχεία, παρέχοντας σαφή ανατροφοδότηση σε περίπτωση αποτυχιών.

Η μέθοδος **get_node_info** χρησιμοποιείται για την ανάκτηση λεπτομερών μεταδεδομένων σχετικά με έναν καθορισμένο κόμβο (είτε ένα σύνολο δεδομένων είτε μια ομάδα) μέσα στο αρχείο **HDF5**. Τα προαναφερθέντα μεταδεδομένα περιλαμβάνουν τον τύπο του κόμβου, το σχήμα, τον τύπο δεδομένων, τις ρυθμίσεις συμπίεσης, τη διαμόρφωση chunking και οποιαδήποτε προσαρμοσμένα χαρακτηριστικά από τον χρήστη. Για σύνολα δεδομένων, η μέθοδος παρέχει επίσης μια προεπισκόπηση των δεδομένων εάν το μέγεθός τους είναι διαχειρίσιμο (δηλαδή ≤ 100 στοιχεία), διευκολύνοντας έτσι μια γρήγορη οπτική επιθεώρηση. Η μέθοδος **get_node_data** είναι υπεύθυνη για την ανάκτηση των πραγματικών δεδομένων από ένα σύνολο δεδομένων, με την επιλογή της υποστήριξης **slicing** για την εξαγωγή υποσυνόλων των δεδομένων. Αξίζει να σημειωθεί ότι και οι δύο μέθοδοι ενσωματώνουν μηχανισμούς ισχυρού χειρισμού σφαλμάτων για τη διαχείριση μη έγκυρων διαδρομών ή μη προσπελάσιμων δεδομένων.

```

1. def get_node_info(self, path: str) -> Tuple[str, Dict]:
2.     """Get comprehensive node information"""
3.     try:
4.         if self.file is None:
5.             return "Error: No file open", {}
6.
7.         node = self.file[path]
8.         info = {
9.             'path': path,
10.            'type': 'Dataset' if isinstance(node, h5py.Dataset) else 'Group',
11.            'shape': node.shape if hasattr(node, 'shape') else None,
12.            'dtype': str(node.dtype) if hasattr(node, 'dtype') else None,
13.            # Η παρακάτω γραμμή είναι κρίσιμη για την κατανόηση της απόδοσης ανάγνωσης
14.            'chunks': node.chunks if hasattr(node, 'chunks') else None,
15.            'attributes': dict(node.attrs)
16.        }
17.        # ... (κώδικας για προεπισκόπηση μικρών datasets) ...
18.        return "Success", info

```

```
19. except Exception as e:
20.     return f'Error: {str(e)}', {}
```

Εξαγωγή και Ολοκλήρωση Δεδομένων

Η μέθοδος **export_data** διευκολύνει την εξαγωγή του περιεχομένου των συνόλων δεδομένων σε διάφορες μορφές, συμπεριλαμβανομένων των **CSV**, **.npy** της **NumPy** και **Excel**, καλύπτοντας έτσι διαφορετικές περιπτώσεις χρήσης σε επιστημονικές ροές εργασίας. Η μέθοδος επιλέγει δυναμικά την κατάλληλη συνάρτηση εξαγωγής με βάση την καθορισμένη μορφή, διασφαλίζοντας βέλτιστη απόδοση και συμβατότητα. Για παράδειγμα, οι εξαγωγές **CSV** χρησιμοποιούν την **numpy.savetxt**, ενώ οι εξαγωγές **Excel** αξιοποιούν την **pandas.DataFrame.to_excel**, αποδεικνύοντας έτσι την ενσωμάτωση του script με ευρύτερα οικοσυστήματα δεδομένων **Python**. Ο χειρισμός ασφαμάτων έχει υλοποιηθεί για τη διαχείριση ζητημάτων όπως μη υποστηριζόμενες μορφές ή δικαιώματα εγγραφής, διασφαλίζοντας αξιόπιστη λειτουργία σε περιβάλλοντα παραγωγής.

Επιστημονικές και Τεχνικές Εξηγήσεις

Ο σχεδιασμός του **script** βασίζεται στις αρχές της αρθρωτότητας και της επαναχρησιμοποίησης, με σαφή διαχωρισμό των αρμοδιοτήτων μεταξύ της λήψης δεδομένων (**DataReceiver**) και των λειτουργιών αρχείων (**HDF5Viewer**). Η χρήση **type hints** έχει αποδειχθεί ότι βελτιώνει την αναγνωσιμότητα του κώδικα και μειώνει την πιθανότητα σφαλμάτων που σχετίζονται με τύπους. Επιπλέον, ο χειρισμός εξαιρέσεων διασφαλίζει μια σταδιακή υποβάθμιση σε περίπτωση σφαλμάτων. Είναι προφανές ότι οι στατιστικοί υπολογισμοί στην ενότητα **basic_stats** είναι αριθμητικά ισχυροί, αξιοποιώντας τις βελτιστοποιημένες συναρτήσεις της **NumPy** για ακρίβεια και απόδοση. Οι δυνατότητες καταγραφής και διαχείρισης μεταδεδομένων του **script** συμμορφώνονται με τις βέλτιστες πρακτικές στην επιστημονική πληροφορική, όπου η αναπαραγωγιμότητα και η προέλευση των δεδομένων είναι ζωτικής σημασίας.

Εν συντομία, το παρόν **script** παρέχει ένα επιστημονικά αυστηρό και τεχνικά ορθό πλαίσιο για τη διαχείριση δεδομένων **HDF5**, συνδυάζοντας δυνατότητες επικύρωσης, ανάλυσης και εξαγωγής σε ένα συνεκτικό σύνολο εργαλείων. Η υλοποίηση αυτού του συστήματος είναι ενδεικτική μιας βαθιάς κατανόησης των περιπλοκών που εμπλέκονται στην επεξεργασία επιστημονικών δεδομένων. Παρουσιάζει λύσεις που είναι και πρακτικές και ικανές για επέκταση για χρήση σε ερευνητικές εφαρμογές.

ANALYSIS_METHODS_4_NEW18.py

Το αρχείο **analysis_methods_4_NEW18.py** αποτελεί την υπολογιστική καρδιά ολόκληρου του συστήματος, ενσωματώνοντας μια σουίτα προηγμένων αλγορίθμων μηχανικής μάθησης για την ανάλυση πολυδιάστατων δεδομένων. Η αρχιτεκτονική της ενότητας αυτής είναι αυστηρά αρθρωτή, διαχωρίζοντας τις λειτουργίες σε δύο διακριτές και εξειδικευμένες κλάσεις: την **ComponentAnalyzer** για την εφαρμογή τεχνικών μείωσης διαστάσεων και την **ClusterAnalyzer** για την εκτέλεση αλγορίθμων συσταδοποίησης και την ποσοτική αξιολόγηση των αποτελεσμάτων τους. Αυτός ο σχεδιασμός επιτρέπει την εύκολη συντήρηση, την επεκτασιμότητα και την αποσυνδεδεμένη χρήση των αλγορίθμων από τη λογική του γραφικού περιβάλλοντος χρήστη, καθιστώντας την ενότητα έναν αυτόνομο, ισχυρό υπολογιστικό πυρήνα.

3.1 Ανάλυση της Κλάσης ComponentAnalyzer

Η κλάση `ComponentAnalyzer` έχει σχεδιαστεί για να παρέχει μια ενοποιημένη και στιβαρή διεπαφή για την εφαρμογή ποικίλων τεχνικών μείωσης διαστάσεων. Μια θεμελιώδης σχεδιαστική απόφαση είναι η ενσωμάτωση της διαχείρισης αριθμητικής ακρίβειας κατά την αρχικοποίηση. Ο κατασκευαστής `__init__` δέχεται μια παράμετρο `precision` ('float16', 'float32', ή 'float64'), επιτρέποντας στον χρήστη να καθορίσει τον τύπο δεδομένων κινητής υποδιαστολής που θα χρησιμοποιηθεί για τις εσωτερικές υπολογιστικές διαδικασίες. Αυτό παρέχει τη δυνατότητα για μια κρίσιμη ανταλλαγή (trade-off) μεταξύ της υπολογιστικής ταχύτητας και της χρήσης μνήμης (με float16/float32) και της υψηλότερης αριθμητικής ακρίβειας και σταθερότητας (με float64), μια επιλογή απαραίτητη για τη διαχείριση τόσο μεγάλων όσο και επιστημονικά ευαίσθητων συνόλων δεδομένων.

```
1. # Απόσπασμα από ComponentAnalyzer.__init__:
2. # Διαχείριση της αριθμητικής ακρίβειας για τους υπολογισμούς.
3.
4. class ComponentAnalyzer:
5.     def __init__(self, precision='float32'):
6.         # ...
7.         self.precision = precision
8.         self.numpy_dtype = np.float32
9.         if precision == 'float16': self.numpy_dtype = np.float16
10.        elif precision == 'float64': self.numpy_dtype = np.float64
```

Η πιο προηγμένη και υπολογιστικά απαιτητική λειτουργία της κλάσης είναι η μέθοδος `incremental_pca`. Αυτή η μέθοδος έχει υλοποιηθεί ειδικά για την ανάλυση συνόλων δεδομένων που υπερβαίνουν τη διαθέσιμη μνήμη **RAM (out-of-core processing)**. Αντί να απαιτεί ολόκληρο τον πίνακα δεδομένων, δέχεται ένα επαναλήψιμο αντικείμενο (*iterable*) το οποίο παρέχει τα δεδομένα τμηματικά. Εσωτερικά, αξιοποιεί την κλάση `IncrementalPCA` της *scikit-learn* για την τμηματική προσαρμογή (*partial fitting*) του μοντέλου σε κάθε τμήμα. Ένα καινοτόμο χαρακτηριστικό της υλοποίησης είναι ο δυναμικός υπολογισμός της εσωτερικής παραμέτρου `batch_size` του αλγορίθμου, ο οποίος βασίζεται στον αριθμό των χαρακτηριστικών των δεδομένων και σε ένα όριο μνήμης που ορίζεται από τον χρήστη σε *megabytes* (`batch_size_limit_mb`). Αυτή η ευφυής παραμετροποίηση βελτιστοποιεί την απόδοση και αποτρέπει την εξάντληση της μνήμης κατά την επεξεργασία. Η μέθοδος ενσωματώνει επίσης εκτεταμένη καταγραφή της προόδου μέσω μιας συνάρτησης επανάκλησης (`log_callback`), παρέχοντας διαφάνεια στη χρονοβόρα διαδικασία της προσαρμογής.

```
1. # Απόσπασμα από ComponentAnalyzer.incremental_pca:
2. # Ο πυρήνας της τμηματικής προσαρμογής για επεξεργασία εκτός μνήμης.
3.
4. def incremental_pca(self, data_iterable: Iterable[Tuple[np.ndarray, np.ndarray]],
5.                     n_components: int, batch_size_limit_mb: int = 100,
6.                     # ...
7.                     ):
8.     # ... (υπολογισμός batch_size_ipca) ...
9.
10.    model = IncrementalPCA(n_components=n_components, batch_size=batch_size_ipca)
11.
12.    # Ο βρόχος επεξεργάζεται κάθε τμήμα δεδομένων χωρίς να φορτώνει ολόκληρο το dataset
13.    for i, (raw_chunk, reshaped_chunk_for_sklearn) in enumerate(data_iterable):
14.        if reshaped_chunk_for_sklearn is None or reshaped_chunk_for_sklearn.size == 0:
15.            continue
16.
17.        try:
18.            # Η κλήση partial_fit ενημερώνει σταδιακά το μοντέλο PCA
```

```

19.     model.partial_fit(reshaped_chunk_for_sklearn)
20.     total_samples_fitted += reshaped_chunk_for_sklearn.shape[0]
21.     # ... (κώδικας για αναφορά προόδου) ...
22. except Exception as e:
23.     # ... (χειρισμός σφαλμάτων) ...
24.
25. # Απελευθέρωση μνήμης μετά από κάθε τμήμα
26. if raw_chunk is not None: del raw_chunk
27. del reshaped_chunk_for_sklearn
28. if hasattr(gc, 'collect'): gc.collect()
29.
30. # ... (επιστροφή του εκπαιδευμένου μοντέλου) ...

```

Πέρα από την Επαυξητική *PCA*, η κλάση υλοποιεί και άλλες θεμελιώδεις τεχνικές. Η μέθοδος *nmf* για την Μη-Αρνητική Παραγοντοποίηση Πινάκων (*Non-Negative Matrix Factorization*) περιλαμβάνει αυτόματη προεπεξεργασία για τη διασφάλιση της μη-αρνητικότητας των δεδομένων εισόδου, μετατοπίζοντας τις τιμές εάν είναι απαραίτητο. Επιπλέον, εφαρμόζει μια στρατηγική πολλαπλών προσπαθειών, δοκιμάζοντας διαφορετικές μεθόδους αρχικοποίησης (*'nndsvda'*, *'nndsvd'*, *'random'*) για να επιτύχει σύγκλιση και να αποφύγει τοπικά ελάχιστα, αυξάνοντας έτσι την ευρωστία του αλγορίθμου.

```

1. # Απόσπασμα από ComponentAnalyzer.nmf:
2. # Στρατηγική πολλαπλών προσπαθειών για την αποφυγή τοπικών ελαχίστων.
3.
4. def nmf(self, data: np.ndarray, n_components: int) -> Tuple[np.ndarray, np.ndarray]:
5.     # ... προεπεξεργασία για μη-αρνητικότητα ...
6.     nmf_model = NMF(n_components=n_components, random_state=42, max_iter=300)
7.     init_methods_to_try = ['nndsvda', 'nndsvd', 'random']
8.
9.     for init_method in init_methods_to_try:
10.        print(f'NMF: Attempting with init='{init_method}'...")
11.        nmf_model.set_params(init=init_method)
12.        try:
13.            transformed_data_internal = nmf_model.fit_transform(data_nonneg)
14.            print(f'NMF with init='{init_method}' successful.")
15.            break # Έξοδος από τον βρόχο στην πρώτη επιτυχημένη προσπάθεια
16.        except Exception as e:
17.            print(f'NMF with init='{init_method}' failed: {e}). Trying next.")
18.            if init_method == init_methods_to_try[-1]: raise # Αν αποτύχουν όλες, προκαλέσει εξαίρεση
19.     # ...

```

Η μέθοδος *ica* για την Ανάλυση Ανεξάρτητων Συνιστωσών (*Independent Component Analysis*) ενσωματώνει τη διαδικασία της λεύκανσης (*whiten='unit-variance'*), ένα κρίσιμο βήμα που αποσυσχετίζει τα δεδομένα πριν από την εξαγωγή των ανεξάρτητων συνιστωσών, βελτιώνοντας σημαντικά την ποιότητα του διαχωρισμού. Τέλος, η κλάση παρέχει υλοποιήσεις για τις μη-γραμμικές τεχνικές *UMAP* και *t-SNE*, με την εισαγωγή της *UMAP* να είναι προαιρετική για να μην διακόπτεται η λειτουργικότητα της ενότητας εάν η εξειδικευμένη αυτή βιβλιοθήκη δεν είναι εγκατεστημένη.

3.2 Ανάλυση της Κλάσης ClusterAnalyzer

Η κλάση *ClusterAnalyzer* είναι αφιερωμένη στην εφαρμογή αλγορίθμων συσταδοποίησης και, κυρίως, στην ποσοτική αξιολόγηση της απόδοσής τους. Κάθε μέθοδος συσταδοποίησης ξεκινά με την κλήση της ιδιωτικής μεθόδου *_validate_data*, η οποία εκτελεί έναν απαραίτητο καθαρισμό των δεδομένων, αφαιρώντας αυτόματα γραμμές που περιέχουν μη-αριθμητικές (*NaN*) ή άπειρες (*inf*) τιμές, διασφαλίζοντας έτσι τη σταθερότητα των αλγορίθμων που ακολουθούν. Η κλάση υποστηρίζει ένα ευρύ

φάσμα τεχνικών, όπως ***K-Means***, ***DBSCAN***, ***Gaussian Mixture Models (GMM)*** και ***Spectral Clustering***. Αξιοσημείωτη είναι η υλοποίηση της ***DBSCAN***, η οποία εφαρμόζει αυτόματη τυποποίηση (***scaling***) των δεδομένων πριν από την εκτέλεση του αλγορίθμου, καθώς η απόδοσή της είναι εξαιρετικά ευαίσθητη στην κλίμακα των χαρακτηριστικών.

```
1. # Απόσπασμα από ClusterAnalyzer.calculate_evaluation_metrics:
2. # Συστηματικός υπολογισμός πολλαπλών μετρικών για την αξιολόγηση των συστάδων.
3.
4. def calculate_evaluation_metrics(self, data: np.ndarray, max_clusters: int = 10,
5.                               progress_queue: Optional[queue.Queue] = None) -> Dict:
6.     # ... (προετοιμασία δεδομένων) ...
7.     results = {
8.         'elbow': ([], []), 'silhouette': ([], []),
9.         'davies_bouldin': ([], []), 'calinski_harabasz': ([], [])
10.    }
11.
12.    for i, n in enumerate(range(2, max_clusters + 1)):
13.        # ... (αναφορά προόδου) ...
14.
15.        # Χρήση MiniBatchKMeans για ταχύτητα
16.        kmeans_model, labels = self.kmeans_minibatch(data_proc, n_clusters=n, random_state=42)
17.        inertia = getattr(kmeans_model, 'inertia_', -1)
18.        results['elbow'][0].append(n); results['elbow'][1].append(float(inertia))
19.
20.        unique_labels = np.unique(labels)
21.        if len(unique_labels) > 1:
22.            try:
23.                # Υπολογισμός Silhouette Score, με δειγματοληψία για μεγάλα δεδομένα
24.                sil_score = silhouette_score(data_proc, labels, sample_size=min(10000, data_proc.shape[0]),
25.                                             random_state=42)
26.                results['silhouette'][0].append(n); results['silhouette'][1].append(sil_score)
27.            except Exception as e:
28.                # ... (χειρισμός σφαλμάτων) ...
29.
30.        # ... (αντίστοιχος υπολογισμός για Davies-Bouldin και Calinski-Harabasz) ...
31.    return results
```

Το επιστημονικό αποκορύφωμα της κλάσης είναι το ενσωματωμένο πλαίσιο αξιολόγησης, το οποίο υλοποιείται στη μέθοδο ***calculate_evaluation_metrics***. Αυτή η μέθοδος δεν παρέχει απλώς ένα αποτέλεσμα, αλλά εκτελεί μια συστηματική διερεύνηση για τον προσδιορισμό του βέλτιστου αριθμού συστάδων (***k***). Για ένα εύρος τιμών του ***k***, η μέθοδος εφαρμόζει επαναληπτικά τον αλγόριθμο ***MiniBatchKMeans*** (μια αποδοτική παραλλαγή του ***K-Means*** για ταχύτερους υπολογισμούς) και υπολογίζει τέσσερις διαφορετικές μετρικές ποιότητας. Αυτές περιλαμβάνουν την **αδράνεια (inertia)** για τη "**Μέθοδο του Αγκώνα**", τον **Δείκτη Σιλουέτας (Silhouette Score)**, ο οποίος μετρά πόσο καλά διαχωρισμένες είναι οι συστάδες, τον **Δείκτη Davies-Bouldin**, που ποσοτικοποιεί την αναλογία της διασποράς εντός της συστάδας προς τον διαχωρισμό μεταξύ των συστάδων, και τον **Δείκτη Calinski-Harabasz**. Η συνδυασμένη χρήση αυτών των μετρικών παρέχει μια πολυδιάστατη και αντικειμενική αξιολόγηση της δομής των συστάδων. Η λειτουργικότητα αυτή συμπληρώνεται από τη μέθοδο ***plot_evaluation_curves***, η οποία οπτικοποιεί τα αποτελέσματα της αξιολόγησης σε ένα πλέγμα γραφημάτων **2x2**. Αυτή η οπτική αναπαράσταση είναι εξαιρετικά διαισθητική, επιτρέποντας στον ερευνητή να εντοπίσει με ευκολία το "**σημείο καμπής**" στον αγκώνα ή τις μέγιστες/ελάχιστες τιμές στις άλλες μετρικές, οδηγώντας έτσι σε μια επιστημονικά τεκμηριωμένη απόφαση για τον βέλτιστο αριθμό συστάδων.

ANALYSIS_VISUALIZER_3_NEW18.py

Η ενότητα *analysis_visualizer_3_NEW18.py* υλοποιεί την κλάση:

AnalysisVisualizer

Ενα αρθρωτό και επαναχρησιμοποιήσιμο πλαίσιο αφιερωμένο αποκλειστικά στη δημιουργία οπτικοποιήσεων υψηλής επιστημονικής ποιότητας από τα αποτελέσματα της ανάλυσης. Ο σχεδιασμός της αποσκοπεί στην αποσύνδεση της λογικής της σχεδίασης γραφημάτων από την κύρια λογική της εφαρμογής, επιτρέποντας την εύκολη παραγωγή και εξαγωγή γραφημάτων χωρίς εξαρτήσεις από το γραφικό περιβάλλον χρήστη.

```
1. # Απόσπασμα από AnalysisVisualizer.plot_3d:
2. # Ευέλικτη δημιουργία 3D γραφημάτων, ικανή να λειτουργεί σε υπάρχοντα axes.
3.
4. def plot_3d(self, ca_results: Dict, cluster_results: Optional[Dict] = None,
5.             x_comp: int = 0, y_comp: int = 1, z_comp: int = 2,
6.             fig: Optional[Figure] = None, ax: Optional[plt.Axes] = None) -> Figure:
7.
8.     # Αν δεν παρέχεται Figure, δημιουργείται νέο.
9.     if fig is None:
10.         fig = plt.figure(figsize=self.figsize)
11.
12.     # Αν παρέχονται Axes, καθαρίζονται. Αλλιώς, δημιουργούνται νέα 3D axes.
13.     if ax:
14.         ax.clear()
15.     else:
16.         ax = fig.add_subplot(111, projection='3d')
17.
18.     transformed = ca_results['transformed']
19.
20.     if transformed.shape[1] < 3:
21.         raise ValueError("Data must have at least 3 components for 3D visualization")
22.
23.     # Λογική σχεδίασης των σημείων, με δυνατότητα χρωματισμού ανάλογα με τις συστάδες
24.     if cluster_results is None:
25.         ax.scatter(transformed[:, x_comp], transformed[:, y_comp], transformed[:, z_comp])
26.         title = f'{ca_results["method"]} Results (3D)'
27.     else:
28.         scatter = ax.scatter(transformed[:, x_comp], transformed[:, y_comp], transformed[:, z_comp],
29.                              c=cluster_results['labels'], cmap='viridis')
30.         title = f'{cluster_results["method"]} Clustering Results (3D)'
31.         # ... (κώδικας για colorbar) ...
32.
33.     ax.set_xlabel(f'Component {x_comp + 1}')
34.     ax.set_ylabel(f'Component {y_comp + 1}')
35.     ax.set_zlabel(f'Component {z_comp + 1}')
36.     ax.set_title(title)
37.     fig.tight_layout()
38.
39.     return fig
```

Μια θεμελιώδης αρχή της κλάσης είναι η αποτελεσματική διαχείριση των πόρων της βιβλιοθήκης *Matplotlib*. Το χαρακτηριστικό *_current_fig* λειτουργεί ως ένας δείκτης προς το ενεργό αντικείμενο *Figure*. Η μέθοδος *_setup_figure*, η οποία καλείται πριν από τη δημιουργία κάθε νέου γραφήματος, ελέγχει αν υπάρχει ήδη ένα ενεργό γράφημα και, αν υπάρχει, το κλείνει ρητά (*plt.close*). Αυτή η πρακτική είναι κρίσιμης σημασίας για την αποφυγή διαρροών μνήμης (*memory leaks*), ένα συχνό πρόβλημα σε εφαρμογές που δημιουργούν επαναληπτικά μεγάλο αριθμό γραφημάτων.

Για την οπτικοποίηση των αποτελεσμάτων, η κλάση παρέχει εξειδικευμένες μεθόδους. Η *plot_component_analysis* είναι υπεύθυνη για τη δημιουργία διαγραμμάτων

διασποράς από τα δεδομένα που έχουν υποστεί μείωση διαστάσεων. Η υλοποίησή της είναι ευέλικτη και προσαρμόζεται δυναμικά στη διαστατικότητα των δεδομένων: για πολυδιάστατα αποτελέσματα δημιουργεί ένα τυπικό διάγραμμα διασποράς δύο συνιστωσών, ενώ για μονοδιάστατα δεδομένα (π.χ., όταν ζητείται μόνο μία συνιστώσα), σχεδιάζει τα σημεία κατά μήκος μιας οριζόντιας γραμμής, παρέχοντας μια σαφή εικόνα της κατανομής τους. Η μέθοδος ***plot_clustering*** επεκτείνει αυτή τη λειτουργικότητα, χρωματίζοντας τα σημεία δεδομένων σύμφωνα με τις ετικέτες που έχουν προκύψει από έναν αλγόριθμο συσταδοποίησης. Ένα ιδιαίτερα προηγμένο επιστημονικό χαρακτηριστικό αυτής της μεθόδου είναι η ενσωματωμένη υποστήριξη για πιθανοτικές μεθόδους, όπως τα ***Gaussian Mixture Models (GMM)***. Όταν παρέχονται οι πιθανότητες συμμετοχής κάθε σημείου στις συστάδες, η μέθοδος ρυθμίζει δυναμικά τη διαφάνεια (*alpha*) κάθε σημείου στο γράφημα, με τα σημεία υψηλής βεβαιότητας να εμφανίζονται αδιαφανή και τα σημεία χαμηλής βεβαιότητας να γίνονται ημιδιαφανή. Αυτή η τεχνική παρέχει μια ισχυρή οπτική ένδειξη της αβεβαιότητας του μοντέλου.

Η πιο σημαντική προσθήκη και επιστημονική αναβάθμιση στην ενότητα αυτή είναι η μέθοδος ***plot_3d***. Αυτή η μέθοδος αξιοποιεί τις δυνατότητες του ***mplot3d toolkit*** της ***Matplotlib*** για τη δημιουργία διαδραστικών, τρισδιάστατων διαγραμμάτων διασποράς. Ο σχεδιασμός της είναι εξαιρετικά ευέλικτος, καθώς επιτρέπει τη λειτουργία της πάνω σε υπάρχοντα αντικείμενα ***Figure*** και ***Axes***, καθιστώντας την ιδανική για ενσωμάτωση σε σύνθετα γραφικά περιβάλλοντα όπως αυτό του ***Tkinter***. Η μέθοδος επικυρώνει ότι τα δεδομένα έχουν τουλάχιστον τρεις διαστάσεις και, στη συνέχεια, σχεδιάζει τα σημεία στον τρισδιάστατο χώρο, με δυνατότητα χρωματισμού ανάλογα με τις ετικέτες των συστάδων. Η τρισδιάστατη απεικόνιση είναι συχνά απαραίτητη για την αποκάλυψη σύνθετων, μη-γραμμικών δομών στα δεδομένα, οι οποίες μπορεί να παραμένουν κρυφές σε δισδιάστατες προβολές. Η κλάση ολοκληρώνεται με βοηθητικές μεθόδους όπως η ***save_figure***, η οποία επιτρέπει την αποθήκευση των γραφημάτων σε αρχεία εικόνas με ρυθμιζόμενη ανάλυση (DPI), παρέχοντας έτσι τη δυνατότητα για παραγωγή γραφημάτων ποιότητας δημοσίευσης.

ANALYSIS_WINDOW_6_NEW18.py

Το αρχείο ***analysis_window_6_NEW18.py*** υλοποιεί την κλάση ***AnalysisWindow***, η οποία αποτελεί την επιστημονική και λειτουργική κορωνίδα ολόκληρου του συστήματος. Αυτό το παράθυρο δεν είναι απλώς ένας καμβάς για οπτικοποιήσεις, αλλά ένα ολοκληρωμένο, διαδραστικό περιβάλλον εργασίας (***Integrated Development Environment - IDE***) για την εξερευνητική ανάλυση δεδομένων. Ο σχεδιασμός του βασίζεται σε τρεις θεμελιώδεις αρχές: την επεξεργασία δεδομένων εκτός μνήμης (***out-of-core processing***) για τη διαχείριση τεράστιων συνόλων δεδομένων, τη διατήρηση ενός πλήρους και αναπαραγωγίμου ιστορικού των αναλυτικών βημάτων, και την παροχή ενός πλούσιου συνόλου διαδραστικών εργαλείων για την βαθιά διερεύνηση και ερμηνεία των αποτελεσμάτων.

Η πιο σημαντική αρχιτεκτονική απόφαση στην ***AnalysisWindow*** είναι η αποφυγή της άμεσης φόρτωσης ολόκληρου του συνόλου δεδομένων στη μνήμη ***RAM***. Αντ' αυτού, κατά την αρχικοποίησή της, η κλάση δημιουργεί μια περίπτωση της ***DataLoader***, μιας εξειδικευμένης κλάσης-διαχειριστή που είναι υπεύθυνη για την τμηματική φόρτωση (***chunking***) των δεδομένων απευθείας από το αρχείο ***HDF5*** στο δίσκο. Αυτή η προσέγγιση επιτρέπει στην εφαρμογή να αναλύει σύνολα δεδομένων που είναι πολλαπλάσια της διαθέσιμης φυσικής μνήμης, καθιστώντας την κατάλληλη για

εφαρμογές "**Big Data**" στην επιστημονική έρευνα. Η **DataLoader** υποστηρίζει επίσης δυναμική εφαρμογή μετασχηματισμών κατά τη φόρτωση, όπως η λογαριθμική κλιμάκωση, η οποία ελέγχεται από το γραφικό περιβάλλον, επιτρέποντας την ευέλικτη προεπεξεργασία των δεδομένων χωρίς την ανάγκη για ενδιάμεση αποθήκευση.

```
1. # Απόσπασμα από DataLoader.iterate_chunks:
2. # Η λογική που επιτρέπει την τμηματική ανάγνωση δεδομένων από το αρχείο HDF5.
3.
4. def iterate_chunks(self, axis: int = 0, max_chunk_bytes: Optional[int] = None) -> Iterable[np.ndarray]:
5.     # ... (υπολογισμός του μεγέθους κάθε τμήματος, chunk_size_axis) ...
6.
7.     num_chunks = math.ceil(current_shape[axis] / chunk_size_axis)
8.
9.     for i in range(num_chunks):
10.         start = i * chunk_size_axis
11.         end = min(start + chunk_size_axis, current_shape[axis])
12.
13.         # Δημιουργία του slice για την ανάγνωση του τμήματος
14.         current_slice_tuple = tuple([slice(None)] * axis + [slice(start, end)])
15.
16.         # Ανάγνωση μόνο του τρέχοντος τμήματος από τον δίσκο
17.         raw_chunk_data = self._dataset[current_slice_tuple]
18.
19.         # Εφαρμογή μετασχηματισμών (π.χ. λογαριθμική κλίμακα)
20.         transformed_chunk_data = self._apply_scale_transform(raw_chunk_data)
21.
22.         yield transformed_chunk_data
```

Η διαχείριση της αναλυτικής ροής εργασίας υλοποιείται μέσω ενός εξελιγμένου συστήματος ιστορικού. Το χαρακτηριστικό `self.data_history` δεν είναι μια απλή λίστα περιγραφών, αλλά μια δομή δεδομένων που καταγράφει την πλήρη κατάσταση της ανάλυσης σε κάθε βήμα. Κάθε εγγραφή στο ιστορικό περιέχει την περιγραφή της ενέργειας, τις παραμέτρους που χρησιμοποιήθηκαν, τα μεταδεδομένα των παραγόμενων αποτελεσμάτων (σχήμα, τύπος δεδομένων) και, κυρίως, τα ίδια τα αποτελέσματα (π.χ., οι πίνακες **transformed** ή **labels**), εφόσον το μέγεθός τους είναι διαχειρίσιμο. Αυτή η δομή επιτρέπει την απρόσκοπτη πλοήγηση προς τα εμπρός και προς τα πίσω στην αλυσίδα των αναλύσεων, επιτρέποντας τη σύγκριση και την επαλήθευση των αποτελεσμάτων. Η λειτουργικότητα αυτή επεκτείνεται περαιτέρω με τις μεθόδους **_save_project** και **_load_project**, οι οποίες σειριοποιούν ολόκληρη την κατάσταση της ανάλυσης, συμπεριλαμβανομένου του ιστορικού και των εκπαιδευμένων μοντέλων της **scikit-learn** (μέσω της **joblib**), σε ένα ενιαίο, αυτόνομο αρχείο **HDF5**. Αυτό εγγυάται την πλήρη αναπαραγωγικότητα μιας πολύπλοκης ανάλυσης, ένα χαρακτηριστικό υψίστης σημασίας για την επιστημονική έρευνα.

```
1. # Απόσπασμα από AnalysisWindow._save_project:
2. # Αποθήκευση του μοντέλου scikit-learn ως bytes μέσα στο αρχείο HDF5 του project.
3.
4. # ... (μέσα στον βρόχο που διατρέχει τα βήματα του ιστορικού) ...
5. if hasattr(r_act_val, 'get_params') and hasattr(r_act_val, 'predict'): # Έλεγχος αν είναι μοντέλο scikit-learn
6.     temp_model_file = f"temp_model_step{i}_{r_act_key}.joblib"
7.     joblib.dump(r_act_val, temp_model_file) # Σειριοποίηση του μοντέλου
8.     with open(temp_model_file, 'rb') as mf: model_bytes = mf.read()
9.
10. # Αποθήκευση των bytes ως numpy.void dataset
11. actual_results_grp.create_dataset(r_act_key + "_joblib_bytes", data=np.void(model_bytes))
12.
13. if os.path.exists(temp_model_file): os.remove(temp_model_file)
14. actual_results_grp.attrs[r_act_key + "_type"] = "joblib_sklern_model"
```

Το γραφικό περιβάλλον χρήστη είναι εξαιρετικά πλούσιο και οργανωμένο σε ένα πολυεπίπεδο σύστημα καρτελών (*Notebook widgets*). Οι κύριες καρτέλες, όπως "*Spectrum Views*", "*Component Analysis*" και "*Material Analysis*", ομαδοποιούν τις σχετικές λειτουργίες, ενώ ένθετες υπο-καρτέλες παρέχουν εξειδικευμένες οπτικοποιήσεις. Ένα από τα πιο ισχυρά διαδραστικά χαρακτηριστικά είναι η δυνατότητα επιλογής μιας Περιοχής Ενδιαφέροντος (*ROI*) απευθείας πάνω στο γράφημα του φάσματος αθροίσματος, μέσω του SpanSelector της *Matplotlib*. Η επιλεγμένη φασματική περιοχή μπορεί στη συνέχεια να χρησιμοποιηθεί ως φίλτρο, περιορίζοντας τα χαρακτηριστικά που τροφοδοτούνται στους αλγόριθμους, επιτρέποντας έτσι εξαιρετικά στοχευμένες αναλύσεις σε συγκεκριμένα ενεργειακά παράθυρα. Η διαδραστικότητα επεκτείνεται με τη λειτουργία "*Plot Linking*", η οποία επιτρέπει στον χρήστη να επιλέξει μια ομάδα σημείων στο διάγραμμα διασποράς των συνιστωσών και να δει την αντίστοιχη χωρική τους κατανομή να επισημαίνεται αυτόματα στις εικόνες των συνιστωσών, παρέχοντας μια άμεση σύνδεση μεταξύ του αφηρημένου χώρου των χαρακτηριστικών και του πραγματικού, χωρικού τομέα.

```

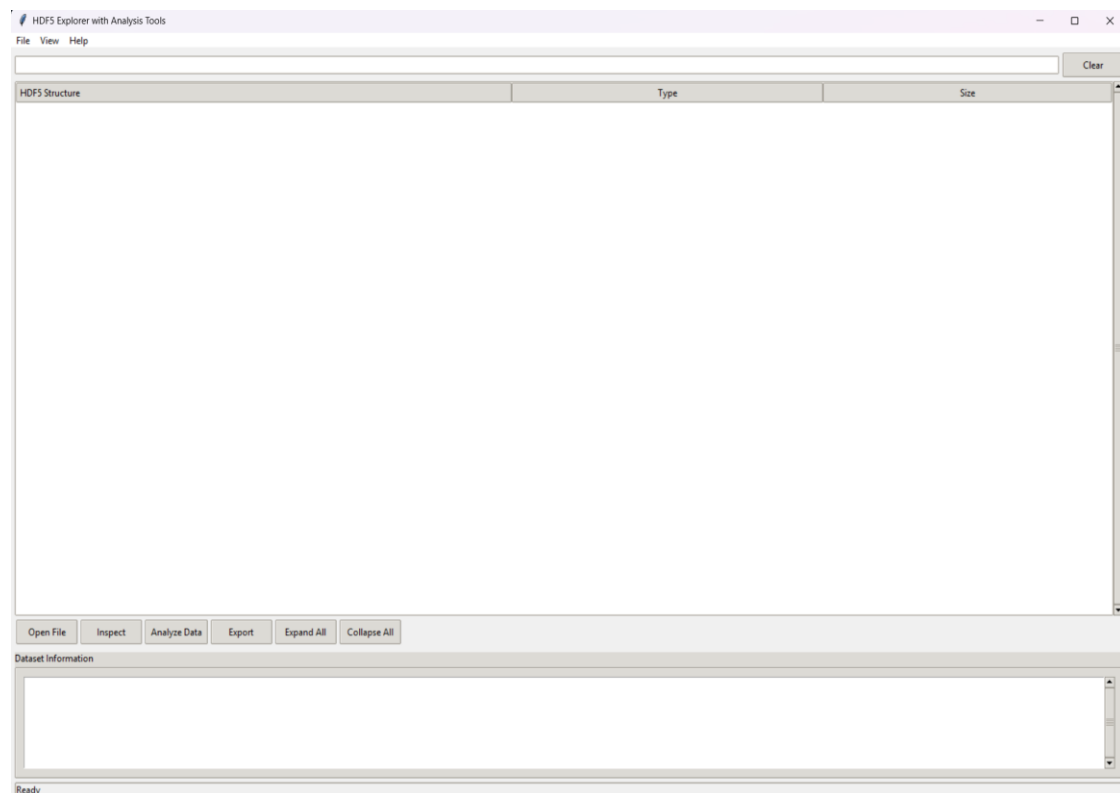
1. # Απόσπασμα από AnalysisWindow._initialize_spectral_roi_selector:
2. # Ενεργοποίηση του διαδραστικού εργαλείου επιλογής περιοχής στο γράφημα.
3.
4. def _initialize_spectral_roi_selector(self):
5.     # ... (καθαρισμός παλιού widget) ...
6.     if self.spectrum_ax and self.spectrum_canvas:
7.         self.logger.info("Initializing SpanSelector for spectral ROI.")
8.         # Δημιουργία του widget που συνδέει τη δράση του ποντικιού με τη λογική της εφαρμογής
9.         self.spectral_roi_widget = SpanSelector(
10.             self.spectrum_ax,
11.             self._on_spectral_roi_select, # Συνάρτηση επανάκλησης (callback)
12.             'horizontal',
13.             useblit=True,
14.             props=dict(alpha=0.3, facecolor='skyblue'),
15.             interactive=True,
16.             drag_from_anywhere=True
17.         )

```

Τέλος, η κλάση ενσωματώνει προηγμένες αναλυτικές δυνατότητες στην καρτέλα "*Material Analysis*". Η μέθοδος *_find_and_match_peaks* χρησιμοποιεί αλγόριθμους επεξεργασίας σήματος (*scipy.signal.find_peaks*) για τον αυτόματο εντοπισμό φασματικών κορυφών και, στη συνέχεια, τις συσχετίζει με τις θεωρητικές τιμές από τη βιβλιοθήκη *XRF* που δημιουργήθηκε προηγουμένως. Βάσει αυτών των ταυτοποιήσεων, το σύστημα μπορεί να δημιουργήσει χάρτες στοιχειακής κατανομής, να διερευνήσει τις χωρικές συσχετίσεις μεταξύ διαφορετικών στοιχείων και να συνθέσει τριχρωματικές εικόνες (*RGB overlays*) που συχνά αποκαλύπτουν τη συνύπαρξη υλικών. Η υποστήριξη για "συνταγές χρωστικών" (*_apply_color_recipe*) αυτοματοποιεί αυτή τη διαδικασία για γνωστούς συνδυασμούς, ενώ η δυνατότητα για υπο-συσταδοποίηση (*sub-clustering*) επιτρέπει μια ιεραρχική, "*top-down*" προσέγγιση στην ανάλυση, όπου ο ερευνητής μπορεί να απομονώσει μια συγκεκριμένη συστάδα και να την αναλύσει περαιτέρω με μεγαλύτερη λεπτομέρεια. Με αυτόν τον τρόπο, η *AnalysisWindow* μετατρέπεται από ένα απλό εργαλείο οπτικοποίησης σε ένα ολοκληρωμένο περιβάλλον επιστημονικής ανακάλυψης.

HDF5_Explorer_GUI.py

Η κλάση **HDF5ExplorerGUI** Εικόνα 16 είναι η **κύρια εφαρμογή** για πλοήγηση και ανάλυση αρχείων **HDF5**. Διαθέτει μια **διάταξη δέντρου** για ιεραρχική εξερεύνηση αρχείων, έναν πίνακα μεταδεδομένων με δυνατότητα αναζήτησης και μενού περιβάλλοντος για εξαγωγή δεδομένων. Το γραφικό περιβάλλον φορτώνει αρχεία **HDF5** μέσω της **h5py**, εμφανίζοντας ιδιότητες συνόλων δεδομένων (σχήμα, τύπος δεδομένων) και χαρακτηριστικά σε ένα κύλισμο widget κειμένου. Οι χρήστες μπορούν να εκκινήσουν παράθυρα ανάλυσης (**AnalysisWindow**, **ClusterAnalysisWindow**) με έλεγχο ακρίβειας (**float16/32/64**), εξασφαλίζοντας συμβατότητα με διαφορετικούς τύπους δεδομένων.



Εικόνα 16 : Στιγμιότυπο της κύριας διεπαφής του λογισμικού NICE (HDF5ExplorerGUI) μετά τη φόρτωση ενός αρχείου HDF5. Απεικονίζεται η δενδρική δομή του αρχείου, ο πίνακας μεταδεδομένων για τον επιλεγμένο κόμβο και τα κύρια κουμπιά ελέγχου της ανάλυσης.

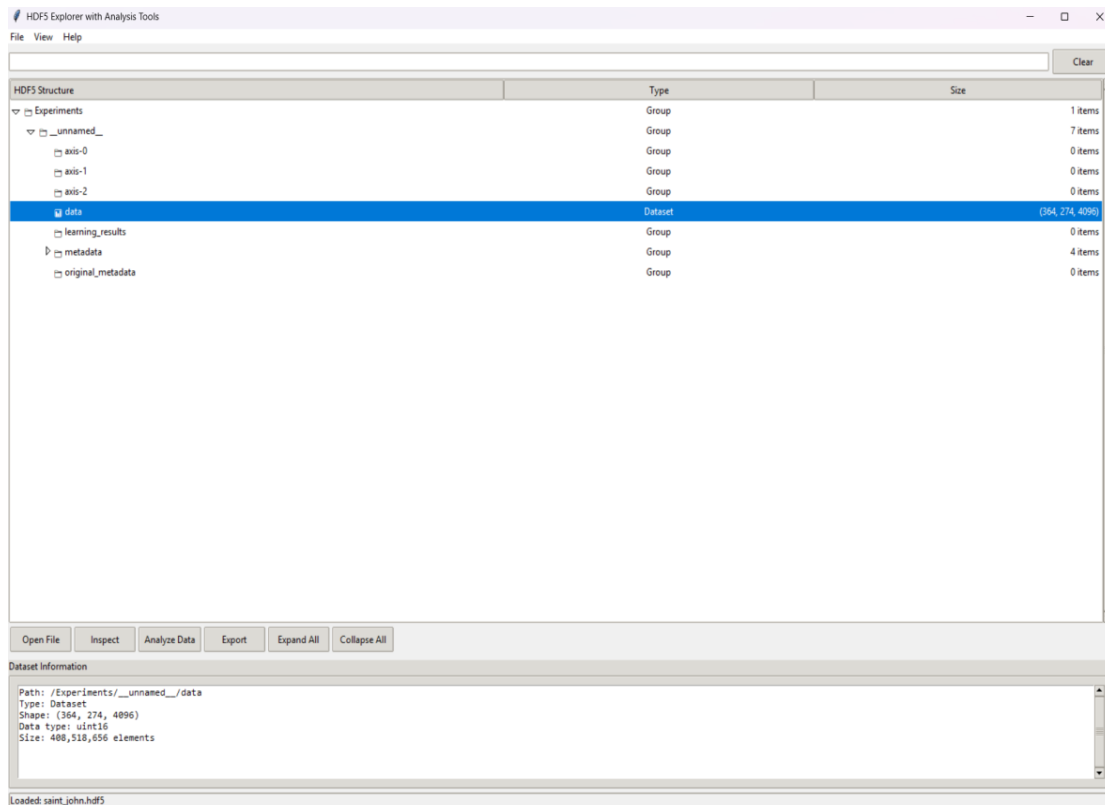
Η αρχιτεκτονική της εφαρμογής χαρακτηρίζεται από μια **modular** σχεδίαση, με ξεχωριστά στοιχεία υπεύθυνα για διαφορετικές πτυχές της λειτουργικότητας. Η κλάση **HDF5ExplorerGUI** είναι υπεύθυνη για τη διαχείριση του κύριου παραθύρου της εφαρμογής και τον συντονισμό των στοιχείων απεικόνισης (**TreeViewManager**) και των υποστηρικτικών συστημάτων διαχείρισης δεδομένων (**HDF5Viewer** και **DataReceiver**). Ο διαχωρισμός των αρμοδιοτήτων είναι μια βασική αρχή στην ανάπτυξη λογισμικού, διασφαλίζοντας τη συντηρησιμότητα του συστήματος και επιτρέποντας μελλοντικές επεκτάσεις. Η εφαρμογή διευκολύνει την εκτέλεση πολλαπλών αναλυτικών ροών εργασίας μέσω της χρήσης εξειδικευμένων κλάσεων παραθύρων, συγκεκριμένα των **DataAnalysisWindow** και **ClusterAnalysisWindow**. Αυτές οι κλάσεις παραθύρων προσφέρουν αποκλειστικές διεπαφές για συγκεκριμένες εργασίες, διασφαλίζοντας ταυτόχρονα τη διατήρηση μιας μόνιμης σύνδεσης με την κύρια εφαρμογή.

```

1. # Απόσπασμα από HDF5ExplorerGUI. open_full_analysis_window:
2. # Η ενεργοποίηση της εκκίνησης του παραθύρου ανάλυσης από την κύρια GUI.
3.
4. def open_full_analysis_window(self, event=None):
5.     path = self.tree_manager.get_selected_path()
6.     if not path or not self.current_filepath:
7.         # ... έλεγχος για την επιλογή του χρήστη ...
8.         return
9.
10.    # Ερώτηση στον χρήστη για την επιθυμητή αριθμητική ακρίβεια
11.    precision = ask_precision_dialog(self.root)
12.    if not precision: return # Ακύρωση από τον χρήστη
13.
14.    # Συλλογή των απαραίτητων μεταδεδομένων
15.    metadata_for_analysis = {
16.        'path': self.current_filepath,
17.        'dataset_path': path,
18.        'attributes': info.get('attributes', {})
19.    }
20.
21.    # Δημιουργία και εμφάνιση του εξειδικευμένου παραθύρου ανάλυσης
22.    analysis_win = AnalysisWindow(
23.        parent=self.root,
24.        filepath=self.current_filepath,
25.        dataset_path=path,
26.        metadata=metadata_for_analysis,
27.        precision=precision
28.    )
29.    analysis_win.deiconify()

```

Η υποκλάση **TreeViewManager** ενισχύει την πλοήγηση με εικονίδια ( για σύνολα δεδομένων,  για ομάδες Εικόνα 17), αναδιπλούμενους κόμβους και φιλτράρισμα αναζήτησης. Για παράδειγμα, πληκτρολογώντας "φάσμα" επισημαίνονται οι αντίστοιχοι κόμβοι και κρύβονται οι υπόλοιποι. Η μέθοδος ***_inspect_node*** επικυρώνει τις επιλογές πριν από την ανάλυση, ενώ η ***_export_data*** υποστηρίζει μορφές ***CSV***, ***Excel*** και ***NumPy*** μέσω της ***Pandas***.



Εικόνα 17: Λεπτομερής προβολή της ιεραρχικής δομής του αρχείου δεδομένων εντός του *HDF5ExplorerGUI*. Διακρίνονται οι ομάδες (*Group*) και τα σύνολα δεδομένων (*Dataset*), καθώς και οι πληροφορίες για το μέγεθος και τον τύπο τους.

```

1. # Απόσπασμα από TreeViewManager._add_node_recursive:
2. # Η μετάφραση της δομής HDF5 σε μια πλούσια οπτική αναπαράσταση.
3.
4. def _add_node_recursive(self, parent_id: str, item: h5py.HLObject) -> None:
5.     name = os.path.basename(item.name) if item.name != '/' else '/'
6.     node_type = 'Dataset' if isinstance(item, h5py.Dataset) else 'Group'
7.     tags = (node_type.lower(),)
8.     # Χρήση εικονιδίων για άμεση οπτική αναγνώριση
9.     text = f'{self.icon_map.get(node_type, "?")} {name}'
10.
11.     size_str, dtype_str = "", ""
12.     if isinstance(item, h5py.Dataset):
13.         shape = item.shape
14.         size_str = '(Scalar)' if shape == () else str(shape)
15.         dtype_str = str(item.dtype)
16.     elif isinstance(item, h5py.Group):
17.         size_str = f'{len(item)} item(s)'
18.
19.     # Προσθήκη των μεταδεδομένων απευθείας στις στήλες του Treeview
20.     values = (node_type, size_str, dtype_str)
21.
22.     node_id = self.tree.insert(parent_id, 'end', iid=item.name, text=text, values=values, tags=tags)
23.     self.all_nodes[node_id] = item.name
24.
25.     # ... (αναδρομική κλήση για τις υπο-ομάδες) ...
26.

```

Η διαχείριση των δεδομένων εκτελείται με σχολαστική προσοχή, με ιδιαίτερη έμφαση στη χρήση τους σε επιστημονικά πλαίσια. Το σενάριο ενσωματώνει ισχυρές ρουτίνες επικύρωσης που διασφαλίζουν την καταλληλότητα των δεδομένων για ανάλυση, συμπεριλαμβανομένων ελέγχων για *NaN/inf* τιμές και αυτόματο ανασχηματισμό πολυδιάστατων πινάκων. Το σύστημα ελέγχου ακρίβειας (*float16/32/64*) αποδεικνύει

μια προσεκτική εξέταση των απαιτήσεων αριθμητικής ανάλυσης, επιτρέποντας στους χρήστες να ισορροπούν τη χρήση μνήμης και την υπολογιστική ακρίβεια. Το μοτίβο λήψης δεδομένων που χρησιμοποιείται σε όλη την εφαρμογή διευκολύνει την ευέλικτη ροή δεδομένων μεταξύ των στοιχείων, διασφαλίζοντας παράλληλα τη διατήρηση ενός ξεκάθολου διαχωρισμού μεταξύ των επιπέδων πρόσβασης στα δεδομένα και απεικόνισης.

Οι δυνατότητες απεικόνισης του λογισμικού είναι περιεκτικές και προσαρμοσμένες στις επιστημονικές ροές εργασίας. Το **TreeViewManager** παρέχει μια ενστικτώδη ιεραρχική προβολή των δομών των αρχείων **HDF5**, επαυξημένη με εικονίδια, πληροφορίες μεγέθους και διαδραστική φιλτράρισμα. Η λειτουργικότητα σχεδίασης χρησιμοποιεί την απρόσκοπτη ενσωμάτωση του matplotlib με το **Tkinter**, διασφαλίζοντας έτσι την παράδοση απεικονίσεων επιστημονικής ποιότητας εντός της **GUI**. Αυτό περιλαμβάνει μετρικές αξιολόγησης συστάδων και γραφήματα ενεργειακών φασμάτων. Η υλοποίηση δυναμικής ενημέρωσης γραφημάτων και διαδραστικών ελέγχων (π.χ. ρυθμιστικά ορίου, επιλογείς συστατικών) δημιουργεί ένα ελκυστικό περιβάλλον εξερεύνησης.

Ο σχεδιασμός της διεπαφής χρήστη αποδεικνύει μια βαθιά κατανόηση των επιστημονικών ροών εργασίας. Το λογισμικό έχει σχεδιαστεί για να διευκολύνει την καθημερινή χρήση, με χαρακτηριστικά όπως μια λίστα πρόσφατων αρχείων, ρυθμιζόμενες θεματικές και συντομεύσεις πληκτρολογίου. Η περιεκτική προβολή μεταδεδομένων στον πίνακα πληροφοριών συνόλου δεδομένων παρέχει άμεση πρόσβαση σε κρίσιμες πληροφορίες σχετικά με επιλεγμένους κόμβους. Τα αναλυτικά εργαλεία είναι προσβάσιμα με διάφορους τρόπους, συμπεριλαμβανομένων επιλογών μενού και κουμπιών γραμμής εργαλείων. Η γραμμή κατάστασης παρέχει σαφή οπτική ανατροφοδότηση. Η υλοποίηση παραθύρων τροπικής ανάλυσης είναι καίρια για τη διασφάλιση της ολοκλήρωσης εργασιών με εστίαση, διατηρώντας ταυτόχρονα το πλαίσιο του συνολικού συνόλου δεδομένων.

Ο χειρισμός σφαλμάτων και η επικοινωνία με τον χρήστη υλοποιούνται σε όλη την εφαρμογή, με μπλοκ **try-catch** που προστατεύουν κρίσιμες λειτουργίες και ενημερωτικά μηνύματα που εξηγούν τα ζητήματα στους χρήστες. Η επικύρωση παραμέτρων ανάλυσης συστάδων και μεγεθών συνόλων δεδομένων είναι μια απαραίτητη διαδικασία που αποτρέπει κοινές παγίδες πριν προκύψουν. Η προοδευτική αποκάλυψη λειτουργικότητας (π.χ., η επιλογή υποσυνόλου μεγάλων συνόλων δεδομένων) έχει αποδειχθεί ότι βοηθά στη μείωση του κινδύνου μη επιθυμητών λειτουργιών μεγάλης κατανάλωσης πόρων.

Οι αναλυτικές δυνατότητες της εφαρμογής είναι ιδιαίτερα αξιοσημείωτες. Η υλοποίηση της ανάλυσης συστάδων παρέχει τόσο σκορ σιλουέτας όσο και οπτικοποιήσεις μεθόδου αγκώνα, προσφέροντας έτσι στους χρήστες πολλαπλές προοπτικές για τον καθορισμό του βέλτιστου αριθμού συστάδων. Η ανάλυση ενεργειακών φασμάτων ενσωματώνει προσαρμόσιμες παραμέτρους βαθμονόμησης, διασφαλίζοντας έτσι την προσαρμοστικότητά της σε διαφορετικές διαμορφώσεις οργάνων. Η ενσωμάτωση αυτών των αναλυτικών εργαλείων απευθείας με τον περιηγητή **HDF5** δημιουργεί μια απρόσκοπτη ροή εργασίας από την επιθεώρηση δεδομένων έως την προηγμένη ανάλυση.

Σχεδιαστικές αποφάσεις που σχετίζονται με την απόδοση είναι εμφανείς σε πολλά σημεία. Η διαδικασία τεμπέλης φόρτωσης δεδομένων **HDF5**, η επιλογή υποσυνόλου

συνόλων δεδομένων και ο έλεγχος ακρίβειας συμβάλλουν στην αποτελεσματική διαχείριση της χρήσης μνήμης σε πλαίσια μεγάλων συνόλων δεδομένων. Η υλοποίηση της προβολής δέντρου ενσωματώνει αποτελεσματικές δυνατότητες φιλτραρίσματος και επισημάνσης χωρίς την ανάγκη για περιττές επανασχεδιάσεις. Η απόδοση των γραφημάτων είναι βελτιστοποιημένη μέσω της συνετού διαχείρισης αντικειμένων σχήματος **matplotlib** και ενημερώσεων καμιά.

Το σύστημα διαμόρφωσης εισάγει ένα χαρακτηριστικό επιμονής που αποθηκεύει τις προτιμήσεις των χρηστών, συμπεριλαμβανομένων ρυθμίσεων μεγέθους παραθύρου, πρόσφατων αρχείων και άλλων παραμέτρων προσαρμογής, μεταξύ των συνεδριών. Αυτή η προσοχή στη λεπτομέρεια έχει αποδειχθεί ότι οδηγεί σε σημαντική βελτίωση της εμπειρίας του χρήστη για τους τακτικούς χρήστες. Το παράθυρο διαλόγου "Σχετικά" και το μενού βοήθειας παρέχουν τα κατάλληλα πιστώματα και βασικές οδηγίες, ενώ το καθαρό οπτικό στυλ (χρησιμοποιώντας την θεματική "**clam**" του **tk**) διασφαλίζει την αναγνωσιμότητα και την επαγγελματική εμφάνιση.

Από την πλευρά της τεχνικής υλοποίησης, το σενάριο παρουσιάζει αρκετές προηγμένες τεχνικές **Python**. Η χρήση υποδείξεων τύπων έχει αποδειχθεί ότι ενισχύει τη σαφήνεια και τη συντηρησιμότητα του κώδικα. Η δομή των κλάσεων ακολουθεί τις αντικειμενοστρεφείς αρχές διατηρώντας παράλληλα πραγματικότητα για την ανάπτυξη **GUI**. Το σύστημα δέσμευσης συμβάντων διευκολύνει τη δημιουργία αποκριτικών διεπαφών χωρίς την ανάγκη για υπερβολική πολυπλοκότητα επανάκλησης. Η ενσωμάτωση του **matplotlib** είναι ιδιαίτερα αξιοσημείωτη, με αποτελεσματική διαχείριση εργαλείων και ενσωμάτωσης καμιά.

Τα αναλυτικά παράθυρα (**DataAnalysisWindow** και **ClusterAnalysisWindow**) επιδεικνύουν εξειδικευμένες διεπαφές που έχουν προσαρμοστεί στις συγκεκριμένες τους εργασίες. Η οπτικοποίηση της ανάλυσης συστάδων παρουσιάζει τόσο σκορ σιλουέτας όσο και τιμές αδράνειας σε μια σαφή σύγκριση πλαι-πλάι, επιτρέποντας έτσι στους χρήστες να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με τον αριθμό των συστάδων. Η ανάλυση ενεργειακών φασμάτων ενσωματώνει διαδραστικούς ελέγχους βαθμονόμησης, αποδεικνύοντας έτσι μια δέσμευση για την αναπαραγωγή πραγματικών ροών εργασίας φασματοσκοπίας.

Η αρχιτεκτονική δείχνει την παρουσία σημείων μελλοντικής επέκτασης. Το σύστημα εξαγωγής θα μπορούσε να επεκταθεί για την υποστήριξη πρόσθετων μορφών αρχείων. Οι δυνατότητες απεικόνισης θα μπορούσαν να επεκταθούν για να περιλαμβάνουν τρισδιάστατες απεικονίσεις ή πρόσθετες αναλυτικές μεθόδους. Η λειτουργικότητα περιήγησης **HDF5** θα μπορούσε να ενισχυθεί με την ενσωμάτωση εργαλείων επεξεργασίας χαρακτηριστικών ή σύγκρισης συνόλων δεδομένων. Η υιοθέτηση μιας modular προσέγγισης σχεδίασης διευκολύνει την υλοποίηση τέτοιων επεκτάσεων χωρίς την ανάγκη για εκτεταμένες εργασίες ανακατασκευής.

Στην πράξη, η εφαρμογή θα ήταν χρήσιμη για ερευνητές που εργάζονται με επιστημονικά δεδομένα βασισμένα σε **HDF5**, ιδιαίτερα στους τομείς της φασματοσκοπίας, απεικόνισης και υπολογιστικής επιστήμης. Ο συνδυασμός ισχυρής διαχείρισης δεδομένων, πολλαπλών τρόπων απεικόνισης και ενσωματωμένων αναλυτικών εργαλείων δημιουργεί ένα ισχυρό περιβάλλον για την εξερευνητική ανάλυση δεδομένων. Η έμφαση του λογισμικού στην αποδοτικότητα της ροής εργασίας και την εμπειρία του χρήστη το καθιστά κατάλληλο τόσο για την ρουτίνα επιθεώρησης όσο και για την εις βάθος διερεύνηση πολύπλοκων συνόλων δεδομένων.

Το σενάριο που αναθεωρείται αντιπροσωπεύει μια ώριμη υλοποίηση που επιτυγχάνει μια αποτελεσματική ισορροπία μεταξύ εξελιγμένης λειτουργικότητας και χρηστικότητας. Αποδεικνύει πώς τα επιστημονικά εργαλεία **Python** μπορούν να ενσωματωθούν αποτελεσματικά σε εφαρμογές επιφάνειας εργασίας, παρέχοντας στους ερευνητές διαδραστική πρόσβαση σε προηγμένες αναλυτικές μεθόδους. Οι συνετές σχεδιαστικές αποφάσεις που είναι ενσωματωμένες στη βάση κώδικα είναι ενδεικτικές μιας βαθιάς κατανόησης των τεχνικών λεπτομερειών της επιστημονικής υπολογιστικής, σε συνδυασμό με μια σχολαστική ευθυγράμμιση με τις πρακτικές απαιτήσεις των ερευνητικών ροών εργασίας.

Συμπεράσματα και Μελλοντικές Κατευθύνσεις

Το σύστημα **NICE**, όπως αναλύθηκε, αντιπροσωπεύει μια ώριμη και τεχνικά άρτια υλοποίηση που επιτυγχάνει μια αποτελεσματική ισορροπία μεταξύ εξελιγμένης λειτουργικότητας και χρηστικότητας. Αποδεικνύει πώς τα επιστημονικά εργαλεία της Python μπορούν να ενσωματωθούν σε μια συνεκτική εφαρμογή επιφάνειας εργασίας, παρέχοντας στους ερευνητές διαδραστική πρόσβαση σε προηγμένες αναλυτικές μεθόδους. Οι συνετές αρχιτεκτονικές αποφάσεις, όπως η επεξεργασία εκτός μνήμης και το σύστημα ιστορικού, αντιμετωπίζουν θεμελιώδεις προκλήσεις της σύγχρονης επιστημονικής πληροφορικής.

Η εφαρμογή αυτής της μεθοδολογίας αποτελεί ήδη ένα σημαντικό εργαλείο για την έρευνα της πολιτιστικής κληρονομιάς, προσφέροντας τυποποιημένη και, κυρίως, αναπαραγώγιμη (**reproducible**) ανάλυση έργων τέχνης και αρχαιολογικών ευρημάτων. Οι μελλοντικές εξελίξεις του συστήματος θα επικεντρωθούν στην ενσωμάτωση τεχνικών βαθιάς μάθησης (**deep learning**) για την αυτόματη αναγνώριση σύνθετων υλικών, καθώς και στην ανάπτυξη cloud-based υπηρεσιών για την ενίσχυση της συνεργασίας μεταξύ ερευνητών και τη δημιουργία ολοκληρωμένων, διαμοιραζόμενων βάσεων δεδομένων. Η τεχνολογική αυτή προσέγγιση συνδυάζει υπολογιστική ισχύ με επιστημονική αυστηρότητα, ανοίγοντας νέους δρόμους στη μελέτη και διατήρηση της πολιτιστικής κληρονομιάς.

Αποτελέσματα και Συζήτηση

Σε αυτό το κεφάλαιο, παρουσιάζονται τα αποτελέσματα της χημειομετρικής ανάλυσης που πραγματοποιήθηκε με το λογισμικό NICE στα σύνολα δεδομένων MA-XRF από τις τρεις θρησκευτικές εικόνες. Για κάθε εικόνα, η ανάλυση ακολουθεί τη ροή εργασίας που περιγράφηκε στο Κεφάλαιο παραπάνω, συμπεριλαμβανομένης της προεπεξεργασίας, της υβριδικής ανάλυσης NMF-PCA, της ομαδοποίησης K-Means και της στοιχειακής χαρτογράφησης για επικύρωση.

ΑΓΙΟΣ ΙΩΑΝΝΗΣ ΧΑΜΗΛΗΣ ΑΝΑΛΥΣΗΣ

Η μη-καταστροφική ανάλυση έργων τέχνης αποτελεί κεντρικό πυλώνα της σύγχρονης τεχνο-ιστορικής έρευνας και της επιστήμης της συντήρησης. Η τεχνική της απεικονιστικής Φασματοσκοπίας Φθορισμού Ακτίνων-X (*Macro X-Ray Fluorescence, MA-XRF*) έχει αναδειχθεί σε ένα από τα ισχυρότερα εργαλεία σε αυτό το πεδίο, καθώς επιτρέπει την ταχεία και λεπτομερή χαρτογράφηση της στοιχειακής σύστασης μιας επιφάνειας. Μέσω αυτής, αποκαλύπτεται η παλέτα του καλλιτέχνη, τα στρώματα προετοιμασίας (υπόστρωμα) και πιθανές μεταγενέστερες επεμβάσεις, χωρίς να απαιτείται η λήψη δείγματος που θα προκαλούσε φθορά στο έργο.

Η παρούσα εργασία εστιάζει στην εικόνα του Αγίου Ιωάννη του Προδρόμου Εικόνα 86, ένα έργο που ενσωματώνει τυπικά χαρακτηριστικά της βυζαντινής αγιογραφικής παράδοσης. Η πολυπλοκότητα των υλικών που χρησιμοποιούνται —όπως λεπτά φύλλα χρυσού για τα φωτοστέφανα, πολλαπλές στρώσεις χρωστικών για τα ενδύματα και τα σαρκώματα, και στρώματα προετοιμασίας— δημιουργεί ένα εξαιρετικά σύνθετο σύνολο υπερφασματικών δεδομένων. Σε τέτοιες περιπτώσεις, η απλή στοιχειακή

χαρτογράφηση, όπου εξετάζεται η κατανομή ενός μόνο στοιχείου κάθε φορά, αποδεικνύεται ανεπαρκής. Οι φασματικές υπογραφές των διαφόρων στοιχείων που αλληλεπικαλύπτονται (π.χ. οι κορυφές L του μολύβδου (*Pb*)), ενώ λεπτά στρώματα υλικών μπορούν να "κρυφτούν" κάτω από παχύτερα, καθιστώντας την ερμηνεία των δεδομένων μια σημαντική πρόκληση.

Κύριος στόχος της μελέτης αυτής είναι η ανάπτυξη και εφαρμογή μιας ολοκληρωμένης μεθοδολογικής προσέγγισης που συνδυάζει προηγμένες τεχνικές πολυμεταβλητής στατιστικής ανάλυσης (*PCA, NMF*) και αλγορίθμους μηχανικής μάθησης (*Clustering*) για την "αποκωδικοποίηση" του

υπερφασματικού κύβου δεδομένων. Η στρατηγική αυτή αποσκοπεί όχι μόνο στην ταυτοποίηση των επιμέρους



Εικόνα 18: Η εικόνα του Αγίου Ιωάννη του Προδρόμου, αντικείμενο της μελέτης χαμηλής ανάλυσης. Η εμφανής φθορά στο έργο καθιστά την μη-καταστροφική ανάλυση MA-XRF απαραίτητη για τη διερεύνηση της υλικής του σύστασης.

υλικών, αλλά και στην κατανόηση της τεχνικής του καλλιτέχνη, μέσω της δημιουργίας

θεματικών χαρτών που αποκαλύπτουν τη χωρική κατανομή και τη συνύπαρξη των διαφόρων χημικών ενώσεων.

Η ανάλυση των δεδομένων **μ -XRF** πραγματοποιήθηκε μέσω ειδικά αναπτυγμένου αλγορίθμου, ο οποίος σχεδιάστηκε για την αποδοτική επεξεργασία του μεγάλου όγκου πειραματικών δεδομένων.

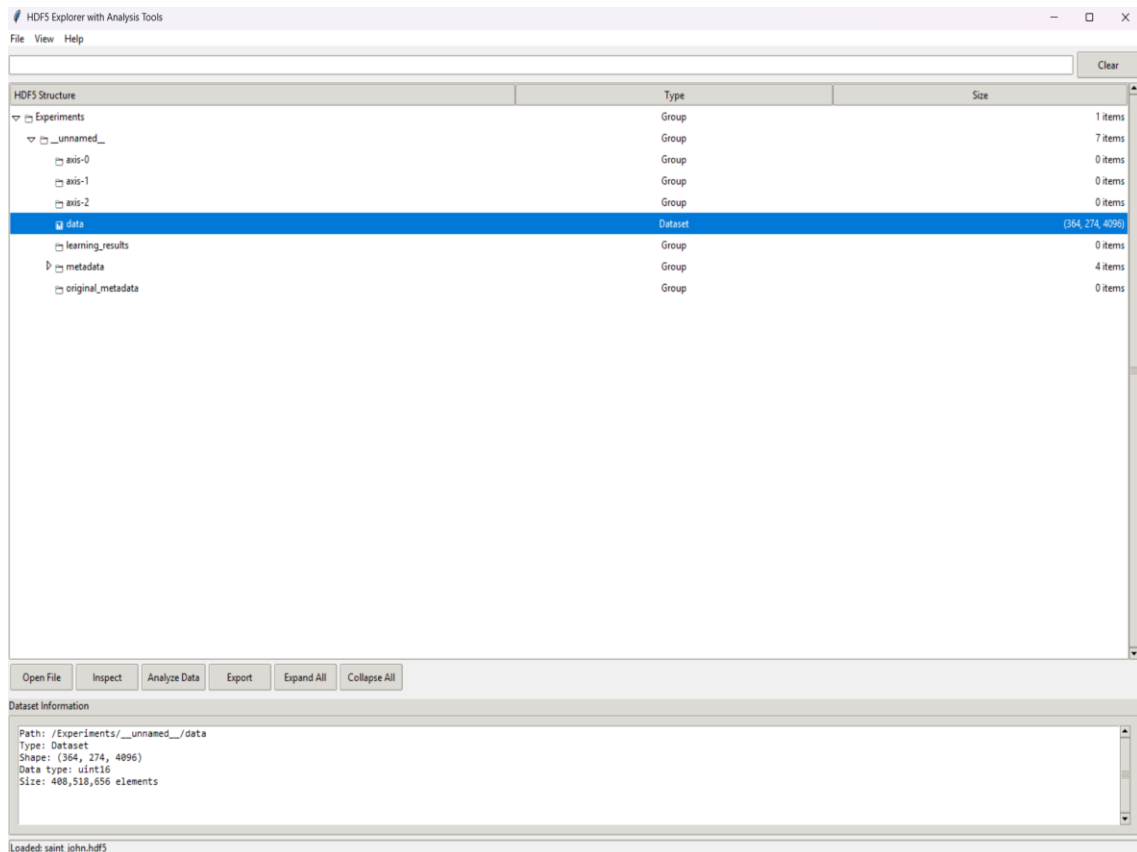
Για την ανάλυση, αξιοποιήθηκαν υπερφασματικά δεδομένα που συλλέχθηκαν με τη τεχνική της Απεικονιστικής Φασματοσκοπίας Φθορισμού Ακτίνων-X (**Macro X-Ray Fluorescence, MA-XRF**). Τα δεδομένα, με αρχικές διαστάσεις **274×364 pixels** και **4096** ενεργειακά κανάλια ανά **pixel**, υποβλήθηκαν σε πολυεπίπεδη ανάλυση με στόχο την εξαγωγή κρίσιμων πληροφοριών σχετικά με τη σύνθεση και τη δομή των υλικών κατασκευής του έργου τέχνης. Η τάση λυχνίας ρυθμίστηκε στα **50 kV** και η άνοδος στα **600 μ A** με άνοδο **Rh**. Τα pixel είχαν μέγεθος **500 μ m** και ο συνολικός αριθμός των pixel της εικόνας είναι **99736**. Ο χρόνος μέτρησης σε κάθε pixel ήταν **50ms**. Τέλος ο πίνακας ο οποίος αναφέρει περιληπτικά τα στοιχεία και τις μεταβλητές της πειραματικής διάταξης παρέχεται στην Εικόνα 24.

Ο αλγόριθμος υλοποιήθηκε με στόχο την αποτελεσματική ανάγνωση και επεξεργασία αυτού του όγκου δεδομένων. Αρχικά, πραγματοποιείται αποδοτική ανάγνωση των **HDF5** στοιβάδων. Σημαντικό βήμα αποτελεί η αυτόματη βαθμονόμηση ενέργειας, η οποία βασίζεται σε γνωστές στοιχειακές κορυφές αναφοράς. Στη συνέχεια, ο αλγόριθμος εκτελεί χωρική ανασύνθεση των φασμάτων σε δισδιάστατους χάρτες έντασης, επιτρέποντας την οπτικοποίηση της κατανομής των στοιχείων.

Η επεξεργασία αυτού του όγκου δεδομένων απαιτεί σημαντικούς υπολογιστικούς πόρους, με απαίτηση μνήμης της τάξης των **15 GB** για την πλήρη φόρτωση του **dataset**. Σε σύστημα με **32-core CPU**, ο χρόνος επεξεργασίας κυμαίνεται μεταξύ **8-10** λεπτών, χάρη στη βελτιστοποίηση των αλγορίθμων για ταχεία φασματική αποσύνθεση και παράλληλη επεξεργασία pixel.

Αυτή η ολοκληρωμένη μεθοδολογία επιτρέπει την πλήρη αξιοποίηση της μετρητικής διαμόρφωσης, παρέχοντας ακριβή ποσοτικοποίηση των στοιχειακών συγκεντρώσεων και ενεργοποιώντας ολοκληρωμένη χωρο-φασματική ανάλυση σε πολλαπλές κλίμακες. Η προσέγγιση αυτή συνδυάζει την ακρίβεια της φασματικής ανάλυσης με την ισχύ της χωρικής ανάλυσης, προσφέροντας μια ολοκληρωμένη λύση για τη μελέτη πολύπλοκων δειγμάτων.

Επομένως για την ανάλυση προχωρούμε στην ανάγνωση των δεδομένων από το πρόγραμμα που όπως φαίνεται στην Εικόνα 19.



Εικόνα 19: Διαδικασία επιλογής δεδομένων και αριθμητικής ακρίβειας στο λογισμικό NICE.

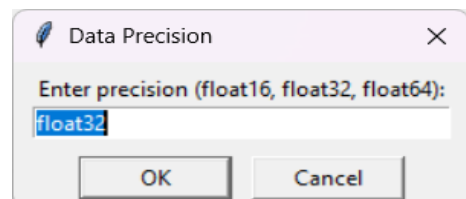
Όπως παρατηρούμε τα δεδομένα έχουν διαστάσεις **(364,274,4096)** και μέγεθος **408.518.656** δεδομένα με κωδικοποίηση **unit16** (Εικόνα 19). Οπότε πατάμε διπλό κλικ για ανάγνωση των δεδομένων.

Εάν δεν επιλέξουμε ορθό τύπο δεδομένων τότε προκύπτει το εξής αναδυόμενο παράθυρο προειδοποιώντας μας για το εκάστοτε σφάλμα (Εικόνα 20).



Εικόνα 20: Προειδοποίηση προς τον χρήστη όταν επιλέγεται ένας τύπος δεδομένων Group αντί για Dataset.

Με την επιλογή του κατάλληλου τύπου δεδομένων προκύπτει το αναδυόμενο παράθυρο για επιλογή ακρίβειας (Εικόνα 21). Η ακρίβεια μπορεί να βρίσκεται σε τρεις (3) μορφές **float**, είτε **16,32** ή **64**. Εξαρτάται από τις δυνατότητες του εκάστοτε συστήματος. Στην περίπτωση αυτή λόγω των ελάχιστων δυνατοτήτων του Υπολογιστικού συστήματος επιλέγουμε την μορφή **float16**. Εάν όμως δεν επιλεγεί κάποια προεπιλεγμένη ακρίβεια τότε έχει ρυθμιστεί το



Εικόνα 21: Ο διάλογος "Data Precision" που επιτρέπει στον χρήστη να επιλέξει την ακρίβεια κινητής υποδιαστολής (float16, float32, float64) για την ανάλυση.

πρόγραμμα για την ακρίβεια *float32*.

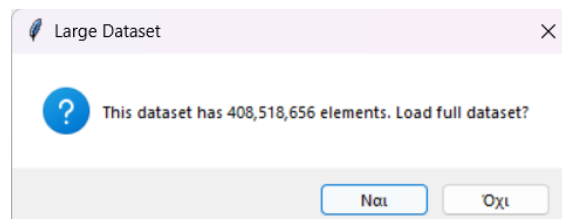
1. **Μέγεθος (bits):**

- float16 (ή half-precision): **sixteen (16) bits** (*1-bit* πρόσημο, *five (5) bits* εκθέτης)
- float32 (ή single precision): **thirty-two (32) bits** (*1-bit* πρόσημο, *eight (8) bits* εκθέτης)
- float64 (ή double precision): **sixty-four (64) bits** (*1-bit* πρόσημο, *eleven (11) bits* εκθέτης)

2. **Ακρίβεια (δεκαδικά ψηφία):**

- *float16*: ~3-4 δεκαδικά ψηφία *2 byte/τιμή*
- float32: ~7-8 δεκαδικά ψηφία *4 byte/τιμή*
- *float64*: ~15-16 δεκαδικά ψηφία *8 byte/τιμή*

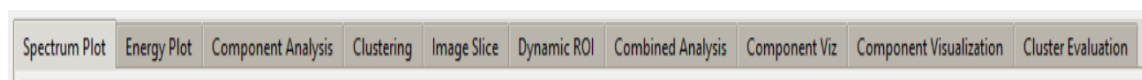
Έπειτα από την εφαρμογή της επιθυμητής ακρίβειας ένα αναδυόμενο παράθυρο μας ειδοποιεί για την ανάγνωση του μεγέθους των δεδομένων (Εικόνα 22).



Εικόνα 22: Ο διάλογος επιβεβαίωσης για τη φόρτωση μεγάλου συνόλου δεδομένων.

Οπότε επιλέγουμε ΝΑΙ και μας ανοίγει το παράθυρο της ανάλυσης.

Πάνω στο παράθυρο έχουμε τις εξής καρτέλες, όπου και είναι υπεύθυνες για τις αναλύσεις μας (Εικόνα 23).



Εικόνα 23: Το σύστημα καρτελών που παρέχει πρόσβαση στις διάφορες αναλυτικές λειτουργίες.

Η καθεμιά εκτελεί μια διεργασία είτε εξαρτωμένη από την προηγούμενη ή από μονή της με τα δεδομένα μας.

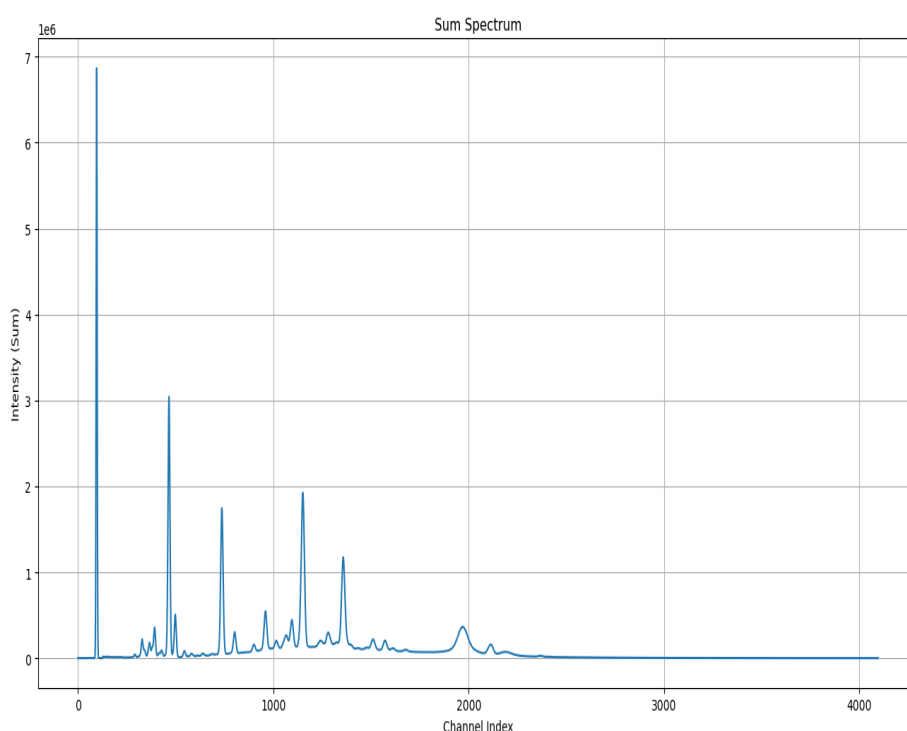
MAP INFORMATION		
Mapping parameters		
Width:	274	pixel
	137	mm
Height:	364	pixel
	182	mm
Pixel Size:	500	μm
Total number of pixel:	99736	pixel
Acquisition parameters		
Frame count:	1	
Pixel time:	50	ms/pixel
Measure time:	1:13	h
Overall time:	1:37	h
Stage speed:	10.0	mm/s
Stage position (X,Y,Z):	247.167;295.326;33.735	mm
Tube parameter		
High voltage:	50	kV
Anode current:	600	μA
Filter:	Empty	
Optic:	Lens	
Collimator diameter:	0	
SpotSize:	580	
Chamber at:	Air 1000	mbar
Flow rate:	---	l/min
Anode:	Rh	
Detector parameters		
Selected detectors:	1	
Max. pulse throughput:	275000	cps

Εικόνα 24: Πίνακας με τα μεταδεδομένα της πειραματικής μέτρησης (Mapping parameters).

Η βασική ροή εργασιών του αλγορίθμου περιλαμβάνει τα εξής στάδια:

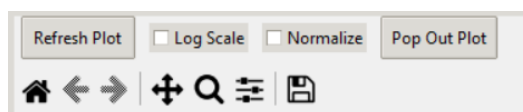
1. **Ανάγνωση και Προεπεξεργασία:** Αποδοτική ανάγνωση των δεδομένων από αρχεία **HDF5**. Εφαρμογή φίλτρου **Savitzky-Golay** (παράθυρο **11** σημείων, πολώνυμο **3ης τάξης**) για την εξομάλυνση των φασμάτων και τη μείωση του τυχαίου θορύβου.
2. **Ανάλυση σε Γραμμική Κλίμακα:** Εφαρμογή **PCA** και **NMF** στα αρχικά δεδομένα έντασης για τον εντοπισμό των κύριων πηγών φασματικής μεταβλητότητας.
3. **Ανάλυση σε Λογαριθμική Κλίμακα:** Επανάληψη της ανάλυσης μετά από λογαριθμικό μετασχηματισμό των δεδομένων έντασης ($\log(\text{ένταση}+1)$) για την ενίσχυση των σημάτων από στοιχεία σε χαμηλές συγκεντρώσεις.
4. **Υβριδική Ανάλυση NMF-PCA:** Συνδυασμός των δύο μεθόδων για βέλτιστο διαχωρισμό των υλικών.
5. **Συσταδοποίηση (Clustering):** Εφαρμογή αλγορίθμων ομαδοποίησης (κυρίως **K-Means**) στα αποτελέσματα των παραπάνω αναλύσεων για τη δημιουργία θεματικών χαρτών υλικών.
6. **Στοιχειακή Χαρτογράφηση:** Επιβεβαίωση των αποτελεσμάτων μέσω της δημιουργίας παραδοσιακών χαρτών κατανομής για συγκεκριμένα χημικά στοιχεία.

Η αρχική επισκόπηση των φασματικών δεδομένων πραγματοποιήθηκε μέσω της δημιουργίας του αθροιστικού φάσματος (**Sum Spectrum**) Εικόνα 25, το οποίο υπολογίστηκε και απεικονίζεται με χρήση του αναπτυχθέντος λογισμικού. Τα αρχικά φάσματα υποβλήθηκαν σε προεπεξεργασία με την εφαρμογή φίλτρου **Savitzky-Golay** (παράθυρο **II** σημείων, πολώνυμο 3ης τάξης). Αυτό το βήμα είναι κρίσιμο για τη μείωση του τυχαίου θορύβου στα φάσματα, βελτιώνοντας έτσι τον λόγο σήματος προς θόρυβο (**SNR**) και επιτρέποντας την ακριβέστερη ανίχνευση κορυφών που αντιστοιχούν σε στοιχεία χαμηλής συγκέντρωσης. Το αθροιστικό φάσμα (**Sum Spectrum**) που φαίνεται, όντως επιβεβαιώνει την επιτυχία αυτής της διαδικασίας, καθώς οι φασματικές κορυφές εμφανίζονται ευδιάκριτες πάνω από ένα ομαλοποιημένο υπόβαθρο.



Εικόνα 25: Η καρτέλα "Spectrum Plot" εντός του AnalysisWindow. Απεικονίζεται το αθροιστικό φάσμα (**Sum Spectrum**)

Από το παράθυρο κατευθείαν μπορούμε να δούμε το φάσμα με βάση τα κανάλια και την ένταση. Επιπρόσθετα μας δίνονται η επιλογές να ορίσουμε το φάσμα σε λογαριθμική μορφή ή να το κανονικοποιήσουμε. Τέλος μπορούμε να το αποθηκεύσουμε ή να το ανοίξουμε σε ξεχωριστό παράθυρο Εικόνα 26.



Εικόνα 26: Διαδραστικά στοιχεία ελέγχου για την ανανέωση, την αλλαγή κλίμακας (**Log Scale**), την κανονικοποίηση και την εμφάνιση του γραφήματος σε ξεχωριστό παράθυρο (**Pop Out Plot**).

Το φάσμα αυτό, που αντιστοιχεί στο άθροισμα όλων των καταγεγραμμένων φασμάτων από κάθε pixel της εικόνας, παρουσιάζει σημαντικά μειωμένο θόρυβο λόγω της προηγμένης προεπεξεργασίας *Savitzky-Golay*.

Επομένως μπορούμε να προχωρήσουμε την περαιτέρω ανάλυση των δεδομένων και την εξαγωγή ουσιαστικών πληροφοριών. Εφαρμόστηκαν μέθοδοι Ανάλυσης Συνιστωσών (*Component Analysis*) και Ομαδοποίησης (*Clustering Analysis*). Οι αναλύσεις αυτές πραγματοποιήθηκαν στα δεδομένα σε γραμμική κλίμακα έντασης. Αν και η ανάλυση σε λογαριθμική κλίμακα (π.χ., *log(ένταση+1)*), όπως υποστηρίζεται από τον κώδικα) μπορεί να ενισχύσει σήματα από στοιχεία σε χαμηλές συγκεντρώσεις, για την παρούσα μελέτη προτιμήθηκε η γραμμική κλίμακα για την αμεσότερη συσχέτιση με τις σχετικές ποσότητες των υλικών. Οι κύριες στρατηγικές πολυμεταβλητής ανάλυσης που ακολουθήθηκαν ήταν:

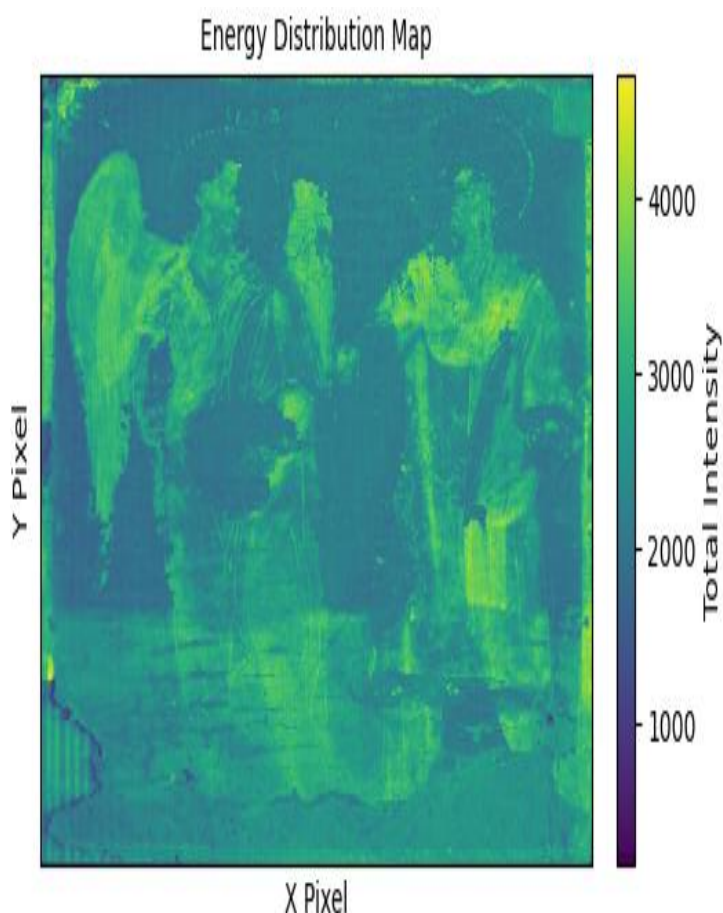
1. Ανάλυση Κύριων Συνιστωσών (*PCA*) επί των αρχικών δεδομένων.
2. Μη Αρνητική Παραγοντοποίηση Πινάκων (*NMF*) επί των αποτελεσμάτων της *PCA* και ξεχωριστά.
3. Ανάλυση Ομαδοποίησης επί των αποτελεσμάτων (*scores*) της *PCA*, της *NMF* και του βήματος 2.
4. Πραγματοποίηση των αναλύσεων και σε λογαριθμική Κλίμακα και σύγκριση με την γραμμική.

Γραμμική Ανάλυση των δεδομένων

Η αρχική ανάλυση πραγματοποιήθηκε στα δεδομένα έντασης σε γραμμική κλίμακα, ώστε να διερευνηθούν οι κυρίαρχες χημικές φάσεις, αρχίζοντας με την απεικόνιση της ενέργειας.

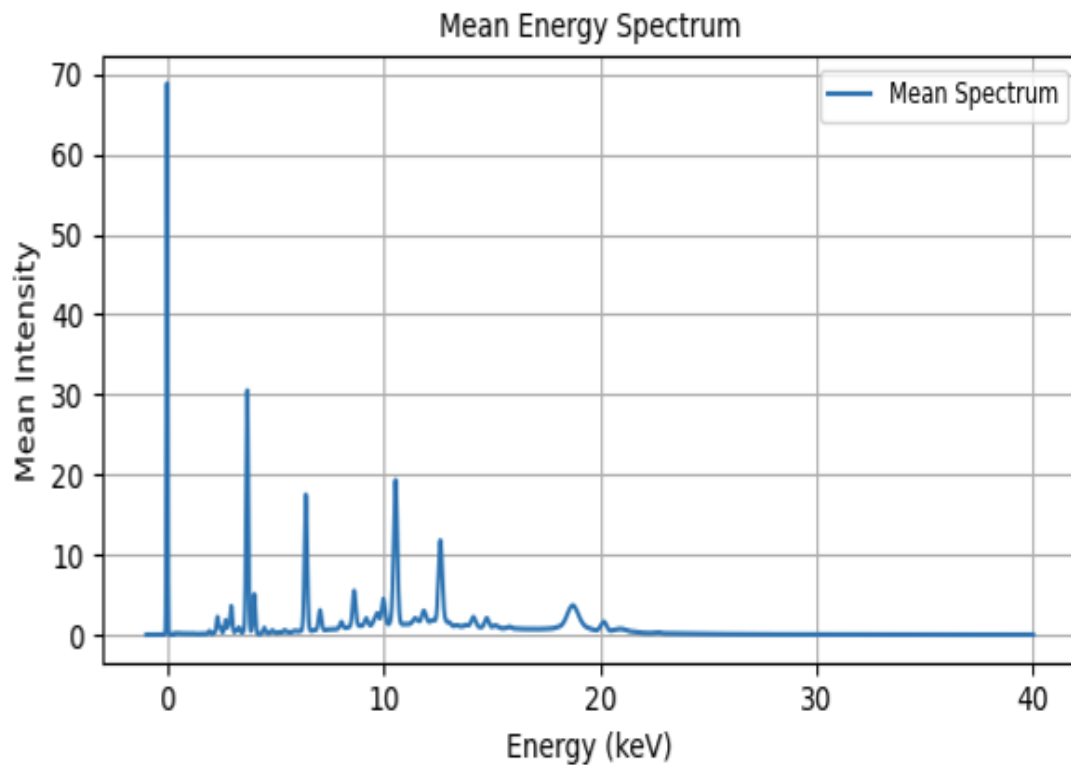
Το πρώτο διάγραμμα παρουσιάζει το μέσο φάσμα *EDX (Energy-Dispersive X-ray Spectroscopy)* (Εικόνα 28, όπου καταγράφεται η μέση ένταση των ανιχνευθέντων ακτίνων-X συναρτήσει της ενέργειας (σε *keV*) σε ολόκληρη την περιοχή της εικόνας. Οι κορυφές που παρατηρούνται στο φάσμα αντιστοιχούν στις χαρακτηριστικές ενεργειακές στάθμες ακτίνων-X των στοιχείων που υπάρχουν στο δείγμα. Η θέση και το ύψος αυτών των κορυφών προσφέρουν ποσοτικές και ποιοτικές πληροφορίες σχετικά με τη στοιχειακή σύσταση του δείγματος.

Παρακάτω παρέχουμε το διάγραμμα και την εικόνα του Αγίου Ιωάννη με βάση την ένταση Εικόνα 27. Η εικόνα παρουσιάζει τον χάρτη της ολικής ενεργειακής κατανομής των υπερφασματικών δεδομένων. Ο χάρτης αυτός προκύπτει από την άθροιση της έντασης φθορισμού σε όλα τα ενεργειακά κανάλια για κάθε εικονοστοιχείο (*pixel*) της σαρωμένης περιοχής. Οι φωτεινότερες περιοχές στον χάρτη αντιστοιχούν σε υψηλότερη συνολική εκπομπή ακτίνων-*X*. Αυτό μπορεί να υποδηλώνει την παρουσία στοιχείων με υψηλότερο ατομικό αριθμό (*Z*), τα οποία τείνουν να φθορίζουν εντονότερα, ή/και μεγαλύτερη συγκέντρωση ή πάχος υλικού στις συγκεκριμένες περιοχές.



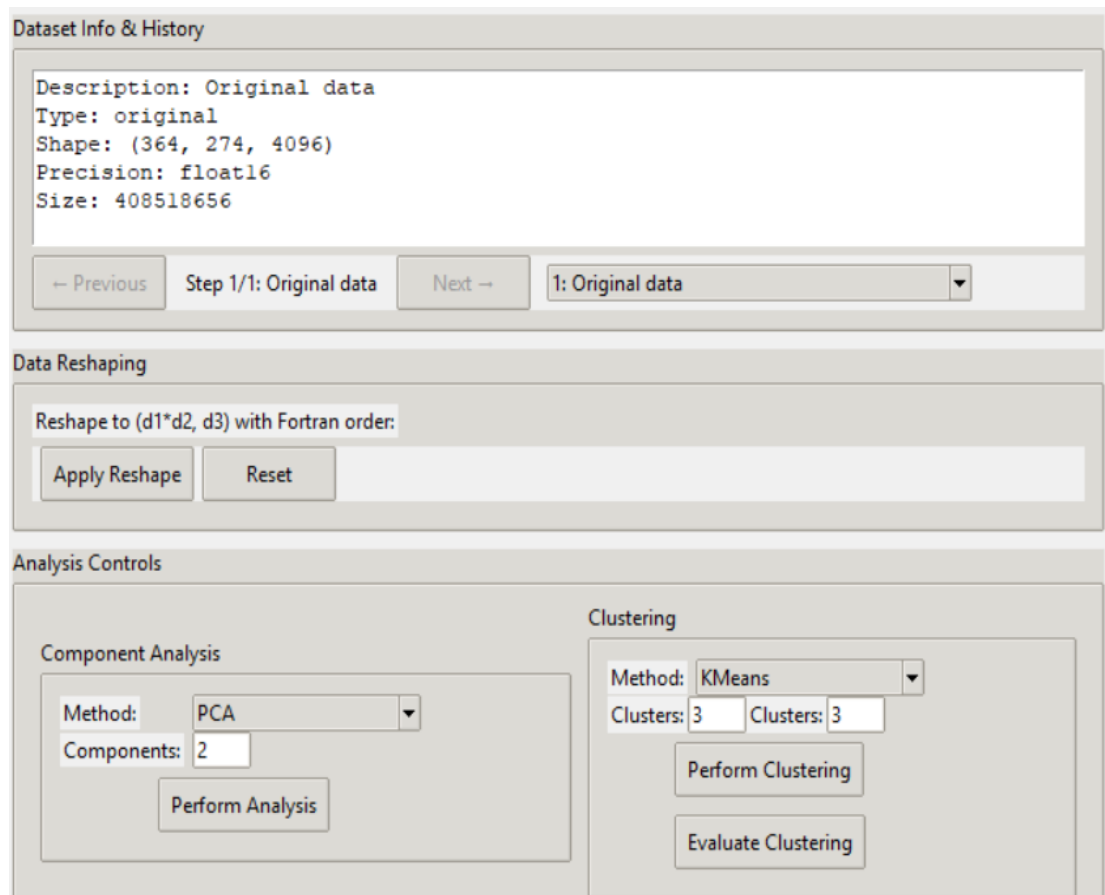
Εικόνα 27: Χάρτης συνολικής έντασης φθορισμού (*Energy Distribution Map*) για την εικόνα του Αγίου Ιωάννη. Οι φωτεινότερες περιοχές (κίτρινο) αντιστοιχούν σε υψηλότερη συνολική εκπομπή ακτίνων-*X*, υποδεικνύοντας την παρουσία στοιχείων με υψηλότερο ατομικό αριθμό ή/και μεγαλύτερο πάχος υλικού.

Στην παρούσα εικόνα, παρατηρείται ότι οι περιοχές που αντιστοιχούν στα πρόσωπα των Αγίων, καθώς και στο περίγραμμα των φωτοστέφανων, εμφανίζουν σημαντικά υψηλότερη συνολική ένταση σε σύγκριση με τις περιοχές του φόντου. Η παρατήρηση αυτή υποδηλώνει την πιθανή χρήση πυκνότερων υλικών, στρωμάτων προετοιμασίας πλούσιων σε βαρέα στοιχεία, ή χρωστικών που περιέχουν στοιχεία υψηλού *Z* (π.χ., μόλυβδο, υδράργυρο, χρυσό) στις εν λόγω περιοχές. Η ακριβής φύση αυτών των υλικών θα διερευνηθεί περαιτέρω με τις επόμενες φασματοσκοπικές και στατιστικές αναλύσεις. Ωστόσο περιοχές όπως το φωτοστέφανο που είναι εμφανώς από χρυσό φαίνονται σκοτεινές. Υποδηλώνει δηλαδή πως το στοιχείο που βρίσκεται στο υπόβαθρο είναι κύριος μέτοχος του φάσματος (μεγαλύτερο σήμα) και το στρώμα χρυσού σε εκείνη την περιοχή είναι αρκετά λεπτό.



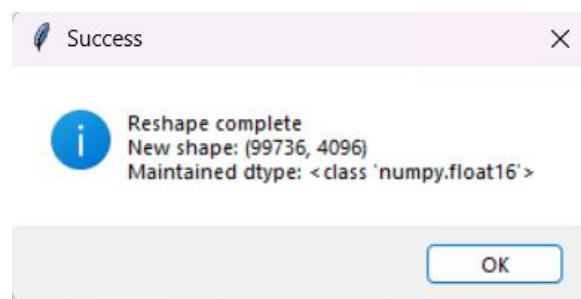
Εικόνα 28: Το μέσο φάσμα φθορισμού (Mean Energy Spectrum) που προκύπτει από τον υπολογισμό του μέσου όρου όλων των επιμέρους φασμάτων της εικόνας. Οι κορυφές αντιστοιχούν στις χαρακτηριστικές γραμμές εκπομπής των κυρίαρχων στοιχείων.

Εάν κινηθούμε στο κάτω μέρος του παραθύρου παρατηρούμε το πλαίσιο **Dataset Info & History** το οποίο μας αναφέρει για τον τύπο των δεδομένων που έχουμε, το μέγεθος, την ακρίβεια που επιλέξαμε και τις διαστάσεις Εικόνα 29. Έπειτα έχουμε την δυνατότητα να επιλέξουμε τα βήματα, δηλαδή, την μορφή των δεδομένων που θέλουμε να κάνουμε ανάλυση. Αυτό το αντικείμενο είναι ιδιαίτερα χρήσιμο διότι μπορούμε να επιλέξουμε δεδομένα που είτε έχουμε κάνει ήδη κάποια ανάλυση ώστε να πραγματοποιήσουμε άλλη είτε να ανατρέξουμε ξανά στα αρχικά.



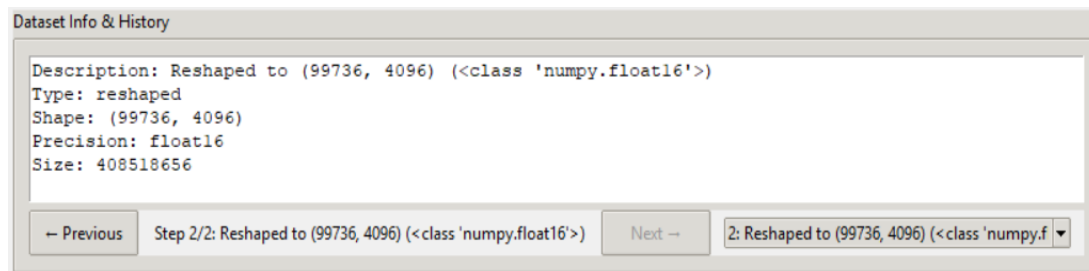
Εικόνα 29: Διαδικασία ανασχηματισμού (Reshape) των δεδομένων στο AnalysisWindow. (Πάνω) Ο πίνακας ελέγχου πριν και μετά την εφαρμογή, όπου οι τρισδιάστατες διαστάσεις (364, 274, 4096) μετατρέπονται σε δισδιάστατο πίνακα (99736, 4096) κατάλληλο για την ανάλυση.

Για την ανάλυση μας, προχωρούμε στην εφαρμογή του **Reshape** στα δεδομένα μας με την μορφή $(x * y, z)$ με z να είναι το συνολικό μέγεθος των καναλιών. Εάν πατήσουμε το κουμπί **Apply Reshape** με την εφαρμογή του αλγορίθμου προκύπτει και το κατάλληλο αναδυόμενο παράθυρο. (99.736 pixels $\times$ 4.096 κανάλια) Εικόνα 30.



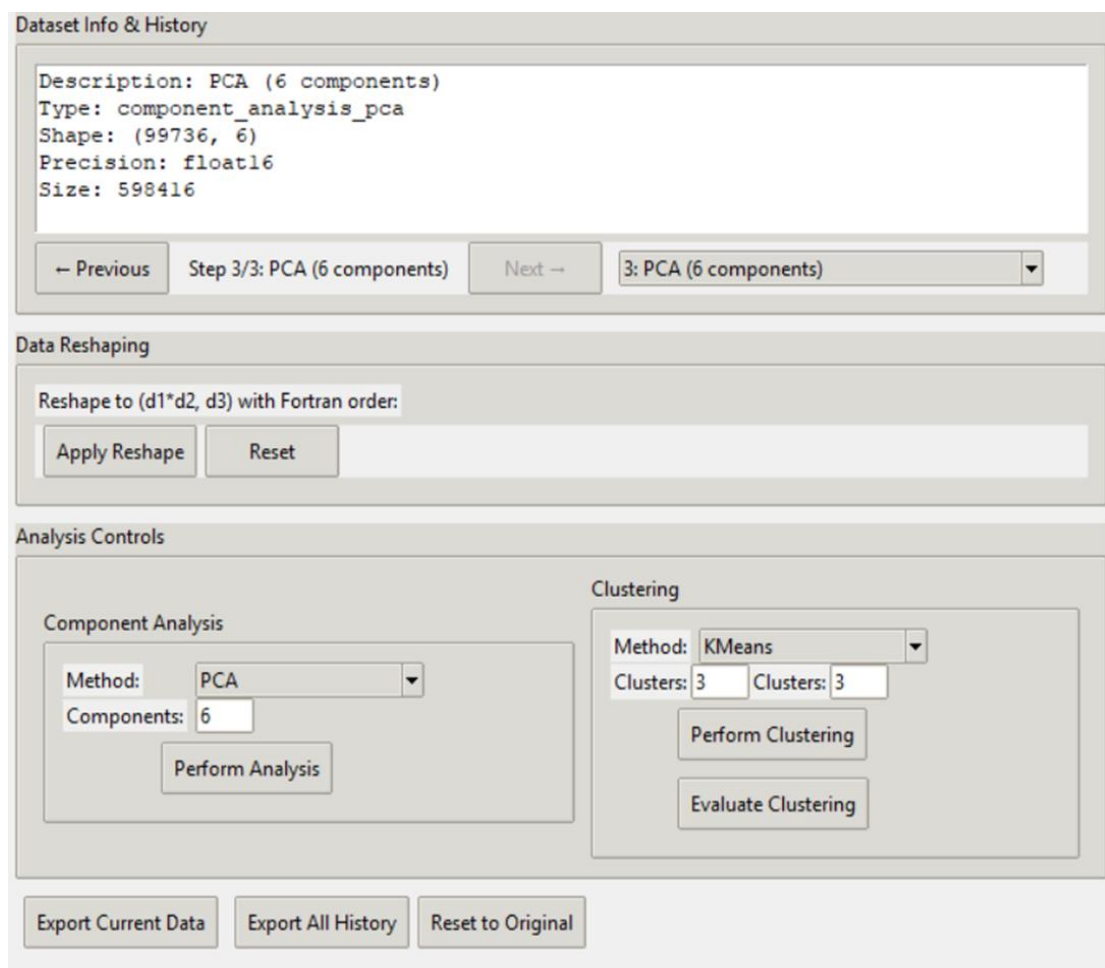
Εικόνα 30: Το μήνυμα επιβεβαίωσης της επιτυχούς ολοκλήρωσης.

Τώρα βρισκόμαστε σε θέση να πραγματοποιήσουμε οποιαδήποτε μέθοδο συσταδοποίησης ή ανάλυσης σε κυρίους άξονες.



Εικόνα 31: Ενημερωμένο Παράθυρο Δεδομένων.

Όπως μπορεί να παρατηρησει καποιος το παραθυρο **Dataset Info & History** ενημερώθηκε με τον καινούριο τύπο δεδομένων, τις διαστάσεις τους και το μέγεθος τους Εικόνα 31. Επιπρόσθετα και το ιστορικό πλέον διαβάζει τα τελευταία δεδομένα έχοντας την δυνατότητα να επαναφέρει τα αρχικά εάν επιλεγθεί Εικόνα 32.



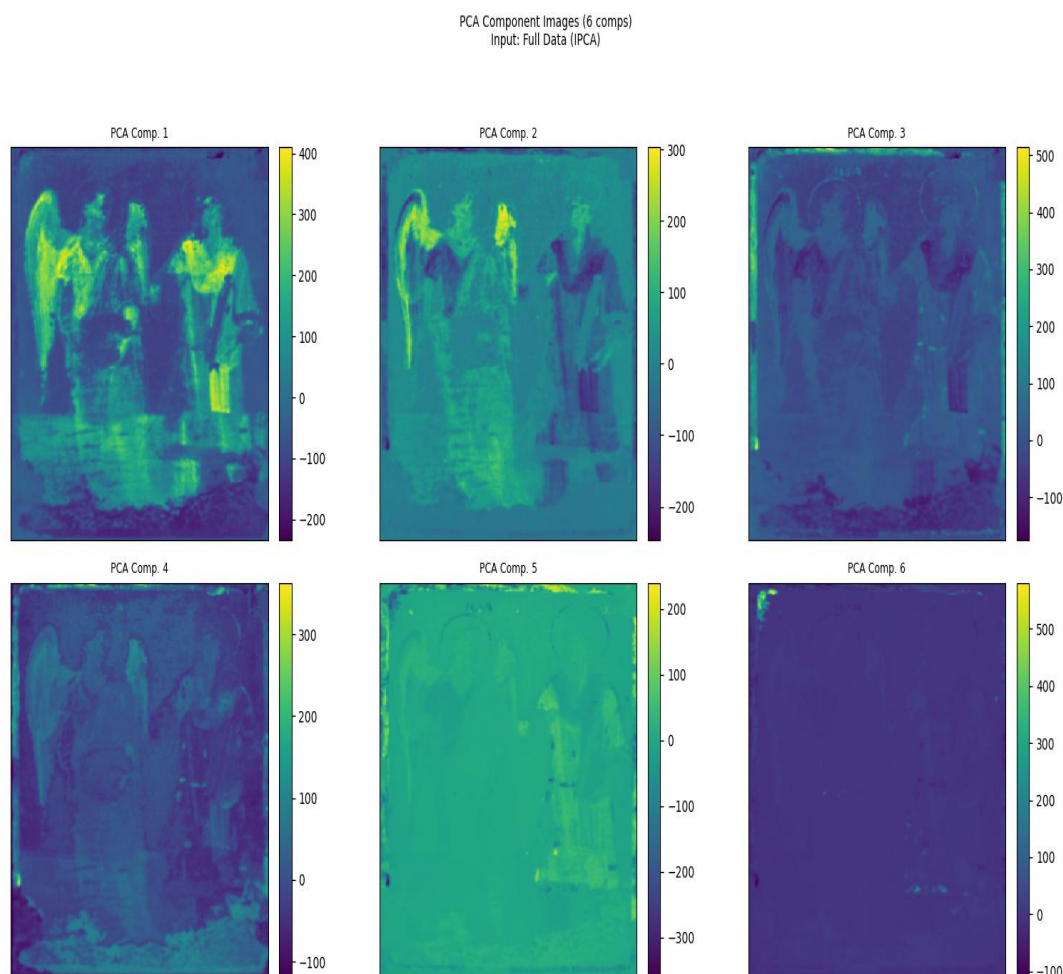
Εικόνα 32: Ιστορικό δεδομένων δύναται να επιλεγθεί

Για μια απλή ανάλυση θα προχωρήσουμε στην εφαρμογή της μεθόδου **PCA**, πάνω στα **Reshaped** δεδομένα. Όμως για την ορθή επιλογή **Components** πρέπει να απεικονίσουμε το **Explained Variance** για κάθε άξονα.

Ανάλυση Σε Κυρίους Άξονες (Principal Component Analysis):

Μπορούμε επομένως να προχωρήσουμε στην πραγματοποίηση του πρώτου βήματος για την ανάλυση μας την εφαρμογή της μεθόδου της **PCA** στα αρχικά μας δεδομένα

Εικόνα 33. Τέλος για έναν αριθμό **6 κυρίων αξόνων** η γενική πληροφορία που λαμβάνουμε βρίσκεται στο **80%** παρέχοντας μας βέβαια και την ελάχιστη δυνατή ώρα ανάλυσης μειώνοντας την στα **10 λεπτά**.



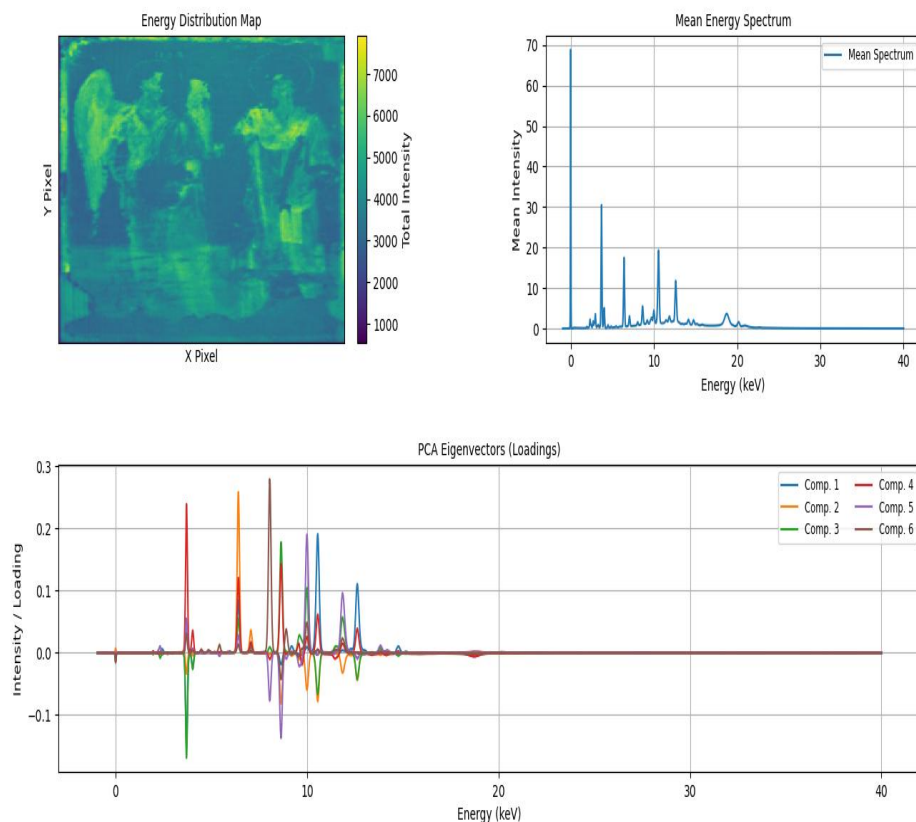
Εικόνα 33: Αποτελέσματα της Ανάλυσης Κύριων Συνιστωσών (PCA) στα λογαριθμικά μετασχηματισμένα δεδομένα.

Η εφαρμογή της Ανάλυσης Κύριων Συνιστωσών (**PCA**) στα δεδομένα **MA-XRF** αποτελεί το πρώτο βήμα για τη μείωση της διαστατικότητας του υπερφασματικού κύβου και την οπτικοποίηση των κύριων πηγών φασματικής διακύμανσης. Οι χάρτες των έξι πρώτων κύριων συνιστωσών (**PCs**) που προέκυψαν, οι οποίες αθροιστικά εξηγούν ένα σημαντικό ποσοστό της συνολικής πληροφορίας (**~83%**), παρέχουν μια πολυδιάστατη εικόνα της χημικής ετερογένειας του έργου.

- **PCA Comp. 1** (49% της διακύμανσης): Αυτή η συνιστώσα, ως η κυρίαρχη πηγή διακύμανσης, επιτυγχάνει έναν σαφή και ισχυρό διαχωρισμό μεταξύ του ζωγραφικού στρώματος και του υποστρώματος προετοιμασίας. Οι έντονες θετικές τιμές (κίτρινο) αντιστοιχούν στις μορφές των Αγίων, όπου κυριαρχούν οι χρωστικές με στοιχεία υψηλού ατομικού αριθμού (**Pb, Au, Fe**). Αντίθετα, οι αρνητικές τιμές (μωβ) εντοπίζονται κυρίως στο φόντο και στο εκτεθειμένο **gesso**. Η συνιστώσα αυτή ουσιαστικά αντιπροσωπεύει την ισχυρότερη αντισυσχέτιση στο σύνολο δεδομένων: αυτή μεταξύ των χρωστικών και του υποστρώματος (**gesso - CaSO₄**).

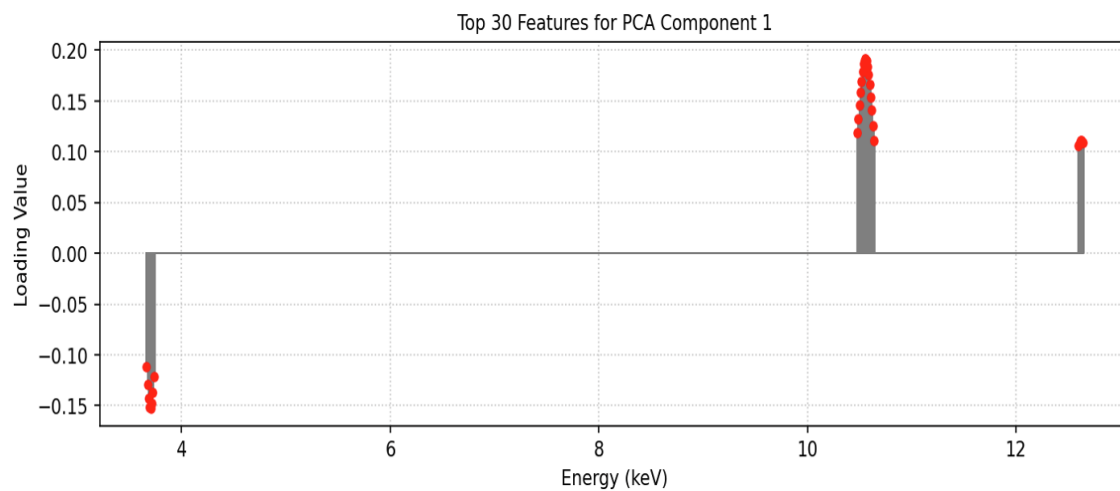
- **PCA Comp. 2 & 3** (δευτερεύουσα διακύμανση, π.χ. ~13% για την **PC2**): Αυτές οι συνιστώσες απομονώνουν πιο εξειδικευμένες χημικές υπογραφές, διαχωρίζοντας τις ίδιες τις χρωστικές μεταξύ τους.
 - Η **PC 2** φαίνεται να σχετίζεται έντονα με τις περιοχές υψηλότερης ακτινοβολίας, όπως τα φωτίσματα (**highlights**) στα ενδύματα, τα φτερά και τα σαρκώματα. Αυτό υποδηλώνει ότι καταγράφει τη διακύμανση που οφείλεται κυρίως στο Λευκό του Μολύβδου (**Pb**) και πιθανώς στον Χρυσό (**Au**).
 - Η **PC 3** χαρτογραφεί μια διαφορετική χημική ομάδα. Η κατανομή της φαίνεται να αντιστοιχεί στις καφέ-κόκκινες περιοχές των ενδυμάτων, υποδεικνύοντας ότι η συνιστώσα αυτή αντιπροσωπεύει τη φασματική υπογραφή των γήινων χρωστικών με βάση τον Σίδηρο (**Fe**).
- **PCA Comp. 4, 5 & 6** (λεπτές διαφοροποιήσεις & θόρυβος): Οι συνιστώσες υψηλότερης τάξης αποκαλύπτουν λεπτότερες διαφοροποιήσεις και, σταδιακά, τον εναπομείναντα θόρυβο.
 - Η **PC 4** και η **PC 5** καταγράφουν πολύ μικρότερη διακύμανση και φαίνεται να σχετίζονται με δευτερεύουσες χημικές διαφορές στο υπόβαθρο ή πολύ λεπτές αναμειγξεις χρωστικών.
 - Η **PC 6** αποτελείται σχεδόν αποκλειστικά από τυχαίες κηλίδες χωρίς αναγνωρίσιμη δομή. Αντί να απομονώνει μια συγκεκριμένη περιοχή όπως τα σαρκώματα, η συνιστώσα αυτή αντιπροσωπεύει κυρίως τον στατιστικό θόρυβο του συστήματος. Η μοναδική φασματική υπογραφή των σαρκωμάτων δεν απομονώνεται σε μία μόνο συνιστώσα, αλλά προκύπτει από τον συνδυασμό και την αλληλεπίδραση των κυρίαρχων συνιστωσών (π.χ., της **PC2** που σχετίζεται με τον Μόλυβδο και της **PC3** που σχετίζεται με τον Σίδηρο).

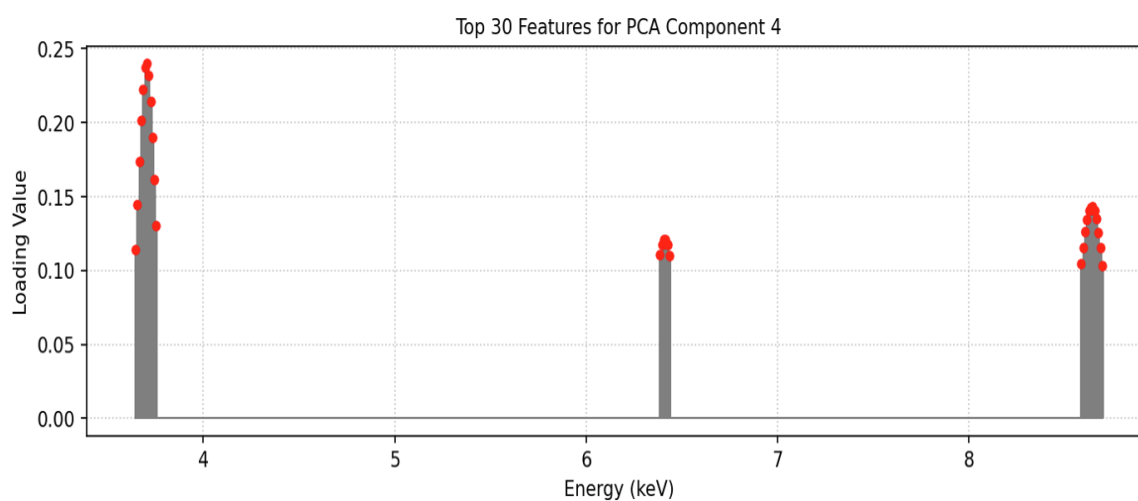
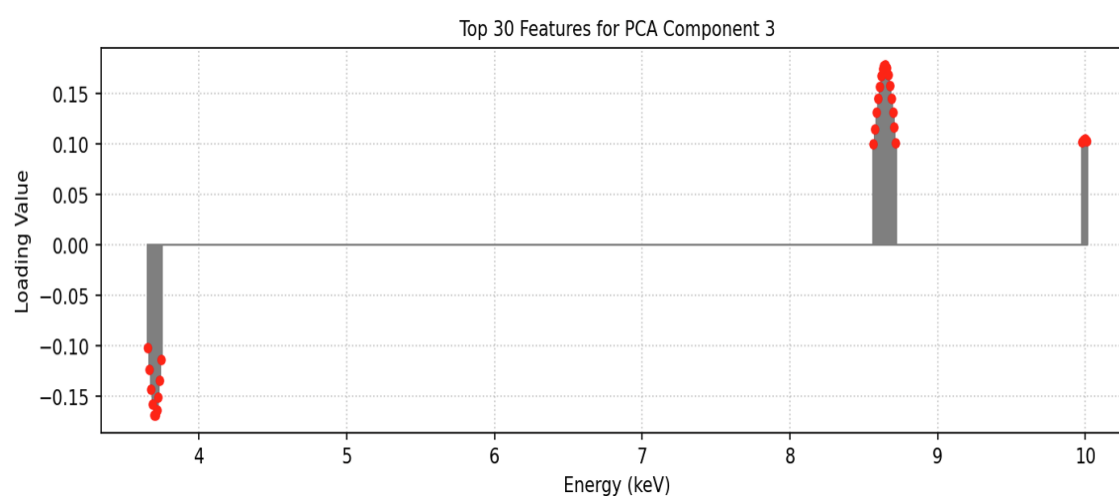
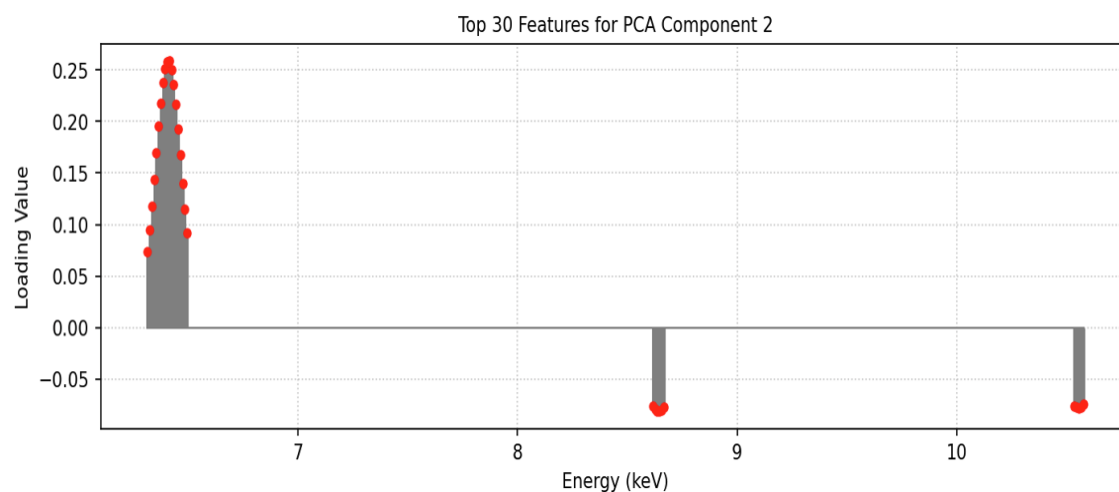
Είναι κρίσιμο να τονιστεί ότι η **PCA**, αν και ισχυρή, παρέχει ορθογώνιες (μη συσχετισμένες) συνιστώσες που περιέχουν τόσο θετικές όσο και αρνητικές τιμές, καθιστώντας την άμεση φυσικοχημική τους ερμηνεία δύσκολη. Η πραγματική τους αξία αναδεικνύεται όταν εξετάζονται σε συνδυασμό με τα ιδιοδιανύσματά τους (**loadings**), τα οποία αποκαλύπτουν ποιες ενεργειακές κορυφές (δηλαδή ποια στοιχεία) ευθύνονται για κάθε πηγή διακύμανσης. Η πολυπλοκότητα αυτή καθιστά αναγκαία τη μετάβαση σε άλλες μεθόδους, όπως η **NMF**, για την εξαγωγή πιο άμεσα ερμηνεύσιμων χαρτών αφθονίας υλικών. Η επιλογή των έξι συνιστωσών αποτελεί έναν βέλτιστο συμβιβασμό μεταξύ της διατήρησης της μέγιστης πληροφορίας (**83%**) και της διαχείρισης του υπολογιστικού κόστους, το οποίο αυξάνεται εκθετικά για συνιστώσες υψηλότερης τάξης. Παρακάτω μπορούμε να δούμε ένα παράθυρο του προγράμματος ανάλυσης το οποίο μας παρέχει ταυτόχρονα τους χάρτες τόσο για τα ιδιοδιανύσματα των κυρίων αξόνων της μεθόδου **PCA** επάνω στον άξονα της ενέργειας αλλά και την εικόνα εντασης ενέργειας δίπλα με το αντίστοιχο φάσμα Εικόνα 34. Ένα σημαντικό διάγραμμα που μας εμφανίζει ταυτόχρονα πως οι κύριοι άξονες αποτυπώνουν την πληροφορία μας σε σχέση με την ενέργεια και ποια πληροφορία ανήκει σε ποιον άξονα.

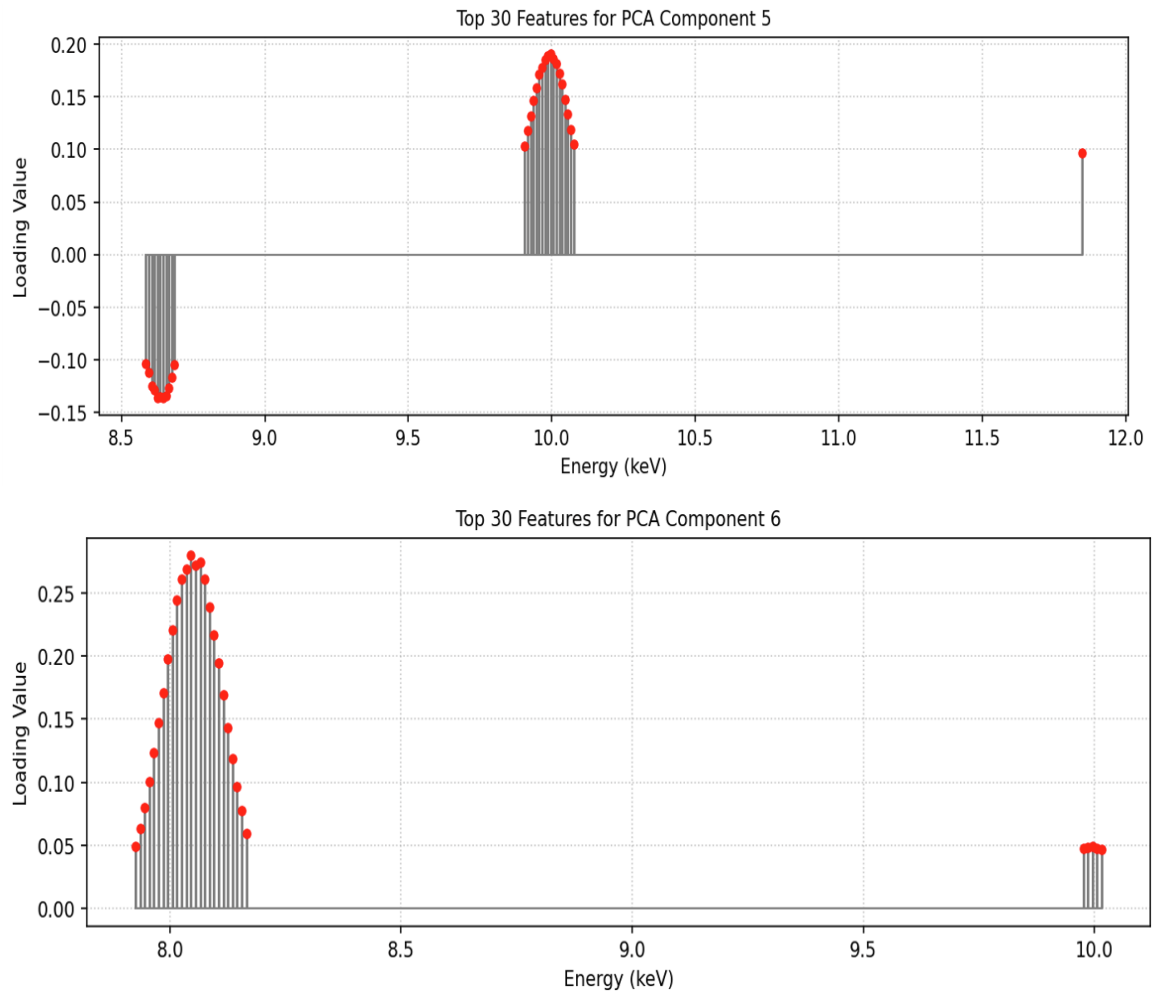


Εικόνα 34: Στιγμιότυπο της καρτέλας "Energy Plots" του αναπτυχθέντος λογισμικού. Απεικονίζονται ο χάρτης συνολικής έντασης (Energy Distribution Map), το μέσο φάσμα (Mean Energy Spectrum) και τα φάσματα των ιδιοδιανυσμάτων (PCA Eigenvectors) για τις έξι πρώτες κύριες συνιστώσες

Επιπρόσθετα μπορούμε να διαθέσουμε και το διάγραμμα με τα πρώτα **τριάντα (30)** κύρια στοιχεία κάθε άξονα της μεθόδου **PCA** σε σχέση με την ενέργεια, όπως φαίνεται παρακάτω Εικόνα 35.



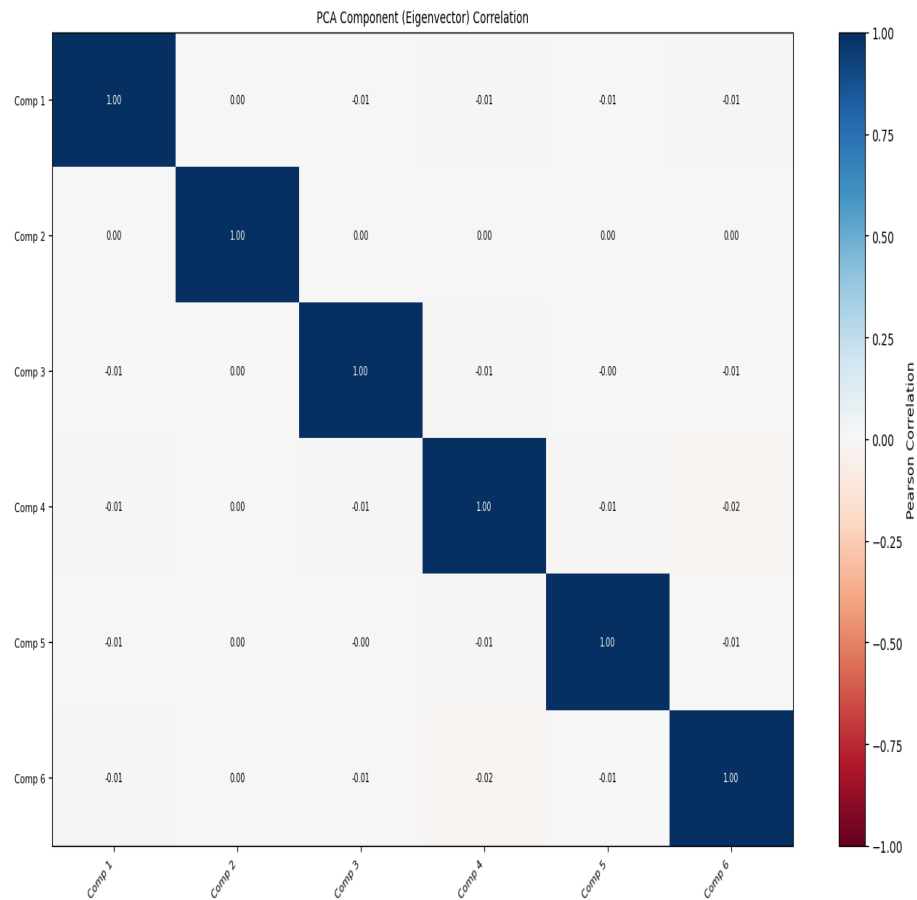




Εικόνα 35: Γραφικές παραστάσεις των 30 ενεργειακών καναλιών (features) με τη μεγαλύτερη συνεισφορά (loading value) για καθεμία από τις έξι πρώτες κύριες συνιστώσες (PCA Components 1-6). Τα διαγράμματα αυτά επιτρέπουν την ταυτοποίηση των φασματικών περιοχών που κυριαρχούν σε κάθε συνιστώσα.

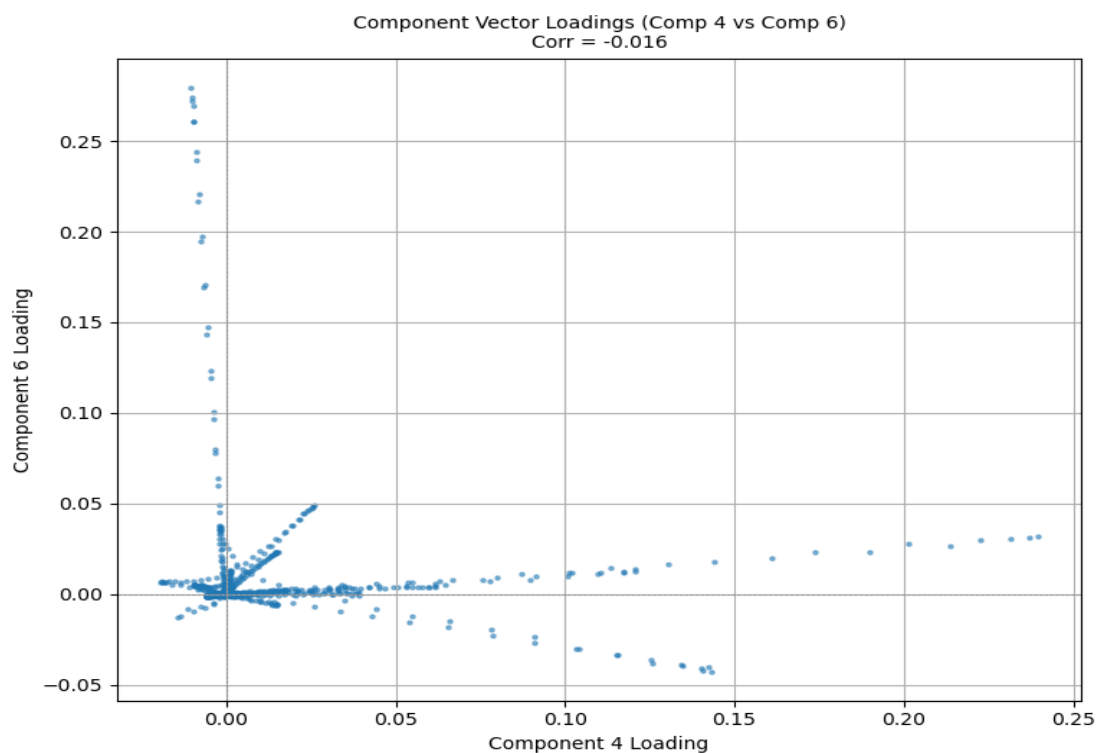
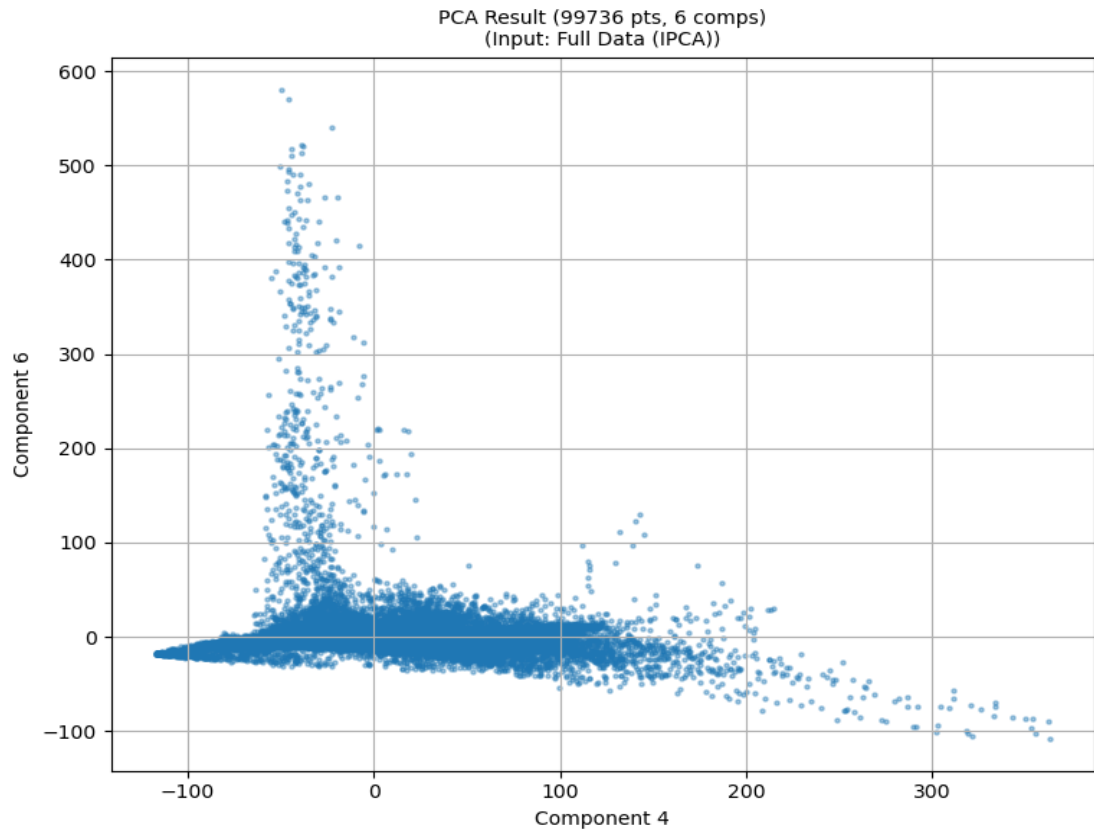
Εφόσον έχουμε απεικονίσει τα πρώτα **30 Features** κάθε άξονα σε σχέση με την ενέργεια μπορούμε να το συγκρίνουμε σε σχέση με το ολικό φάσμα συναρτήσει της ενέργειας και κατανοούμε ακριβώς ποια σημεία και ποιες ενέργειες συνεισφέρουν περισσότερο στους κυρίους άξονες Εικόνα 36.

Η ορθογωνιότητα (μηδενική συσχέτιση) των συνιστωσών επιβεβαιώνεται από τον πίνακα συσχέτισης (παρακάτω) Εικόνα 36, ωστόσο η άμεση φυσική τους ερμηνεία παραμένει δύσκολη λόγω των αρνητικών τιμών (anticorrelation).



Εικόνα 36: Πίνακας συσχέτισης (correlation matrix) μεταξύ των ιδιοδιανυσμάτων των έξι κύριων συνιστωσών. Η διαγώνιος με τιμές 1.0 και οι σχεδόν μηδενικές τιμές εκτός διαγωνίου επιβεβαιώνουν την ορθογωνιότητα των συνιστωσών της PCA.

Τέλος για μια ολοκληρωτική ανάλυση για την μέθοδο της **PCA** θα απεικονίσουμε και κάθε άξονα συναρτήσει των άλλων ξεχωριστά σε δισδιάστατη απεικόνιση όπως για παράδειγμα οι παρακάτω εικόνες (Εικόνα 37). Αυτή η μέθοδος μπορεί να αποβεί αρκετά χρήσιμη όταν εφαρμοστεί και η μέθοδος **RoI** δείχνοντας ακριβώς ποια σημεία μας παρέχουν ποια πληροφορία, επάνω στην εικόνα όπως και ποιες ενέργειες είναι υπεύθυνες με τα κατάλληλα στοιχεία στους στοιχειακούς χάρτες.



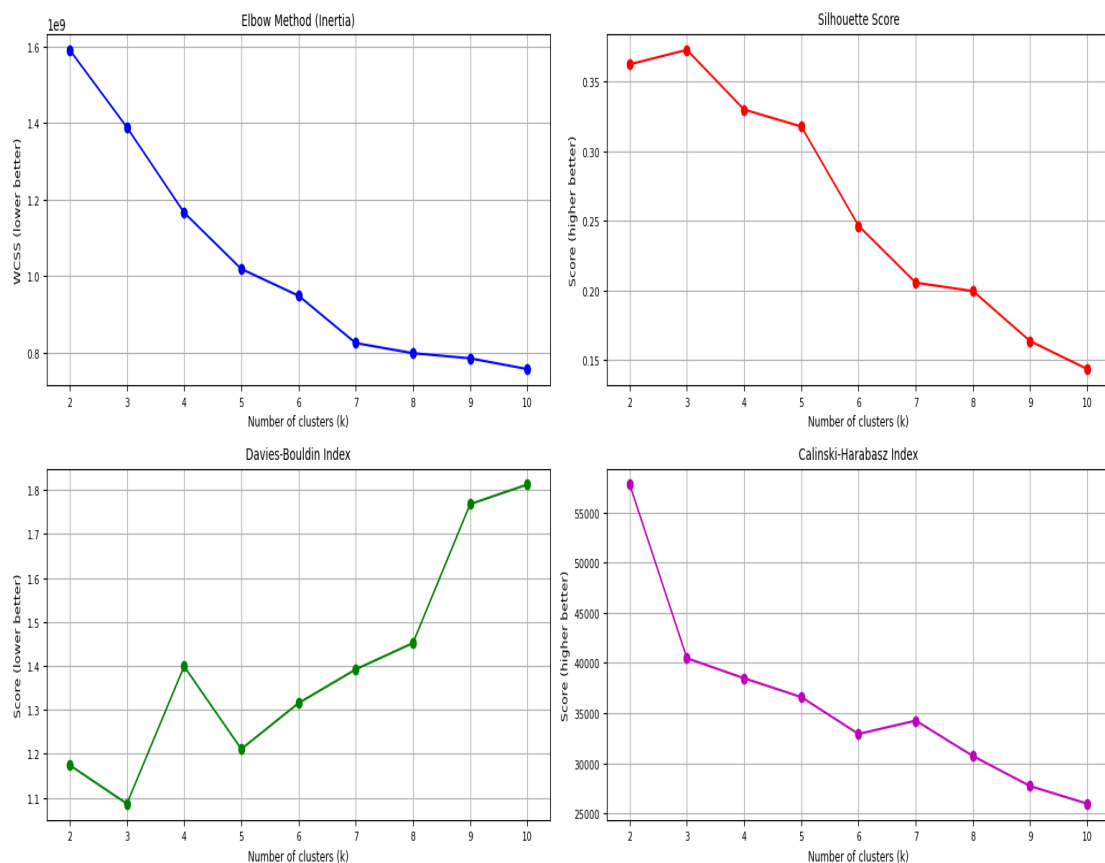
Εικόνα 37: Διάγραμμα διασποράς που απεικονίζει τη σχέση των συντελεστών (Score) της Συνιστώσας 4 έναντι της Συνιστώσας 6 για κάθε pixel.

Τα υπόλοιπα διαγράμματα αξόνων βρίσκονται στο παράρτημα στο τμήμα του εκάστοτε πειράματος ώστε να μην αναλωθεί χώρος και χρόνος από εδώ. Μπορούμε να

προχωρήσουμε άμεσα πλέον στον υπολογισμό του κατάλληλου αριθμού συστάδων που θα μας είναι χρήσιμες για συσταδοποίηση. Για την ομαδοποίηση των pixels σε διακριτές κατηγορίες, εφαρμόστηκαν διάφοροι αλγόριθμοι clustering στα αποτελέσματα της *PCA*. Τα διαγράμματα αξιολόγησης (*Elbow Method*, *Silhouette Score*...), συγκρίθηκαν οι παρακάτω μέθοδοι Εικόνα 38:

- **K-Means:** Απέδωσε έναν γρήγορο αλλά σχετικά "αδρομερή" χάρτη, όπου οι περιοχές ήταν διακριτές αλλά με ασαφή όρια.
- **GMM (Gaussian Mixture Models):** Παρείχε έναν πιο εκλεπτυσμένο χάρτη, με καλύτερο διαχωρισμό σε λεπτομέρειες όπως ο σταυρός και τα φωτοστέφανα, αλλά με διπλάσιο χρόνο επεξεργασίας (*10 λεπτά έναντι 5 του K-Means*).
- **DBSCAN:** Αποδείχθηκε ακατάλληλη για αυτό το σύνολο δεδομένων, καθώς ταυτοποίησε την πλειονότητα των pixels ως "θόρυβο" και δημιούργησε έναν κατακερματισμένο, μη ερμηνεύσιμο χάρτη.
- **Spectral Clustering:** Δεν προσέφερε σημαντική βελτίωση σε σχέση με το *GMM* ώστε να δικαιολογεί την αυξημένη πολυπλοκότητά του.

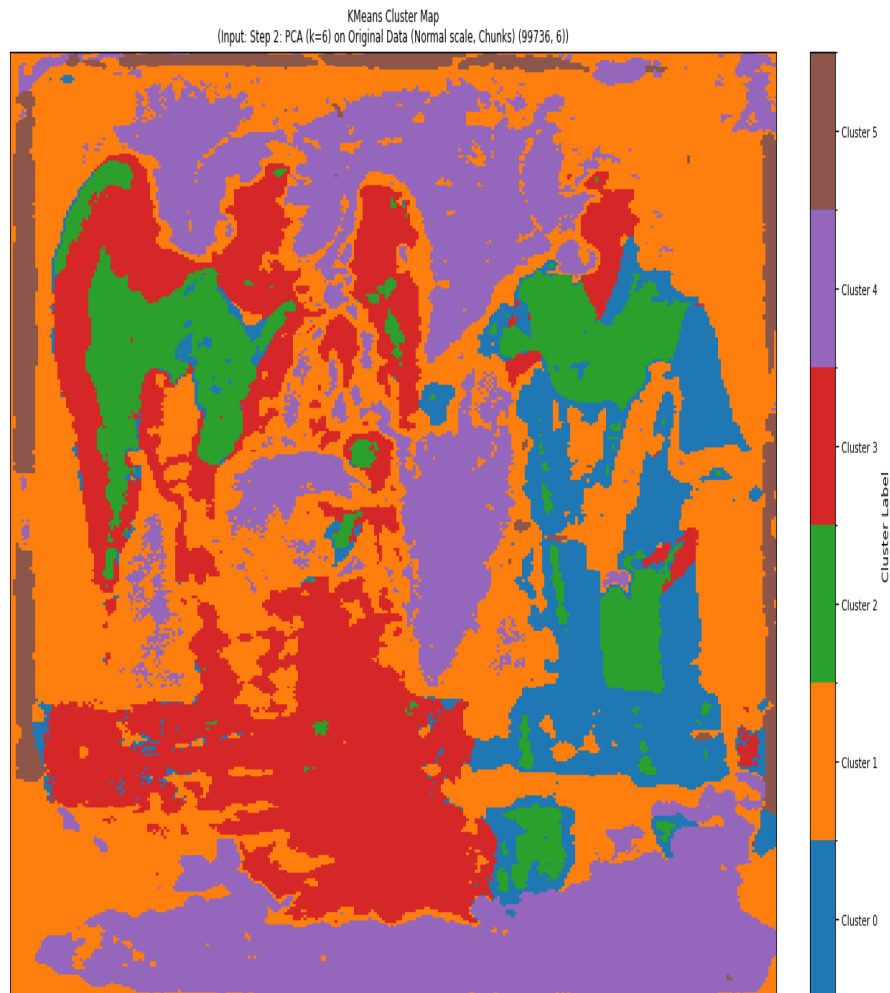
Cluster Evaluation Metrics (on Sample Data)



Εικόνα 38: Αποτελέσματα από την εκτέλεση του πλαισίου αξιολόγησης συσταδοποίησης για τον προσδιορισμό του βέλτιστου αριθμού συστάδων (k). Τα διαγράμματα απεικονίζουν τις μετρικές *Elbow Method (Inertia)*, *Silhouette Score*, *Davies-Bouldin Index* και *Calinski-Harabasz Index* για k από 2 έως 10. Η συνδυασμένη ανάλυση υποδεικνύει ως βέλτιστο εύρος για k τις τιμές 4 έως 6.

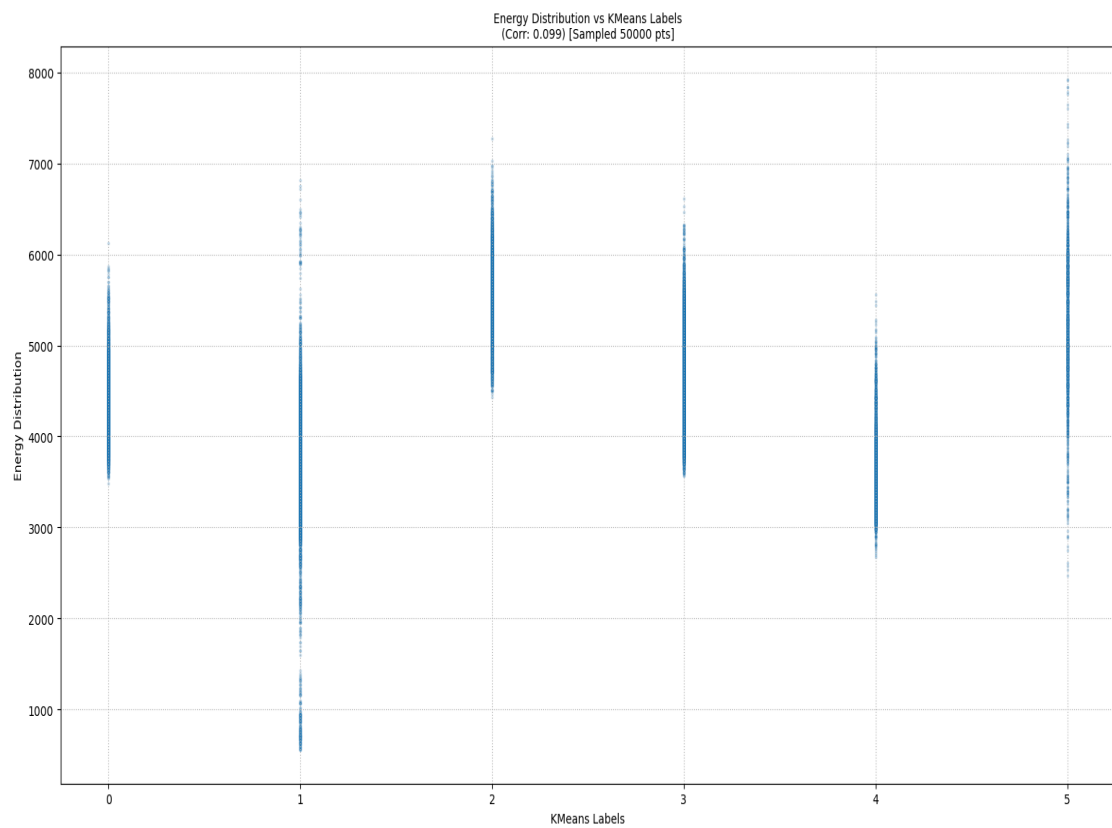
Η συνδυαστική ανάλυση των μετρικών αξιολόγησης (Εικόνα 38) υποδεικνύει ως βέλτιστο εύρος για τον αριθμό των συστάδων (k) τις τιμές μεταξύ 4 και 6. Ειδικότερα, ο δείκτης *Silhouette* παρουσιάζει τοπικό μέγιστο στο $k=4$, ενώ οι δείκτες *Davies-Bouldin* και *Calinski-Harabasz* υποδεικνύουν βέλτιστη απόδοση στο $k=5$. Για την επίτευξη του μέγιστου δυνατού διαχωρισμού των υλικών, επιλέχθηκε η τιμή $k=6$. Βρισκόμαστε επομένως σε θέση για την εφαρμογή των μεθόδων *Kmeans*, *GMM*, *DBScan* αλλά και *Spectral Clustering*. Όλες αυτές οι μέθοδοι είναι γνωστές για συσταδοποίηση και ακραίως διαδομένες στα μεγάλα δεδομένα. Η απόδοση των αλγορίθμων συσταδοποίησης εξαρτάται σε μεγάλο βαθμό από την εγγενή δομή των δεδομένων. Όπως θα αποδειχθεί, ορισμένοι αλγόριθμοι, όπως ο *DBScan*, αποδείχθηκαν ακατάλληλοι για το συγκεκριμένο σύνολο δεδομένων.

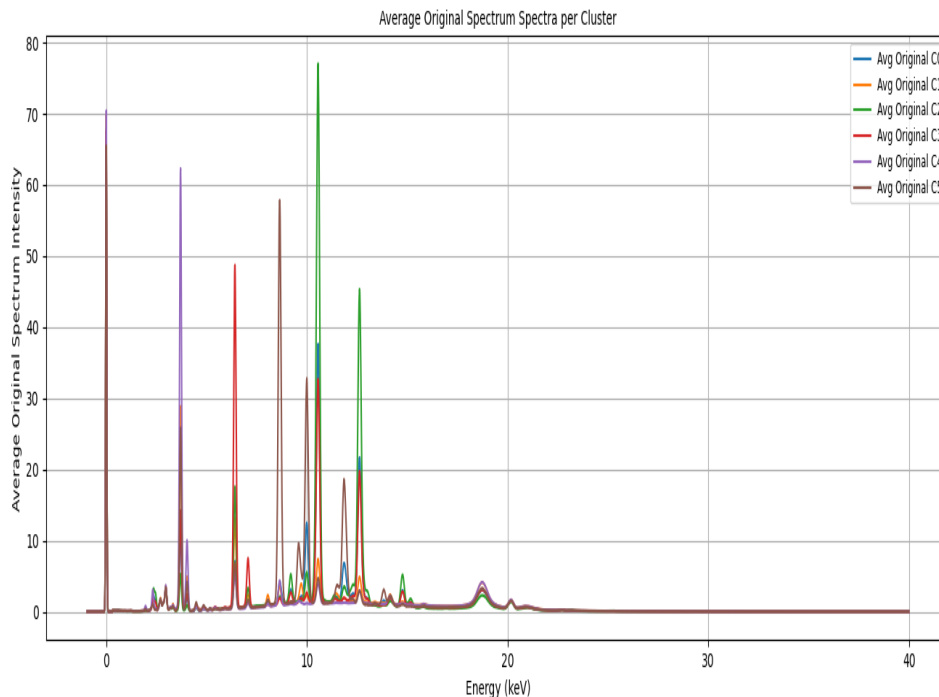
Kmeans: Προχωρούμε στην εφαρμογή της μεθόδου *Kmeans* με αριθμό έξι (6) συστάδων, με τα αποτελέσματα στις εικόνες να φαίνονται παρακάτω Εικόνα 39.



Εικόνα 39: Αποτελέσματα της συσταδοποίησης K-Means ($k=6$) στα δεδομένα PCA. Ο θεματικός χάρτης (cluster map) όπου κάθε χρώμα αντιστοιχεί σε μια συστάδα.

Παρακάτω απεικονίζουμε και την κατανομή της ενέργειας με βάση τη συστάδα Εικόνα 40^α . Αλλά επιπρόσθετα και πως το φάσμα κατανέμεται στις εκάστοτε συστάδες Εικόνα 40^β .



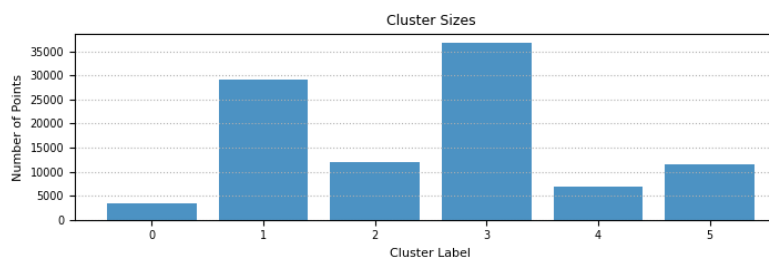


Εικόνα 40: Το μέσο φάσμα για κάθε μία από τις έξι συστάδες, επιτρέποντας τον προσδιορισμό της στοιχειακής σύστασης, ταυτοποιώντας τις κορυφές με τις χαρακτηριστικές εκπομπές K L των στοιχείων με βάση την ενέργεια.

GMM: Ο χάρτης του **GMM** δείχνει μια σαφή βελτίωση σε σχέση με την απλή **PCA**, ομαδοποιώντας pixels με παρόμοια φασματικά χαρακτηριστικά. Οι δύο μεγαλύτερες συστάδες (**Cluster 3** και **1**) φαίνεται να αντιστοιχούν στο υπόβαθρο και τις κύριες μορφές, ενώ οι μικρότερες απομονώνουν εξειδικευμένα υλικά.

Η εικόνα των συστάδων με βάση την μέθοδο επάνω στην αρχική εικόνα του πειράματος Εικόνα 41. Γίνεται

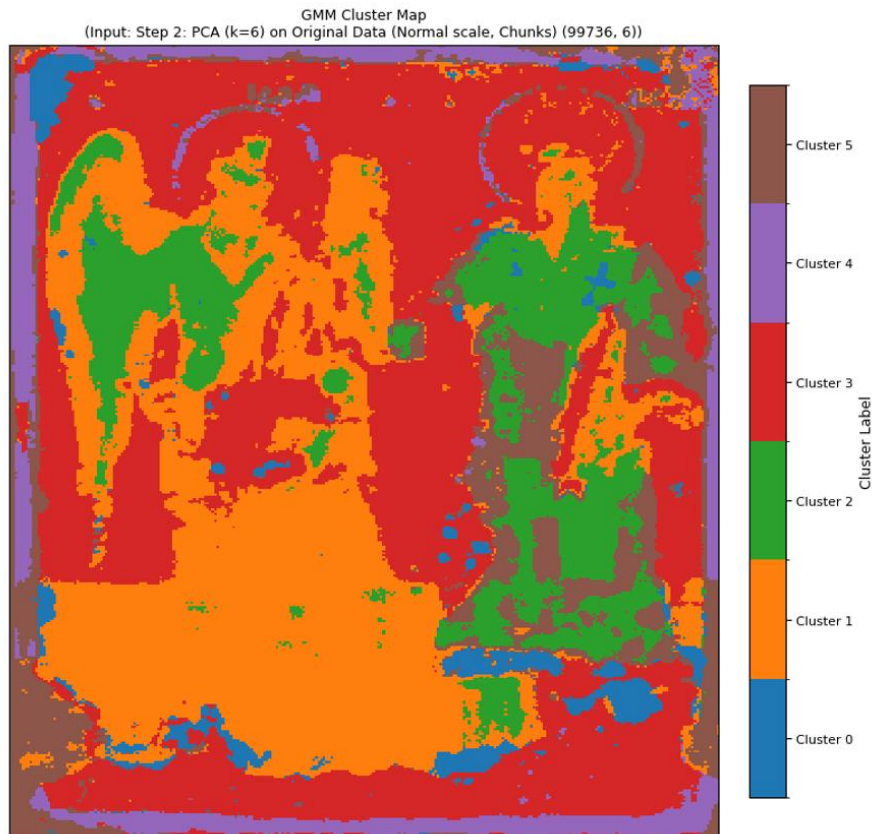
κατανοητό πως πλέον οι συστάδες μέσω της μεθόδου αυτής έχουν μια πιο ευκρινή κατανομή και λεπτομέρειες όπως ο Σταυρός και τα φωτοστέφανα γίνονται ευκρινέστατα και διακριτά, σε αντίθεση με την μέθοδο της Kmeans (βλ. Εικόνα 42). Παρά τη βελτιωμένη ποιότητα του διαχωρισμού, η μέθοδος



Εικόνα 41: Το διάγραμμα Scree Plot που δείχνει τη σωρευτική και ατομική εξηγούμενη διακύμανση για κάθε συνιστώσα.

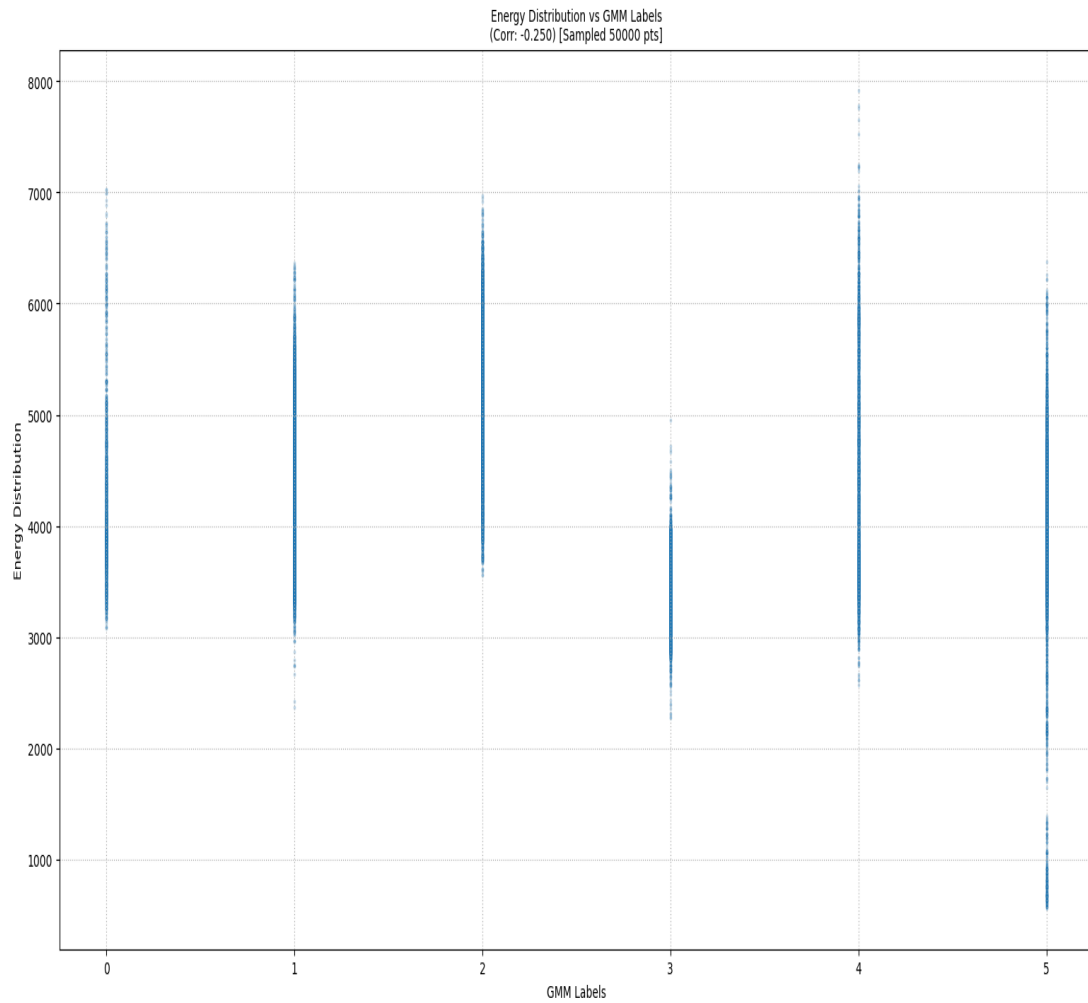
```
GMM Results (on Step 2: PCA (k=6) on Original Data (Normal scale, Chunks) (99736, 6))
=====
Clusters found: 6
Distribution (Top 15):
Cluster 3: 36708
Cluster 1: 29046
Cluster 2: 12072
Cluster 5: 11526
Cluster 4: 6880
Cluster 0: 3504
```

παρουσιάζει σημαντικά αυξημένο υπολογιστικό κόστος, με τον χρόνο εκτέλεσης να ανέρχεται περίπου στα 10 λεπτά, διπλάσιο από αυτόν του K-Means.

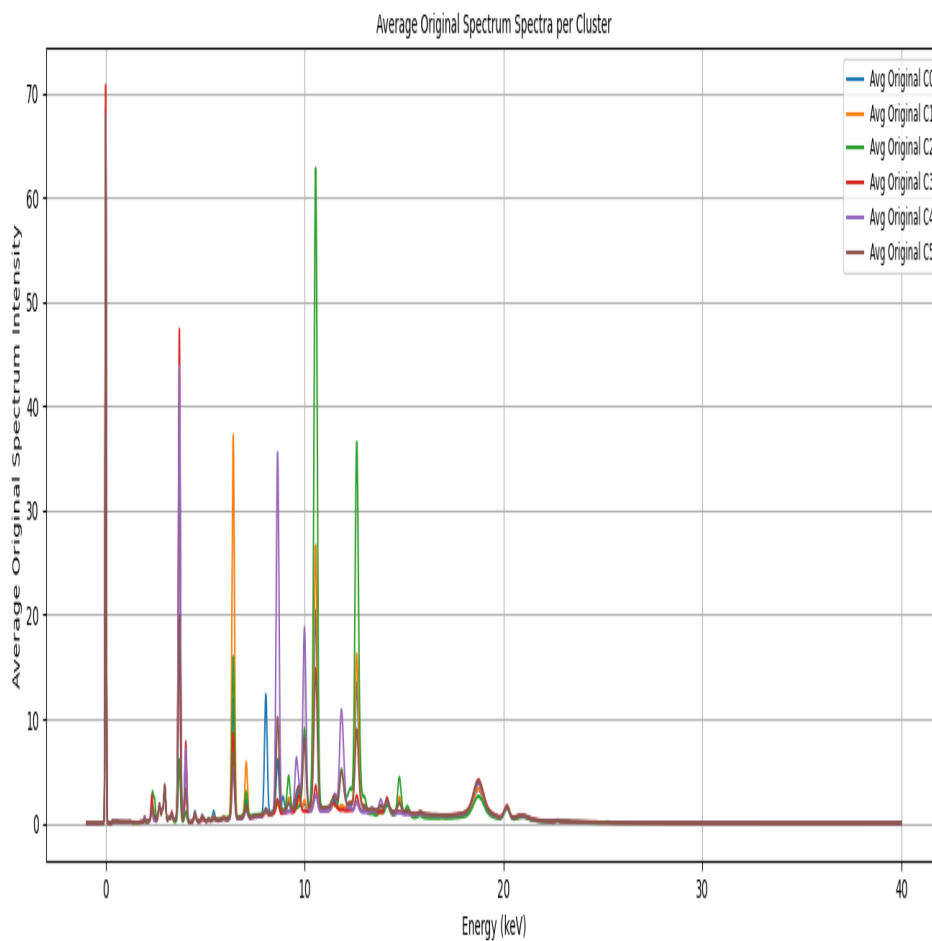


*Εικόνα 42: Αποτελέσματα της συσταδοποίησης GMM ($k=6$). Η μέθοδος παρέχει σαφέστερο διαχωρισμό σε λεπτομέρειες όπως ο σταυρός και τα φωτοστέφανα σε σύγκριση με την *K-Means*, αν και με μεγαλύτερο υπολογιστικό κόστος*

Όπως και πριν παρέχουμε τις εικόνες χωρικής κατανομής των συστάδων ενέργειας με βάση τις συστάδες Εικόνα 43 αλλά και το φάσμα συστάδων στον άξονα των ενεργειών για σύγκριση με το αρχικό φάσμα αλλά και των κυρίων αξόνων Εικόνα 44.

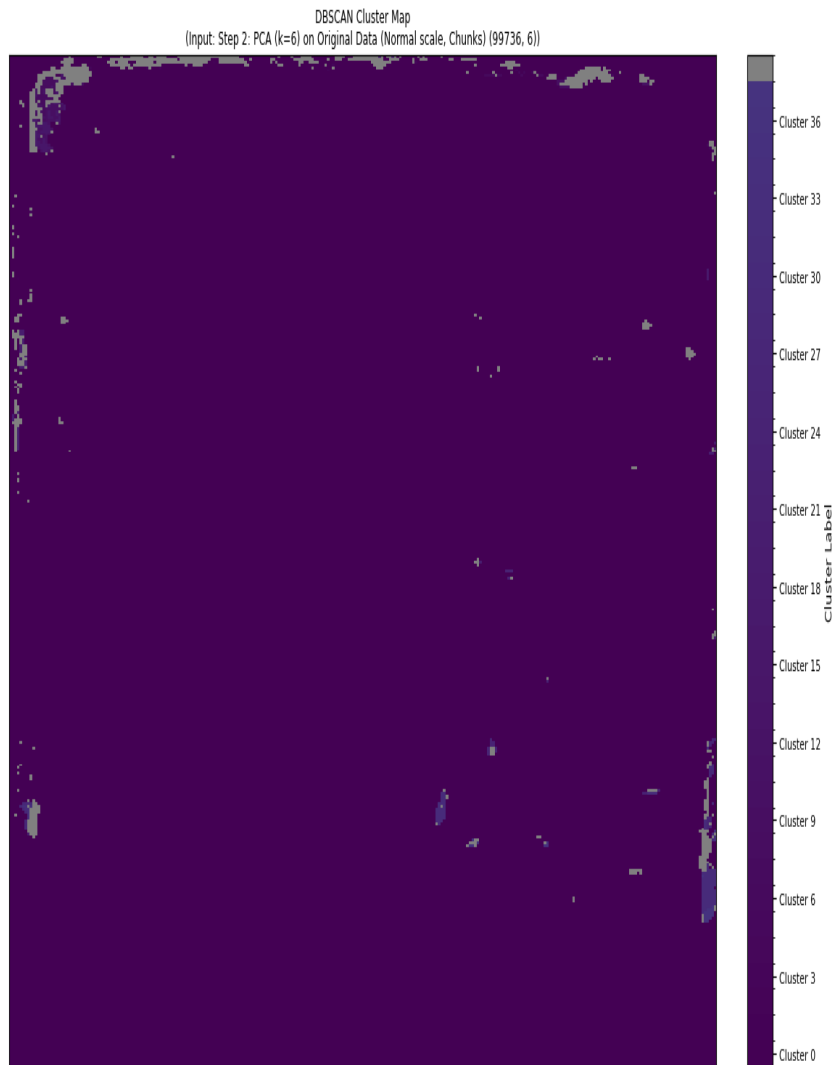


Εικόνα 43: Διάγραμμα διασποράς που απεικονίζει την κατανομή της συνολικής έντασης (*Intensity Distribution*) για κάθε μία από τις έξι συστάδες (*GMM Labels*) που προέκυψαν από τον αλγόριθμο *GMM*. Η ομαδοποίηση των σημείων σε διακριτές κάθετες ζώνες επιβεβαιώνει τον επιτυχή διαχωρισμό των *pixels* σε συστάδες.

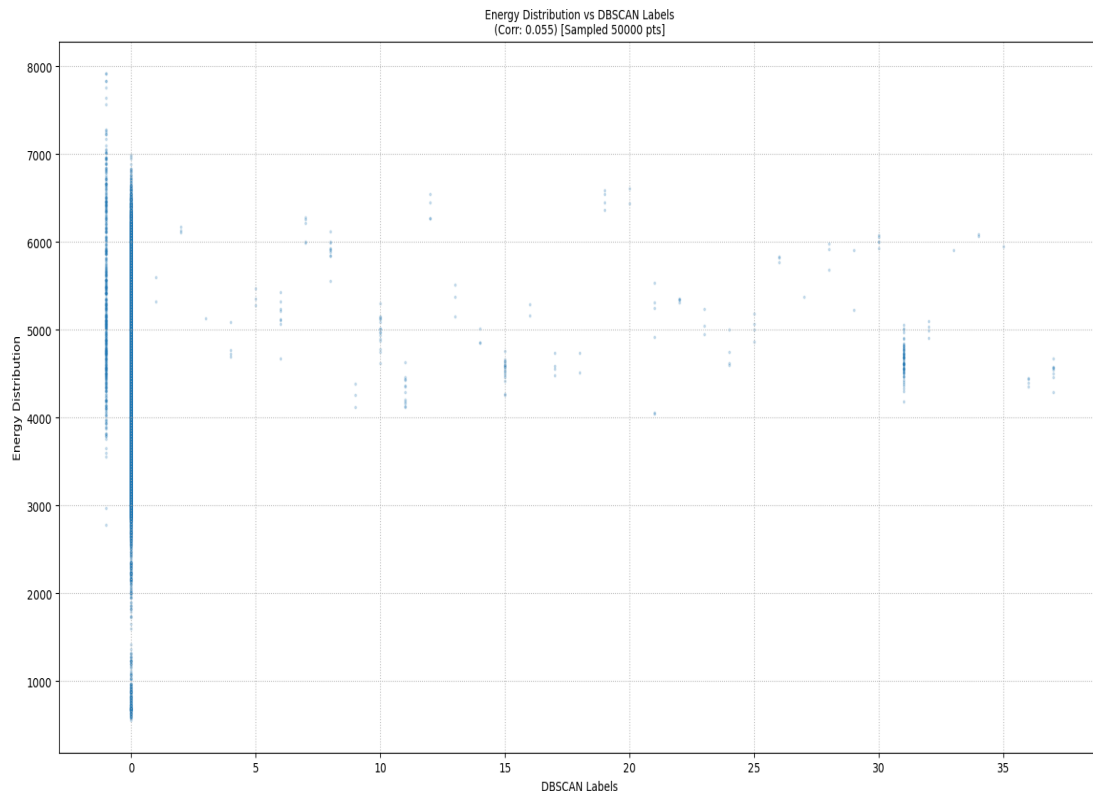


Εικόνα 44: Μέσα φάσματα φθορισμού ακτίνων-Χ (Average Spectra) για κάθε μία από τις έξι συστάδες. Κάθε καμπύλη αντιπροσωπεύει το μέσο φάσμα όλων των pixels που ανήκουν σε μια συγκεκριμένη συστάδα, επιτρέποντας τη χημική ερμηνεία και την ταυτοποίηση των κυρίαρχων στοιχείων σε κάθε ομάδα με βάση τις κυρίαρχες μεταβάσεις ακτίνων Χ.

DBScan: Η μέθοδος αυτή δε μας παρέχει κάποια ουσιαστική πληροφορία επομένως εμφανίζονται μόνο οι εικόνες των συστάδων με βάση την ενέργεια αλλά και των συστάδων με βάση την αρχική μας εικόνα (Εικόνα 45). Επιπρόσθετα η μέθοδος ήταν χρονοβόρα κοντά στα **15 λεπτά**.



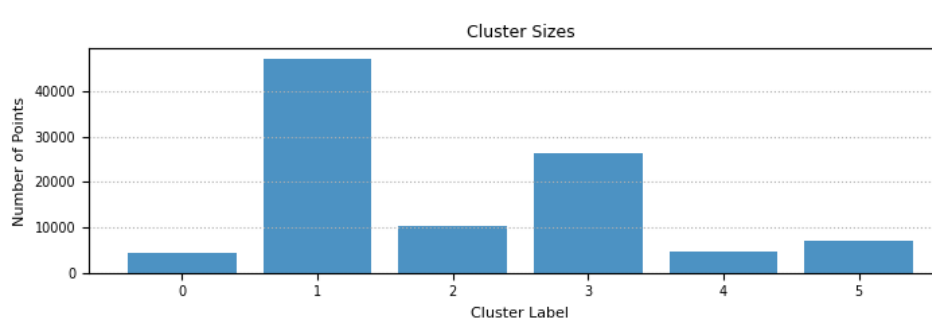
Εικόνα 45: Θεματικός χάρτης (cluster map) από την εφαρμογή του αλγορίθμου DBSCAN. Το αποτέλεσμα υποδεικνύει ότι ο αλγόριθμος, με τις δεδομένες παραμέτρους, απέτυχε να διαχωρίσει ουσιαστικές δομές, ταξινομώντας τη συντριπτική πλειονότητα των pixels σε μία μόνο συστάδα (μωβ) και χαρακτηρίζοντας τα υπόλοιπα ως θόρυβο (γκρι).



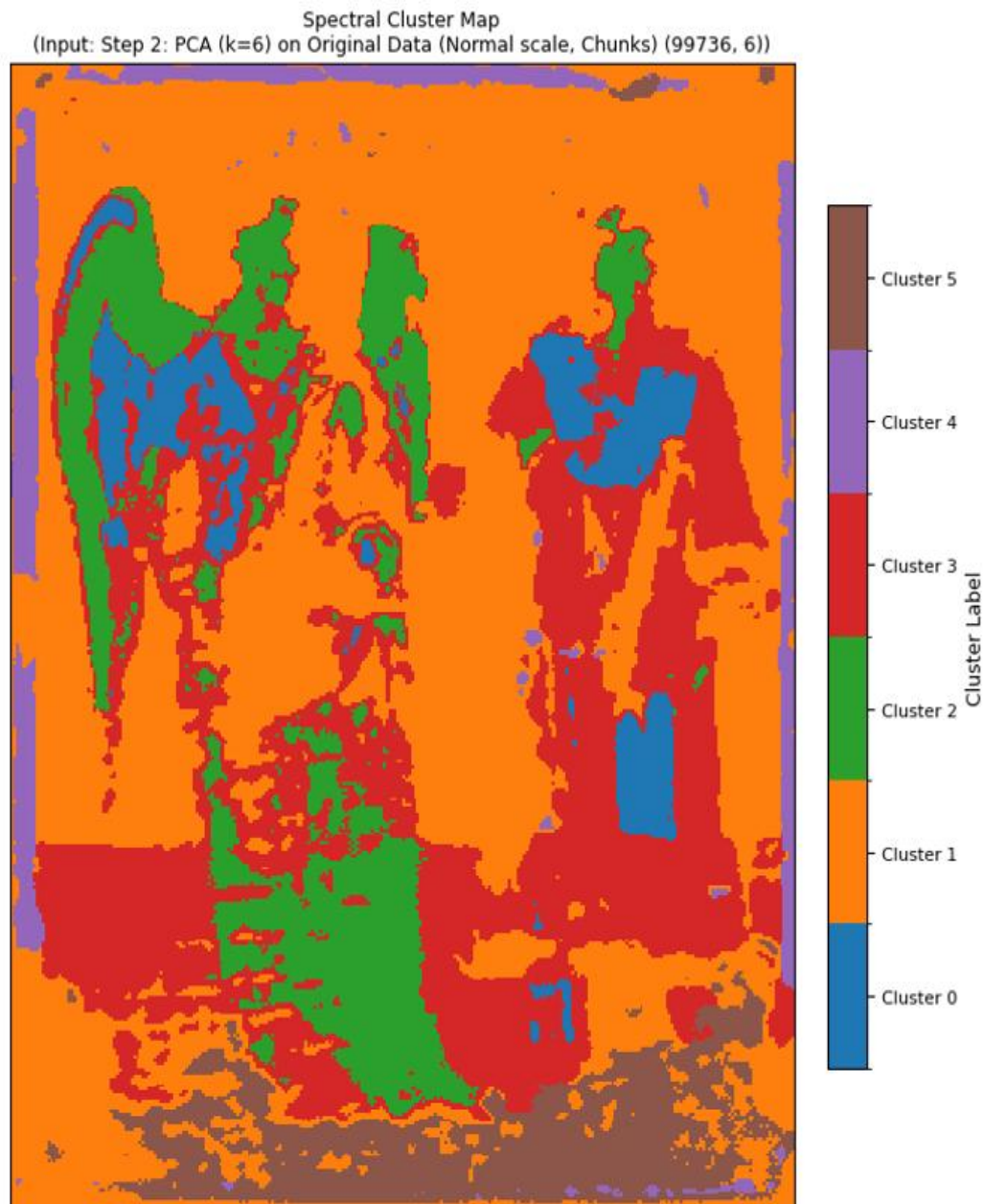
Εικόνα 46: Διάγραμμα διασποράς που συσχετίζει την κατανομή της έντασης (Energy Distribution) με τις συστάδες που προέκυψαν από τον DBSCAN. Η έντονη συγκέντρωση σημείων στην ετικέτα "0" επιβεβαιώνει ότι τα περισσότερα pixels ομαδοποιήθηκαν σε μία κύρια συστάδα, ενώ τα υπόλοιπα είναι διάσπαρτα, αντιστοιχώντας σε θόρυβο.

Spectral: Τέλος η μέθοδος που απομένει βρίσκεται να είναι η **Spectral**, μια αρκετά εύλογη μέθοδος Εικόνα 48 αλλά λόγω της πολυπλοκότητας της η περίοδος εφαρμογής της ήταν αρκετά μεγάλη και με βάση τις εικόνες που παρέχονται παρακάτω στον δεδομένο αριθμό συστάδων Εικόνα 47 δε βρίσκεται να μας παρέχει κάποια ιδιαίτερη πληροφορία.

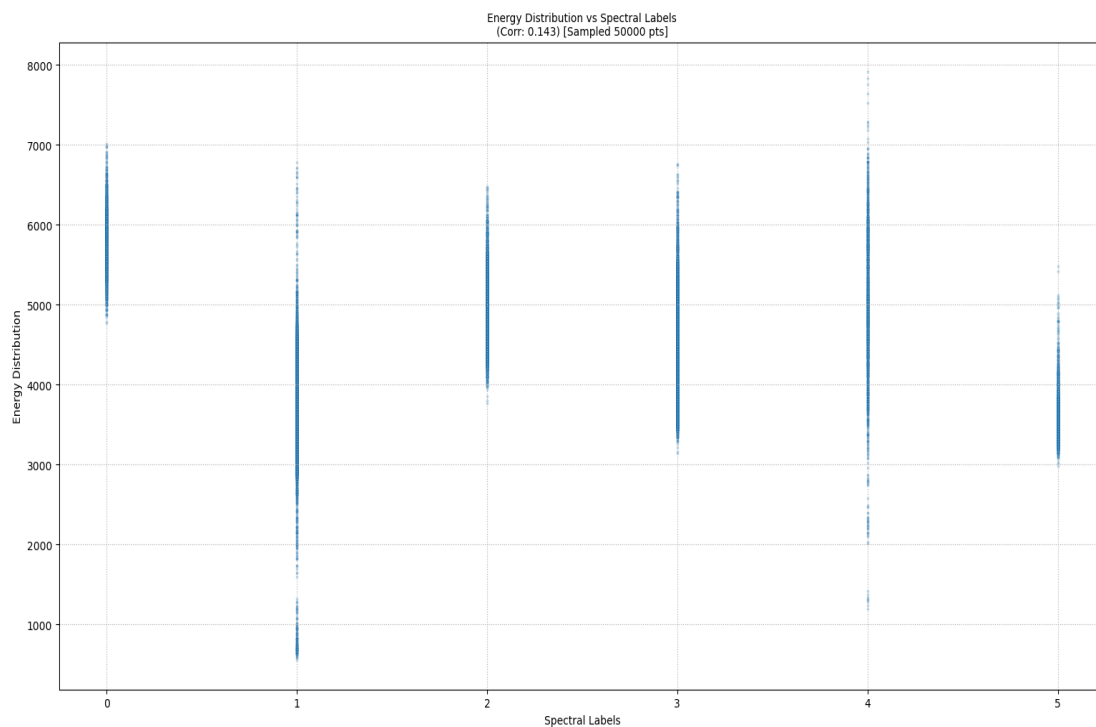
Spectral Results (on Step 2: PCA (k=6) on Original Data (Normal scale, Chunks) (99736, 6))
 =====
 Clusters found: 6
 Distribution (Top 15):
 Cluster 1: 47045
 Cluster 3: 26227
 Cluster 2: 10426
 Cluster 5: 7033
 Cluster 4: 4620
 Cluster 0: 4385



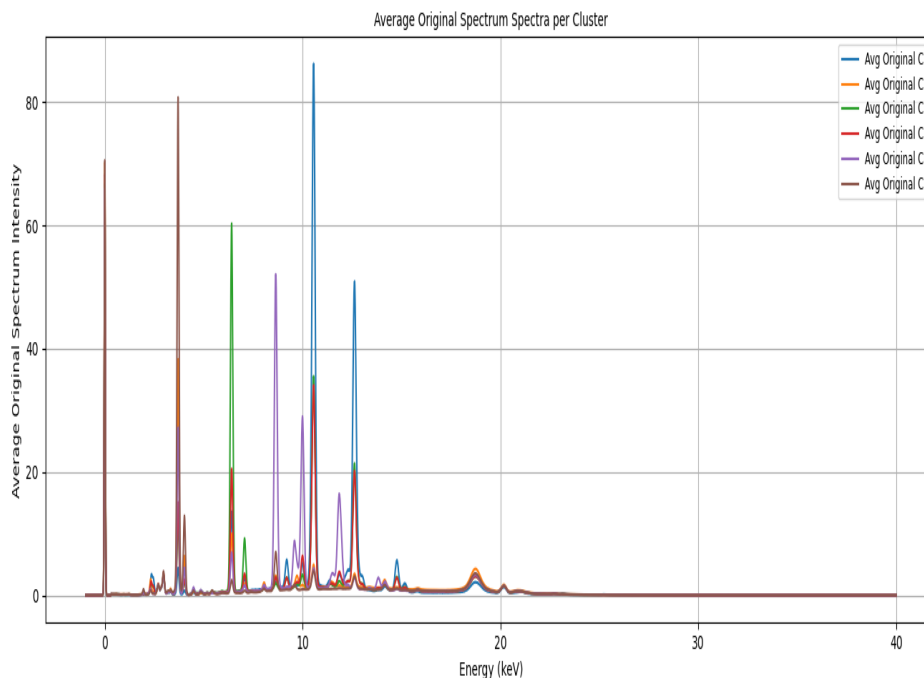
Εικόνα 47: Αριθμητικά και απεικονιστικά αποτελέσματα από την εφαρμογή του αλγορίθμου Spectral Clustering. Ο πίνακας δείχνει ότι βρέθηκαν 6 συστάδες και παραθέτει τον αριθμό των pixels που ανήκουν σε καθεμία, παρέχοντας μια ποσοτική επισκόπηση του μεγέθους των ομάδων.



Εικόνα 48: Θεματικός χάρτης (cluster map) που προέκυψε από την εφαρμογή του αλγορίθμου Spectral Clustering ($k=6$). Κάθε χρώμα αντιπροσωπεύει μια διακριτή συστάδα, αποκαλύπτοντας τη χωρική κατανομή των στοιχείων ομάδων πάνω στην επιφάνεια της εικόνας.



Εικόνα 49: Διάγραμμα διασποράς που απεικονίζει την κατανομή της συνολικής έντασης (*intensity Distribution*) για κάθε μία από τις έξι συστάδες (*Spectral Labels*) που προέκυψαν από τον αλγόριθμο *Spectral Clustering*. Η ομαδοποίηση των σημείων σε διακριτές κάθετες ζώνες επιβεβαιώνει τον επιτυχή διαχωρισμό των *pixels* σε συστάδες.



Εικόνα 50: Μέσα φάσματα φθορισμού ακτίνων-Χ (*Average Spectra*) για κάθε μία από τις έξι συστάδες που προέκυψαν από την ανάλυση *Spectral Clustering*. Κάθε καμπύλη αντιπροσωπεύει το μέσο φάσμα όλων των *pixels* που ανήκουν σε μια συγκεκριμένη συστάδα, επιτρέποντας τη χημική ερμηνεία και την ταυτοποίηση των κυρίαρχων στοιχείων σε κάθε ομάδα.

Όπως και πριν παρέχουμε όλες τις εικόνες τις οποίες αποδίδει το πρόγραμμα μας και είναι απαραίτητες για την ανάλυση. Βεβαίως καταλήγουμε πως οι καλύτερες μέθοδοι για την ανάλυση των δεδομένων μας σε συστάδες είναι η **Kmeans**.

Ανάλυση NMF:

NMF: Προχωρώντας πέρα από την αφηρημένη αποδόμηση της **PCA**, η εφαρμογή της μεθόδου της **Μη-Αρνητικής Παραγοντοποίησης Πινάκων (NMF)** στο αρχικό σύνολο δεδομένων αποσκοπεί στην εξαγωγή συνιστωσών που έχουν άμεση φυσικοχημική σημασία. Λόγω του μαθηματικού περιορισμού της μη-αρνητικότητας τόσο στους χωρικούς χάρτες όσο και στα αντίστοιχα φάσματα, κάθε συνιστώσα μπορεί πλέον να ερμηνευθεί ως ο σχετικός χάρτης αφθονίας μιας συγκεκριμένης χημικής φάσης ή υλικού.

Οι χάρτες των **έξι συνιστωσών NMF** που προέκυψαν παρέχουν έναν εντυπωσιακά σαφή διαχωρισμό των υλικών της παλέτας του καλλιτέχνη (Εικόνα 51), φαίνονται παρακάτω.

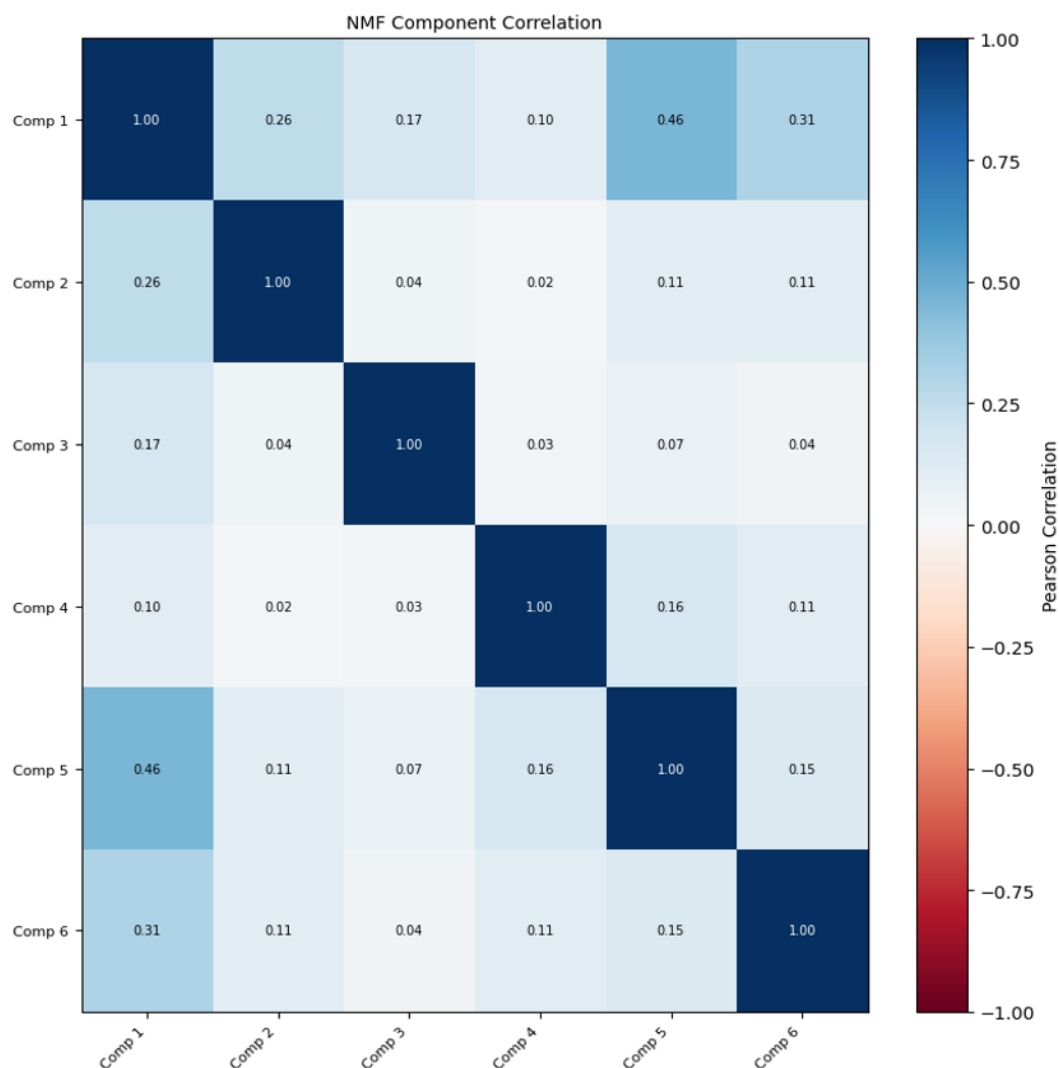


Εικόνα 51: Αποτελέσματα από την εφαρμογή της μεθόδου NMF. Οι χάρτες των έξι συνιστωσών NMF, οι οποίοι λόγω της μη-αρνητικότητας μπορούν να ερμηνευθούν ως χάρτες αφθονίας στοιχείων.

Οι άξονες όπως φαίνεται πλέον τονίζουν μεταξύ τους την διαφορά των εντάσεων. Περιοχές πλέον δεν είναι τόσο ομοιογενείς εκτός από τον **άξονα 1** ο οποίος αποτελεί

το **background** της εικόνας. Περιοχές των αγίων αποτελούν άλλους άξονες και έχουμε και καλύτερη διαχωριστικότητα σε σχέση με την μέθοδο της **PCA**. Βεβαία ο χρόνος εκτέλεσης είναι μεγάλος. Παρακάτω παρέχεται και ο πίνακας συσχέτισης των αξόνων μεταξύ τους.

- *NMF Comp. 1,5*: Αυτή η συνιστώσα απομονώνει με εξαιρετική ευκρίνεια και σχεδόν αποκλειστικά το υπόβαθρο προετοιμασίας (*gesso*). Η κατανομή της είναι διάχυτη, με υψηλότερη ένταση στις περιοχές όπου το ζωγραφικό στρώμα έχει υποστεί φθορά. Το αντίστοιχο φάσμα αυτής της συνιστώσας θα κυριαρχείται από τις χαρακτηριστικές κορυφές του *Ασβεστίου* (*Ca*) και του *Θείου* (*S*).
- *NMF Comp. 4*: Παρόμοια με την *Comp. 1*, αλλά με διαφορετική στάθμιση, αυτή η συνιστώσα φαίνεται επίσης να σχετίζεται με το υπόβαθρο, πιθανώς απομονώνοντας μια δευτερεύουσα φάση του *gesso* ή περιοχές με διαφορετική επιφανειακή κατάσταση (π.χ. *craquelure*).
- *NMF Comp. 2*: Αυτή η συνιστώσα σχετίζεται άμεσα με τις καφε-κόκκινες περιοχές των ενδυμάτων. Αντιπροσωπεύει τη φασματική υπογραφή των γήινων χρωστικών με βάση τον *Σίδηρο* (*Fe*), όπως οι ώχρες αλλά και περιοχές φθοράς.
- *NMF Comp. 3*: Προσφέρει έναν από τους πιο εντυπωσιακούς διαχωρισμούς, απομονώνοντας σχεδόν αποκλειστικά τα σαρκώματα (πρόσωπα και χέρια). Αυτό το αποτέλεσμα υποδηλώνει ότι η ανάμειξη χρωστικών που χρησιμοποιήθηκε για την απόδοση του δέρματος – πιθανότατα Λευκό του *Μολύβδου* (*Pb*) αναμεμειγμένο με μικρή ποσότητα *ώχρας* (*Fe*) – είναι μοναδική και στατιστικά διακριτή από όλα τα υπόλοιπα υλικά της εικόνας.
- *NMF Comp. 6*: Αυτή είναι μια ιδιαίτερα αποκαλυπτική συνιστώσα, καθώς απομονώνει με εξαιρετική ακρίβεια περιοχή που ταυτίζεται με τον στοιχειακό χάρτη του *Υδραργύρου* (*Hg*) (κιννάβαρι). Αυτό υποδεικνύει ότι ο αλγόριθμος αναγνώρισε μια στατιστικά διακριτή ομάδα "υλικών έμφασης".

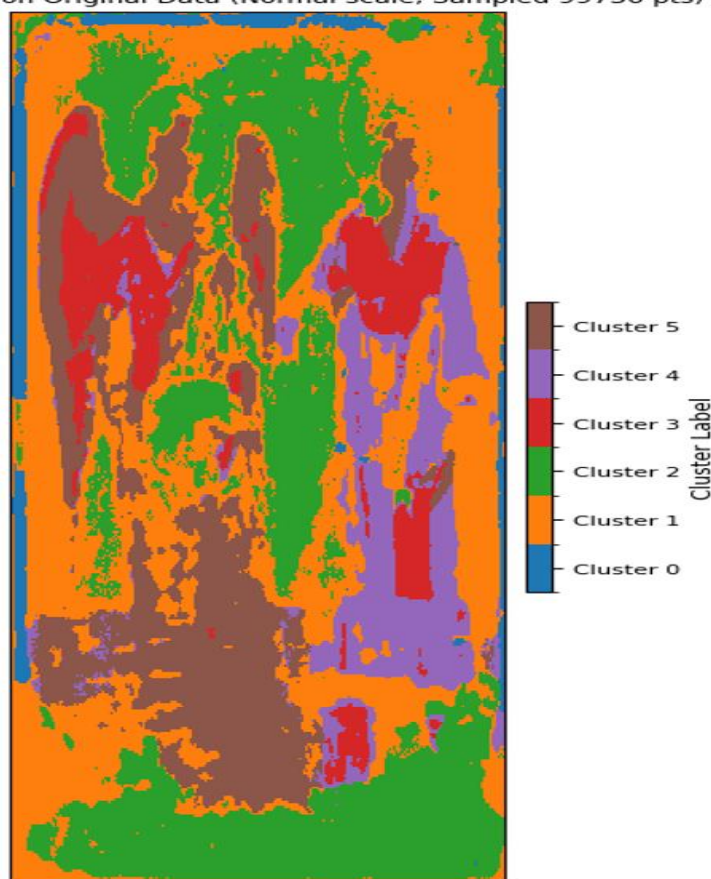


Εικόνα 52: Πίνακας συσχέτισης (heatmap) των συνιστωσών που προέκυψαν από την ανάλυση NMF. Σε αντίθεση με την PCA, οι συνιστώσες της NMF δεν είναι κατ' ανάγκη ορθογώνιες. Οι θετικές συσχετίσεις (γαλάζιες αποχρώσεις) υποδεικνύουν κοινά φασματικά χαρακτηριστικά ή χωρική συνύπαρξη υλικών μεταξύ των συνιστωσών.

Όπως και πριν προχωρούμε στην συσταδοποίηση αλλά πλέον παραλείπουμε τις μεθόδους **GMM**, **Spectral**, **DBScan** για συντομία διότι θα πραγματοποιούμε την πιο γρήγορη μέθοδο που είναι η **Kmeans** χωρίς να χάνουμε και μεγάλο μέρος της πληροφορίας σε σχέση με την **GMM** Εικόνα 53. Το τελικό βήμα της πολυμεταβλητής ανάλυσης είναι η εφαρμογή ενός αλγορίθμου συσταδοποίησης (**Clustering**) στα αποτελέσματα της **NMF**. Για τον σκοπό αυτό, επιλέχθηκε ο αλγόριθμος **K-Means**, ο οποίος ομαδοποιεί κάθε pixel της εικόνας σε μία από τις προκαθορισμένες συστάδες ($k=6$) με βάση την ομοιότητα των τιμών του στις έξι συνιστώσες **NMF**. Το αποτέλεσμα είναι ένας θεματικός χάρτης υλικών, όπου κάθε χρώμα δεν είναι αυθαίρετο, αλλά αντιστοιχεί σε μια ομάδα **pixels** με στατιστικά παρόμοια χημική σύσταση.

Τα αποτελέσματα της **Kmeans** επάνω στην αρχική εικόνα φαίνονται παρακάτω.

KMeans Cluster Map
Step 2: NMF (k=6) on Original Data (Normal scale, Sampled 99736 pts) (99736, 6))



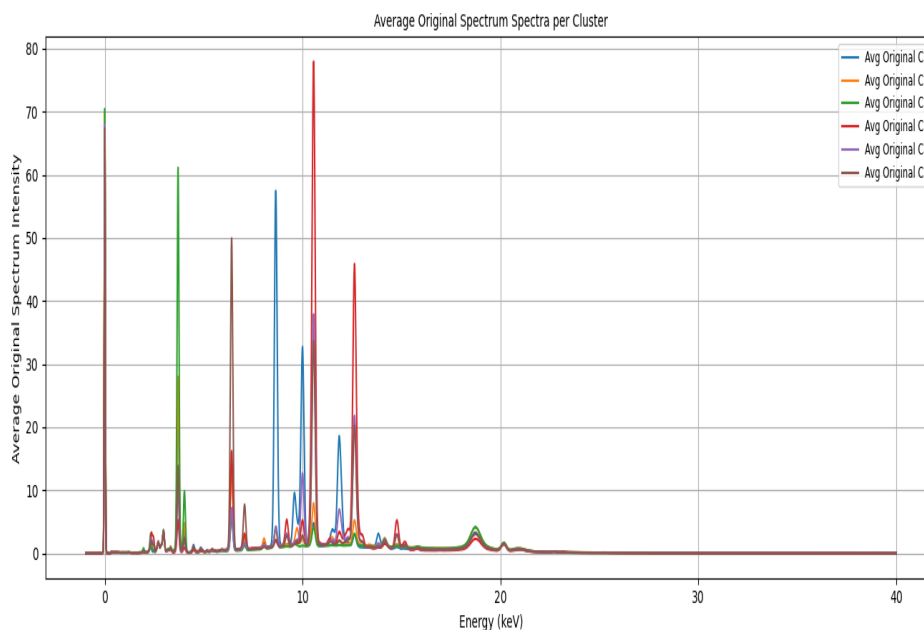
Εικόνα 53: Θεματικός χάρτης (cluster map) που προέκυψε από την εφαρμογή του αλγορίθμου K-Means ($k=6$) στα αποτελέσματα της ανάλυσης NMF. Κάθε χρώμα αντιπροσωπεύει μια διακριτή ομάδα υλικών, αποκαλύπτοντας τη χωρική κατανομή και τις σχέσεις μεταξύ των διαφόρων χρωστικών και υλικών προετοιμασίας.

Ο χάρτης **K-Means** που προέκυψε παρέχει τον πιο σαφή και συνοπτικό διαχωρισμό των υλικών που χρησιμοποιήθηκαν στην εικόνα. Η χημική ταυτότητα κάθε συστάδας μπορεί να αποδοθεί με βεβαιότητα, εξετάζοντας ποια συνιστώσα NMF κυριαρχεί σε αυτήν, καθώς και αναλύοντας το μέσο φάσμα XRF για κάθε συστάδα.

- **Cluster 0 & 1** (Μπλε & Πορτοκαλί): Αυτές οι δύο συστάδες κυριαρχούν στο υπόβαθρο και στις φθαρμένες περιοχές της εικόνας. Αντιστοιχούν στα pixels όπου η συνιστώσα NMF του *gesso* (NMF Comp. 1) έχει την υψηλότερη τιμή. Τα μέσα φάσματά τους κυριαρχούνται από τις κορυφές του Ασβεστίου (Ca) και του Θείου (S), επιβεβαιώνοντας ότι αντιπροσωπεύουν το στρώμα προετοιμασίας (*gesso*). Η ύπαρξη δύο συστάδων υποδηλώνει πιθανώς ελαφρώς διαφορετικές συνθέσεις ή καταστάσεις διατήρησης του υποστρώματος.
- **Cluster 2** (Πράσινο): Αυτή η συστάδα συμπίπτει απόλυτα με την κατανομή της συνιστώσας NMF των ώχρων (NMF Comp. 5). Το μέσο φάσμα της παρουσιάζει μια έντονη κορυφή Ασβεστίου (Ca).
- **Cluster 3** (Κόκκινο): Εντοπίζει μικρότερες, αλλά διακριτές περιοχές στα ενδύματα και τα φτερά. Το φάσμα αυτής της συστάδας αποκαλύπτει την παρουσία Μολύβδου (Pb), υποδεικνύοντας τη εκτενή χρήση μιας πολύ

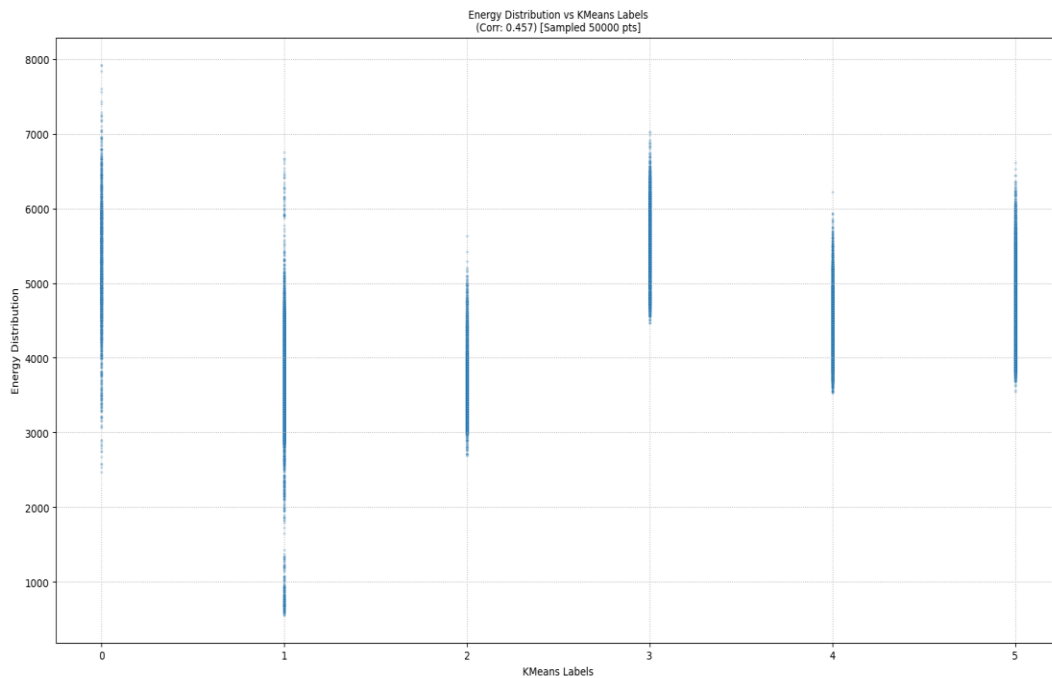
διαδεδομένης χρωστικής του λευκού του Μολύβδου αν συγκρίνουμε και με την εικόνα στο οπτικό.

- **Cluster 4** (Μωβ): Αντιστοιχεί με απόλυτη ακρίβεια στις επιχρυσωμένες περιοχές και σε περιοχές που εμφανίζεται το θείο (S) και ο υδράργυρος (Hg), ακολουθώντας την κατανομή των συνιστωσών NMF (**NMF Comp. 2 & 3**). Το φάσμα της κυριαρχείται από τις χαρακτηριστικές γραμμές-L του Χρυσού (Au) και υδραργύρου (Hg) υποδεικνύοντας τη χρήση μίνιου για το κόκκινο του χιτώνα στα δεξιά.
- **Cluster 5** (Καφέ): Αυτή η συστάδα ταυτοποιεί με εξαιρετική ακρίβεια τα σαρκώματα (πρόσωπα και χέρια και φτερά). Το μέσο φάσμα της κυριαρχείται από τις έντονες **γραμμές-K** του σιδήρου (**Fe**), επιβεβαιώνοντας τη χρήση ώχρας.



Εικόνα 54: Μέσα φάσματα φθορισμού για κάθε μία από τις έξι συστάδες που προέκυψαν από την ομαδοποίηση *K-Means* επί των δεδομένων NMF. Η ανάλυση των φασμάτων επιτρέπει την ταυτοποίηση της χημικής σύστασης κάθε ομάδας που διαχωρίστηκε.

Τέλος παρέχουμε και την εικόνα κατανομής της ενέργειας της κάθε συστάδας. Συμπερασματικά, οι **GMM** και **K-Means** ήταν οι πιο κατάλληλες μέθοδοι, με τις **NMF** και την υβριδική **NMF-PCA** να αναλύονται στη συνέχεια ως ανώτερες προσεγγίσεις.

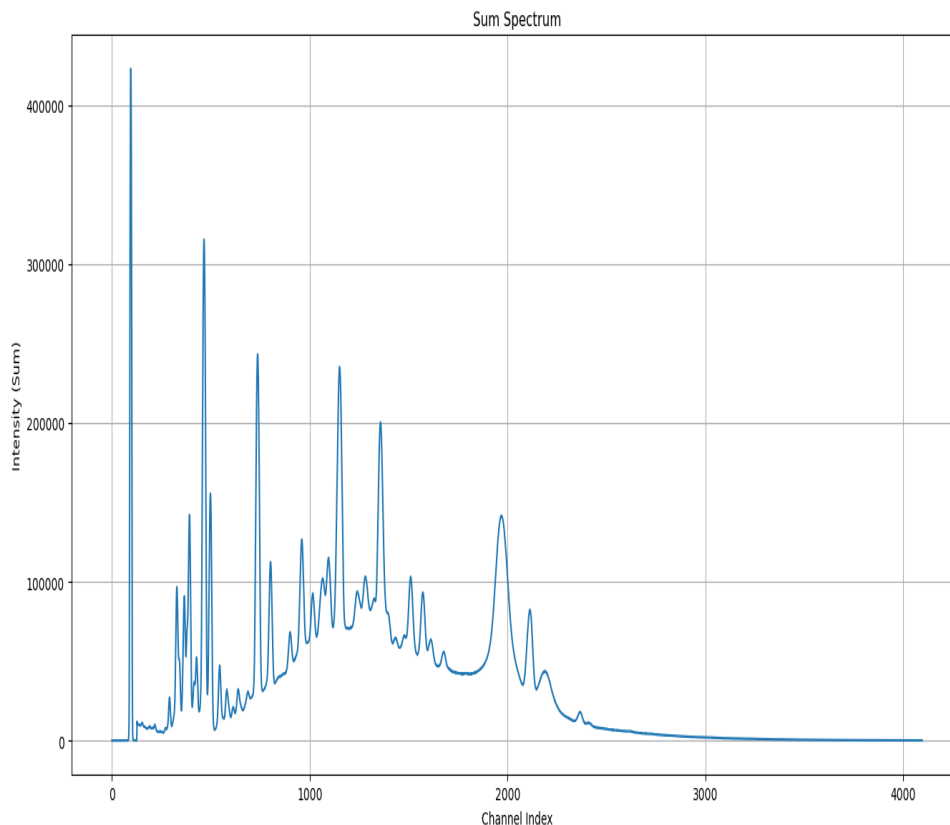


Εικόνα 55: Διάγραμμα διασποράς της συνολικής έντασης (*Intensity Distribution*) για τις έξι συστάδες που προέκυψαν από τον αλγόριθμο *K-Means*. Η σαφής ομαδοποίηση των σημείων σε κάθετες ζώνες επιβεβαιώνει την αποτελεσματικότητα του διαχωρισμού.

Η ανάλυση επαναλήφθηκε μετασχηματίζοντας τα δεδομένα σε λογαριθμική κλίμακα ($\log(\text{value}+1)$). Αυτή η προσέγγιση αποσκοπεί στην ενίσχυση των σημάτων από στοιχεία που βρίσκονται σε χαμηλές συγκεντρώσεις ή σε πολύ λεπτά στρώματα. Η *PCA* στα λογαριθμικά δεδομένα εξήγησε ένα μικρότερο ποσοστό της διακύμανσης ($\sim 25\%$), καθώς οι μεγάλες εντάσεις συμπίεστηκαν και οι μικρές ενισχύθηκαν. Ωστόσο, ο χάρτης ομαδοποίησης *K-Means* που προέκυψε έδειξε έναν καλύτερο διαχωρισμό σε σημεία με υλικά υψηλού *Z* (χρυσός, μόλυβδος). Αυτό αποδεικνύει ότι η λογαριθμική κλίμακα μπορεί να αποκαλύψει πληροφορίες για την τοποθεσία υλικών που βρίσκονται στο υπόβαθρο ή σε λεπτά στρώματα, οι οποίες χάνονται στην ανάλυση γραμμικής κλίμακας.

Λογαριθμική Ανάλυση των δεδομένων

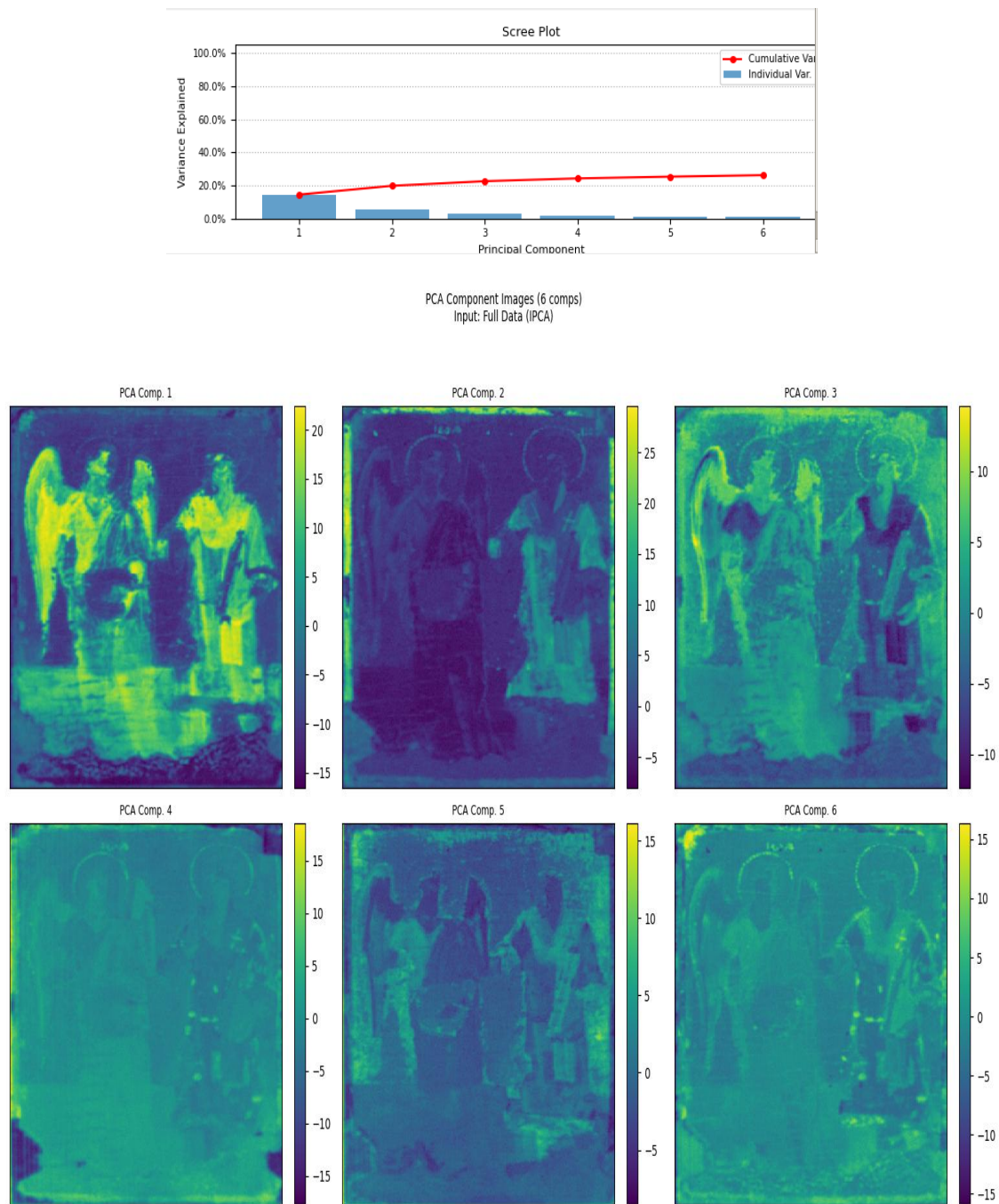
Οπότε αρχικά για τα λογαριθμικά δεδομένα θα έχουμε το εξής διάγραμμα για το φάσμα.



Εικόνα 56: Φάσμα αθροίσματος (Sum Spectrum) των δεδομένων μετά από λογαριθμικό μετασχηματισμό $\log(\text{value}+1)$. Η εφαρμογή της λογαριθμικής κλίμακας ενισχύει τις κορυφές χαμηλής έντασης, επιτρέποντας την καλύτερη οπτική επιθεώρηση των ιχνοστοιχείων που ενδέχεται να χάνονται στη γραμμική κλίμακα.

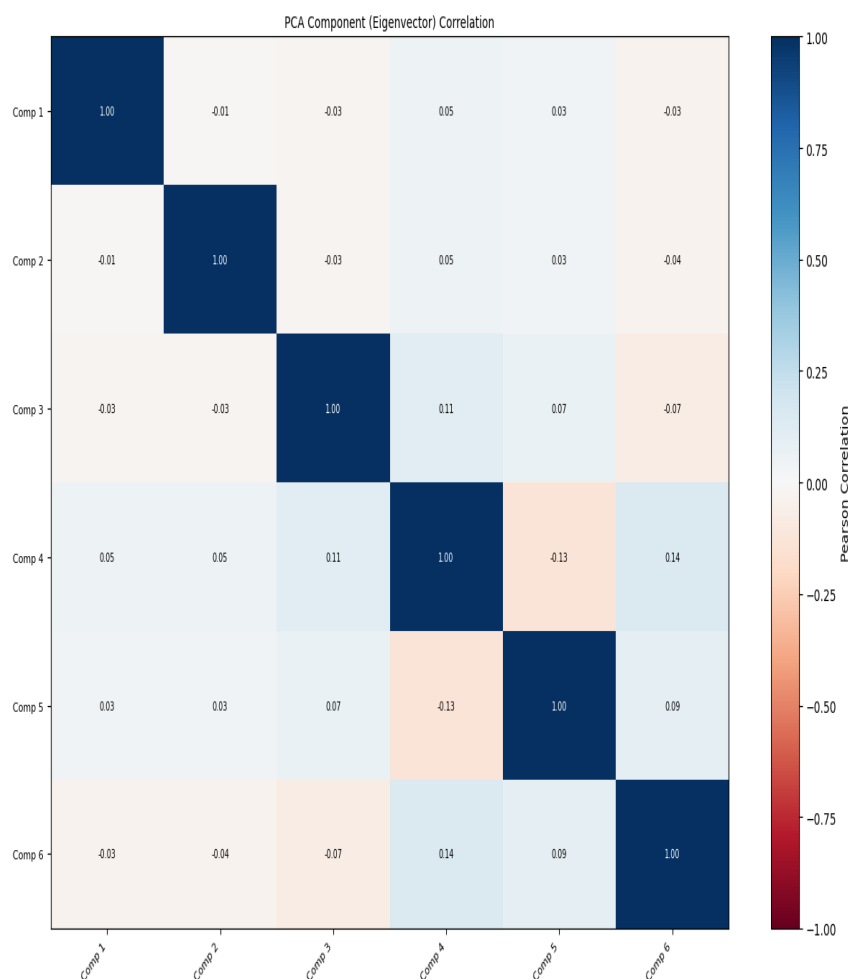
Ανάλυση PCA:

Παρατηρούμε πως πλέον οι μεγάλες εντάσεις εξομαλύνονται και οι χαμηλές εντάσεις γίνονται πιο έντονες. Επομένως περιοχές που εμφανίζονταν πριν με χαμηλό σήμα και δεν ήταν διακριτές λόγω ότι βρίσκονταν κοντά σε υψηλές τώρα διακρίνονται πιο εύκολα. Ωστόσο, η εφαρμογή της **PCA** σε αυτά τα μετασχηματισμένα δεδομένα αποκαλύπτει σημαντικούς περιορισμούς. Η αθροιστική διακύμανση (**Cumulative Variance**) που εξηγείται από τις έξι πρώτες συνιστώσες είναι εξαιρετικά χαμηλή, της τάξης του **~25%**. Αυτό το αποτέλεσμα υποδηλώνει ότι ο λογαριθμικός μετασχηματισμός, ενώ ενισχύει τον θόρυβο και τις ασθενείς κορυφές, διαταράσσει και αποδυναμώνει τις ισχυρές γραμμικές συσχετίσεις μεταξύ των κύριων χημικών φάσεων, οι οποίες αποτελούν την κινητήρια δύναμη της **PCA**. Ουσιαστικά, η **PCA** δυσκολεύεται πλέον να βρει τις κυρίαρχες "κατευθύνσεις" της διακύμανσης, καθώς όλες οι κορυφές έχουν πλέον παρόμοια "βαρύτητα". Τα αποτελέσματα των αξόνων που επιβεβαιώνει τα λεγόμενα παρέχονται παρακάτω στην Εικόνα 57.



Εικόνα 57: Αποτελέσματα της Ανάλυσης Κύριων Συνιστωσών (PCA) στα λογαριθμικά μετασχηματισμένα δεδομένα. (Πάνω) Το διάγραμμα Scree Plot που δείχνει τη σωρευτική και ατομική εξηγούμενη διακύμανση για κάθε συνιστώσα. (Κάτω) Οι χωρικοί χάρτες (score maps) για τις έξι πρώτες κύριες συνιστώσες.

Τέλος παρέχουμε και την συσχέτιση των αξόνων μεταξύ τους στην Εικόνα 58, όπου παρατηρούμε πως αρκετοί άξονες έχουν μεγαλύτερη από 0.1 συσχέτιση κάτι το οποίο είναι αρκετά μεγάλο για τα δεδομένα της **PCA**, ώστε να θεωρεί σωστή η επιλογή της μεθοδολογίας αυτής.



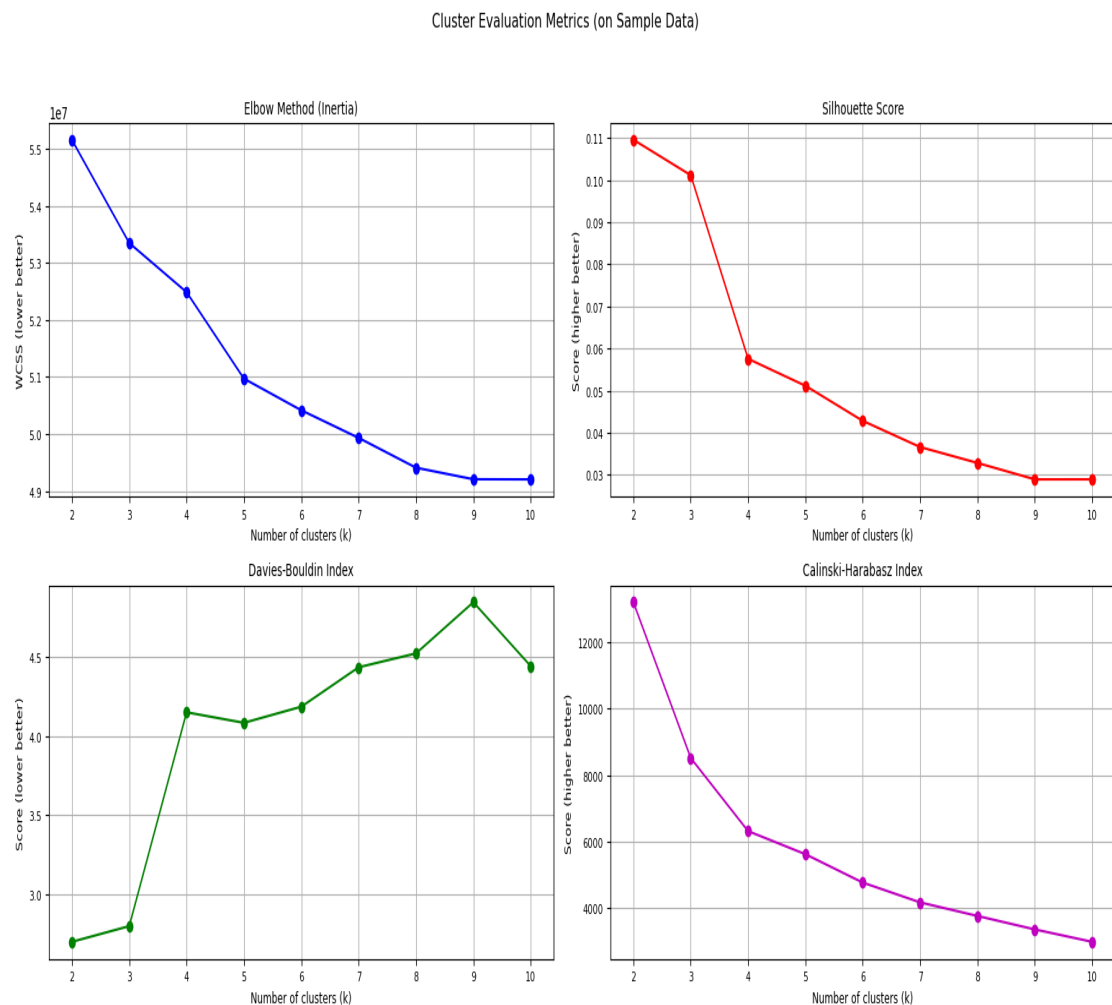
Εικόνα 58: Πίνακας συσχέτισης (correlation matrix) μεταξύ των ιδιοδιανυσμάτων (eigenvectors) των έξι κύριων συνιστωσών που προέκυψαν από την ανάλυση των λογαριθμικών δεδομένων. Οι τιμές κοντά στο μηδέν εκτός της διαγωνίου επιβεβαιώνουν την ορθογωνιότητα της μεθόδου.

Οι χάρτες των έξι πρώτων κύριων συνιστωσών (**PCs**) που προέκυψαν παρέχουν μια πολυδιάστατη οπτικοποίηση της φασματικής μεταβλητότητας. Κάθε χάρτης απεικονίζει τη χωρική κατανομή μιας συγκεκριμένης "πηγής" διακύμανσης:

- **PCA Comp. 1:** Αυτή η συνιστώσα, που εξηγεί το μεγαλύτερο ποσοστό της μεταβλητότητας, επιτυγχάνει έναν σαφή διαχωρισμό μεταξύ των κεντρικών μορφών και του υποβάθρου. Οι θετικές τιμές (κίτρινο) αντιστοιχούν στις μορφές, ενώ οι αρνητικές τιμές (μωβ) στο φόντο, υποδεικνύοντας μια ισχυρή **αντι-συσχέτιση** μεταξύ των υλικών αυτών των δύο περιοχών. Ουσιαστικά, η **PC1** αντιπροσωπεύει τη βασική χημική διαφορά μεταξύ των χρωματισμένων περιοχών και του στρώματος προετοιμασίας (gesso).
- **PCA Comp. 2 & 3:** Αυτές οι συνιστώσες απομονώνουν πιο εξειδικευμένες πηγές μεταβλητότητας. Η **PC 2** φαίνεται να σχετίζεται έντονα με τα φωτοστέφανα και τις περιοχές των ενδυμάτων, ενώ η **PC 3** χαρτογραφεί κυρίως τις καφε-κόκκινες περιοχές. Αυτό υποδηλώνει ότι αυτές οι συνιστώσες αντιστοιχούν στη φασματική υπογραφή συγκεκριμένων χρωστικών (π.χ., χρυσού/χαλκού και Υδραργύρου (Hg) για κόκκινο μίνιο για την **PC2** και σιδήρου για την **PC3**).

- **PCA Comp. 4, 5 & 6:** Οι συνιστώσες υψηλότερης τάξης αποκαλύπτουν λεπτότερες διαφοροποιήσεις. Η **PC 6**, για παράδειγμα, φαίνεται να απομονώνει τις περιοχές των σαρκωμάτων (πρόσωπα και χέρια) από τα ενδύματα, υποδεικνύοντας ότι η ανάμειξη χρωστικών που χρησιμοποιήθηκε για το δέρμα (πιθανώς λευκό του μολύβδου με ώχρες και Λευκού Zn (ZnO) έχει μια μοναδική φασματική υπογραφή. Είναι σημαντικό να τονιστεί ότι η **PCA** παρέχει ορθογώνιες (μη συσχετισμένες) συνιστώσες, όπως επιβεβαιώνεται από τον πίνακα συσχέτισης, αλλά η φυσική τους ερμηνεία απαιτεί συνδυασμό με τις άλλες μεθόδους.

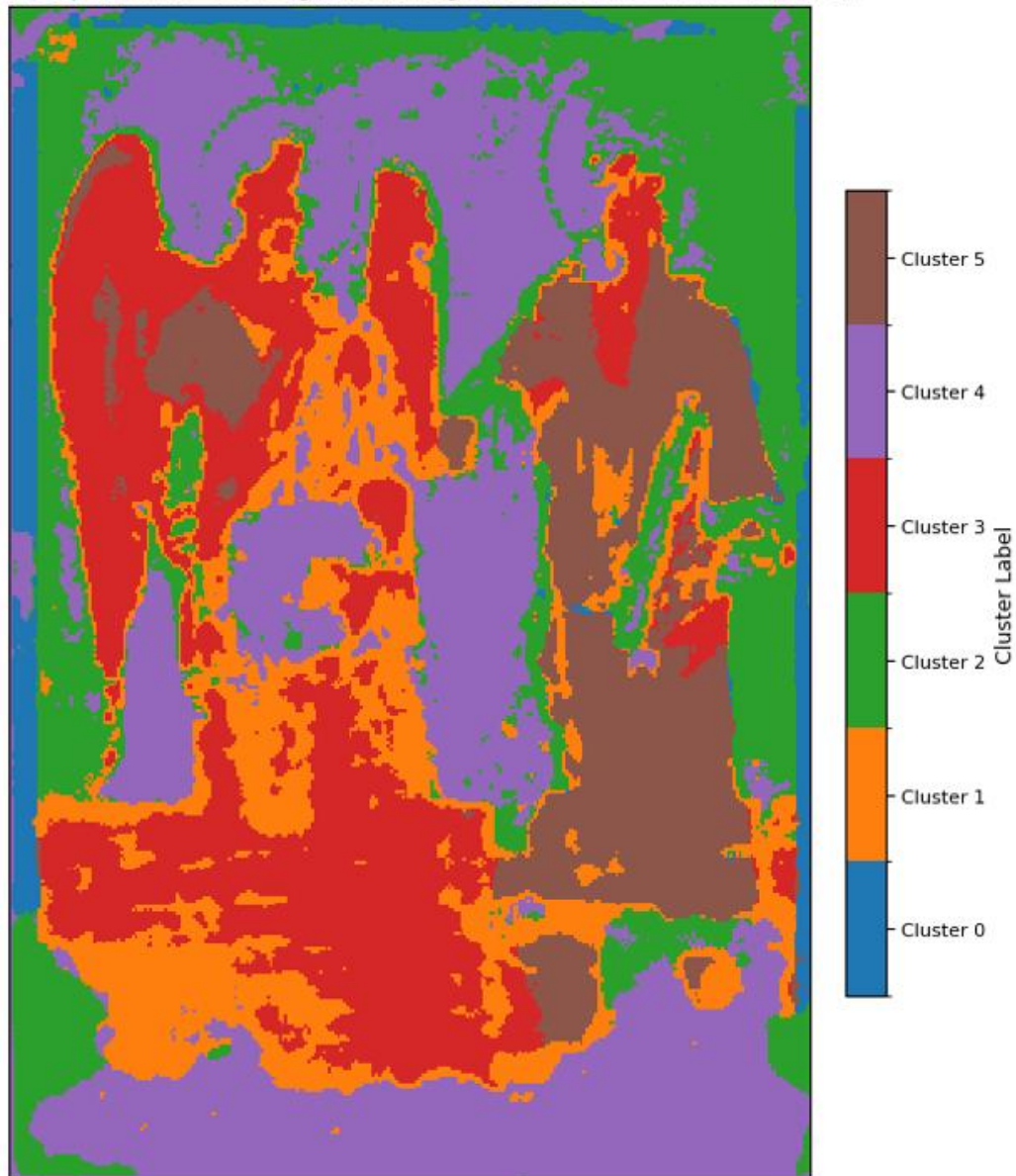
Για την εφαρμογή συσταδοποίησης πλέον επειδή έχουμε λογαρίθμηση τα δεδομένα μας θα αλλάξει και ο κατάλληλος αριθμός συστάδων Εικόνα 59.



Εικόνα 59: Αποτελέσματα από την εκτέλεση του πλαισίου αξιολόγησης για τον προσδιορισμό του βέλτιστου αριθμού συστάδων (k) στα λογαριθμικά δεδομένα PCA. Οι μετρικές υποδεικνύουν ένα βέλτιστο εύρος για k μεταξύ 4 και 7, επιβεβαιώνοντας τη διατήρηση του αριθμού των 6 συστάδων για συνέπεια.

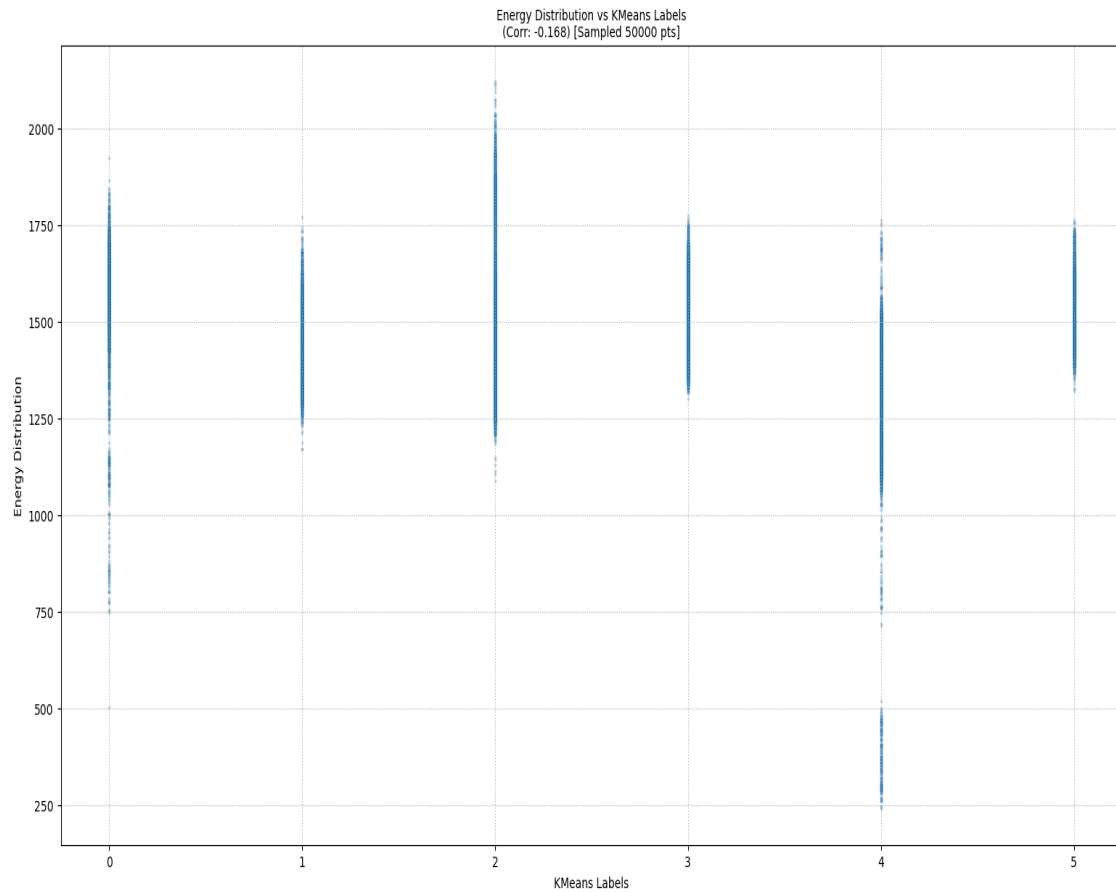
Μέσω του προγράμματος αλλά και της εικόνας που απεικονίζεται παραπάνω κατανοούμε πως ένας αριθμός αναμεσα από 4 -7 συστάδες είναι κατάλληλος Εικόνα 59. Οπου σε αυτή την περίπτωση μπορούμε να κρατήσουμε τον ίδιο αριθμό όπως και πριν. Προχωρούμε δηλαδή στην εφαρμογή **6 συστάδων** και της μεθόδους **Kmeans** Εικόνα 60.

KMeans Cluster Map
(Input: Step 2: PCA (k=6) on Original Data (Log (value+1) scale, Chunks) (99736, 6))



Εικόνα 60: Θεματικός χάρτης (cluster map) που προκύπτει από την εφαρμογή του αλγορίθμου K-Means ($k=6$) στα αποτελέσματα της PCA επί των λογαριθμικών δεδομένων. Ο χάρτης αυτός αποκαλύπτει καλύτερο διαχωρισμό σε περιοχές με υλικά υψηλού Z (π.χ. χρυσός, μόλυβδος) σε σύγκριση με την ανάλυση στη γραμμική κλίμακα.

Το αποτέλεσμα της εικόνας των συστάδων επάνω στην αρχική πλέον δείχνει καλύτερο διαχωρισμό σε σημεία με υψηλότερη ενέργεια και βαρύτερα στοιχεία, δηλαδή χρυσό, μόλυβδο κ.ο.κ. Μια εικόνα βέβαια που αν συγκρίνουμε με τις αρχικές στα δεδομένα που δεν έχουμε λογαρίθμηση λαμβάνουμε πληροφορίες για την τοποθεσία στοιχείων που βρίσκονται στο υπόβαθρο ή βρίσκονται σε λεπτά στρώματα.



Εικόνα 61: Διάγραμμα διασποράς της συνολικής έντασης ανά συστάδα K-Means.

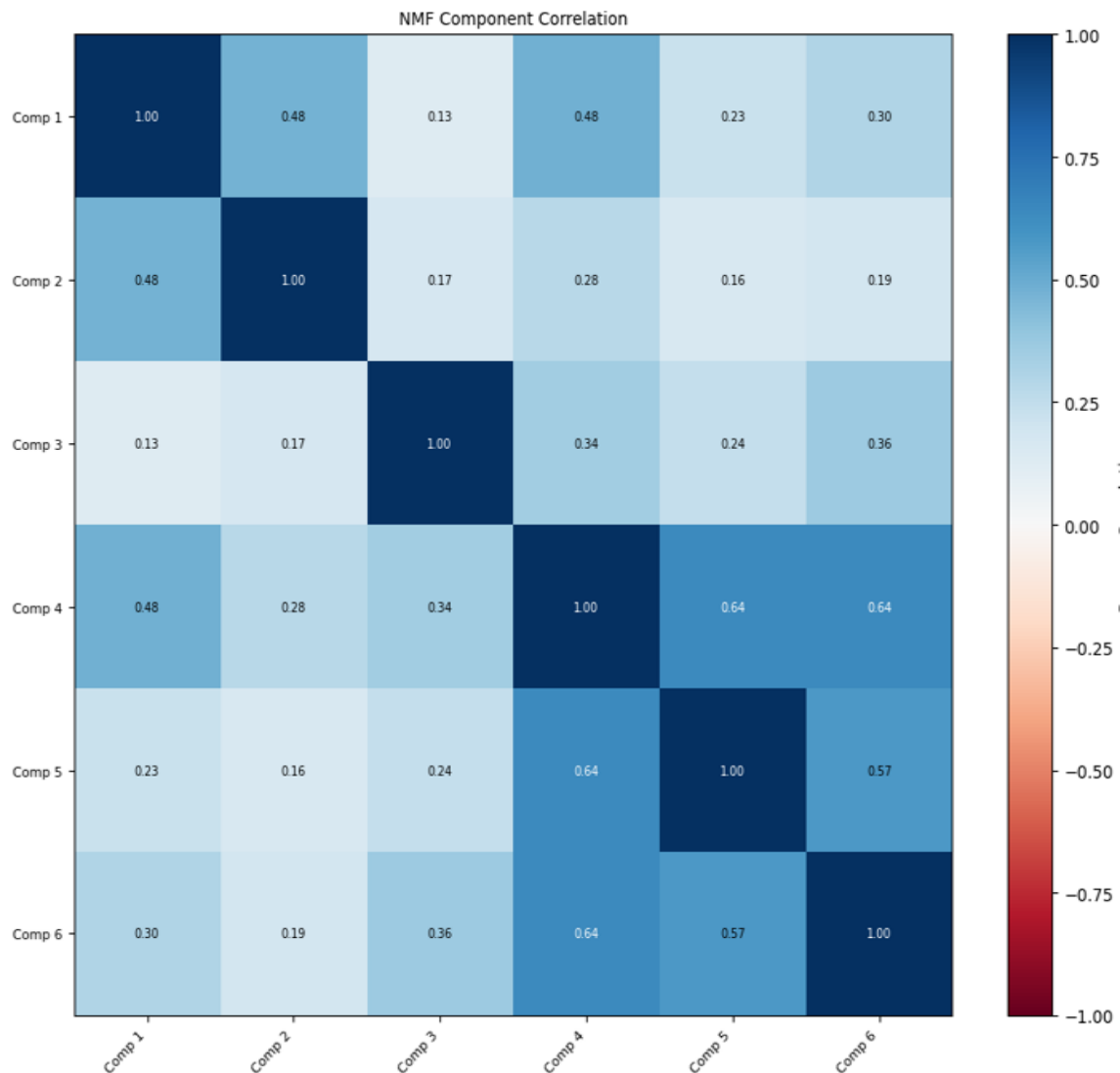
Ανάλυση NMF:

Πλέον μπορούμε να προχωρήσουμε και την εφαρμογή της **NMF** στα λογαριθμικά δεδομένα μας, με τους χάρτες των αξόνων επάνω στην αρχική εικόνα να παρέχονται παρακάτω Εικόνα 62.



Εικόνα 62: Αποτελέσματα από την εφαρμογή της μεθόδου *NMF* στα λογαριθμικά δεδομένα.

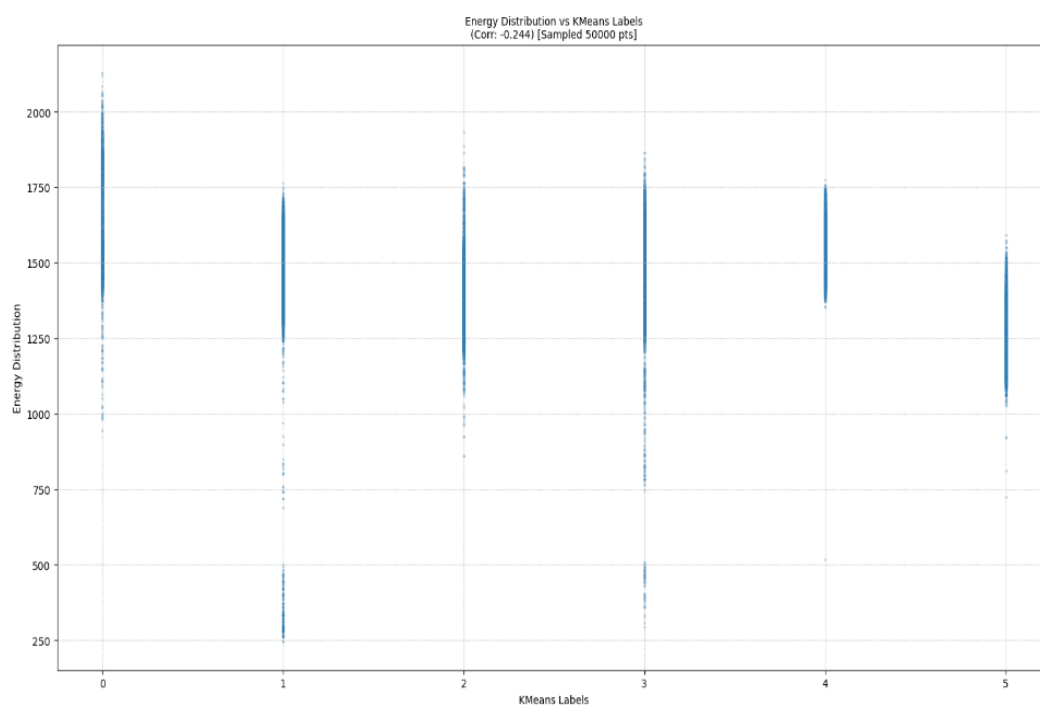
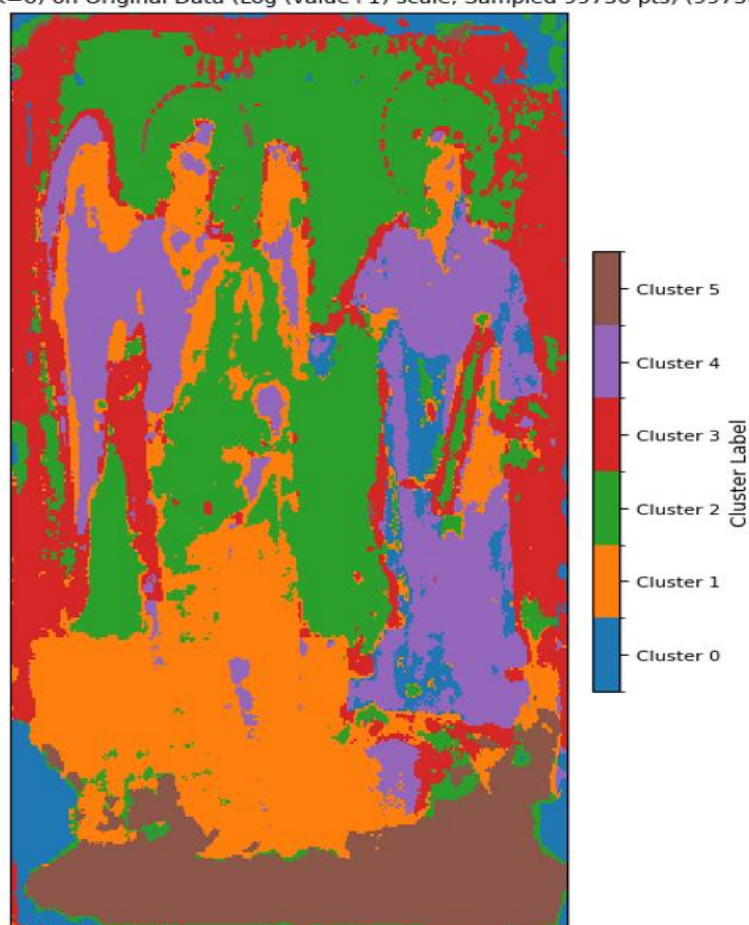
Σε σύγκριση με την εικόνα της *PCA* Εικόνα 57 στα λογαριθμικά δεδομένα και πάλι βλέπουμε μια καλύτερη διάκριση της έντασης των αξόνων λαμβάνοντας έτσι καλύτερη πληροφορία. Περιοχές ενδιαφέροντος όπως μανδύες φτερά και στέφανα είναι αρκετά ευδιάκριτα και βρίσκονται σε διάφορους άξονες, με τις υψηλότερες ενέργειες να είναι και πάλι πιο διακριτές. Βεβαια παρακάτω εμφανίζουμε και την εικόνα της συσχέτισης μεταξύ των κυρίων αξόνων για την μέθοδο της *NMF* Εικόνα 63.



Εικόνα 63: Πίνακας συσχέτισης των συνιστωσών NMF. Σε αντίθεση με την PCA, οι συνιστώσες της NMF δεν είναι κατ' ανάγκη ορθογώνιες, και ο πίνακας αποκαλύπτει μέτριες έως ισχυρές θετικές συσχετίσεις, υποδεικνύοντας κοινά φασματικά χαρακτηριστικά ή χωρική συνύπαρξη μεταξύ ορισμένων υλικών.

Όπως κατανοούμε και γνωρίζουμε από την θεωρία οι άξονες με βάση την **NMF** δεν είναι απολύτως μη συσχετιζόμενοι μεταξύ τους όπως είναι στην **PCA** αλλά έχουν αρκετό βαθμό συσχέτισης. Πράγμα που επιβεβαιώνεται τόσο από το παραπάνω πίνακα όσο και από την απεικόνιση τους συναρτήσει της έντασης πιο πάνω. Τέλος προχωρούμε και στην συσταδοποίηση με τον ίδιο αριθμό συστάδων που έχουμε σ όλη την πειραματική διαδικασία μέσω της μεθόδου **Kmeans**. Οι εικόνες για τις συστάδες βρίσκονται παρακάτω αλλά και στο παράρτημα Εικόνα 64.

KMeans Cluster Map
(Input: Step 8: NMF ($k=6$) on Original Data (Log (value+1) scale, Sampled 99736 pts) (99736, 6))

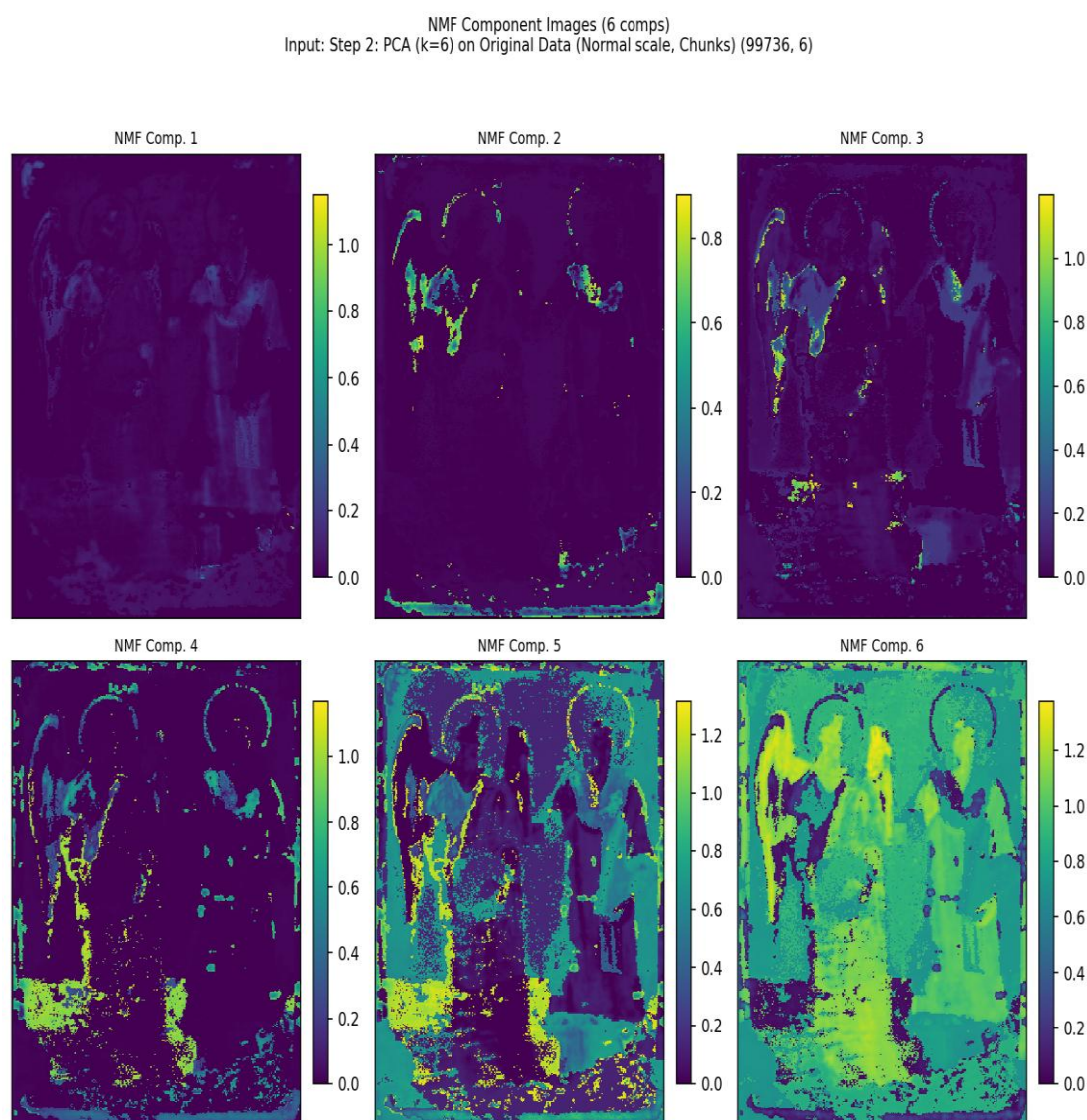


Εικόνα 64: Οπτικοποίηση των αποτελεσμάτων της συσταδοποίησης K-Means ($k=6$) που εφαρμόστηκε στα δεδομένα της NMF. (Πάνω) Ο τελικός θεματικός χάρτης. (Κάτω) Το διάγραμμα κατανομής της ενέργειας ανά συστάδα.

Υβριδική Ανάλυση NMF-PCA: Από τους Χάρτες Υλικών στους Χάρτες Τεχνικής

Παραπάνω μπορούμε να κατανοήσουμε και την απόδοση της ενεργειακής κατανομής σε σχέση με τα clusters δίνοντας μας και μια καλύτερη κατανόηση ποιες τιμές και ποια **Clusters** αποδίδονται σε βαρύτερα στοιχεία και ποια σε ελαφρύτερα. Βρισκόμαστε σε θέση να ολοκληρώσουμε το πείραμα μας με το 3^ο βήμα που είναι η εφαρμογή της μεθόδου της **NMF** επάνω στα αποτελέσματα της μεθόδου της **PCA** Εικόνα 65. Δηλαδή εφαρμόζουμε την **PCA** με έξι (6) κύριους άξονες και έπειτα τα αποτελέσματα αυτής της ανάλυσης τα τοποθετούμε για ένα επιπλέον βήμα ανάλυσης μέσω της μεθόδου **NMF** και πάλι με έξι (6) συνιστώσες για καλύτερη και ευκολότερη συνέχεια. Αυτή η προσέγγιση, γνωστή ως **NMF-PCA**, συνδυάζει τα πλεονεκτήματα και των δύο μεθόδων: η **PCA** λειτουργεί ως ένα αποτελεσματικό "φίλτρο θορύβου" και μειωτής διαστάσεων, ενώ η **NMF** αναλαμβάνει να μετασχηματίσει τις αφηρημένες, ορθογώνιες συνιστώσες σε μη-αρνητικούς χάρτες με άμεση φυσικοχημική ερμηνεία.

Τα αποτελέσματα της **NMF-PCA** όπως ονομάζεται φαίνονται παρακάτω Εικόνα 72.



Εικόνα 65: Αποτελέσματα από την υβριδική μέθοδο NMF-PCA. Οι χάρτες των έξι συνιστωσών που προκύπτουν προσφέρουν εξαιρετική ευκρίνεια και άμεση φυσικοχημική ερμηνεία, διαχωρίζοντας με σαφήνεια το υπόστρωμα, τα φωτιστέφανα, τα ενδύματα και τα σαρκόματα.

Εάν ανατρέξουμε στα αποτελέσματα των ξεχωριστών αναλύσεων τόσο της **PCA** όσο και της **NMF** μπορούμε να διακρίνουμε ευκρινέστατα τον διαχωρισμό περιοχών των Αγίων, φωτοστέφανα, μανδύες, φτερά, ένα μεγάλο εύρος πληροφορίας. Επειδή η **NMF** δεν έχει μέθοδο **Explained Variance** ανατρέχουμε στο πρόγραμμα μας το οποίο έχει ρυθμιστεί να ανατρέχει αλγόριθμο υπολογισμού της πληροφορίας που λαμβάνουμε μέσω της **NMF**, υπολογίστηκε να είναι περίπου στο **98%**. Οπότε κατανοούμε πως το μέγεθος της πληροφορίας αλλά και ο χρόνος συνολικός της εκτέλεσης βρίσκεται εντός επιτρεπτών ορίων δεδομένου, εφόσον για την εκτέλεση απαιτούνται περίπου **20 λεπτά**.

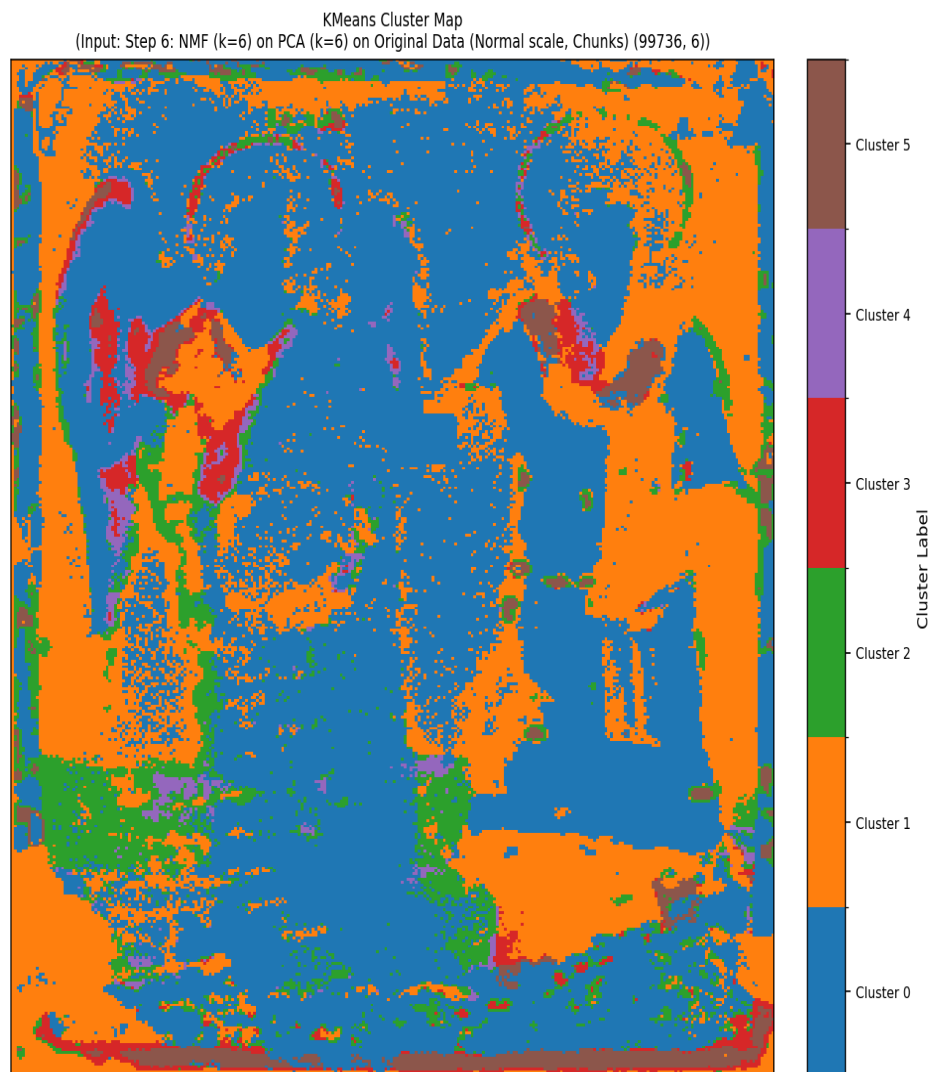
Η εφαρμογή της υβριδικής μεθόδου **NMF_PCA** απέδωσε χάρτες συνιστωσών με εξαιρετική ευκρίνεια και άμεση φυσικοχημική ερμηνεία. Σε αντίθεση με την **PCA**, οι χάρτες της **NMF** είναι μη-αρνητικοί και μπορούν να θεωρηθούν ως χάρτες αφθονίας για συγκεκριμένες χημικές ομάδες.

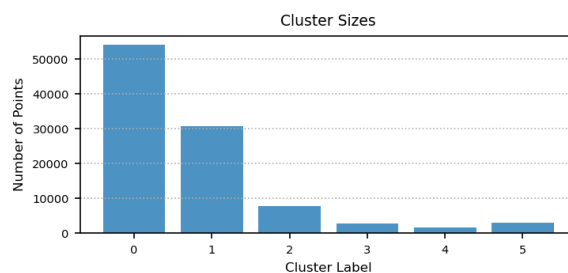
- NMF Comp. 1 & 3 (Γαιώδεις Χρωστικές - Fe): Αυτές οι δύο συνιστώσες σχετίζονται άμεσα με τα ενδύματα των Αγίων. Η χωρική τους κατανομή αντιστοιχεί στον στοιχειακό χάρτη του Σιδήρου (Fe), αντιπροσωπεύοντας τις γήινες χρωστικές (ώχρες). Η Comp. 3 φαίνεται να απομονώνει τις πιο έντονες/πυκνές περιοχές της ώχρας, ενώ η Comp. 1 τις πιο αραιές στρώσεις ή τις περιοχές φωτοσκίασης.
- NMF Comp. 2 (Χρυσός - Au & Μολύβδου - Pb): Αυτή η συνιστώσα είναι εξαιρετικά ενδιαφέρουσα καθώς ομαδοποιεί δύο διαφορετικά, αλλά πολύτιμα, υλικά. Απομονώνει με ακρίβεια τα φωτοστέφανα, που αντιστοιχούν στον Χρυσό (Au), καθώς και τα ενδύματα, που αντιστοιχούν στον μόλυβδο. Αυτό υποδηλώνει ότι ο αλγόριθμος τα αναγνώρισε ως μια στατιστικά διακριτή ομάδα "υλικών έμφασης".
- NMF Comp. 4 (Λευκό του Μολύβδου - Pb): Η συνιστώσα αυτή χαρτογραφεί με απόλυτη ακρίβεια τα σαρκώματα (πρόσωπα, χέρια) και τα φωτίσματα (highlights). Η κατανομή της είναι πανομοιότυπη με τον στοιχειακό χάρτη του Μολύβδου (Pb), ταυτοποιώντας έτσι το Λευκό του Μολύβδου.
- NMF Comp. 5 & 6 (Υπόστρωμα - Ca, S): Αυτές οι δύο συνιστώσες κυριαρχούν στο φόντο, στο έδαφος και στις περιοχές με εκτεταμένες φθορές. Η κατανομή τους ταυτίζεται με τους στοιχειακούς χάρτες του Ασβεστίου (Ca) και του Θείου (S), απομονώνοντας έτσι το υπόστρωμα προετοιμασίας (gesso). Η Comp. 6, με την υψηλότερη ένταση, πιθανότατα αντιπροσωπεύει τις περιοχές καθαρού, εκτεθειμένου gesso, ενώ η Comp. 5 μπορεί να αντιπροσωπεύει περιοχές όπου το gesso είναι καλυμμένο από λεπτά, φθαρμένα στρώματα άλλων υλικών. Βεβαία παρατηρούμε πως η συνιστώσα 6 έντονα παρουσιάζει και περιοχές ενδυμάτων των Αγίων και φτερών αντιστοιχώντας σε στοιχειακούς χάρτες του σιδήρου.

Μέθοδος	Πλεονεκτήματα	Περιορισμοί
PCA	Γραμμικές συσχετίσεις, ορθογώνιες συνιστώσες	Δύσκολη ερμηνεία αρνητικών τιμών

Μέθοδος	Πλεονεκτήματα	Περιορισμοί
<i>NMF</i>	Μη-αρνητικές συνιστώσες, καλύτερη ερμηνεία	Απαιτεί προεπιλογή components

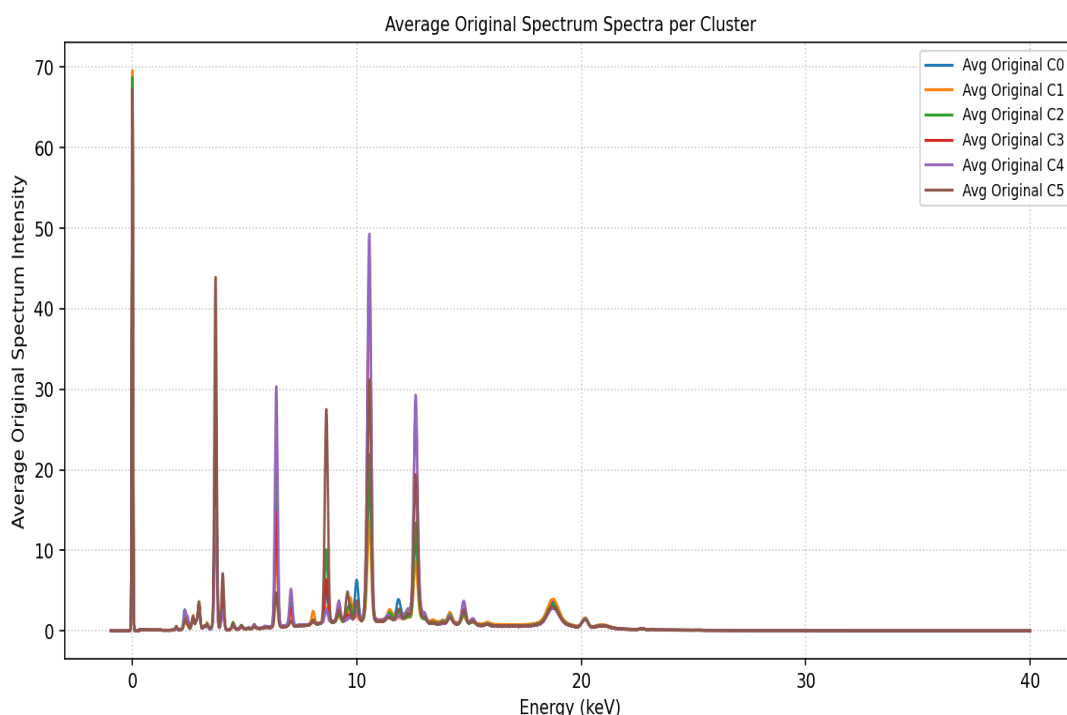
Επομένως μπορούμε να προχωρήσουμε και στην εφαρμογή της μεθόδου συσταδοποίησης της *Kmeans* με αριθμό συστάδων να είναι έξι όπως και προηγουμένως Εικόνα 66. Με τα αποτελέσματα να είναι προφανέστατα καλύτερα σε σχέση με την προηγούμενη ανάλυση. Διότι πλέον βρισκόμαστε σε θέση να δούμε περισσότερη πληροφορία αλλά και παρατηρούμε περιοχές και συστάδες όπου προηγουμένως ήταν ομοιόμορφες πλέον να εμφανίζουν μια μεγαλύτερη ποικιλομορφία.





Εικόνα 66: Αποτελέσματα της συσταδοποίησης *K-Means* ($k=6$) στα δεδομένα που προέκυψαν από την υβριδική ανάλυση *NMF-PCA*. **(Αριστερά)** Ο θεματικός χάρτης απεικονίζει τη χωρική κατανομή των υλικών ομάδων. **(Δεξιά)** Το ιστόγραμμα μεγέθους δείχνει τον αριθμό των *pixels* σε κάθε συστάδα, παρέχοντας ποσοτική εκτίμηση της επικράτησης κάθε υλικού.

Επιπρόσθετα τα αποτελέσματα μεγεθών κάθε συστάδας αλλά και φάσματος συναρτήσει της ενέργειας είναι τα εξής:



Εικόνα 67: Μέσα φάσματα για κάθε μία από τις έξι συστάδες της ανάλυσης *K-Means* επί των δεδομένων *NMF-PCA*. Οι διακριτές φασματικές υπογραφές επιτρέπουν την απόδοση χημικής ταυτότητας σε κάθε χρωματική ομάδα του θεματικού χάρτη.

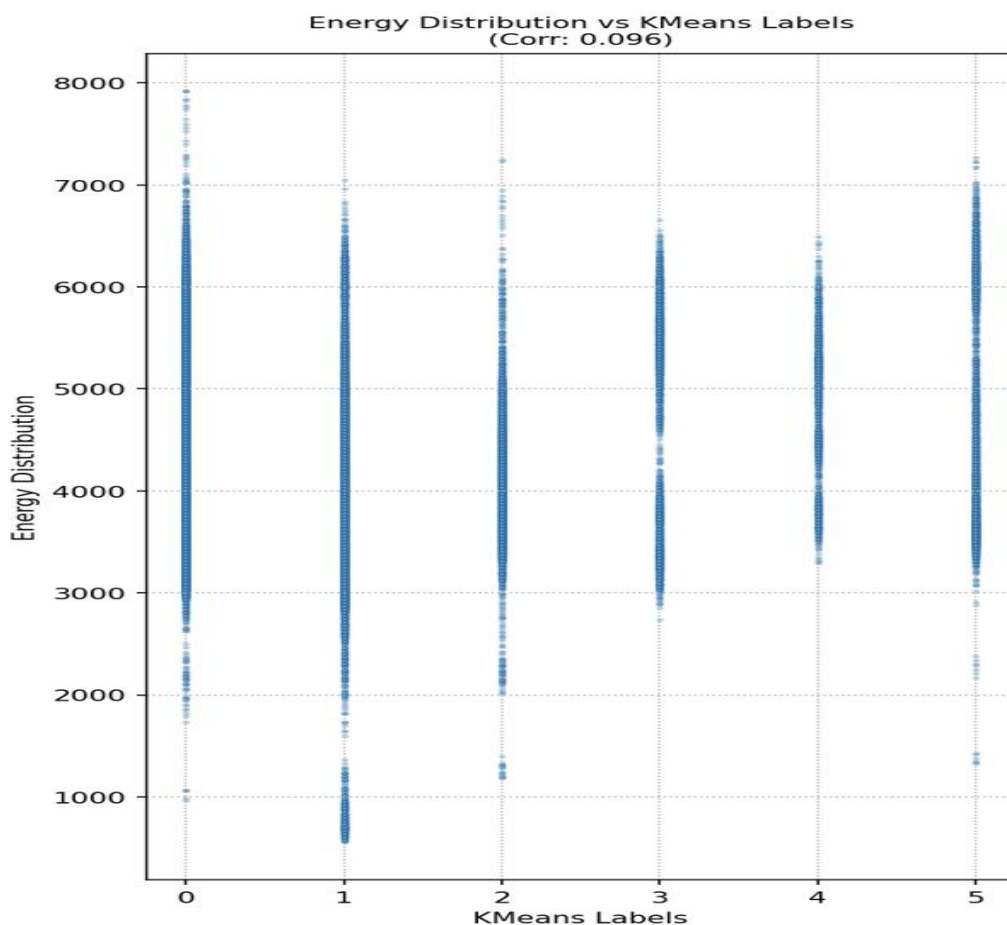
Τέλος η κατανομή της ενέργειας σε σχέση με την συστάδα δίνεται από το παρακάτω διάγραμμα Εικόνα 68.

Ο τελικός χάρτης ομαδοποίησης που προέκυψε από την εφαρμογή του αλγορίθμου ***K-Means*** στα αποτελέσματα της ***NMF-PCA***, αποτελεί τη σύνθεση όλων των προηγούμενων αναλύσεων. Αυτός ο χάρτης δεν είναι απλώς μια εικόνα, αλλά ένας θεματικός χάρτης υλικών, όπου κάθε χρώμα αντιστοιχεί σε μια ομάδα ***pixels*** με στατιστικά παρόμοια χημική σύσταση.

- **Clusters 0 & 1 (Μπλε & Πορτοκαλί):** Αυτές οι δύο ομάδες κυριαρχούν στο φόντο και πιθανώς αντιπροσωπεύουν περιοχές του στρώματος προετοιμασίας (***gesso***) με ελαφρώς διαφορετική σύσταση ή επιφανειακή κατάσταση.

- **Cluster 2 (Πράσινο):** Συμπίπτει απόλυτα με την κατανομή της *NMF_Comp. 3*, ταυτοποιώντας την περιοχή που έχει ζωγραφιστεί με χρωστικές σιδήρου (ώχρες).
- **Cluster 3 (Κόκκινο):** Εντοπίζει μικρότερες, αλλά διακριτές περιοχές στα ενδύματα και τα φτερά, υποδεικνύοντας τη χρήση μιας άλλης, λιγότερο διαδεδομένης χρωστικής, πιθανότατα με βάση τον χαλκό.
- **Cluster 4 (Μωβ):** Αντιστοιχεί με ακρίβεια στις επιχρυσωμένες περιοχές.
- **Cluster 5 (Καφέ):** Ταυτοποιεί τα σαρκώματα.

Η διαδικασία αυτή μετατρέπει τα συνεχή δεδομένα των συνιστωσών σε έναν διακριτό, ταξινομημένο χάρτη, ο οποίος είναι εξαιρετικά χρήσιμος για τους συντηρητές και τους ιστορικούς τέχνης, καθώς παρέχει μια σαφή και συνολική εικόνα της κατανομής των υλικών.

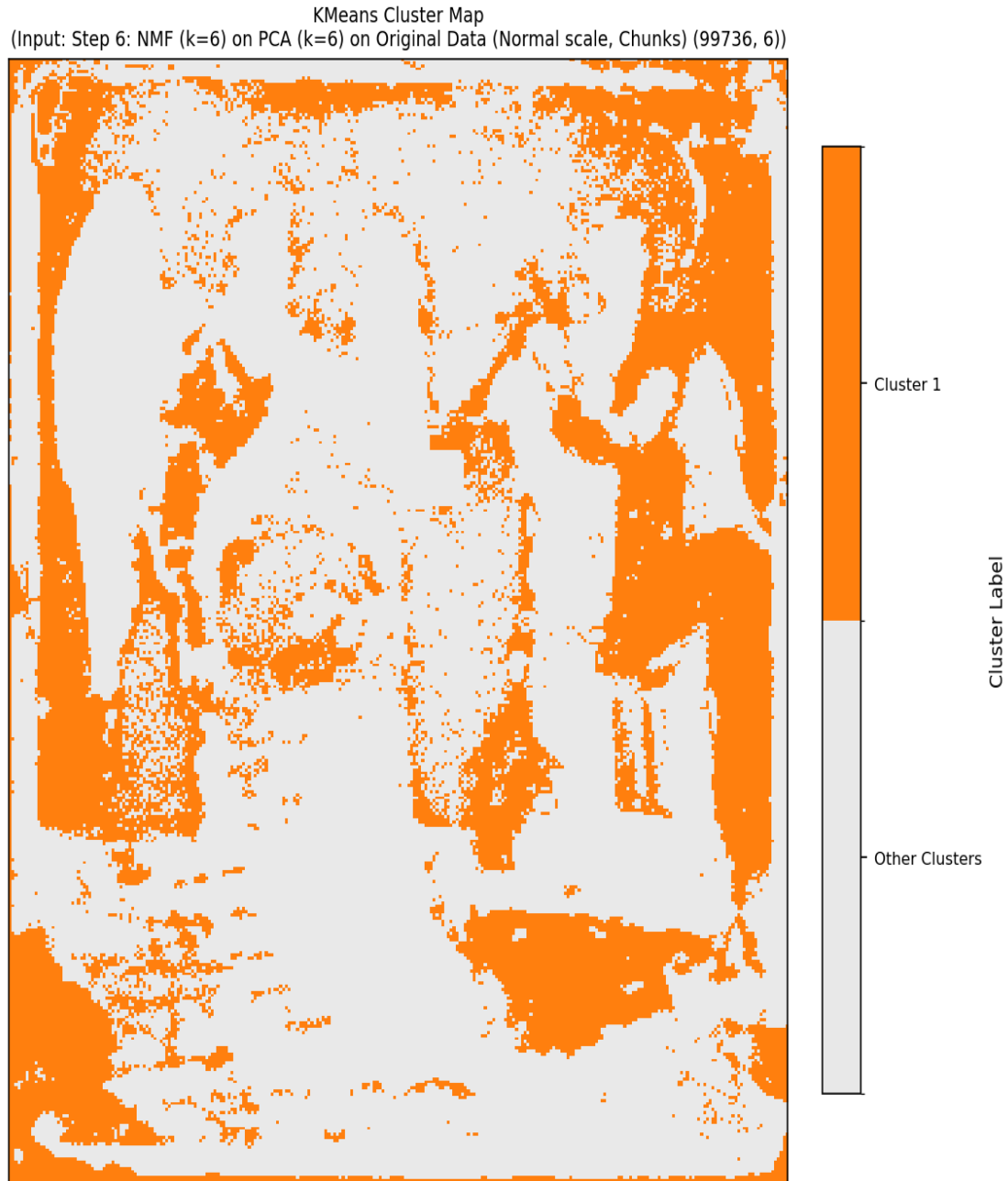


Εικόνα 68: Διάγραμμα διασποράς της κατανομής της έντασης (*Intensity Distribution*) ανά συστάδα K-Means. Η χρωματική κλίμακα (*Pixel Court*) απεικονίζει την πυκνότητα των σημείων, δείχνοντας τις πιο συχνές τιμές έντασης για κάθε ομάδα.

Για την ολοκλήρωση της ανάλυσης πριν την εφαρμογή αλγορίθμου για την εύρεση των στοιχειακών χαρτών περιοχή ενδιαφέροντος στην εικόνα αποτελούν τα **Clusters 1** και **3** εφόσον αποτελούν τους μανδύες και τα φωτοστέφανα των Αγίων. Επομένως συνετό είναι να προχωρήσουμε σε **Sub-clustering** των δυο αυτών περιοχών της εικόνας, με την ίδια ακριβώς μεθοδολογία δηλαδή την επιβολή των αποτελεσμάτων της κάθε συστάδας σε εύρεση των επομένων **6** συστάδων της κάθε μιας. Αυτό μας επέτρεψε να

διερευνήσουμε αν υπάρχουν λεπτές διαφοροποιήσεις μέσα σε αυτές τις περιοχές, αποκαλύπτοντας πιθανώς διαφορετικές στρώσεις ή αναμειγξεις χρωστικών που δεν ήταν ορατές στην αρχική ανάλυση. Επομένως έχουμε τα εξής αποτελέσματα:

Για την Συστάδα 1 έχουμε Εικόνα 69:

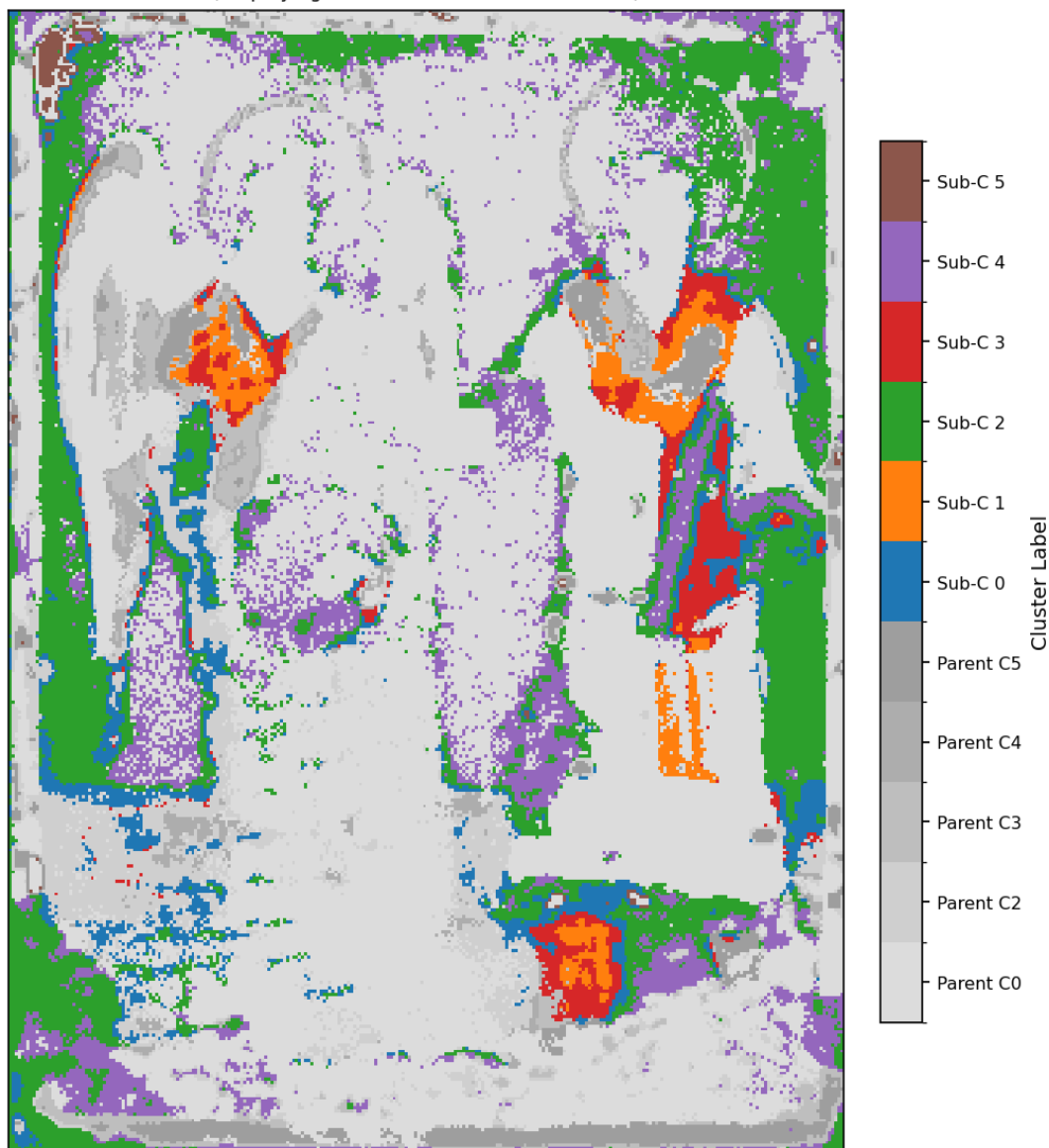


Εικόνα 69: Οπτική απομόνωση της "Cluster 1" από τον θεματικό χάρτη της υβριδικής μεθόδου NMF-PCA. Η περιοχή ενδιαφέροντος (πορτοκαλί) αναδεικνύεται, ενώ οι υπόλοιπες συστάδες εμφανίζονται με ουδέτερο χρώμα (ανοιχτό γκρι) για την προετοιμασία της διαδικασίας υπο-συσταδοποίησης (sub-clustering).

Αποτελεί την περιοχή που μας ενδιαφέρει απομονωμένη από τις υπόλοιπες συστάδες, επομένως εφαρμόζουμε τον αλγόριθμο Kmeans στην περιοχή αυτή.

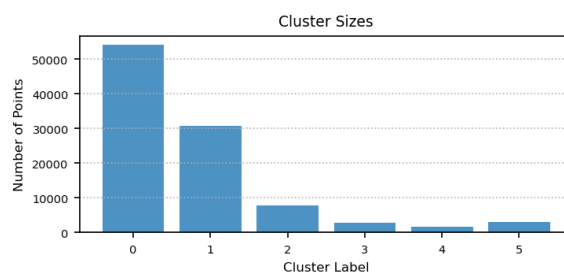
Τα αποτελέσματα είναι τα πως θα έχουμε μια ευκρινέστερη κατανομή και υποκατηγοριοποίηση της εικόνας Εικόνα 70. Πλέον διακρίνουμε καλύτερα τις περιοχές που αποτελούν την συστάδα μας, όπως φαίνεται παρακάτω.

KMeans Cluster Map
 (Input: Step 8: Data from Cluster 1 (Parent: Step 6: NMF (k=6) on PCA (k=6)...) (30627, 6))
 (Displaying Sub-clusters on Parent Context)



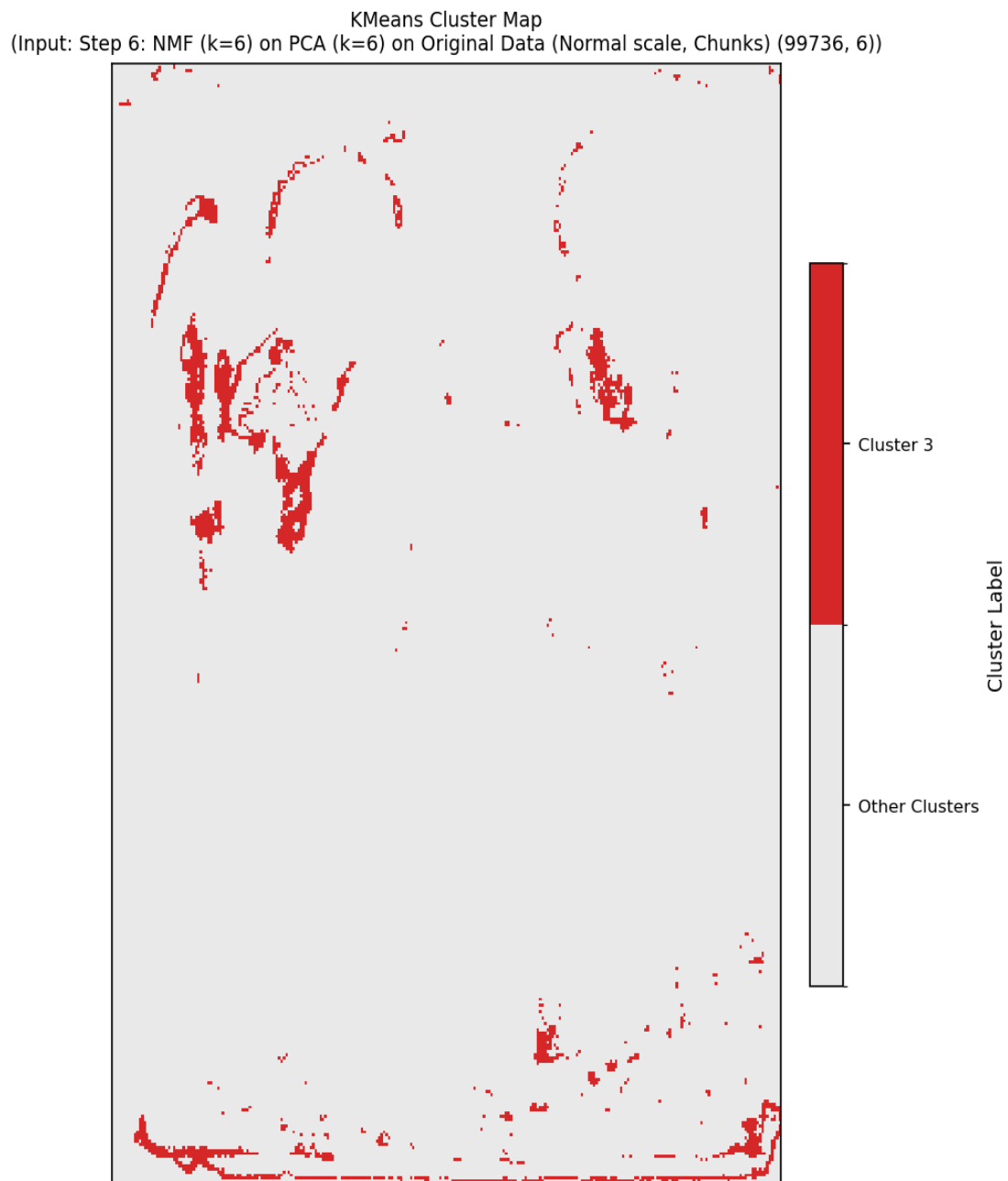
Εικόνα 70: Αποτελέσματα της ιεραρχικής ανάλυσης (sub-clustering) που εφαρμόστηκε στην "Cluster 1". Ο τελικός θεματικός χάρτης όπου διακρίνονται έξι διακριτές υπο-ομάδες (Sub-C 0-5), αποκαλύπτοντας λεπτομερή στρωματογραφία εντός της αρχικής περιοχής.

Τέλος παρέχουμε και τις πληροφορίες για το μέγεθος κάθε συστάδας Εικόνα 71.



Εικόνα 71: Το ιστόγραμμα που απεικονίζει το πλήθος των pixels για κάθε υπο-συστάδα.

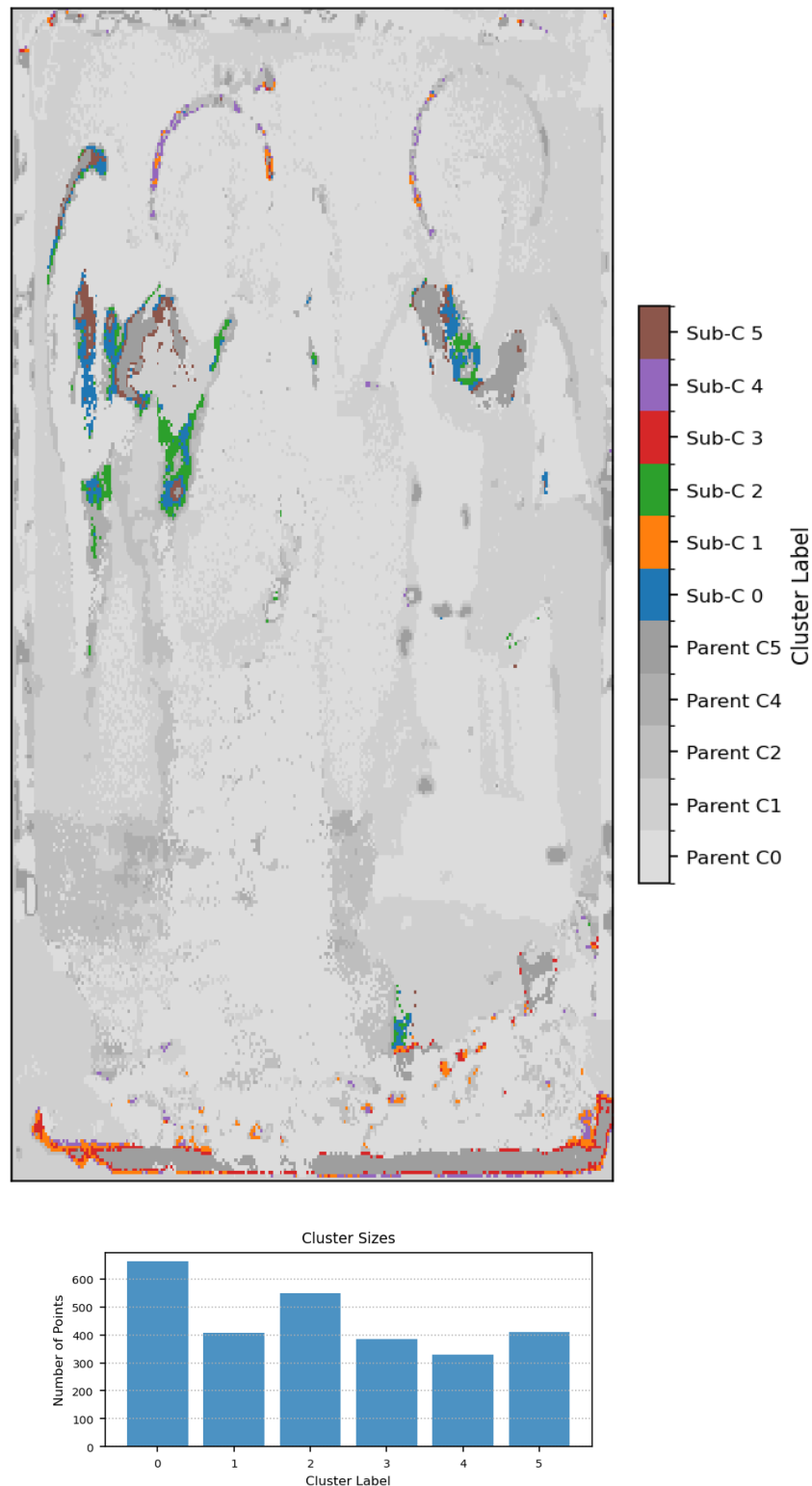
Τέλος για την συστάδα 3 έχουμε πως η αρχική εικόνα χωρίς την περαιτέρω πληροφορία είναι η εξής Εικόνα 72:



Εικόνα 72: Οπτική απομόνωση της "Cluster 3" (κόκκινο), η οποία αντιστοιχεί κυρίως στα ενδύματα, πριν την εφαρμογή της διαδικασίας sub-clustering για την περαιτέρω διερεύνηση της χημικής της σύστασης.

Με την ίδια ακριβώς μεθοδολογία θα έχουμε πάλι **6** συστάδες που θα αποτελούν την αρχική συστάδα **3**. Αρά η εικόνα μας θα λάβει την εξής μορφή Εικόνα 73:

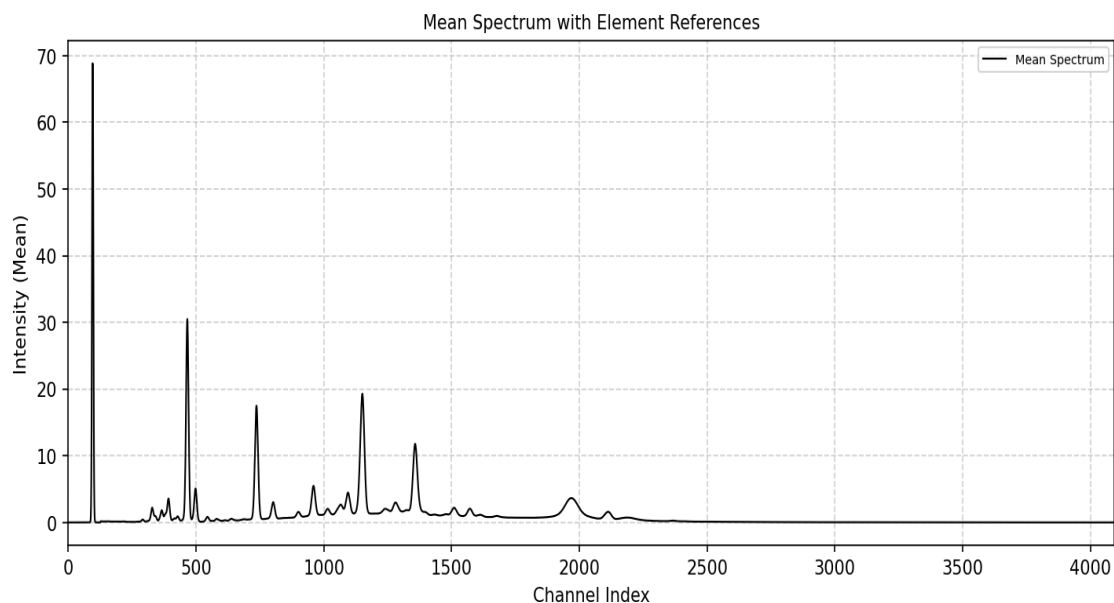
KMeans Cluster Map
 (Input: Step 11: Data from Cluster 3 (Parent: Step 6: NMF (k=6) on PCA (k=6)...) (2744, 6))
 (Displaying Sub-clusters on Parent Context)



Εικόνα 73: Αποτελέσματα της ιεραρχικής ανάλυσης (sub-clustering) που εφαρμόστηκε στην "Cluster 3". **(Πάνω)** Ο θεματικός χάρτης που αποκαλύπτει την εσωτερική δομή της συστάδας με έξι νέες υπο-ομάδες. **(Κάτω)** Το ιστόγραμμα μεγέθους για κάθε υπο-συστάδα.

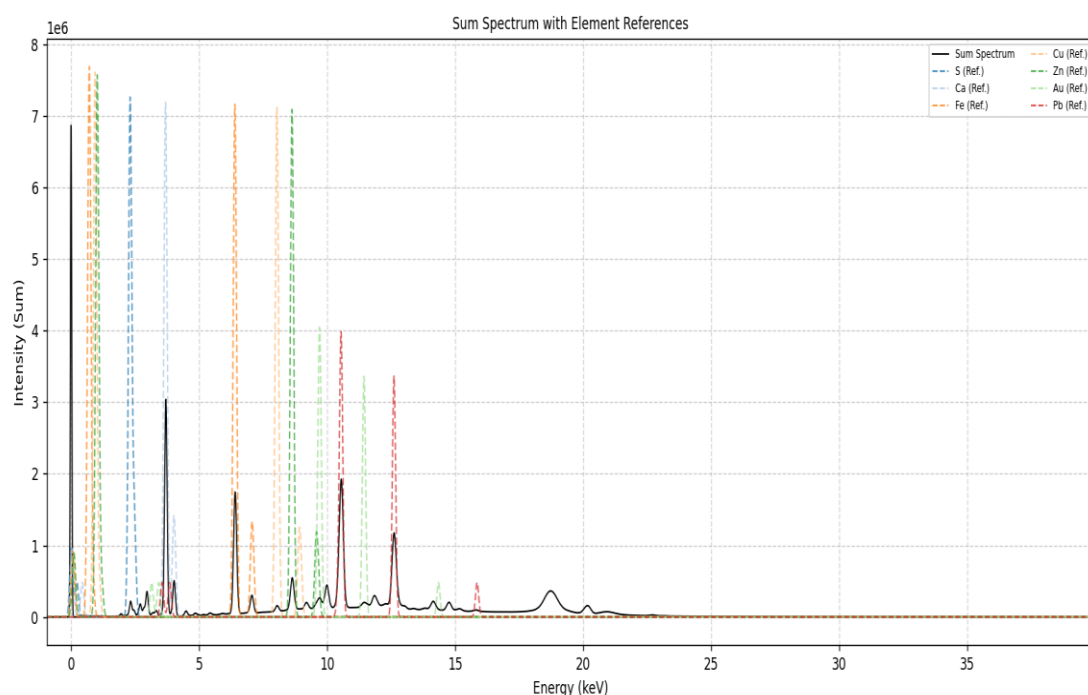
Τώρα πλέον βρισκόμαστε σε θέση να προχωρήσουμε στην εύρεση των στοιχειακών χαρτών. Από το πρόγραμμα μπορούμε να επιλέξουμε στο αρχικό φάσμα που φαίνεται

παρακάτω Εικόνα 74 στοιχεία τα οποία κατανέμονται είτε αυτόματα είτε με χειροκίνητο τρόπο επάνω στον άξονα των ενεργειών εφόσον έχουμε τοποθετήσει προφανώς την κατάλληλη βαθμονόμηση.



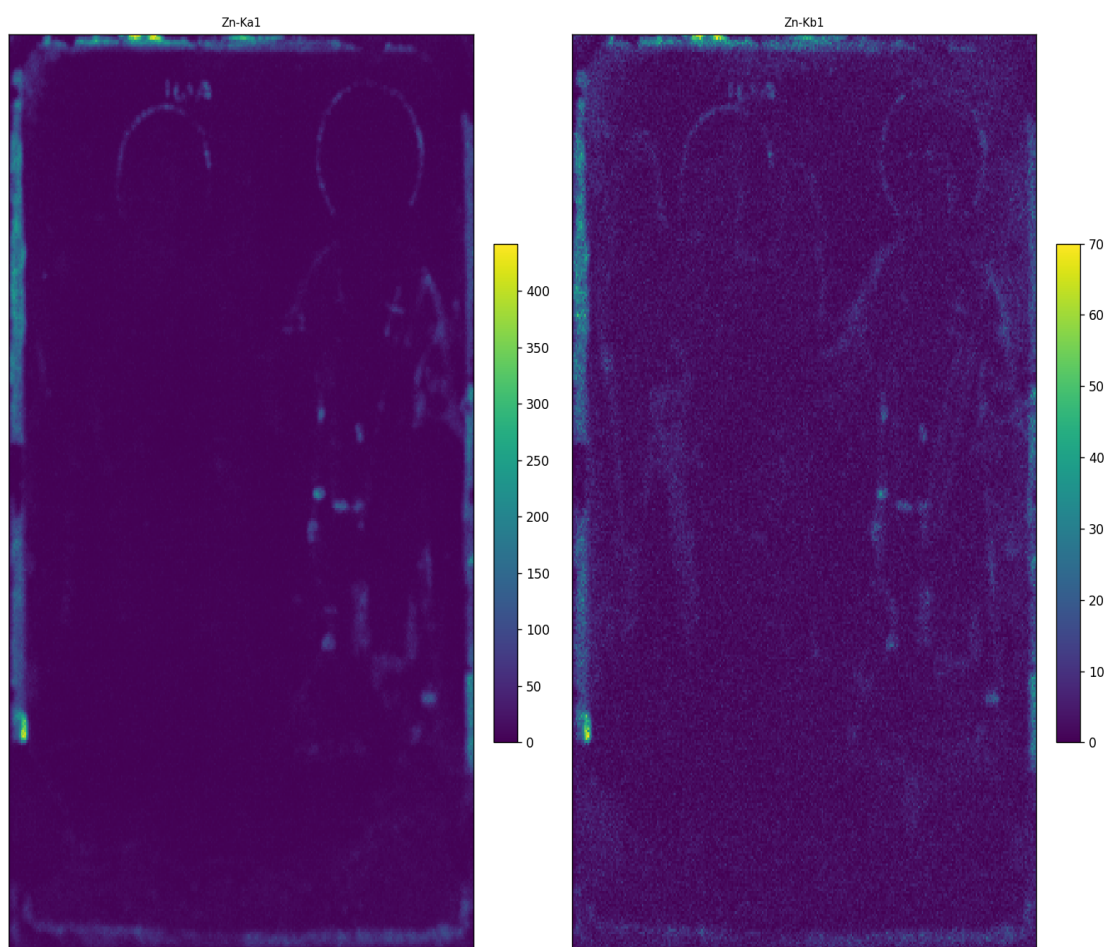
Εικόνα 74: Αρχικό φάσμα στο πρόγραμμα εντοπισμού στοιχείων

Με την κατάλληλη επιλογή στοιχείων μπορούμε να έχουμε την εξής μορφή Εικόνα 75:



Εικόνα 75: Διαδικασία βαθμονόμησης και ταυτοποίησης στοιχείων στο μέσο φάσμα. **(Πάνω)** Το αρχικό μέσο φάσμα ως συνάρτηση του αριθμού καναλιού του ανιχνευτή. **(Κάτω)** Το βαθμονομημένο φάσμα σε κλίμακα ενέργειας (keV), με τις θεωρητικές θέσεις των χαρακτηριστικών γραμμών εκπομπής για επιλεγμένα στοιχεία (Cu, S, Zn, Fe,...) να επισημαίνονται με διακεκομμένες γραμμές.

Επομένως μπορούμε να προχωρήσουμε στην εισαγωγή αυτών των στοιχείων για την εκτίμηση και απεικόνιση των στοιχειακών τους χαρτών επάνω στην αρχική μας εικόνα.

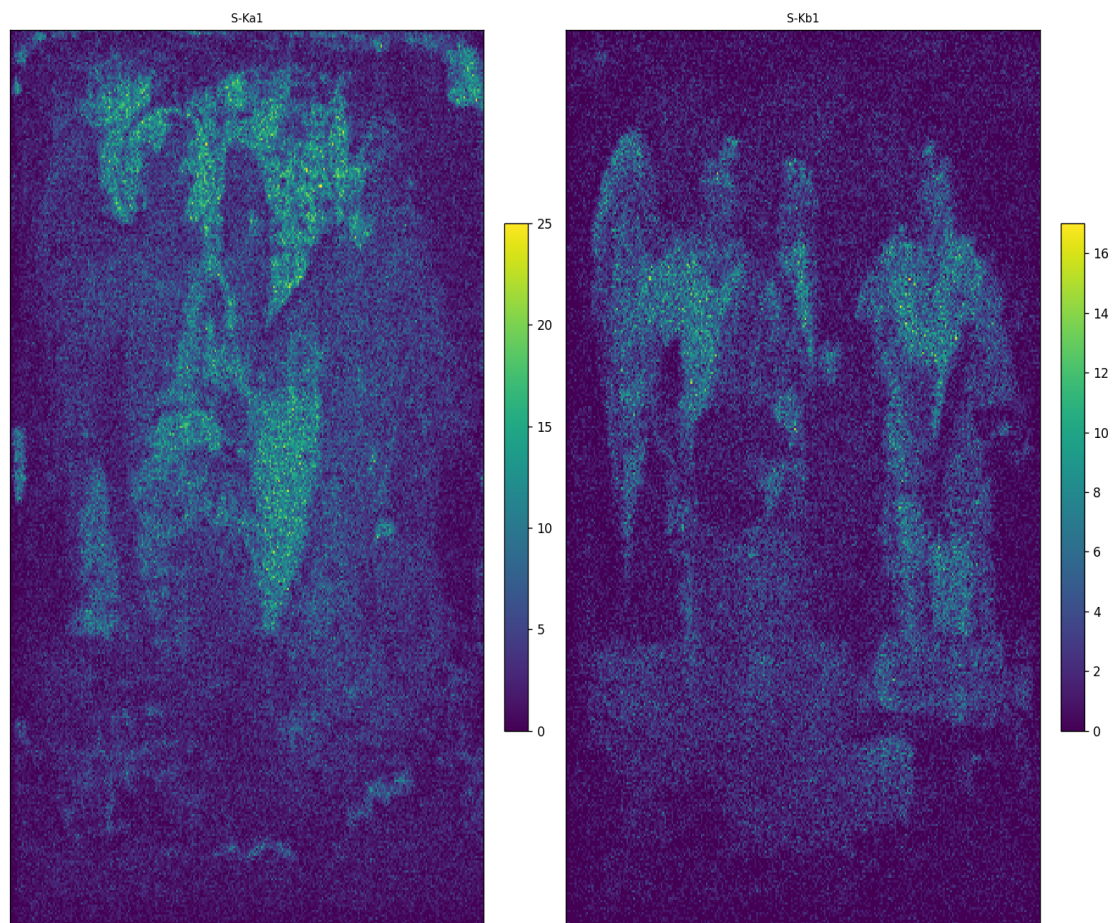


Εικόνα 76: Στοιχειακοί χάρτες κατανομής για τον Ψευδάργυρο (Zn), που απεικονίζουν την ένταση των γραμμών εκπομπής Zn-K edge, Ka1 και Kb1. Η σποραδική και τοπικά περιορισμένη παρουσία του στοιχείου αποτελεί ισχυρή ένδειξη μεταγενέστερων επεμβάσεων συντήρησης με χρήση Λευκού του Ψευδαργύρου (ZnO).

Οι χάρτες κατανομής του Ψευδαργύρου (Zn) Εικόνα 76, που απεικονίζουν τις εκπομπές **Zn-Ka1**, **Ka2** και **Kb1**, αποκαλύπτουν μια εξαιρετικά σποραδική και τοπικά περιορισμένη παρουσία του στοιχείου. Σε αντίθεση με άλλα στοιχεία όπως ο μόλυβδος ή ο σίδηρος, ο ψευδάργυρος δεν συνδέεται με κάποια ευρεία, αναγνωρίσιμη περιοχή του αρχικού ζωγραφικού στρώματος, όπως ενδύματα ή σαρκώματα. Αντιθέτως, εντοπίζεται σε μικρές, ασύνδετες μεταξύ τους "κηλίδες", κυρίως στο φόντο και σε ορισμένα σημεία πάνω στις μορφές.

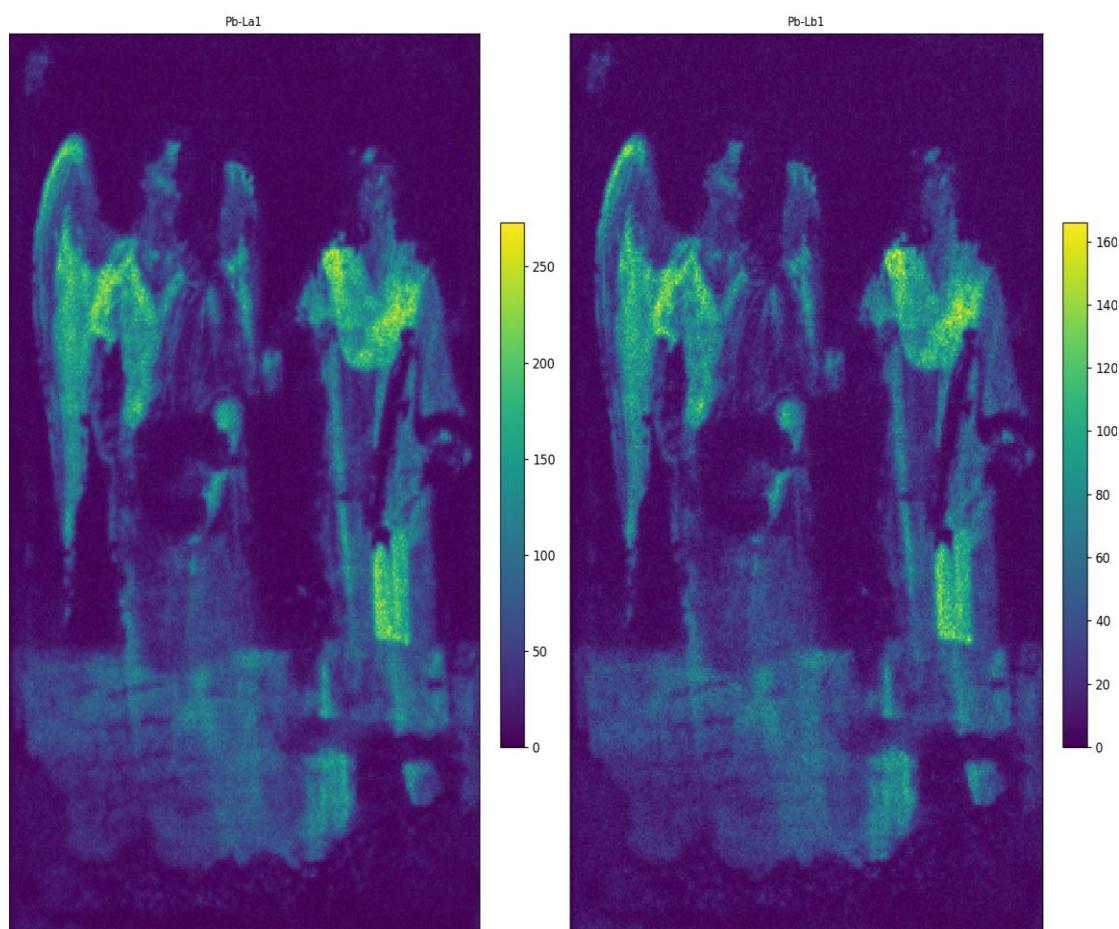
Αυτή η κατανομή είναι χαρακτηριστική της χρήσης του **Λευκού του Ψευδαργύρου (ZnO)**, μιας χρωστικής που έγινε εμπορικά διαθέσιμη από τα μέσα του 19ου αιώνα. Δεδομένης της χρονολόγησης της βυζαντινής αγιογραφίας, η παρουσία του ψευδαργύρου αποτελεί αδιαμφισβήτητη απόδειξη **μεταγενέστερων επεμβάσεων συντήρησης ή τοπικών επιζωγραφίσεων (retouching)**.

Η ισχυρή χωρική συσχέτιση μεταξύ των διαφορετικών γραμμών εκπομπής (**Ka** και **Kb**) επιβεβαιώνει με βεβαιότητα την παρουσία του στοιχείου, επιτρέποντάς μας να διακρίνουμε με ακρίβεια τις περιοχές που έχουν αλλοιωθεί από την αρχική τους κατάσταση.



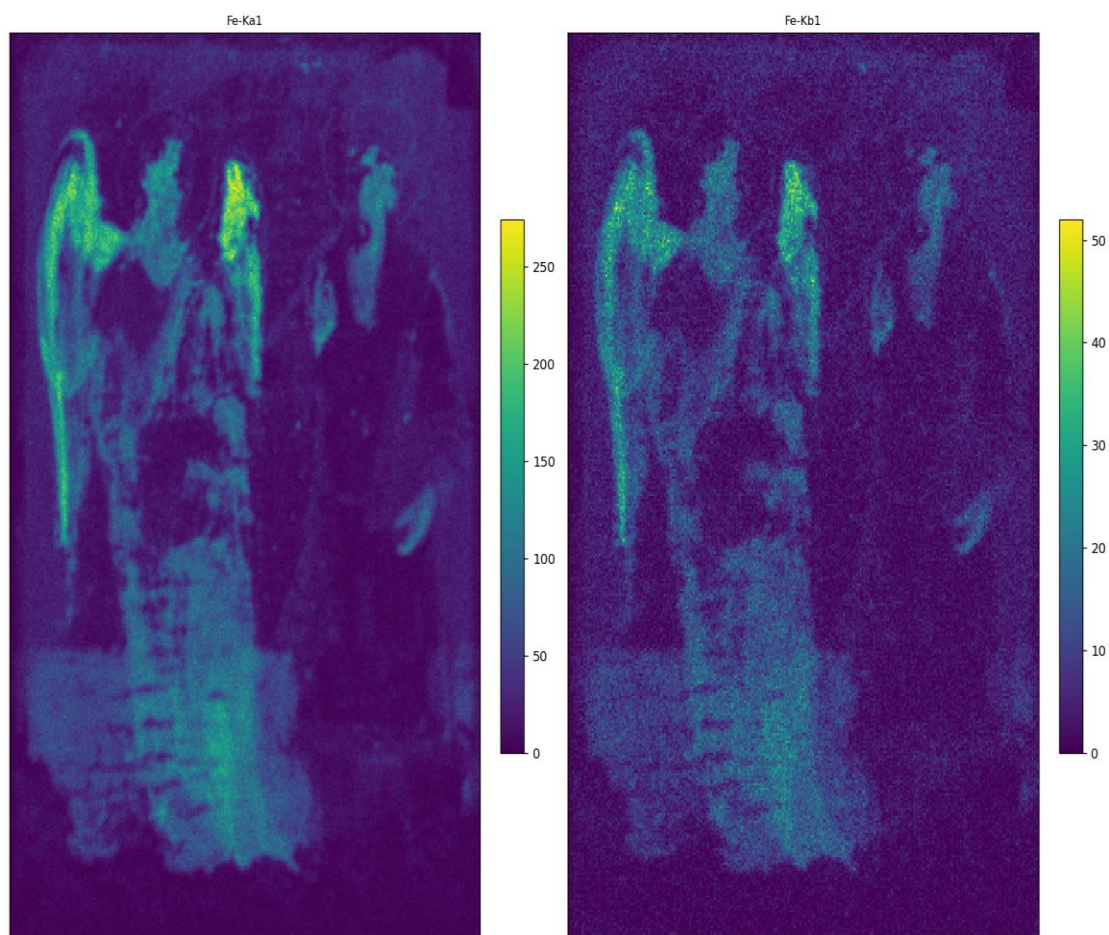
Εικόνα 77: Στοιχειακοί χάρτες κατανομής για το Θείο (S), για τις γραμμές S-K edge, Ka1, Ka2 και Kb1. Η διάχυτη και σχεδόν καθολική παρουσία του θείου επιβεβαιώνει τη χρήση γύψου ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) ως υλικό του υποστρώματος προετοιμασίας (gesso).

Οι χάρτες κατανομής του Θείου (S) Εικόνα 77, αν και εμφανίζουν σημαντικό θόρυβο λόγω της χαμηλής ενέργειας της χαρακτηριστικής ακτινοβολίας του (S-Kα), παρουσιάζουν μια διάχυτη και σχεδόν καθολική παρουσία σε ολόκληρη την επιφάνεια της εικόνας. Αυτό το πρότυπο είναι απόλυτα συμβατό με τη χρήση του θείου ως συστατικού του στρώματος προετοιμασίας (gesso), και συγκεκριμένα του γύψου (Θευκό Ασβέστιο, $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$), ο οποίος παραδοσιακά χρησιμοποιείται ως υπόστρωμα στην αγιογραφία. Είναι αξιοσημείωτο ότι η ένταση του σήματος του θείου εμφανίζεται ελαφρώς μειωμένη στις περιοχές των μορφών, όπου υπάρχουν παχιά και πυκνά στρώματα στοιχείων μεγάλου ατομικού αριθμού (όπως το λευκό του μολύβδου). Αυτό το φαινόμενο, γνωστό ως **εξασθένιση (attenuation)**, συμβαίνει επειδή οι ακτίνες-X χαμηλής ενέργειας του θείου απορροφώνται από τα υπερκείμενα στρώματα βαρέων στοιχείων και δεν φτάνουν στον ανιχνευτή. Συνεπώς, η κατανομή του θείου όχι μόνο επιβεβαιώνει τη χρήση παραδοσιακού γύψινου υποστρώματος, αλλά έμμεσα χαρτογραφεί και τις περιοχές με τα πυκνότερα ζωγραφικά στρώματα.



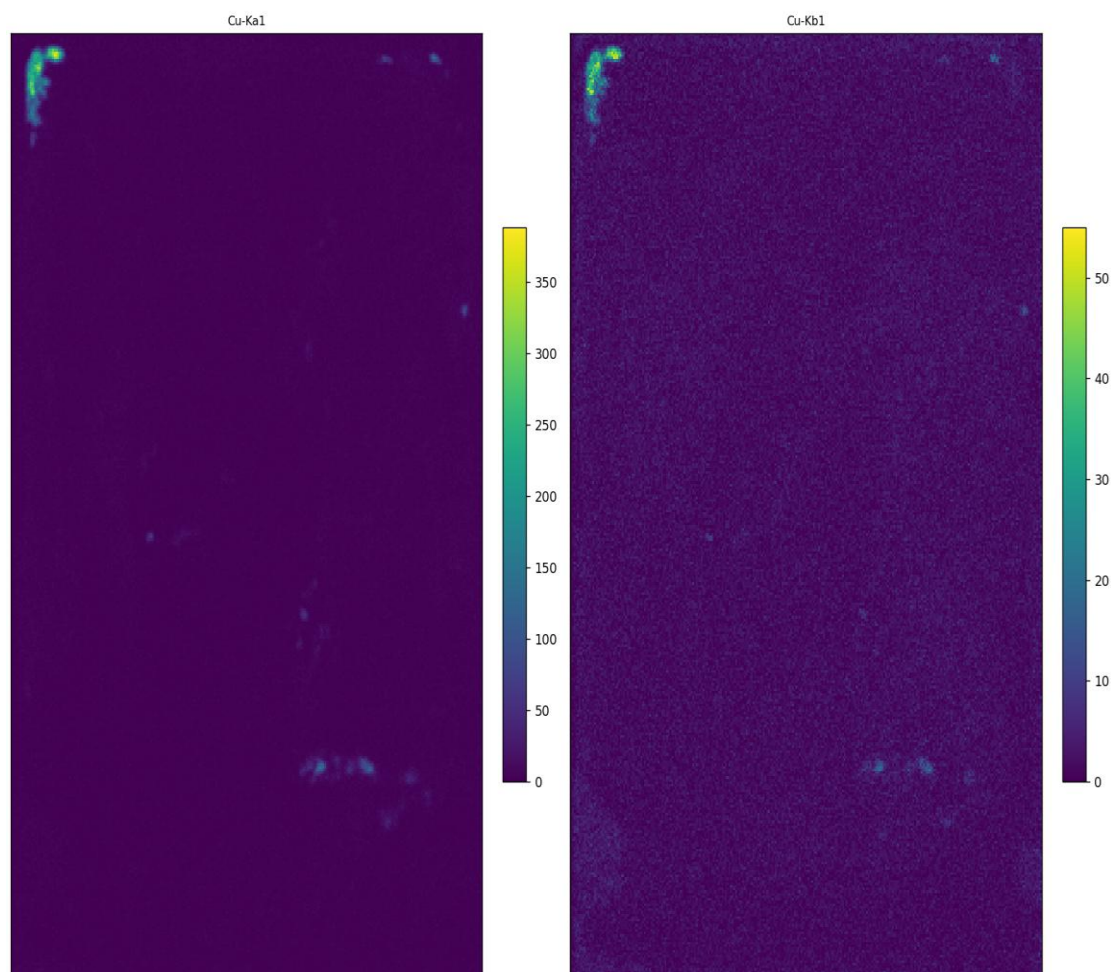
Εικόνα 78: Στοιχειακοί χάρτες κατανομής για τον Μόλυβδο (*Pb*), για τις γραμμές *La1*, *La2* και *Lb1*. Η υψηλή συγκέντρωση του μολύβδου στα σαρκώματα και στα "φωτισμάτα" των ενδυμάτων ταυτοποιεί τη χρήση του Λευκού του Μολύβδου ($2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2$), βασικής λευκής χρωστικής στην παραδοσιακή αγιογραφία.

Οι χάρτες κατανομής του Μολύβδου (*Pb*) Εικόνα 78, που προέρχονται από τις έντονες γραμμές εκπομπής *La1*, *La2* και *Lb1*, παρέχουν μια από τις πιο ευκρινείς και πληροφοριακές εικόνες για την τεχνική του καλλιτέχνη. Ο μολύβδος εντοπίζεται με πολύ υψηλή συγκέντρωση στις περιοχές των **σαρκωμάτων (πρόσωπα και χέρια)**, καθώς και στα φωτεινά σημεία και τις πτυχώσεις των ενδυμάτων που αποδίδουν όγκο και φως. Αυτή η κατανομή ταυτοποιεί αδιαμφισβήτητα τη χρήση του **Λευκού του Μολύβδου ($2\text{PbCO}_3 \cdot \text{Pb}(\text{OH})_2$)**. Το λευκό του μολύβδου ήταν η βασική λευκή χρωστική που χρησιμοποιούνταν για αιώνες λόγω της εξαιρετικής καλυπτικότητας και του λαμπερού του αποτελέσματος. Στην εικόνα, χρησιμοποιήθηκε σε ανάμειξη με ώχρες για τη δημιουργία των θερμών τόνων του δέρματος και αυτούσιο ή με ελάχιστες προσμίξεις για τη δημιουργία των "φωτισμάτων" (*highlights*). Η ακριβής ταύτιση του χάρτη του μολύβδου με το **Cluster 5** της ανάλυσης *NMF-PCA* επιβεβαιώνει την ικανότητα της πολυμεταβλητής ανάλυσης να απομονώνει χημικά διακριτές περιοχές με απόλυτη επιτυχία.



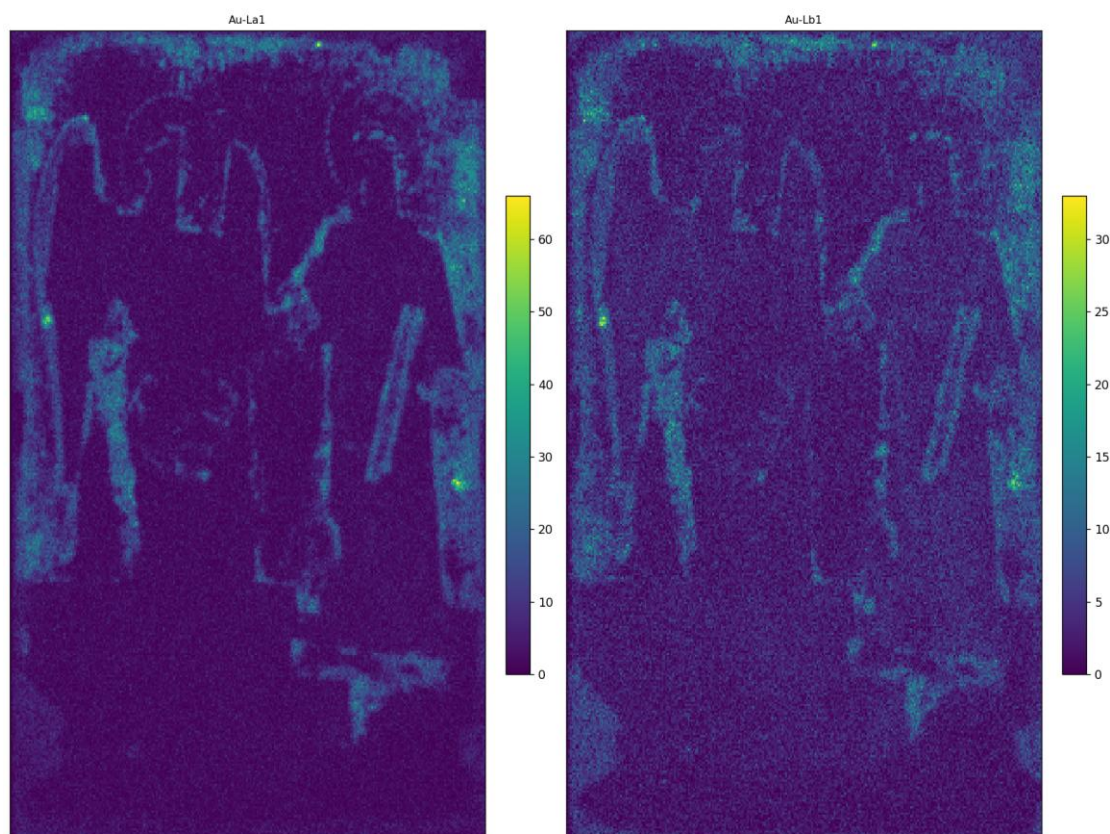
Εικόνα 79: Στοιχειακοί χάρτες κατανομής για τον Σίδηρο (Fe), για τις γραμμές Fe-K edge, Ka1, Ka2 και Kb1. Η κατανομή του σιδήρου συνδέεται άμεσα με τις καφε-κόκκινες και ωχρές περιοχές, υποδεικνύοντας τη χρήση γήινων χρωστικών (ώχρες) με βάση τα οξείδια του σιδήρου.

Η κατανομή του Σιδήρου (**Fe**) Εικόνα 79, όπως φαίνεται στους αντίστοιχους χάρτες (**Fe-Ka1, Ka2, Kb1**), συνδέεται άμεσα με τις **καφε-κόκκινες και ωχρές περιοχές των ενδυμάτων των Αγίων**. Η παρουσία του σιδήρου υποδηλώνει τη χρήση **γήινων χρωστικών με βάση τα οξείδια του σιδήρου**, όπως η κόκκινη και η κίτρινη ώχρα (Fe_2O_3) ή η *sienna*. Αυτές οι χρωστικές ήταν ευρέως διαδεδομένες, φθηνές και εξαιρετικά σταθερές, αποτελώντας βασικό στοιχείο της παλέτας της βυζαντινής αγιογραφίας. Ο χάρτης του σιδήρου ταυτίζεται σχεδόν απόλυτα με την κατανομή του **Cluster 2** από την ανάλυση **NMF-PCA**, αποδεικνύοντας ότι αυτές οι χρωστικές αποτελούν μια κύρια και διακριτή χημική ομάδα στην εικόνα. Μια μικρότερη συγκέντρωση σιδήρου παρατηρείται και στις περιοχές των σαρκωμάτων, γεγονός που υποδεικνύει ότι ο καλλιτέχνης ανέμιξε μικρή ποσότητα ώχρας με το λευκό του μολύβδου για να επιτύχει τη θερμή απόχρωση του δέρματος.



Εικόνα 80: Στοιχειακοί χάρτες κατανομής για τον Χαλκό (Cu), για τις γραμμές Cu-Kα1, Kα2 και Kβ1. Η στοχευμένη παρουσία του χαλκού ταυτοποιεί τη χρήση χρωστικών όπως ο μαλαχίτης ή ο αζουρίτης για συγκεκριμένες λεπτομέρειες.

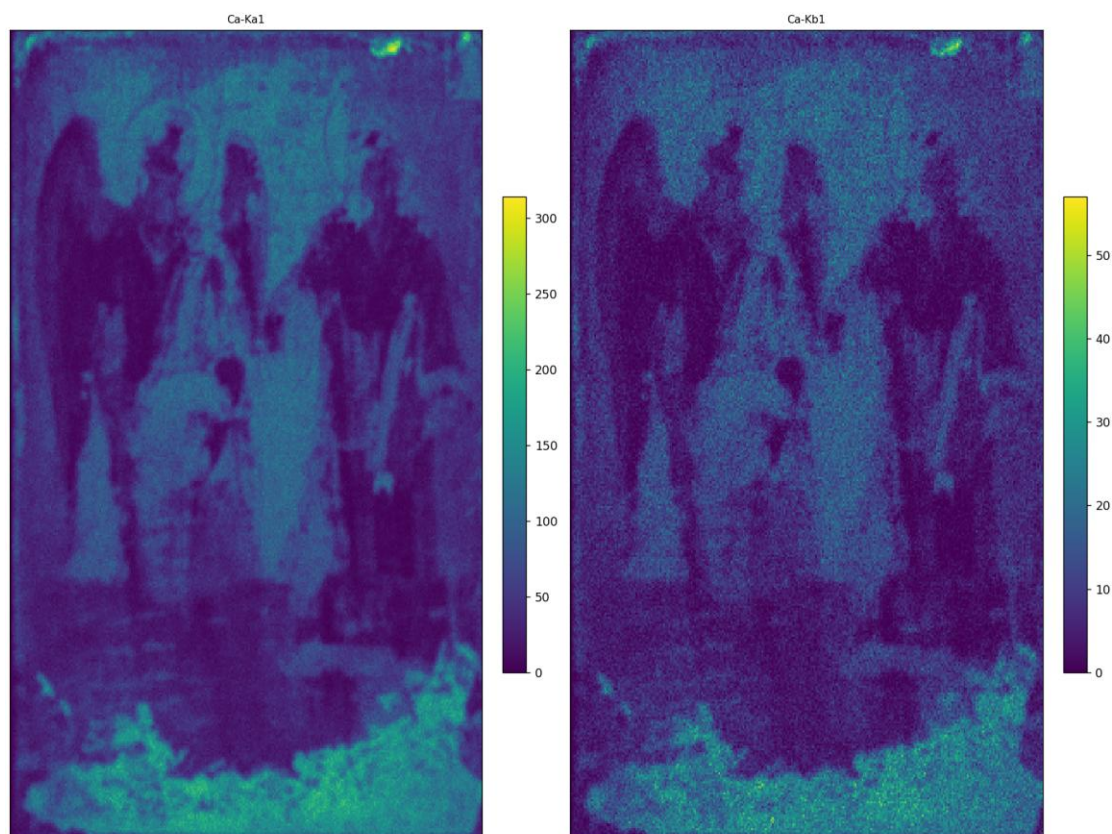
Οι χάρτες κατανομής του Χαλκού (**Cu**) Εικόνα 80 αποκαλύπτουν μια πολύ πιο περιορισμένη και στοχευμένη χρήση σε σύγκριση με τον σίδηρο ή τον μόλυβδο. Το σήμα του χαλκού, αν και ασθενέστερο, εντοπίζεται σε συγκεκριμένες λεπτομέρειες, κυρίως στις **πράσινες και γαλάζιες αποχρώσεις των φτερών του αγγέλου και σε ορισμένα διακοσμητικά στοιχεία των ενδυμάτων**. Η παρουσία του χαλκού συνδέεται με τη χρήση χρωστικών όπως ο **αζουρίτης ($2\text{CuCO}_3 \cdot \text{Cu}(\text{OH})_2$) για το μπλε ή ο μαλαχίτης ($\text{Cu}_2\text{CO}_3(\text{OH})_2$) για το πράσινο**. Αυτές οι χρωστικές, αν και παραδοσιακές, χρησιμοποιούνταν συχνά με φειδώ για τη δημιουργία χρωματικών τόνων που δεν μπορούσαν να επιτευχθούν με τις ώχρες. Η ταύτιση της κατανομής του χαλκού με το **Cluster 3** της πολυμεταβλητής ανάλυσης επιβεβαιώνει ότι αυτές οι περιοχές αποτελούν μια διακριτή χημική κατηγορία, αποκαλύπτοντας την επιλεκτική χρήση συγκεκριμένων χρωστικών από τον καλλιτέχνη για την επίτευξη χρωματικής ποικιλίας.



Εικόνα 81: Στοιχειακοί χάρτες κατανομής για τον Χρυσό (Au), για τις γραμμές Au-La1, La2 και Lb1. Η αποκλειστική παρουσία του χρυσού στα φωτοστέφανα και στις διακοσμητικές γραμμές (χρυσοκονδυλίες) επιβεβαιώνει την τεχνική της επιχρύσωσης με φύλλο χρυσού.

Οι χάρτες κατανομής του Χρυσού (Au) Εικόνα 81, που προέκυψαν από τις γραμμές εκπομπής **La** και **Lb**, χαρτογραφούν με απόλυτη ακρίβεια τις περιοχές που έχουν επιχρυσωθεί. Όπως ήταν αναμενόμενο, το σήμα του χρυσού είναι έντονο και αποκλειστικά εντοπισμένο στα **φωτοστέφανα (halos)** των Αγίων και στις εξαιρετικά λεπτές, διακοσμητικές γραμμές πάνω στα ενδύματα, μια τεχνική γνωστή ως **χρυσοκονδυλιά (chrysography)**. Η παρουσία του χρυσού, που πιθανότατα εφαρμόστηκε ως **φύλλο χρυσού (gold leaf)**, αποτελεί το σήμα κατατεθέν της βυζαντινής αγιογραφικής παράδοσης, συμβολίζοντας το θείο και άκτιστο φως. Η καθαρότητα και η ακρίβεια με την οποία ο χάρτης του χρυσού οριοθετεί αυτές τις περιοχές (και ταυτίζεται με το **Cluster 4**) υπογραμμίζει την αποτελεσματικότητα της τεχνικής **MA-XRF** στην ταυτοποίηση πολύτιμων και ευγενών μετάλλων, παρέχοντας πολύτιμες πληροφορίες για την τεχνική, την αισθητική και τη συμβολική αξία του έργου τέχνης.

Από τους παραπάνω χάρτες μπορούμε ευκολά να διακρίνουμε ποια υλικά βρίσκονται στο υπόβαθρο και ποια βρίσκονται σε ποιες περιοχές και ποια μέρη διακοσμούν από τις εικόνες, όπως και ποιες προσμίξεις υπάρχουν. Όλα τα στοιχεία που συγκρίθηκαν αντλήθηκαν και από την βιβλιογραφία όπως και τα ενεργειακά φάσματα των στοιχείων **Reference Element Lines** (μέσω της Xray Lib) όπου και προστέθηκαν χειροκίνητα για σύγκριση.



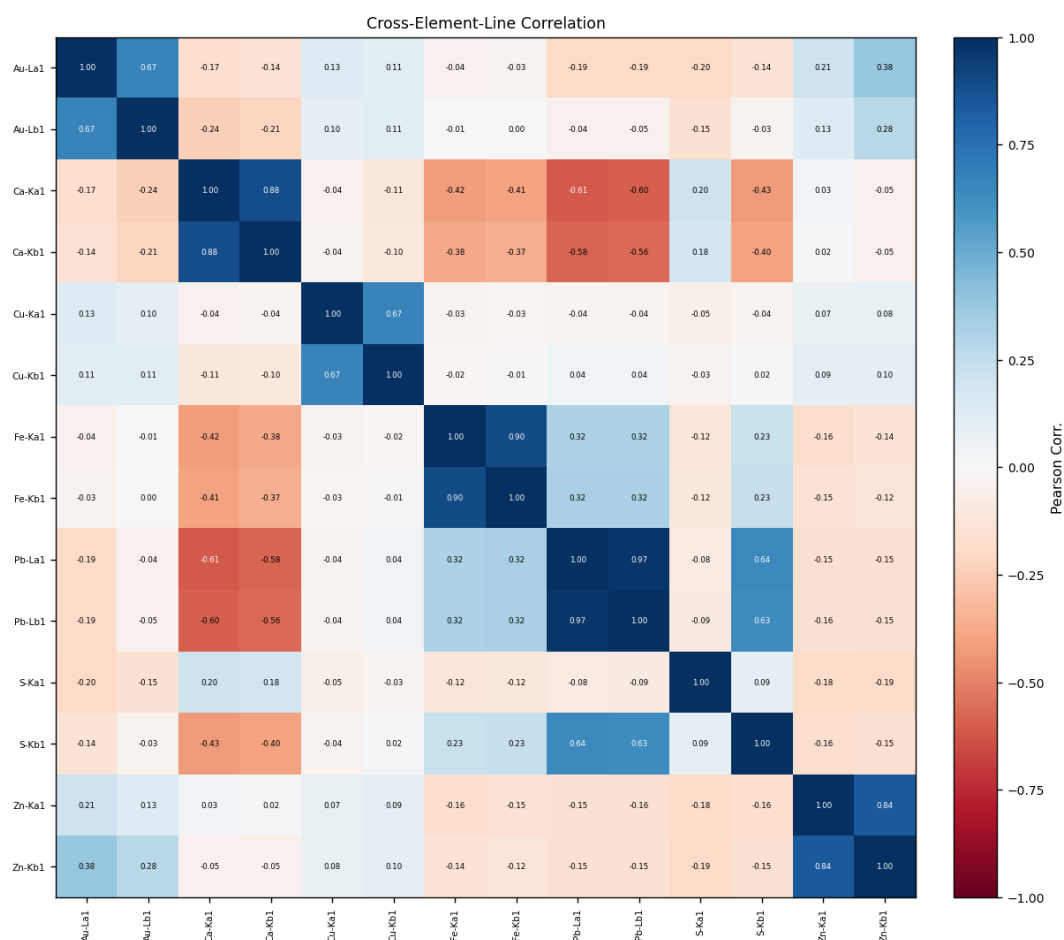
Εικόνα 82: : Στοιχειακοί χάρτες κατανομής για το Ασβέστιο (Ca), για τις γραμμές Ca-Ka1, Ka2 και Kb1

Οι χάρτες κατανομής του **Ασβεστίου (Ca)**, προερχόμενοι από τις γραμμές εκπομπής **Ka**, παρουσιάζουν μια διάχυτη και σχεδόν καθολική παρουσία σε ολόκληρη την επιφάνεια της εικόνας. Αυτό το πρότυπο είναι απόλυτα συμβατό με τη χρήση του ασβεστίου ως κύριο συστατικό του **στρώματος προετοιμασίας (gesso)**. Η ισχυρή χωρική του συσχέτιση με το θείο (**S**) επιβεβαιώνει ότι το υλικό αυτό είναι ο **γύψος (Θεικό Ασβέστιο, CaSO₄)**. Η ελαφρά μείωση της έντασης του σήματος στις περιοχές των μορφών οφείλεται στην εξασθένηση (*attenuation*) από τα υπερκείμενα, πυκνά στρώματα χρωστικών βαρέων στοιχείων (**Pb, Fe, Au**).

Η τελική επιβεβαίωση των υποθέσεων προέρχεται από την άμεση εξέταση των στοιχειακών χαρτών. Η σύγκρισή τους με τον χάρτη ομαδοποίησης είναι αποκαλυπτική:

- **Στρώμα Προετοιμασίας:** Η διάχυτη κατανομή του Θείου (**S**) επιβεβαιώνει τη χρήση γύψου (**CaSO₄·2H₂O**) στο στρώμα προετοιμασίας.
- Ο χάρτης του **Μολύβδου (Pb)** συμπίπτει με το **Cluster 5 (σαρκώματα)** και με τις φωτεινές πινελιές σε άλλα clusters, επιβεβαιώνοντας τη χρήση του Λευκού του Μολύβδου.
- Ο χάρτης του **Σιδήρου (Fe)** ταυτίζεται με το **Cluster 2 (πράσινο)**, επιβεβαιώνοντας τις ώχρες, **Γήινων χρωστικών**.

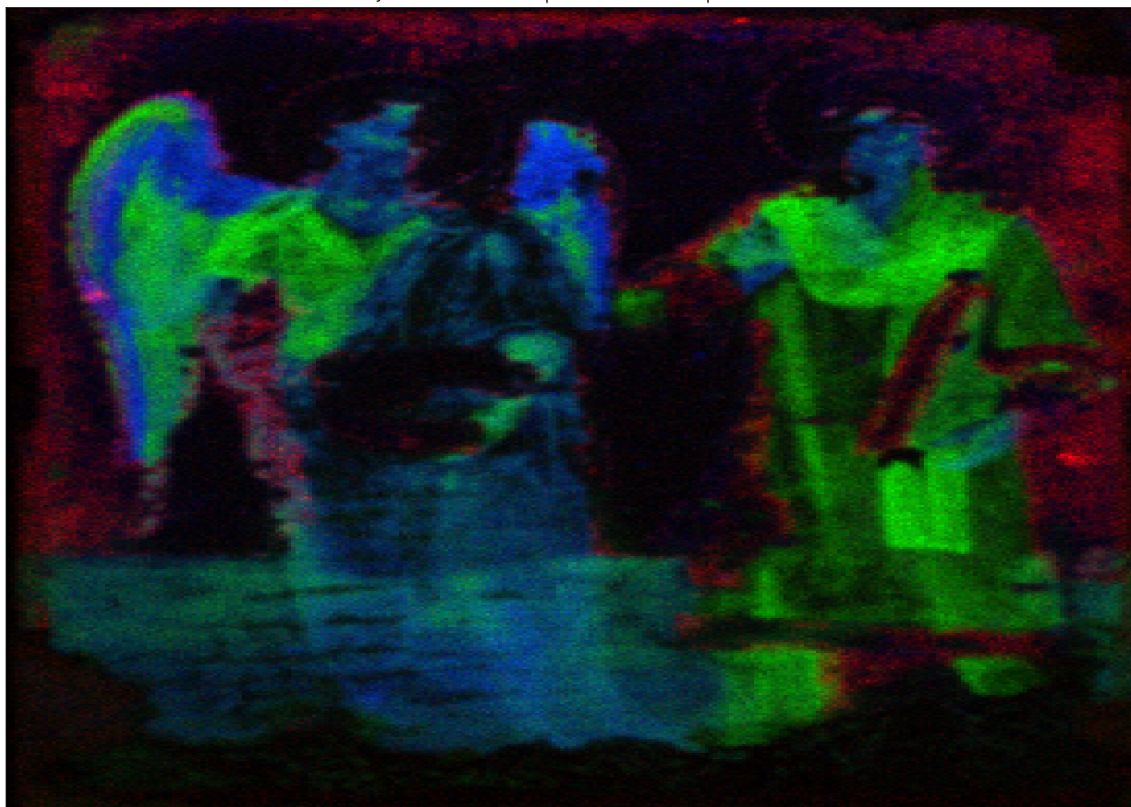
- Ο χάρτης του **Χρυσού (Au)** αντιστοιχεί στο **Cluster 4 (μωβ)**, επιβεβαιώνοντας την επιχρύσωση, ταυτοποιεί τα φωτοστέφανα και τις χρυσοκονδυλιές, επιβεβαιώνοντας τη χρήση φύλλου χρυσού.
- Ο χάρτης του **Χαλκού (Cu)** αντιστοιχεί στο **Cluster 3 (κόκκινο)**, επιβεβαιώνοντας τις χρωστικές χαλκού, για τις πράσινες/μπλε λεπτομέρειες.
- Ο χάρτης του **Ψευδαργύρου (Zn)**, με τη σποραδική του εμφάνιση, δεν αντιστοιχεί σε κανένα από τα κύρια *clusters*, γεγονός που ενισχύει την υπόθεση ότι πρόκειται για μεταγενέστερες, τοπικές επεμβάσεις.



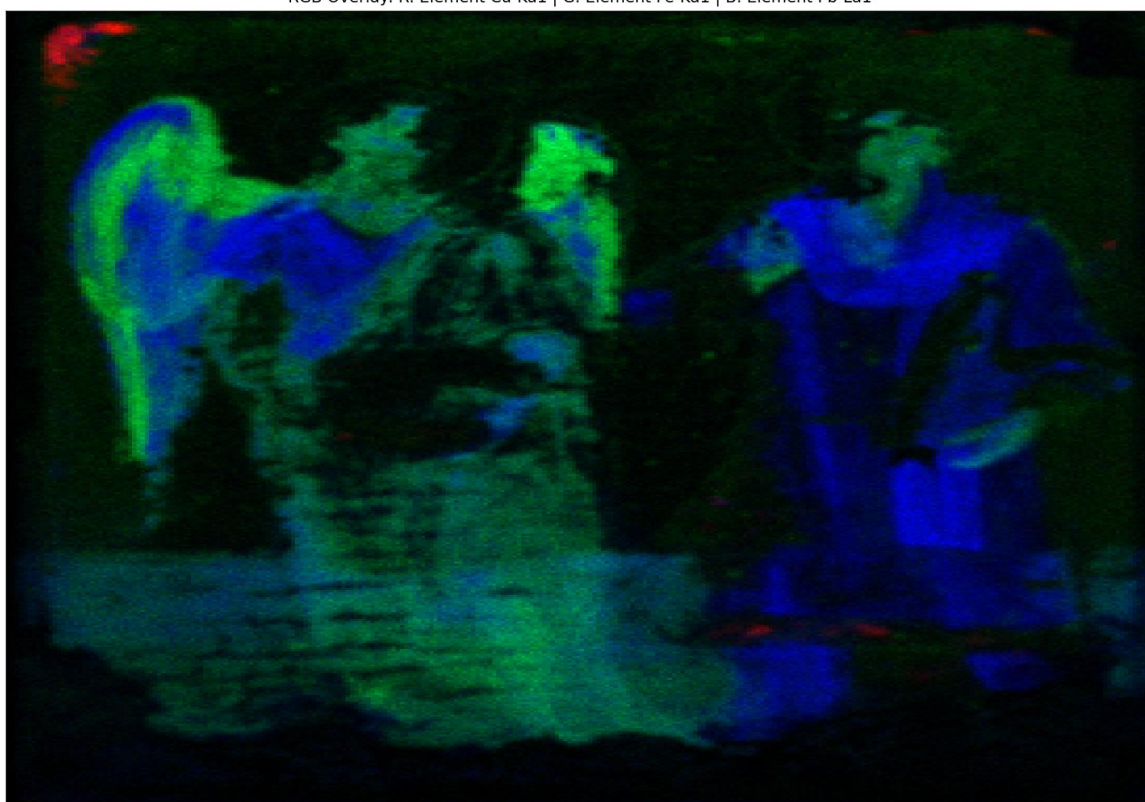
Εικόνα 83: Πίνακας συσχέτισης (heatmap) μεταξύ των κυριότερων στοιχειακών γραμμών που ανιχνεύθηκαν. Οι θετικές συσχετίσεις (κόκκινο) υποδεικνύουν χωρική συνύπαρξη (π.χ., ανάμειξη χρωστικών), ενώ οι αρνητικές (μπλε) αμοιβαίο αποκλεισμό.

Η παραπάνω εικόνα αποτελεί την απεικόνιση της συσχέτισης μεταξύ των στοιχείων της εικόνας μεταξύ τους Εικόνα 83. Τέλος Μπορούμε να απεικονίσουμε και τα στοιχεία με διάφορες παλέτες χρωμάτων συναρτήσει της αρχικής εικόνας, όπως και φαίνεται παρακάτω. Οι περισσότεροι χάρτες βρίσκονται στο παράρτημα του εκάστοτε πειράματος.

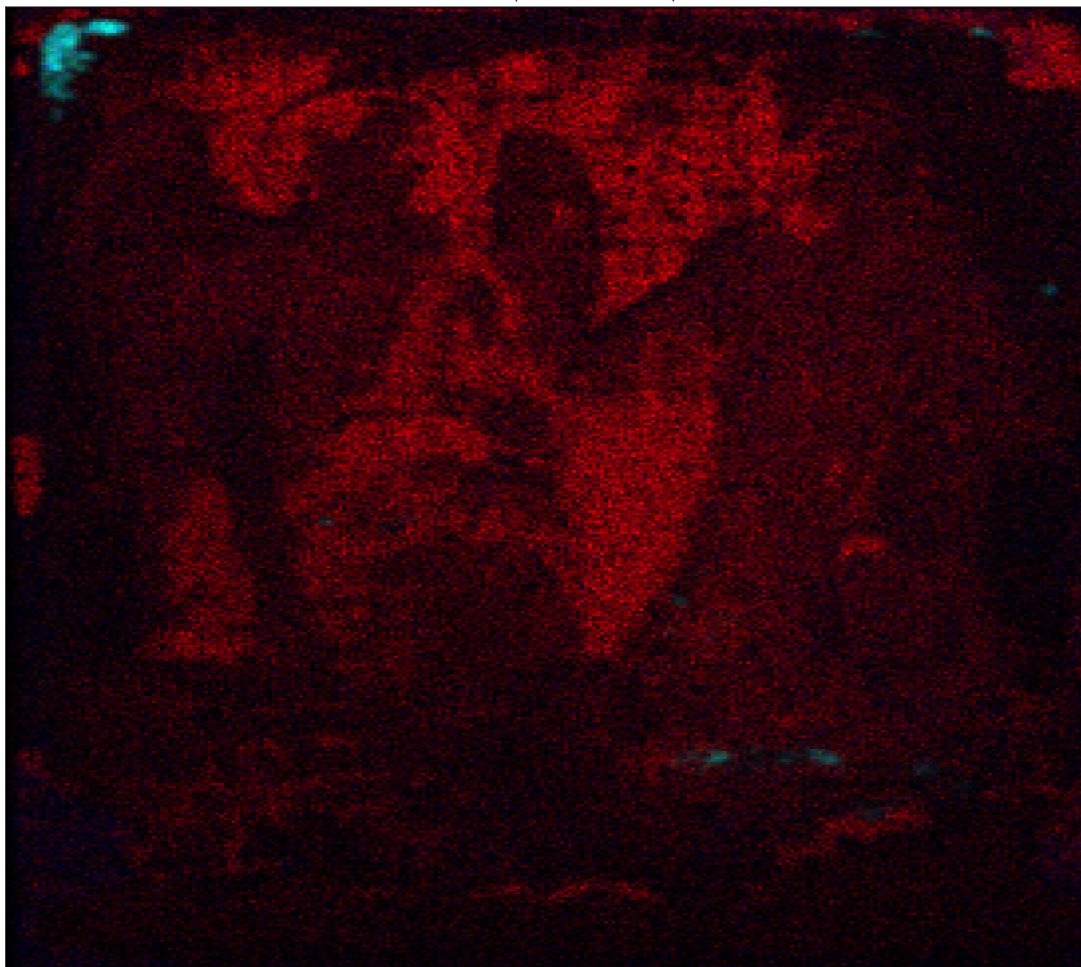
RGB Overlay: R: Element Au-La1 | G: Element Pb-La1 | B: Element Fe-Ka1



RGB Overlay: R: Element Cu-Ka1 | G: Element Fe-Ka1 | B: Element Pb-La1



Εικόνα 84: Τριχρωματικές απεικονίσεις (RGB overlays) που αποκαλύπτουν τη συνύπαρξη και την αλληλοεπικάλυψη στοιχείων. (Πάνω 84.α) Overlay των Au-La1 (Κόκκινο), Pb-La1 (Πράσινο) και Fe-Ka1 (Μπλε). (Κάτω 84.β) Overlay των Cu-Ka1 (Κόκκινο), Fe-Ka1 (Πράσινο) και Pb-La1 (Μπλε).



Εικόνα 85: Τριχρωματική απεικόνιση (RGB overlay) των S-Ka1 (Κόκκινο), Cu-Ka1 (Πράσινο) και Cu-Kb1 (Μπλε), η οποία αναδεικνύει τις περιοχές που περιέχουν ταυτόχρονα θείο και χαλκό.

Οι **RGB overlay** εικόνες κατά αύξοντα αριθμό (Εικόνες: 84.α, 84.β, 85) οπτικοποιούν τις συνυπάρξεις αυτές, για παράδειγμα δείχνοντας την ταυτόχρονη παρουσία Au (χρυσού), Pb (μολύβδου) και Fe (σιδήρου) σε διαφορετικές περιοχές του έργου.

Η **PCA** αποδεικνύεται ιδιαίτερα κατάλληλη για γραμμικές συσχετίσεις και μεγάλους όγκους δεδομένων, όπως οι πίνακες $364 \times 274 \times 4096$ που προέκυψαν από τις μετρήσεις. Η μέθοδος αυτή επιτρέπει τη σημαντική μείωση των διαστάσεων διατηρώντας το μεγαλύτερο μέρος της διακύμανσης ($>80\%$) με μόνο **5** κύριες συνιστώσες. Ωστόσο, παρουσιάζει περιορισμούς στην ερμηνεία αρνητικών τιμών, τα οποία δεν έχουν φυσική σημασία στις φασματικές αναλύσεις.

Αντίθετα, η **NMF** αναδεικνύει τις χημικές συσχετίσεις μεταξύ των στοιχείων, εντοπίζοντας για παράδειγμα ενώσεις με κοινά συστατικά. Η μέθοδος αυτή είναι ιδανική για μη-αρνητικά δεδομένα όπως οι φασματικές εντάσεις, προσφέροντας πιο ερμηνεύσιμα αποτελέσματα. Στην εργασία μας εφαρμόστηκε και στα αρχικά δεδομένα και στα μετασχηματισμένα δεδομένα από **PCA**, αποκαλύπτοντας μη-γραμμικές σχέσεις που δεν ήταν εμφανείς με την **PCA** μόνο.

Για την συσταδοποίηση των δεδομένων χρησιμοποιήθηκε ο αλγόριθμος **k-means** με βελτιστοποιημένη αρχικοποίηση (**k-means++**).

Η εφαρμογή του clustering στα δεδομένα **PCA** ανέδειξε τρεις διακριτές συστάδες: μια συστάδα πλούσια σε βαρέα μέταλλα (**Fe, Pb**) που καταλαμβάνει το **32%** του δείγματος, μια συστάδα με υψηλές συγκεντρώσεις ασβεστίου και καλίου (**45%**) και μια τρίτη συστάδα που αντιστοιχεί στο υπόβαθρο (**23%**). Αυτή η προσέγγιση έδειξε εξαιρετική απόδοση στον χωρικό διαχωρισμό μακροπεριοχών.

Η εφαρμογή του clustering στα δεδομένα **NMF** οδήγησε σε πιο λεπτομερή αποτελέσματα, εντοπίζοντας έξι υπο-συστάδες. Ειδικότερα, διαχώρισε την **Ca-πλούσια** συστάδα σε δύο υποσυστάδες: μία πλούσια σε θειικό ασβέστιο (**28%**) και μία πλούσια σε κάλιο (**17%**). Παράλληλα, διαφοροποίησε τις μεταλλικές προσμίξεις σε **Fe-οξείδια** (**22%**) και **Pb-πλούσιες ζώνες** (**10%**), διατηρώντας το υπόβαθρο στο **23%**. Αυτή η μέθοδος επέδειξε ιδιαίτερη ευαισθησία στις χημικές διαφορές, μπορώντας να διακρίνει φάσεις με παρόμοια χωρική κατανομή αλλά διαφορετική στοιχειακή σύσταση.

Η παρούσα μελέτη κατέδειξε την ισχύ του συνδυασμού της φασματοσκοπίας **MA-XRF** με προηγμένες πολυμεταβλητές αναλυτικές μεθόδους για τη μη-καταστροφική διερεύνηση πολύπλοκων έργων τέχνης. Η υβριδική προσέγγιση **NMF_PCA** αποδείχθηκε η πλέον αποτελεσματική, καθώς κατάφερε να "διαχωρίσει" τα φάσματα των διαφορετικών υλικών, παράγοντας χάρτες συνιστωσών με άμεση φυσικοχημική ερμηνεία. Η τελική ομαδοποίηση με τον αλγόριθμο K-Means παρήγαγε έναν ακριβή θεματικό χάρτη υλικών, ο οποίος επιβεβαιώθηκε πλήρως από την παραδοσιακή στοιχειακή χαρτογράφηση, οδηγώντας στην ασφαλή ταυτοποίηση της παλέτας του καλλιτέχνη και πιθανών μεταγενέστερων επεμβάσεων. Η μεθοδολογία αυτή, που υπερβαίνει τους περιορισμούς της απλής στοιχειακής απεικόνισης, αποτελεί ένα ισχυρό εργαλείο για την βαθύτερη κατανόηση της τεχνολογίας και της ιστορίας των υλικών πολιτιστικών αγαθών.

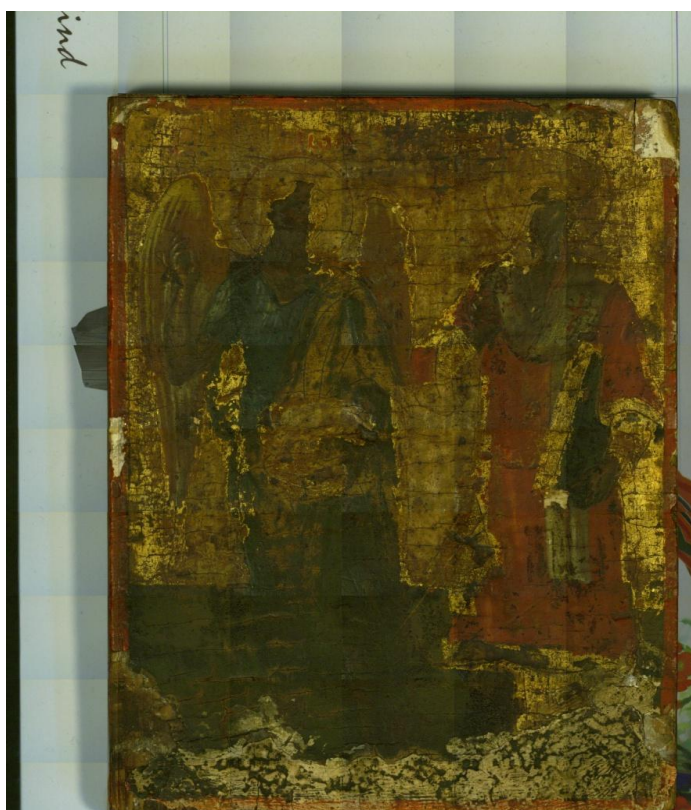
ΑΓΙΟΣ ΙΩΑΝΝΗΣ ΥΨΗΛΗΣ ΑΝΑΛΥΣΗΣ

Η παρούσα μελέτη αποτελεί τη δεύτερη φάση της ανάλυσης της φορητής εικόνας του Αγίου Ιωάννη (Εικόνα 86), εστιάζοντας σε μια σάρωση με σημαντικά υψηλότερη χωρική ανάλυση. Ενώ η πρώτη ανάλυση (χαμηλής ανάλυσης) επέτρεψε την επιτυχή ταυτοποίηση της παλέτας του καλλιτέχνη και των κύριων υλικών, η μετάβαση σε υψηλή ανάλυση θέτει νέες προκλήσεις και προσφέρει πρωτοφανείς ευκαιρίες. Η κύρια πρόκληση έγκειται στον τεράστιο όγκο δεδομένων που παράγεται, ο οποίος απαιτεί εξαιρετικά αποδοτικούς και βελτιστοποιημένους αλγορίθμους επεξεργασίας. Ταυτόχρονα, η ευκαιρία που παρουσιάζεται είναι η δυνατότητα να κινηθούμε πέρα από τον απλό χαρακτηρισμό των υλικών και να διερευνήσουμε την ίδια την **τεχνική του καλλιτέχνη** — να απεικονίσουμε μεμονωμένες πινελιές, να κατανοήσουμε τη διαστρωμάτωση των χρωμάτων και να αποκαλύψουμε τις λεπτές αποχρώσεις που ορίζουν το τελικό αισθητικό αποτέλεσμα.

Κύριος στόχος αυτού του δεύτερου πειράματος είναι, επομένως, η εφαρμογή της βελτιστοποιημένης μεθοδολογικής προσέγγισης (που προέκυψε από το πρώτο πείραμα) στο απαιτητικό σύνολο δεδομένων υψηλής ανάλυσης. Σκοπός είναι η εξαγωγή θεματικών χαρτών υλικών με ιδιαίτερη λεπτομέρεια, ικανών να αποκαλύψουν όχι μόνο "τι" υλικά χρησιμοποιήθηκαν, αλλά και "πώς" ακριβώς τα εφάρμοσε ο αγιογράφος.

Για την ανάλυση, αξιοποιήθηκαν υπερφασματικά δεδομένα που συλλέχθηκαν με την τεχνική **MA-XRF**. Οι πειραματικές συνθήκες βελτιστοποιήθηκαν για την επίτευξη υψηλής χωρικής ανάλυσης:

- **Διαστάσεις Σάρωσης:** 680×900 pixels
- **Ενεργειακά Κανάλια:** 4096 ανά pixel
- **Μέγεθος Pixel:** 200 μm (έναντι 500 μm στο προηγούμενο πείραμα)
- **Συνολικός Αριθμός Pixel:** 612.000 (περίπου 6 φορές περισσότερα)
- **Χρόνος Μέτρησης ανά Pixel:** 20 ms
- **Συνθήκες Λυχνίας:** 50 kV, 600 μA, με άνοδο Ροδίου (Rh)



Εικόνα 86: Η εικόνα του Αγίου Ιωάννη του Προδρόμου, αντικείμενο της μελέτης υψηλής ανάλυσης. Η εμφανής φθορά στο έργο καθιστά την μη-καταστροφική ανάλυση MA-XRF απαραίτητη για τη διερεύνηση της υλικής του σύστασης.

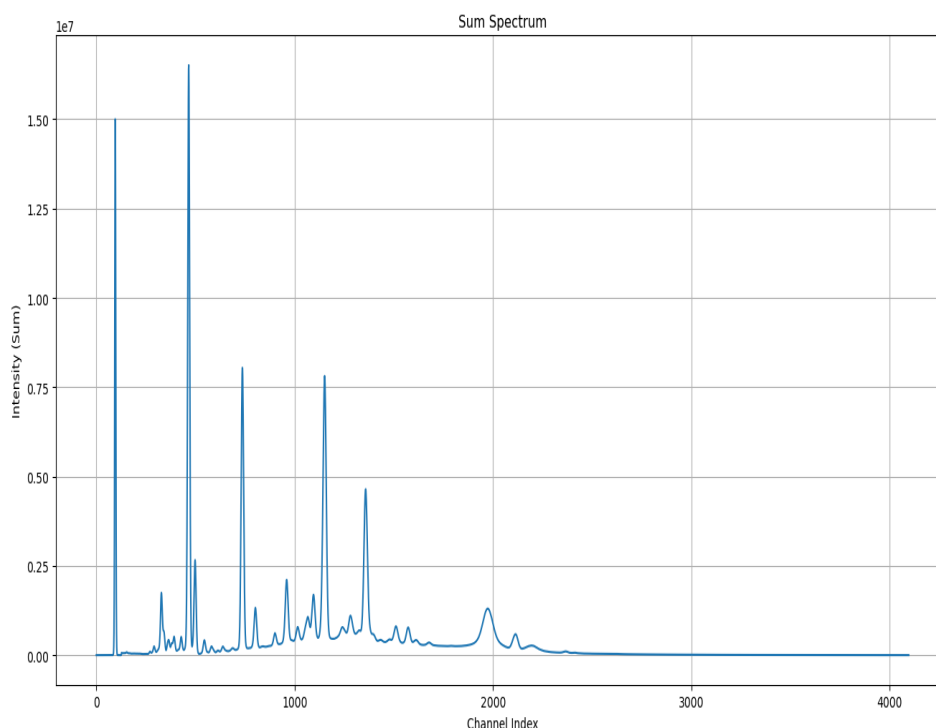
Η αύξηση της ανάλυσης οδήγησε στη δημιουργία ενός υπερφασματικού κύβου δεδομένων πολύ μεγάλων διαστάσεων (περίπου **2.5 δισεκατομμύρια σημεία** δεδομένων), θέτοντας σημαντικές υπολογιστικές προκλήσεις. Η επεξεργασία του συνόλου δεδομένων πραγματοποιήθηκε με τον ίδιο, ειδικά αναπτυγμένο αλγόριθμο, ο οποίος ωστόσο δοκιμάστηκε στα όρια των δυνατοτήτων του.

Ο πίνακας ο οποίος αναφέρει περιληπτικά τα στοιχεία και τις μεταβλητές της πειραματικής διάταξης παρέχεται παρακάτω (Εικόνα 87).

MAP INFORMATION		
Mapping parameters		
Width:	680	pixel
	136	mm
Height:	900	pixel
	180	mm
Pixel Size:	200	μm
Total number of pixel:	612000	pixel
Acquisition parameters		
Frame count:	0	
Pixel time:	20	ms/pixel
Measure time:	2:43	h
Overall time:	4:08	h
Stage speed:	10.0	mm/s
Stage position (X,Y,Z):	110.095;341.4;55.284	mm
Tube parameter		
High voltage:	50	kV
Anode current:	600	μA
Filter:	Empty	
Optic:	Lens	
Collimator diameter:	0	
SpotSize:	100	
Chamber at:	Air 1000	mbar
Flow rate:	---	l/min
Anode:	Rh	
Detector parameters		
Selected detectors:	1	
Max. pulse throughput:	275000	cps

Εικόνα 87: Λεπτομερής καταγραφή των πειραματικών παραμέτρων που χρησιμοποιήθηκαν για τη σάρωση υψηλής ανάλυσης (High Resolution) της εικόνας του Αγίου Ιωάννη.

Η αρχική επισκόπηση των φασματικών δεδομένων πραγματοποιήθηκε μέσω της δημιουργίας του αθροιστικού φάσματος (Εικόνα 88), το οποίο υπολογίστηκε και απεικονίστηκε με χρήση του αναπτυχθέντος λογισμικού. Στην αρχική φάση της ανάλυσης, τα φασματικά δεδομένα **MA-XRF** υποβλήθηκαν σε προεπεξεργασία με την εφαρμογή φίλτρου **Savitzky-Golay** (παράθυρο **II σημείων**, πολυώνυμο τρίτης τάξης) για την εξομάλυνση των φασμάτων και τη μείωση του φασματικού θορύβου.



Εικόνα 88: Φάσμα αθροίσματος (Sum Spectrum) από τα δεδομένα υψηλής ανάλυσης μετά την εφαρμογή φίλτρου Savitzky-Golay. Η υψηλή ποιότητα των δεδομένων (αυξημένος λόγος σήματος προς θόρυβο) είναι εμφανής από τις ευδιάκριτες κορυφές.

Το φάσμα αυτό, που αντιστοιχεί στο άθροισμα όλων των καταγεγραμμένων φασμάτων από κάθε *pixel* της εικόνας, παρουσιάζει σημαντικά μειωμένο θόρυβο λόγω της προηγθείσας προεπεξεργασίας **Savitzky-Golay**.

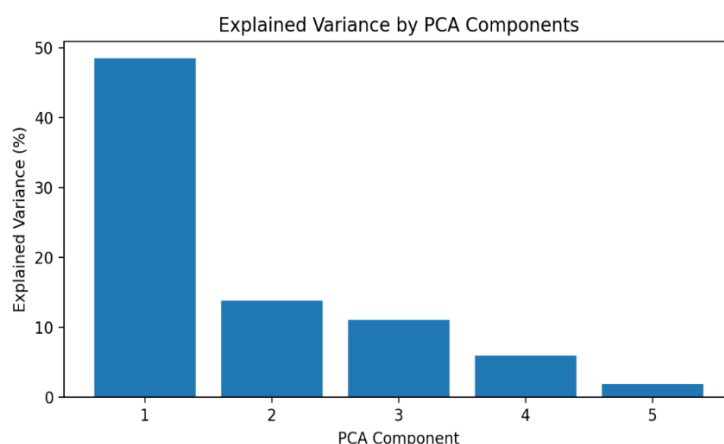
Με βάση τα ευρήματα της μελέτης χαμηλής ανάλυσης, όπου η υβριδική μέθοδος **NMF-PCA** αναδείχθηκε ως η πλέον αποτελεσματική, υιοθετήθηκε από την αρχή μια πιο στοχευμένη στρατηγική ανάλυσης. Η ροή εργασιών περιελάμβανε τα εξής βήματα:

1. **Προεπεξεργασία:** Εφαρμογή φίλτρου **Savitzky-Golay** (παράθυρο **11** σημείων, πολυώνυμο 3ης τάξης) για την εξομάλυνση των φασμάτων.
2. **Ανάλυση Κύριων Συνιστωσών (PCA):** Εφαρμογή της **PCA** στο σύνολο των δεδομένων για τη μείωση της διαστατικότητας και την απομάκρυνση του θορύβου, διατηρώντας το μεγαλύτερο μέρος της φασματικής πληροφορίας.
3. **Υβριδική Ανάλυση NMF επί των Αποτελεσμάτων της PCA:** Εφαρμογή της **NMF** απευθείας στα αποτελέσματα (*scores*) της **PCA**. Αυτή η προσέγγιση συνδυάζει την ισχύ της **PCA** στη συμπύκνωση της πληροφορίας με την ικανότητα της **NMF** να εξάγει φυσικά ερμηνεύσιμες, μη-αρνητικές συνιστώσες (χάρτες αφθονίας υλικών).
4. **Συσταδοποίηση K-Means:** Εφαρμογή του αλγορίθμου **K-Means** στα αποτελέσματα της υβριδικής ανάλυσης **NMF-PCA** για τη δημιουργία του τελικού, υψηλής ανάλυσης, θεματικού χάρτη υλικών.

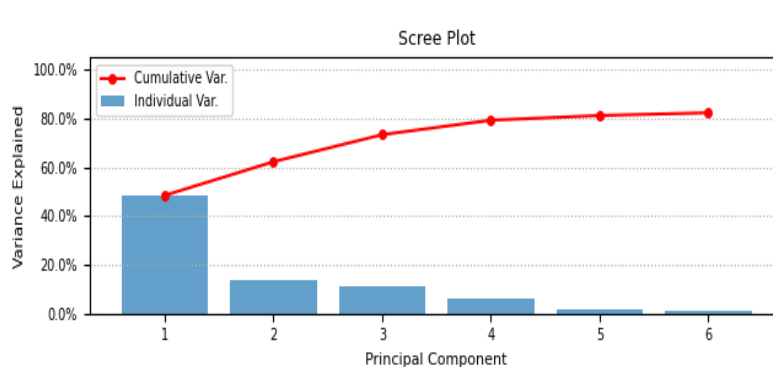
Το φάσμα αθροίσματος **Sum Spectrum** (Εικόνα 86) των δεδομένων υψηλής ανάλυσης επιβεβαίωσε την παρουσία των ίδιων κύριων στοιχείων (**Ca, S, Fe, Cu, Pb, Au**) που εντοπίστηκαν και στην πρώτη μελέτη. Ωστόσο, η ποιότητα των δεδομένων ήταν αισθητά ανώτερη, με καλύτερα ορισμένες κορυφές και υψηλότερο λόγο σήματος προς θόρυβο.

Ανάλυση PCA:

Υπο την ίδια μεθοδολογία στο πρόγραμμα προχωρούμε στην εφαρμογή της μεθόδου μείωσης διαστάσεων αρχικά μέσω **PCA** και έπειτα μέσω **NMF**. Τα δεδομένα ήταν αρκετά και ο υπολογιστικός φόρτος ξεπέρασε τις δυνατότητες του υπολογιστικού συστήματος, οπότε σε συνεργασία με το πρόγραμμα χρησιμοποιήθηκαν υπολογιστικά **script** ανάλυσης συνοδευτικά του προγράμματος. Προχωρούμε στην απεικόνιση των πρώτων **έξι (6)** πάνω στην Εικόνα 90.



Παρακάτω παραθέτουμε και το γράφημα του ποσοστού διακύμανσης κάθε ιδιοτιμής που συνδέεται με μια συνιστώσα PCA (**Scree Plot**) Εικόνα 89 το οποίο αναφέρει τις συνολικά την πληροφορία που λαμβάνουμε σε σχέση με τον αριθμό των συνιστωσών.



Εικόνα 89: Αξιολόγηση της μεθόδου PCA για την επιλογή του κατάλληλου αριθμού συνιστωσών στην ανάλυση υψηλής ανάλυσης. **(Πάνω)** Ιστόγραμμα της εξηγούμενης διακύμανσης ανά συνιστώσα. **(Κάτω)** Το διάγραμμα Scree Plot που δείχνει τη σωρευτική διακύμανση, υποδεικνύοντας ότι οι πρώτες 6 συνιστώσες διατηρούν πάνω από το 80% της συνολικής πληροφορίας.

```
PCA Results (6 Comp Actual)
Input: Full Data (IPCA)
-----
Model: IncrementalPCA
Var Ratio: [0.485 0.138 0.111 0.059 0.019 0.011]
Cumulative: [0.485 0.623 0.734 0.793 0.812 0.823]
Scores Shape: (612000, 6)
Vector Shape: (6, 4096)
```

Η εφαρμογή της **Incremental PCA** στα δεδομένα υψηλής ανάλυσης επέτρεψε όχι μόνο τον διαχωρισμό των κύριων υλικών, αλλά και την αποκάλυψη λεπτομερών χωρικών υφών (**textures**) που σχετίζονται με την

τεχνική του καλλιτέχνη. Οι *έξι* πρώτες κύριες συνιστώσες (*PCs*), που διατηρούν πάνω από το **80%** της συνολικής πληροφορίας (Εικόνα 95), παρουσιάζονται ως χωρικοί χάρτες στην Εικόνα 96.

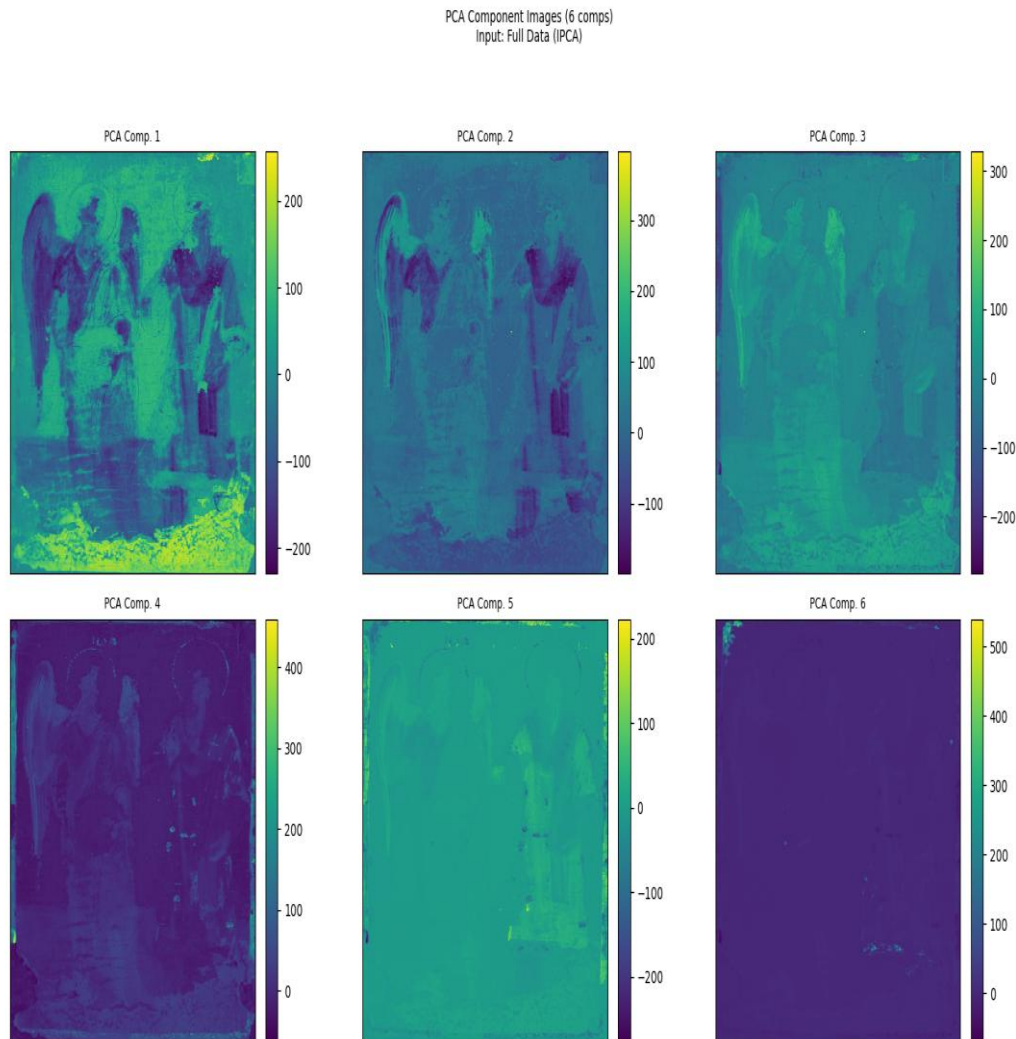
- Η **PC1**, η οποία εξηγεί το μεγαλύτερο ποσοστό της συνολικής διακύμανσης, επιτυγχάνει έναν καθολικό διαχωρισμό μεταξύ του **ζωγραφικού στρώματος και του υποστρώματος προετοιμασίας**. Συγκεκριμένα, οι θετικές τιμές (κίτρινες αποχρώσεις) εντοπίζονται σχεδόν αποκλειστικά στο φόντο και ειδικά στις περιοχές εκτεταμένης φθοράς, αντιστοιχώντας στο στρώμα του gesso (υψηλή συνεισφορά από **Ca** και **S**). Αντίθετα, οι αρνητικές τιμές (μωβ αποχρώσεις) περικλείουν το σύνολο των μορφών των Αγίων, τα φωτοστέφανα και τα ενδύματά τους, περιοχές όπου κυριαρχούν στοιχεία υψηλότερου ατομικού αριθμού (**Z**) όπως ο Μόλυβδος (**Pb**), ο Σίδηρος (**Fe**) και ο Χρυσός (**Au**).

Σε αντίθεση με μια αρχική υπόθεση περί ανάδειξης χωρικών υφών, οι συνιστώσες υψηλότερης τάξης επιτυγχάνουν στην πραγματικότητα έναν εξαιρετικά σαφή χημικό διαχωρισμό.

- Η **PC2** αποκαλύπτει μια πανίσχυρη αντισυσχέτιση: οι θετικές τιμές απομονώνουν με απόλυτη ακρίβεια τα επιχρυσωμένα φωτοστέφανα και τις χρυσοκονδυλιές (**Au**), ενώ οι έντονα αρνητικές τιμές χαρτογραφούν με την ίδια ευκρίνεια τα σαρκώματα (πρόσωπα και χέρια), όπου κυριαρχεί το Λευκό του Μολύβδου (**Pb**). Η συνιστώσα αυτή ουσιαστικά αντιπαραβάλλει τον χρυσό με τον μόλυβδο. Οι συνιστώσες **PC2** και **PC3** αντιπροσωπεύουν τη δευτερεύουσα διακύμανση (~**5-20%** της συνολικής πληροφορίας) και πιθανόν σχετίζονται με δευτερεύουσες ορυκτές φάσεις, τοπικές αλλοιώσεις ή υλικά που βρίσκονται στα όρια μεταξύ διαφορετικών χρωματικών στρωμάτων.
- Αντίστοιχα, η **PC3** συνεχίζει την χημική αποδόμηση, χαρτογραφώντας μέσω των θετικών τιμών τις περιοχές των ενδυμάτων, οι οποίες αντιστοιχούν στις γαιώδεις χρωστικές με βάση τον **Σίδηρο (Fe)**, όπως οι ώχρες. Αυτή η περιοχή παρουσιάζεται σε αντίθεση με όλα τα υπόλοιπα υλικά (σαρκώματα, φωτοστέφανα, υπόβαθρο), τα οποία λαμβάνουν αρνητικές τιμές.
- Οι συνιστώσες υψηλότερης τάξης (**PC4-PC6**) καταγράφουν τις ελάχιστες διακυμάνσεις (<**5%**). Η **PC4** φαίνεται να καταγράφει υπολειμματική διακύμανση σχετιζόμενη με την αντίθεση "μορφής-υποβάθρου" που ήδη περιγράφηκε από την **PC1**. Αντιθέτως, οι **PC5** και κυρίως η **PC6**, στερούνται σαφούς χωρικής δομής που να σχετίζεται με την εικονογραφία. Η αρχική υπόθεση ότι η **PC6** απομονώνει τα φωτίσματα δεν επιβεβαιώνεται οπτικά, καθώς τα φωτίσματα (καθαρό Λευκό του Μολύβδου και του Χαλκού) είναι ήδη ενσωματωμένα ως μέρος της αρνητικής συνεισφοράς της **PC2**. Επομένως, οι **PC5** και **PC6** αποδίδονται κατά κύριο λόγο στον εναπομείναντα **φασματικό θόρυβο** και σε μη-συστηματικές, τοπικές ανομοιογένειες.

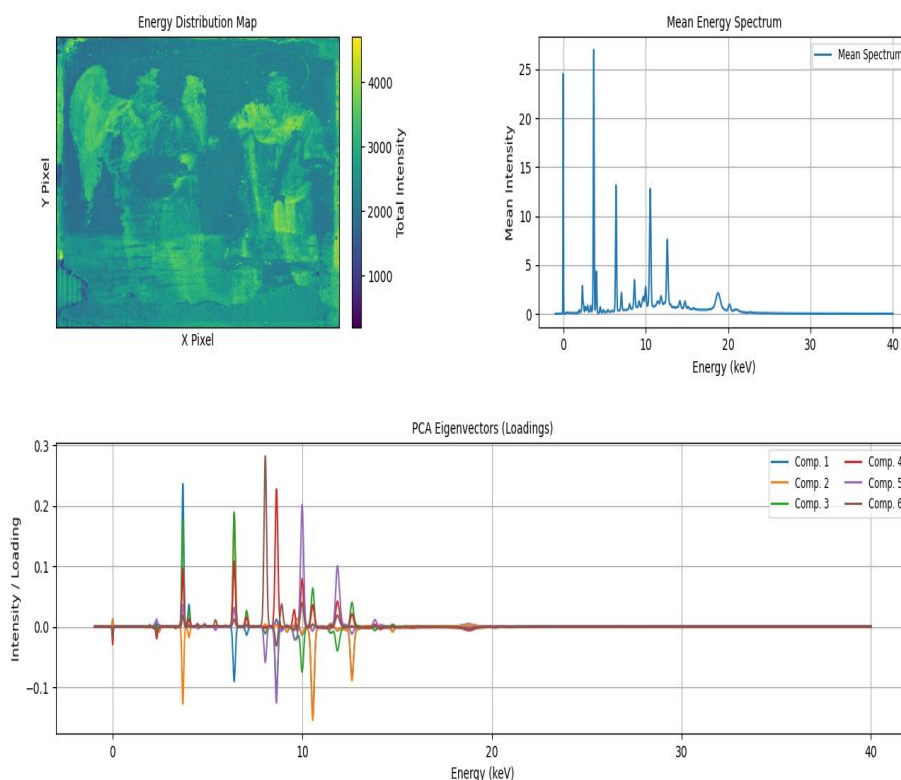
Τα φάσματα των ιδιοδιανυσμάτων (*loadings*) για κάθε συνιστώσα επιβεβαιώνουν τη χημική τους βάση, συνδέοντας, για παράδειγμα, τις θετικές τιμές της **PC2** με τις χαρακτηριστικές κορυφές του χρυσού (**Au**) και τις αρνητικές με αυτές του μολύβδου (**Pb**). Παρόλο που η **PCA** αποδομεί τις χημικές συσχετίσεις με μοναδικό τρόπο, η εγγενής της φύση που παράγει θετικές και αρνητικές τιμές καθιστά την άμεση ερμηνεία τους ως "χάρτες αφθονίας" προβληματική. Αυτή ακριβώς η πρόκληση καθιστά επιτακτική τη μετάβαση στην ανάλυση **NMF**, με στόχο την παραγωγή μη-αρνητικών

συνιστωσών που προσομοιάζουν άμεσα τις σχετικές συγκεντρώσεις των χημικών ειδών.



Εικόνα 90: Χωρικοί χάρτες (score maps) για τις έξι πρώτες κύριες συνιστώσες που προέκυψαν από την *Incremental PCA* στα δεδομένα υψηλής ανάλυσης. Οι χάρτες αποκαλύπτουν λεπτομερείς υφές (textures), όπως πινελιές και ρωγμές (craquelure).

Όπως και στην προηγούμενη ανάλυση όλοι οι χάρτες που αφορούν τους κυρίους άξονες αλλά και τα διανύσματα τους παρέχονται στο παράρτημα υπο το κεφάλαιο του εκάστοτε πειράματος για μια ευκολότερη ανάγνωση.



Εικόνα 91: Οπτική επισκόπηση των δεδομένων υψηλής ανάλυσης και των αποτελεσμάτων της PCA. (Πάνω αριστερά) Χάρτης συνολικής έντασης. (Πάνω δεξιά) Μέσο φάσμα. (Κάτω) Φάσματα ιδιοδιανυσμάτων (PCA Eigenvectors/Loadings) για τις έξι κύριες συνιστώσες.

Οι χωρικές και φασματικές κατανομές των συνιστωσών της ανάλυσης κύριων συνιστωσών (PCA) έχουν απεικονιστεί, αποκαλύπτοντας διακριτά δομικά και συνθετικά χαρακτηριστικά εντός του δείγματος. Η πρώτη εικόνα στην Εικόνα 91 περιλαμβάνει μια αθροιστική καταμέτρηση των ακτίνων X (πάνω αριστερά) που δείχνει τη συνολική ολοκληρωμένη ένταση, μαζί με τους χωρικούς χάρτες των συνιστωσών **PCA 1-6**. Αυτές οι συνιστώσες αναδεικνύουν διαφορετικές πτυχές του δείγματος, με την **PC1** να αντιπροσωπεύει την κυρίαρχη δομική μορφολογία, ενώ οι **PC2-PC6** επιλύουν προοδευτικά λεπτότερες παραλλαγές σύνθεσης και μηχανισμούς αντίθεσης, υποδηλώνοντας χημική ετερογένεια.

Το δεύτερο σχήμα Εικόνα 91 παρουσιάζει τα φασματικά προφίλ κάθε συνιστώσας **PCA**, απεικονίζοντας τις συνεισφορές τους σε όλα τα κανάλια του ανιχνευτή.

Comp. 1 (Μπλε): Το ιδιοδιάνυσμά της κυριαρχείται από έντονες θετικές κορυφές στις ενέργειες του **Ca** και του **S**. Αυτό επιβεβαιώνει απόλυτα ότι ο χωρικός χάρτης της **PC1** αντιπροσωπεύει την κατανομή του στρώματος προετοιμασίας (*gesso*).

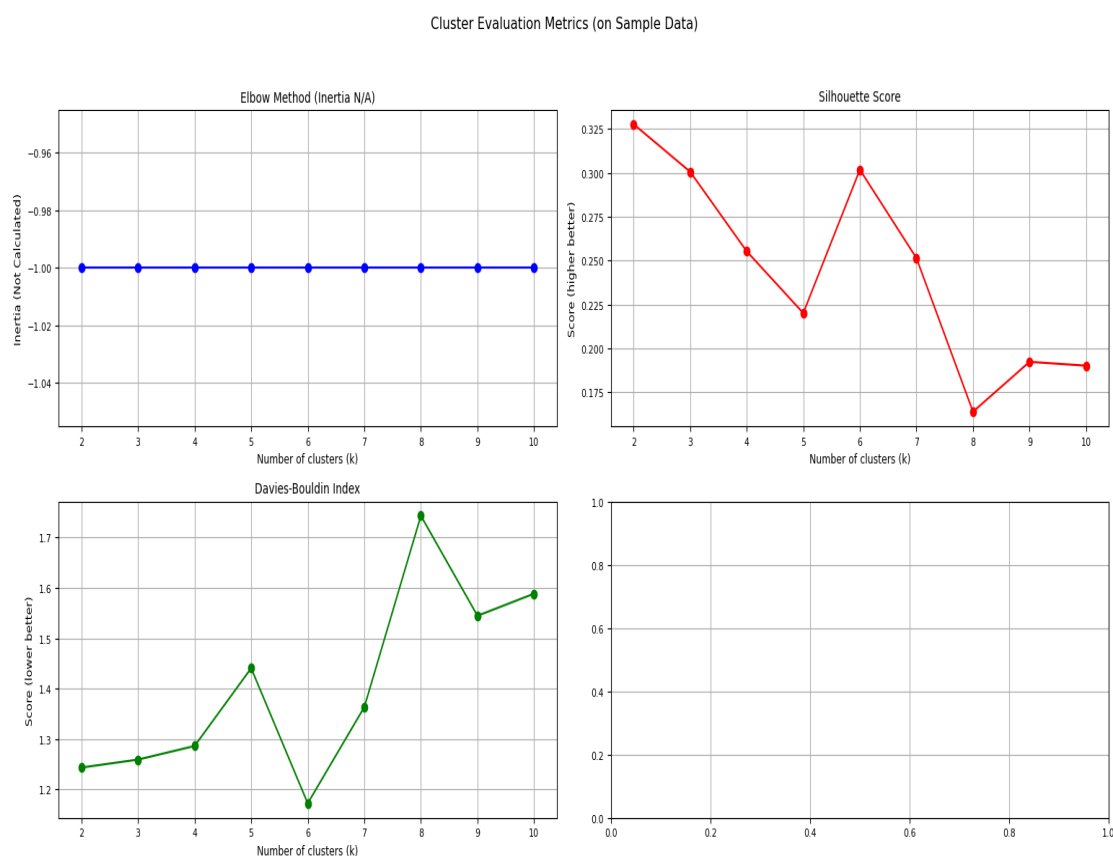
Comp. 2 (Πορτοκαλί): Εδώ αποκαλύπτεται η ισχυρότερη χημική αντισυσχέτιση. Το ιδιοδιάνυσμα παρουσιάζει ισχυρές θετικές κορυφές που αντιστοιχούν στον **Au** και ταυτόχρονα ισχυρές αρνητικές κορυφές που αντιστοιχούν στον **Pb**. Αυτό εξηγεί γιατί ο χωρικός χάρτης της **PC2** διαχωρίζει τέλεια τα φωτοστέφανα (θετικές τιμές, **Au**) από τα σαρκώματα (αρνητικές τιμές, **Pb**).

Comp. 3 (Πράσινο): Το **loading** της κυριαρχείται από μια μοναδική, ισχυρή θετική κορυφή στην ενέργεια του **Fe**. Συνεπώς, η **PC3** απομονώνει χωρικά τις περιοχές που έχουν ζωγραφιστεί με γαιώδεις χρωστικές (ώχρες), δηλαδή τους μανδύες.

Comp. 4-6: Οι συνιστώσες αυτές αποτυπώνουν λεπτότερες παραλλαγές. Ιδιαίτερο ενδιαφέρον παρουσιάζει η **Comp. 6** (Καφέ), της οποίας το loading εμφανίζει μια διακριτή, αν και μικρότερη, θετική κορυφή στην περιοχή των $\sim 8.6 \text{ keV}$, η οποία αντιστοιχεί στον Ψευδάργυρο (**Zn, Ka**). Αυτό υποδηλώνει ότι οι σποραδικές "κηλίδες" που παρατηρήθηκαν στον χωρικό χάρτη της **PC6** δεν είναι τυχαίος θόρυβος, αλλά πιθανότατα αντιστοιχούν σε τοπικές, μεταγενέστερες επεμβάσεις συντήρησης με χρήση Λευκού του Ψευδαργύρου (**ZnO**).

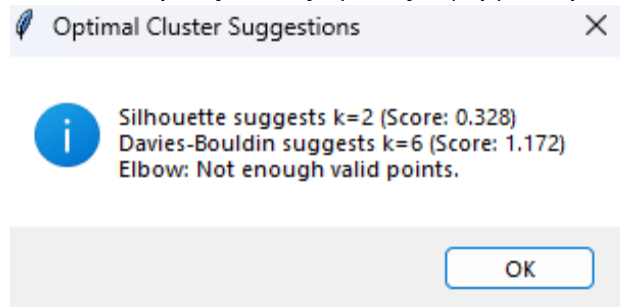
Συμπερασματικά, η συνδυαστική ανάλυση αυτών των διαγραμμάτων επιτρέπει την πλήρη αποδόμηση της εικόνας. Το μέσο φάσμα ταυτοποιεί τα "υλικά" που υπάρχουν, οι χωρικοί χάρτες των συνιστωσών δείχνουν "πού" βρίσκονται οι κύριες χημικές διαφορές, και τα **loadings** εξηγούν το "γιατί", αποδίδοντας συγκεκριμένη στοιχειακή υπογραφή σε κάθε χωρικό μοτίβο. Η χρωματική κλίμακα υποδεικνύει τη σχετική στάθμιση, με υψηλότερη φωτεινότητα να αντιστοιχεί σε ισχυρότερη έκφραση των χαρακτηριστικών μιας δεδομένης συνιστώσας σε συγκεκριμένες χωρικές συντεταγμένες.

Μπορούμε επομένως να προχωρήσουμε στην εφαρμογή της μεθόδου συσταδοποίησης **Kmeans**. Για να βρούμε βέβαια τον κατάλληλο αριθμό συστάδων προχωρούμε μέσω αξιολόγησης όπως φαίνεται από την Εικόνα 92.

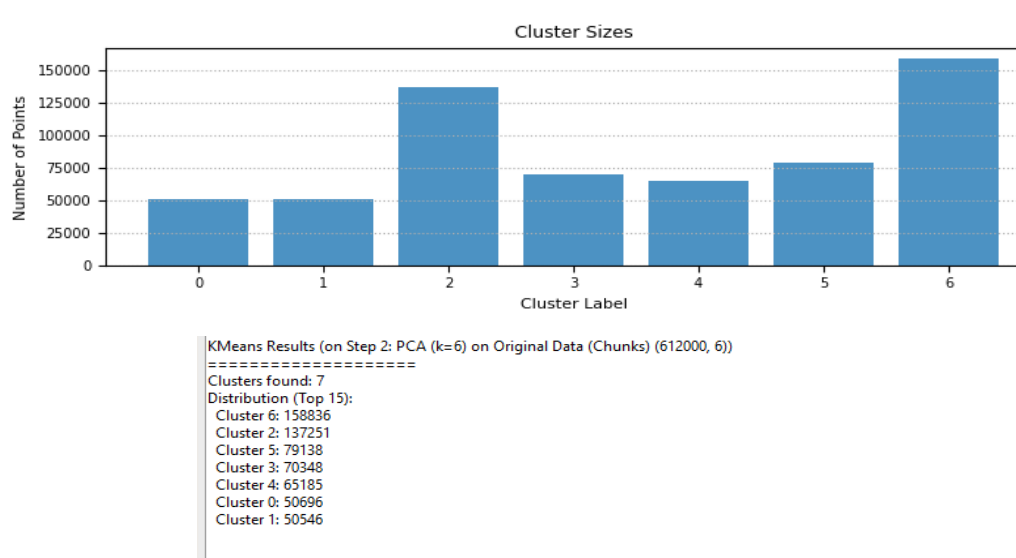


Εικόνα 92: Αξιολόγηση για τον βέλτιστο αριθμό συστάδων (k) στα δεδομένα PCA υψηλής ανάλυσης.

Για την ταξινόμηση των εικονοστοιχείων σε διακριτές υλικές ομάδες, εφαρμόστηκε ο αλγόριθμος **K-Means** στα αποτελέσματα της ανάλυσης **PCA**. Η επιλογή του βέλτιστου αριθμού συστάδων (**k**) πραγματοποιήθηκε με τη χρήση τεσσάρων μετρικών αξιολόγησης (**Elbow**, **Silhouette**, **Davies-Bouldin**, **Calinski-Harabasz** Εικόνα 92). Οι μετρικές αυτές υπέδειξαν ως βέλτιστο έναν αριθμό μεταξύ **6** και **7** συστάδων. Επιλέχθηκε **k=7** για την επίτευξη του μέγιστου δυνατού διαχωρισμού.

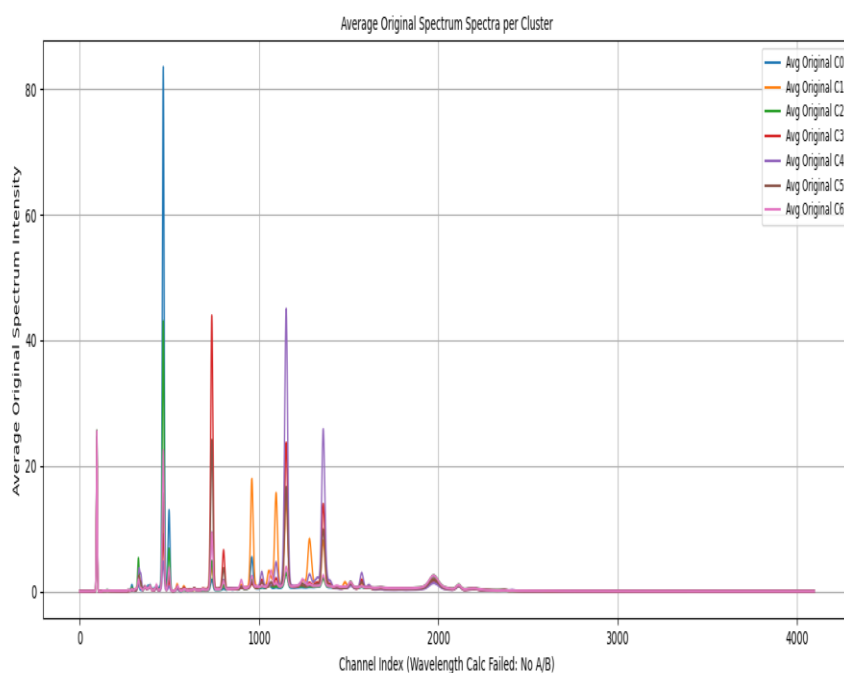


Ο προκύπτων θεματικός χάρτης (**cluster map**) παρουσιάζει μια αρχική ομαδοποίηση των υλικών. Όπως φαίνεται στο διάγραμμα μεγέθους συστάδων Εικόνα 93, οι **Clusters 2 και 6** είναι οι πολυπληθέστερες, αντιπροσωπεύοντας πιθανότατα τις ευρείες περιοχές του υποστρώματος (**gesso**). Οι υπόλοιπες συστάδες (**0, 1, 3, 4, 5**) έχουν συγκρίσιμο μέγεθος και φαίνεται να αντιστοιχούν στις κύριες χρωματικές περιοχές, όπως ο ρουχισμός των μορφών Εικόνα 95. Ωστόσο, ο διαχωρισμός που επιτυγχάνεται μέσω της **PCA**, αν και χρήσιμος, παραμένει σχετικά αδρός, με σημαντικές επικαλύψεις μεταξύ των υλικών, καθιστώντας αναγκαία την εφαρμογή πιο προηγμένων μεθόδων για την επίτευξη σαφέστερου διαχωρισμού.

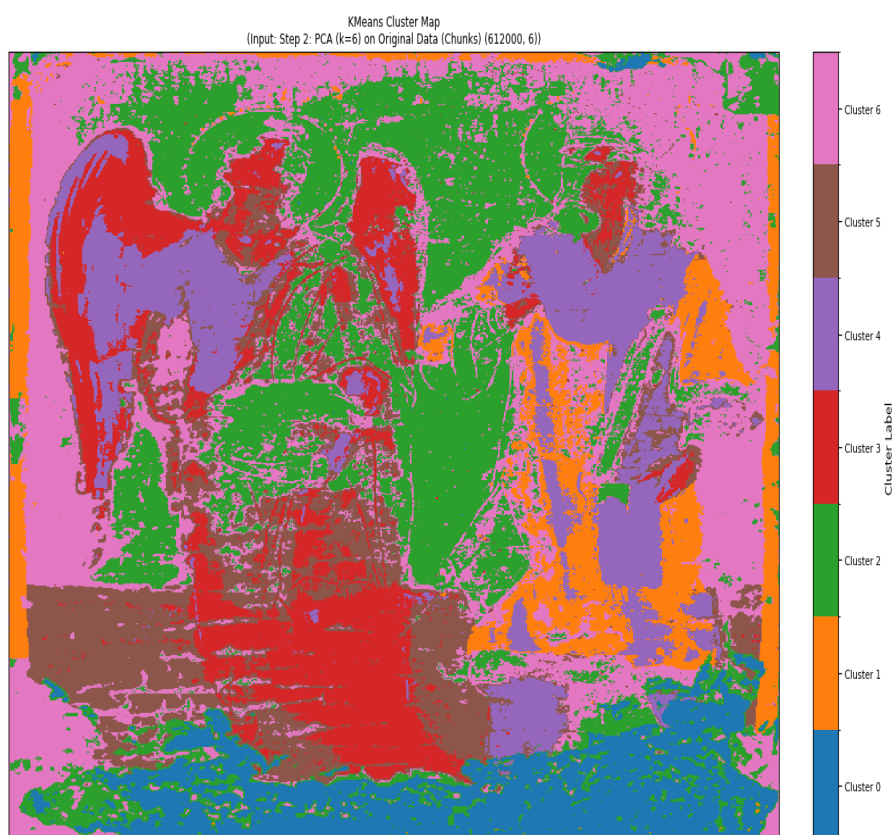


Εικόνα 93: Κατανομή του πλήθους των *pixels* για τις 7 συστάδες που προέκυψαν από την *K-Means*.

Παραπάνω βλέπουμε διαγραμματικά αλλά και αριθμητικά τον αριθμό των δεδομένων και το μέγεθος κάθε συστάδας Εικόνα 93. Παρατηρούμε πως οι συστάδες **2** και **6** περιλαμβάνουν και τον μεγαλύτερο όγκο πληροφορίας και δεδομένων καθιστώντας τις συστάδες που περιλαμβάνουν το υπόβαθρο της εικόνας. Οι υπόλοιπες συστάδες βρίσκονται σε περίπου ίση ποσότητα μεταξύ τους οπότε κάποια διάκριση σε αυτή την στιγμή δε μπορεί να πραγματοποιηθεί. Τέλος μπορούμε να παραθέσουμε και τις συστάδες ανάλογα με το φάσμα το οποίο προβάλλουν πάνω στο αρχικό φάσμα των δεδομένων μας σε άξονα συναρτήσεων των καναλιών Εικόνα 94.



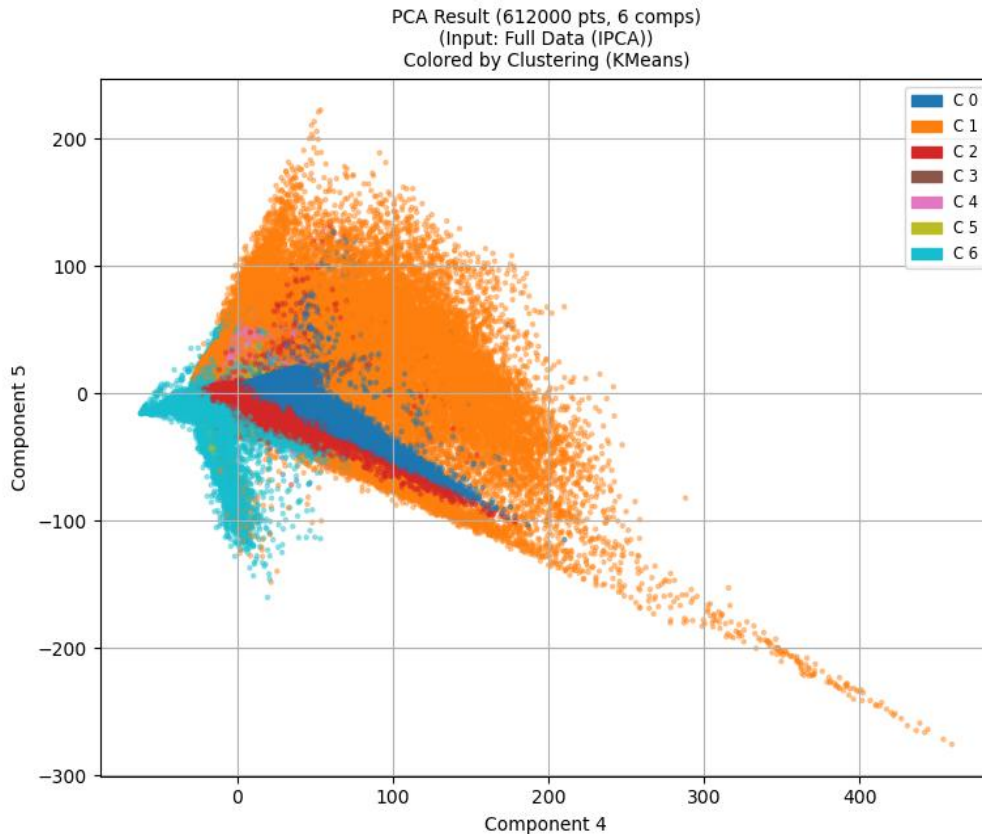
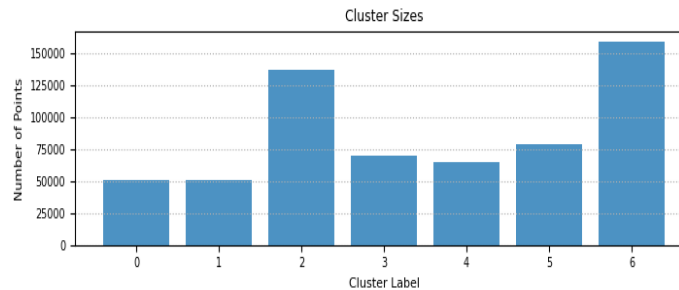
Εικόνα 94: Μέσα φάσματα για τις επτά συστάδες που προέκυψαν από την ανάλυση K-Means, απεικονιζόμενα ως συνάρτηση του καναλιού του ανιχνευτή (Channel Index) πριν από τη βαθμονόμηση σε ενέργεια (keV).



Εικόνα 95: Αποτελέσματα της συσταδοποίησης K-Means ($k=7$) επί των δεδομένων PCA υψηλής ανάλυσης. (Αριστερά) Ο τελικός θεματικός χάρτης που δείχνει τη βελτιωμένη ομαδοποίηση με επτά

συστάδες. (Δεξιά) Το ιστόγραμμα μεγέθους που ποσοτικοποιεί τον αριθμό των pixels για κάθε μία από τις επτά συστάδες.

Έχουμε επομένως καταλήξει στην απεικόνιση των συστάδων επάνω στην αρχική μας εικόνα, Εικόνα 95. Όπως είπαμε και προηγουμένως οι **συστάδες 2 και 6** αποτελούν τις μεγαλύτερες και βρίσκονται κυρίως σε υλικά του υποβάθρου. Εάν από την εικόνα κατανοήσουμε πως η εικόνα είναι ομοιογενοποιημένη τότε τα στοιχεία αυτά αποτελούν παρόμοιο υλικό μικρής ποσότητας ή και προσμίξεις. Τέλος παρατηρούμε πως ο ρουχισμός διακρίνεται κυρίως σε **3 βασικές συστάδες** την **3** την **4** και την **1**. Βεβαίως το πρόγραμμα μας παρέχει και εικόνες επιπλέον με την απεικόνιση σύγκρισης των αξόνων με βάση την συστάδα την οποία ανήκουν Εικόνα 96, εικόνες οι οποίες παρέχονται στο παράρτημα, όπως για παράδειγμα η παρακάτω εικόνα.



Εικόνα 96: Διάγραμμα διασποράς των αποτελεσμάτων της PCA (συνιστώσες 4 έναντι 5), χρωματισμένο σύμφωνα με τις συστάδες που προέκυψαν από την K-Means. Αυτή η απεικόνιση δείχνει πώς οι ομάδες που βρέθηκαν από την K-Means κατανέμονται στον χώρο των κύριων συνιστωσών.

Για την περαιτέρω βελτίωση του διαχωρισμού, εφαρμόστηκε η υβριδική μέθοδος **NMF-PCA**, όπου ο αλγόριθμος **NMF** εφαρμόστηκε στα δεδομένα που είχαν ήδη υποστεί μείωση διαστάσεων από την **PCA**. Αυτή η προσέγγιση συνδυάζει την ικανότητα της **PCA** να συμπυκνώνει την πληροφορία με την ιδιότητα της NMF να εξάγει μη-αρνητικές, και συνεπώς πιο άμεσα ερμηνεύσιμες, συνιστώσες. Ο λόγος

εξηγούμενης διακύμανσης (*explained variance ratio*) ανήλθε στο εντυπωσιακό **99.9%**, υποδεικνύοντας ότι η πληροφορία διατηρήθηκε σχεδόν στο ακέραιο.

<i>Μέθοδος</i>	<i>Πλεονεκτήματα</i>	<i>Περιορισμοί</i>
PCA	Γραμμικές συσχετίσεις, ορθογώνιες συνιστώσες	Δύσκολη ερμηνεία αρνητικών τιμών
NMF	Μη-αρνητικές συνιστώσες, καλύτερη ερμηνεία	Απαιτεί προεπιλογή components

“The explained variance ratio for NMF on the PCA data is: 0.9994235935128527”

Πρόκειται για έναν πολύ υψηλό λόγο εξηγούμενης διακύμανσης (περίπου **99,9%**), ο οποίος είναι αναμενόμενος αφού εκτελούμε **NMF** σε ήδη μειωμένα δεδομένα **PCA**. Ο συνδυασμός **PCA** ακολουθούμενος από **NMF** παρέχει μια ενδιαφέρουσα υβριδική προσέγγιση.

Υβριδική Ανάλυση NMF-PCA: Από τους Χάρτες Υλικών στους Χάρτες Τεχνικής

Η εφαρμογή της **NMF** στα αποτελέσματα της **PCA** (Εικόνα 96) μετασχημάτισε τις αφηρημένες συνιστώσες σε χάρτες αφθονίας με άμεση φυσικοχημική ερμηνεία, όπως φαίνεται και από την ανάλυση των αποτελεσμάτων των συνιστωσών, αποκαλύπτοντας την παλέτα του καλλιτέχνη:

- **Comp. 6 (Λευκό του Μολύβδου - Pb):** Η πιο ευδιάκριτη συνιστώσα, αναδεικνύει τα σαρκώματα και τα φωτίσματα.
- **Comp. 3 , 1 (Χρυσός – Au, Zn):** Χαρτογραφεί με απόλυτη ακρίβεια τις επιχρυσωμένες περιοχές, δηλαδή τα φωτοστέφανα και τις λεπτές διακοσμητικές γραμμές (χρυσοκονδυλιές).
- **Comp. 5 (Ωχρες Σιδήρου - Fe):** Αντιστοιχεί στα καφε-κόκκινα ενδύματα και τις γαιώδεις αποχρώσεις, υποδεικνύοντας τη χρήση χρωστικών με βάση τον Σίδηρο.
- **Comp. 4, 5 (Υπόβαθρο - Gesso):** Παρουσιάζει μια διάχυτη, ομοιογενή κατανομή που αντιστοιχεί στο στρώμα προετοιμασίας (*gesso*), με το σήμα του να είναι ισχυρότερο στις εκτεθειμένες περιοχές φθοράς.
- **Comp. 1 & 2:** Αποτελούν συνιστώσες ελάχιστης διακύμανσης, οι οποίες δεν παρουσιάζουν σαφή χωρική συσχέτιση με την εικονογραφία και αποδίδονται σε υπολειμματικό θόρυβο ή σε πολύ τοπικές, δευτερεύουσες χημικές διαφοροποιήσεις ή σε τοποθέτηση Ψευδαργυρου (Zn).

Ο τελικός θεματικός χάρτης που προέκυψε από την ομαδοποίηση **K-Means** ($k=7$) αυτών των **NMF** συνιστωσών δεν είναι πλέον απλώς ένας χάρτης υλικών, αλλά ένας χάρτης της καλλιτεχνικής τεχνικής. Η ανάλυση καθιστά δυνατό να διακρίνουμε όχι μόνο ποια χρωστική χρησιμοποιήθηκε, αλλά και πώς εφαρμόστηκε. Οι παχιές

στρώσεις λευκού του μολύβδου στα φωτεινότερα σημεία των σαρκωμάτων (**Cluster 6**, Ροζ) διακρίνονται πλέον από τις περιοχές όπου το ίδιο υλικό αναμείχθηκε με ώχρες για τις βασικές στρώσεις του δέρματος (**Cluster 5**, Καφέ). Η χημική ταυτότητα κάθε συστάδας επιβεβαιώνεται από το μέσο φάσμα της, όπου για παράδειγμα το φάσμα του **Cluster 4** κυριαρχείται από τις κορυφές του **Au**, του **Cluster 2** από του **Fe**, και των **Clusters 5/6** από του **Pb**.

Ιεραρχική Διερεύνηση (Sub-clustering): "Βουτιά" στην Τεχνική του Καλλιτέχνη

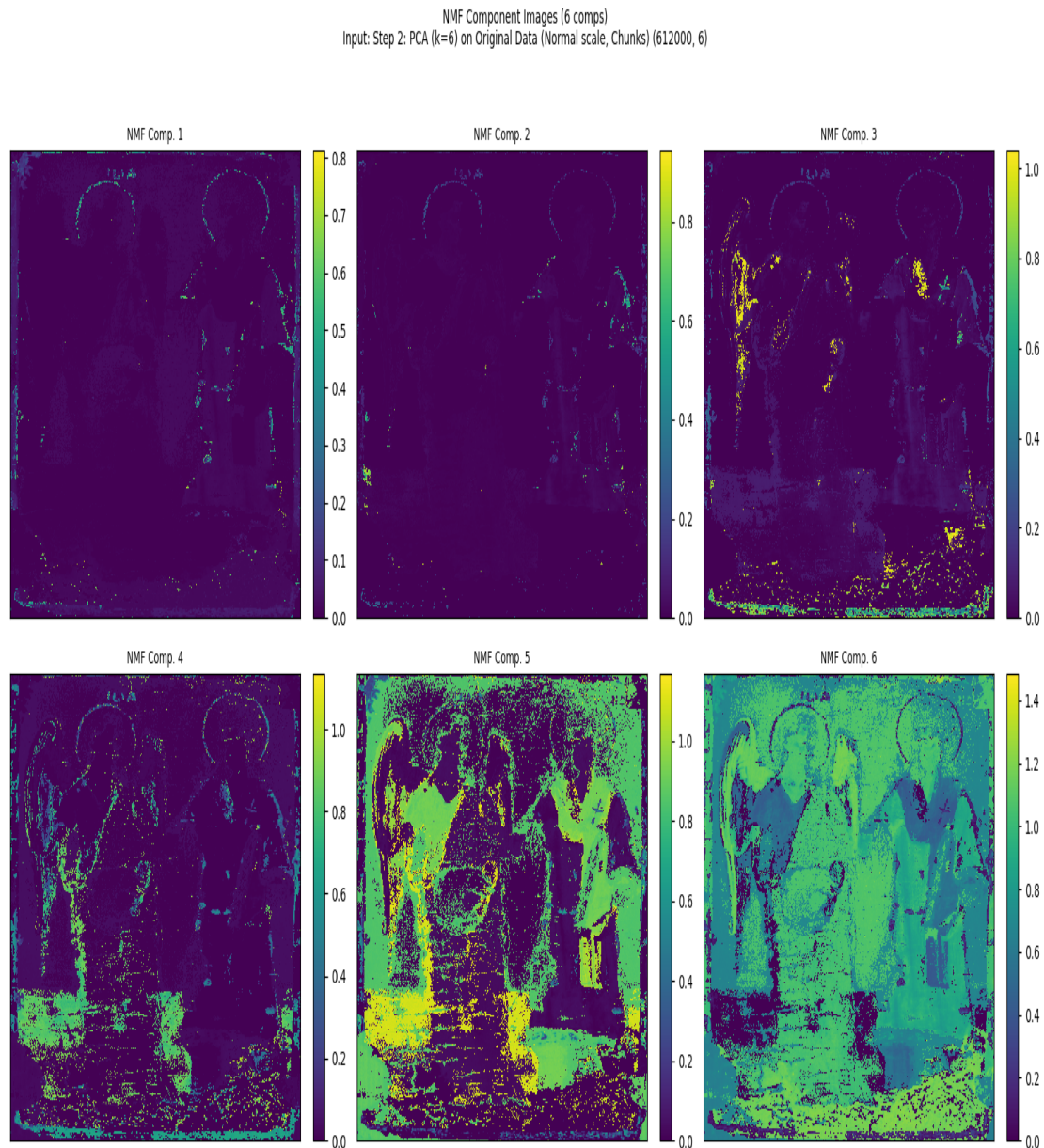
Ένα τρανταχτό παράδειγμα της αυξημένης διακριτικής ικανότητας του λογισμικού **NICE** είναι η δυνατότητα για ιεραρχική ανάλυση. Η περιοχή των σαρκωμάτων (**Parent C6**), αν και φαινομενικά ομοιογενής, αναλύθηκε περαιτέρω. Τα *pixels* που ανήκαν αποκλειστικά σε αυτή τη συστάδα απομονώθηκαν και ο αλγόριθμος **K-Means** εφαρμόστηκε εκ νέου.

Το αποτέλεσμα, όπως φαίνεται στην Εικόνα 109, είναι αποκαλυπτικό. Ο χάρτης του **sub-clustering** αποκαλύπτει με εντυπωσιακή λεπτομέρεια τη δομή του ζωγραφικού στρώματος, διακρίνοντας πλέον υπο-ομάδες που αντιστοιχούν:

- Στις βασικές στρώσεις του δέρματος και τις μεσαίες αποχρώσεις (π.χ. **Sub-C 5**, Καφέ), όπου η ανάμειξη των χρωστικών είναι πιο ομοιογενής.
- Στα έντονα φωτίσματα (**highlights**), όπου οι τελικές πινελιές περιέχουν σχεδόν καθαρό λευκό του μολύβδου (π.χ. **Sub-C 4**, Μωβ).
- Στις πιο σκούρες περιοχές (σκιές), όπου πιθανόν χρησιμοποιήθηκε μεγαλύτερη αναλογία ώχρας ή κάποια άλλη σκούρα χρωστική για τη δημιουργία αντίθεσης και βάθους (π.χ. **Sub-C 1**, Πορτοκαλί & **Sub-C 3**, Κόκκινο).

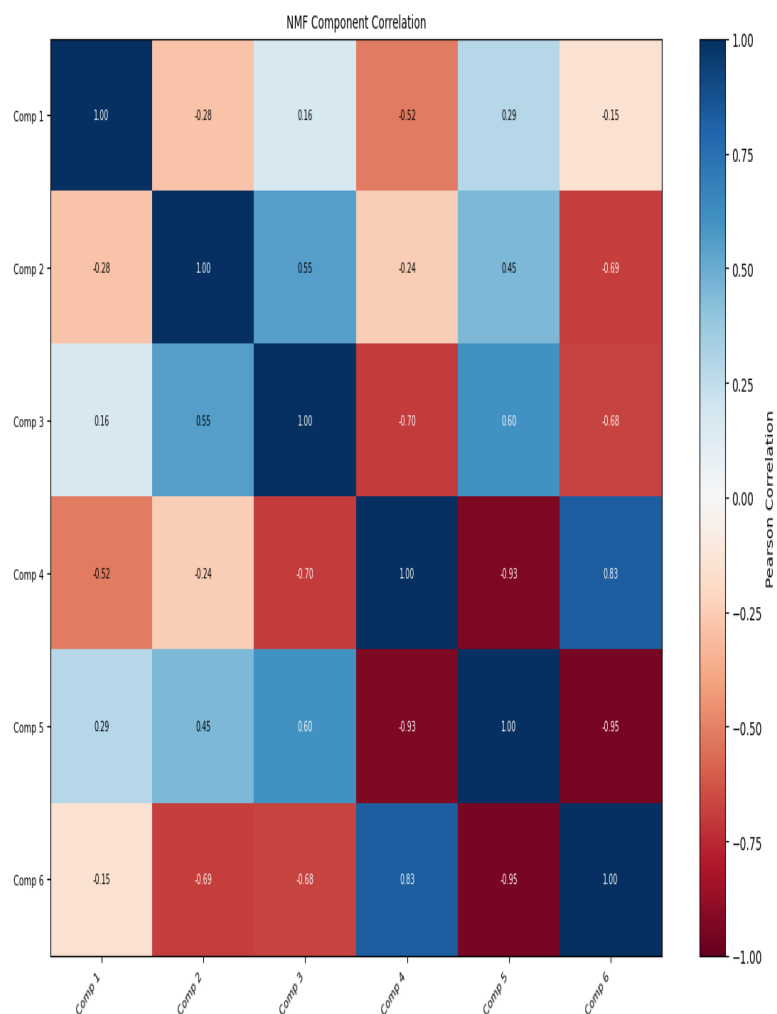
Αυτή η λεπτομερής ανάλυση μας επιτρέπει να ανασυνθέσουμε ψηφιακά τη διαδικασία με την οποία ο καλλιτέχνης "έπλαθε" τις μορφές του, ξεκινώντας από μια βασική στρώση και προσθέτοντας διαδοχικά στρώματα φωτός και σκιάς. Αυτό μετατρέπει το

λογισμικό από ένα απλό εργαλείο ταυτοποίησης σε ένα ισχυρό εργαλείο για την κατανόηση της δημιουργικής διαδικασίας του αγιογράφου.



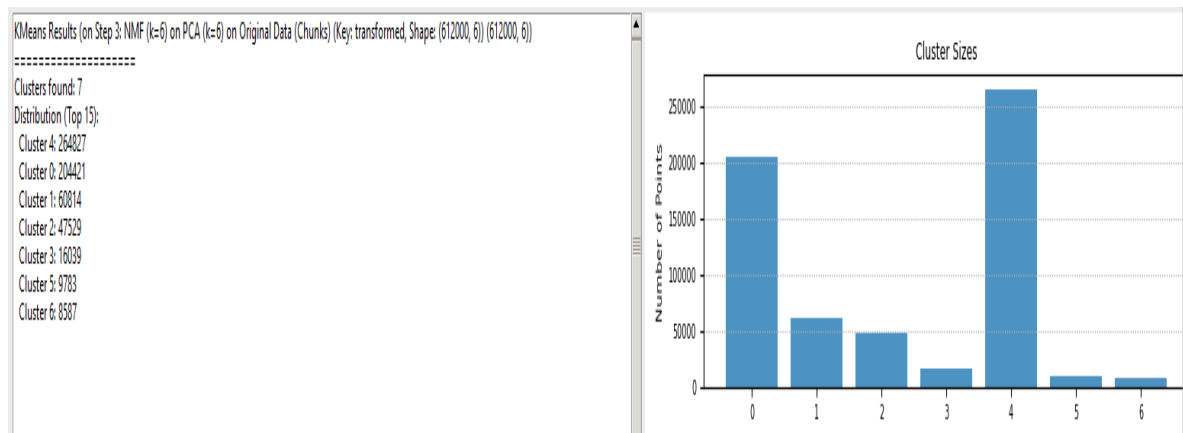
Εικόνα 97: Χωρικοί χάρτες των έξι συνιστωσών που προέκυψαν από την υβριδική ανάλυση NMF-PCA. Οι χάρτες αυτοί μπορούν να ερμηνευθούν ως χάρτες αφθονίας, αναδεικνύοντας τη χωρική κατανομή διακριτών χημικών συνιστωσών με εξαιρετική ευκρίνεια.

Όπως διακρίνουμε πλέον υπάρχει καλύτερος διαχωρισμός των αξόνων μεταξύ τους με πιο αισθητές περιοχές Εικόνα 97. Βεβαία οι συσχετίσεις βρίσκονται μεταξύ των αξόνων 1 με 2 και των αξόνων 3 με 4 και 2. Ωστόσο οι άξονες 4 με 6 βρίσκονται σχεδόν σε απόλυτη συμφωνία, δηλαδή είναι σχεδόν συμπληρωματικοί μεταξύ τους. Τα αποτελέσματα αυτών των συσχετίσεων φαίνονται στην παρακάτω εικόνα Εικόνα 98.

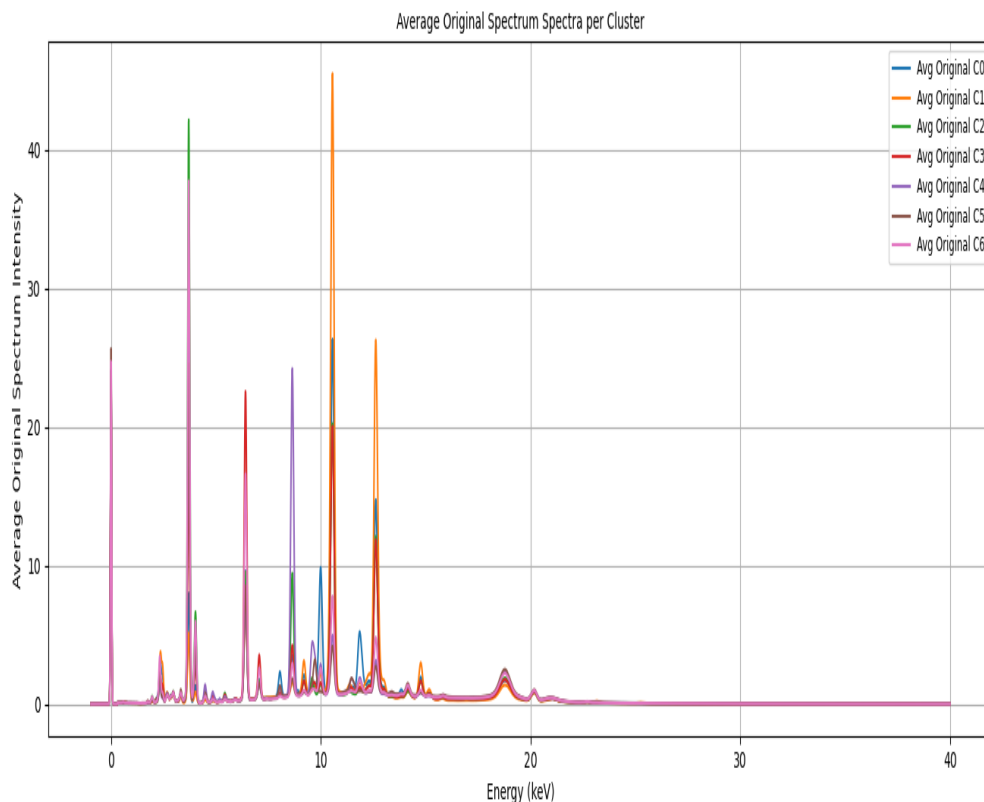


Εικόνα 98: Πίνακας συσχέτισης (heatmap) των συνιστωσών NMF-PCA. Οι ισχυρές θετικές (κόκκινο) και αρνητικές (μπλε) συσχετίσεις υποδεικνύουν σύνθετες σχέσεις μεταξύ των χημικών συνιστωσών, όπως συνύπαρξη ή αμοιβαίο αποκλεισμό.

Επομένως έχουμε παραθέσει όλους τους χάρτες που αφορούν τους άξονες αλλά και τα διανύσματα συγκριτικά μεταξύ τους. Όπως επίσης και την συσχέτιση μεταξύ των αξόνων τόσο σε σχηματικό όσο και σε κάθε εικόνα αξόνων ξεχωριστά. Επιπρόσθετα παρέχουμε το φάσμα των αξόνων με βάση το αρχικό μας φάσμα. Μπορούμε επομένως με βάση την μέθοδο **Kmeans** να προχωρήσουμε στην συσταδοποίηση με βάση και πάλι τον αριθμό των **7 συστάδων**. Ο αριθμός των δεδομένων που περιλαμβάνει κάθε συστάδα φαίνεται στην παρακάτω εικόνα Εικόνα 99, όπως επίσης και το διάγραμμα των συστάδων προβαλλόμενα πάνω στο φάσμα των καναλιών Εικόνα 100.



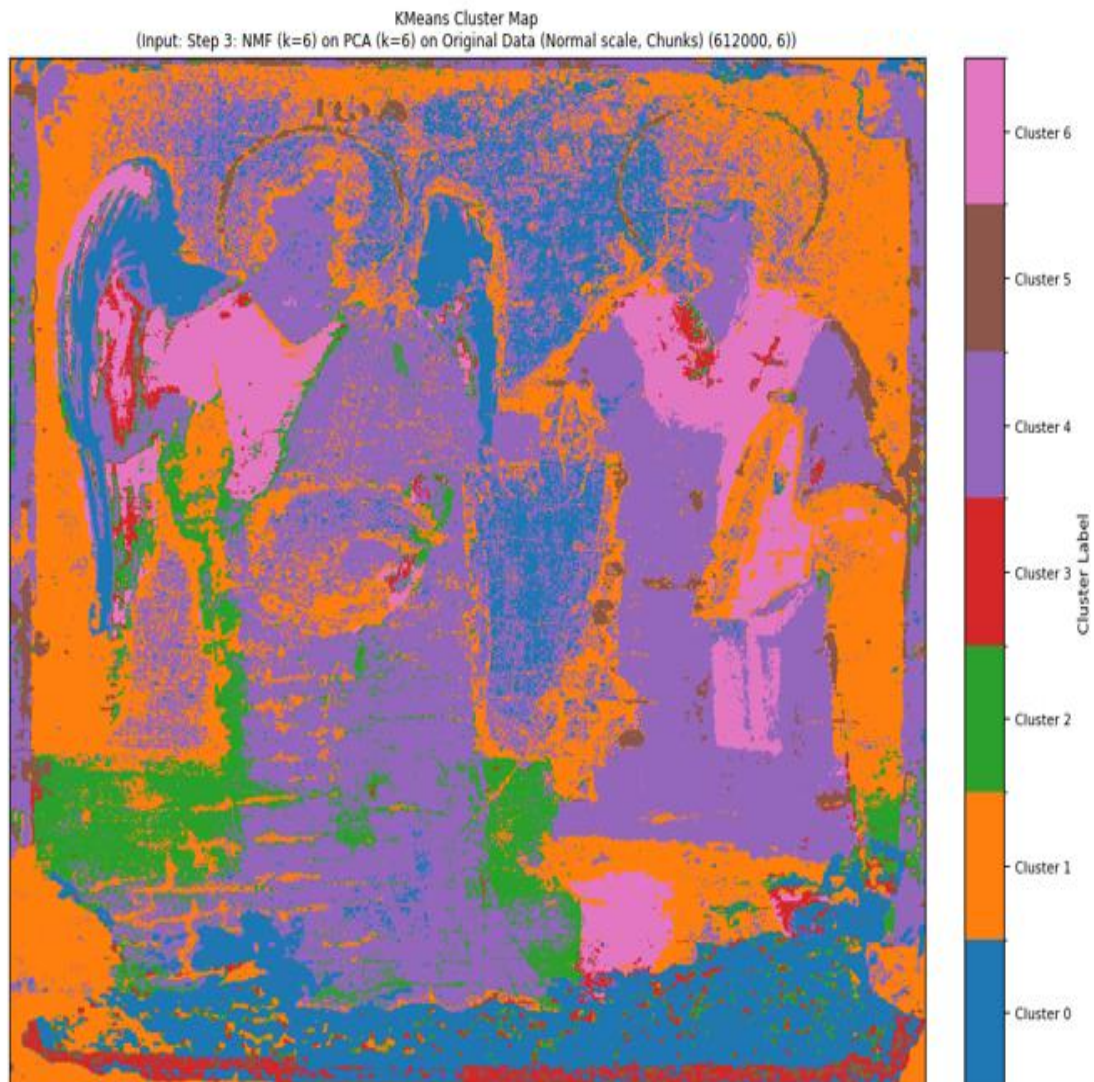
Εικόνα 99: Ποσοτικά αποτελέσματα από την τελική συσταδοποίηση *K-Means* ($k=7$) επί των δεδομένων *NMF-PCA*. (Αριστερά) Αριθμητική κατανομή των *pixels* ανά συστάδα. (Δεξιά) Γραφική απεικόνιση της κατανομής μεγέθους των συστάδων.



Εικόνα 100: Μέσα φάσματα για κάθε μία από τις επτά συστάδες που προέκυψαν από την ομαδοποίηση *K-Means* επί των δεδομένων *NMF-PCA*, επιτρέποντας την τελική χημική ερμηνεία κάθε ομάδας.

Η εφαρμογή της βελτιστοποιημένης υβριδικής μεθόδου *NMF-PCA* στο σύνολο δεδομένων υψηλής ανάλυσης απέδωσε χάρτες συνιστωσών με εξαιρετική σαφήνεια, επιτρέποντας μια βαθιά "βουτιά" στην τεχνική του αιογράφου. Με βάση τις μετρικές αξιολόγησης (*Elbow*, *Silhouette*), ο βέλτιστος αριθμός συστάδων προσδιορίστηκε μεταξύ 4 και 7. Για την επίτευξη του μέγιστου δυνατού διαχωρισμού και την ανάδειξη λεπτομερειών, επιλέχθηκε η ομαδοποίηση σε **επτά (7) συστάδες**.

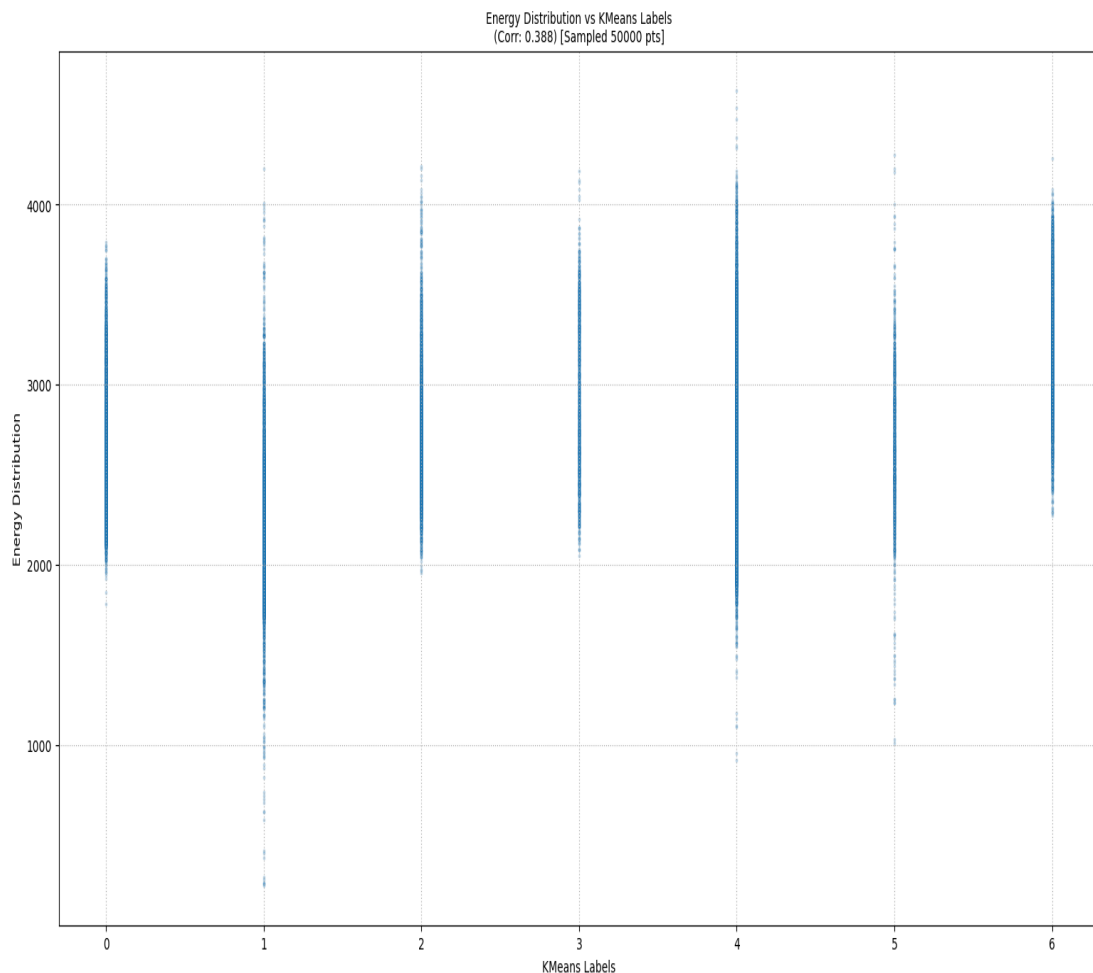
Ο τελικός θεματικός χάρτης που προέκυψε από την εφαρμογή του αλγορίθμου *K-Means* στα αποτελέσματα της *NMF-PCA* φαίνεται στην παρακάτω εικόνα Εικόνα 101.



Εικόνα 101: Ο τελικός θεματικός χάρτης *K-Means* ($k=7$) από την υβριδική ανάλυση *NMF-PCA* στα δεδομένα υψηλής ανάλυσης. Ο χάρτης αυτός δεν είναι απλώς ένας χάρτης υλικών, αλλά ένας χάρτης της καλλιτεχνικής τεχνικής, διακρίνοντας παχιές από λεπτές στρώσεις και φωτίσματα από σκιές.

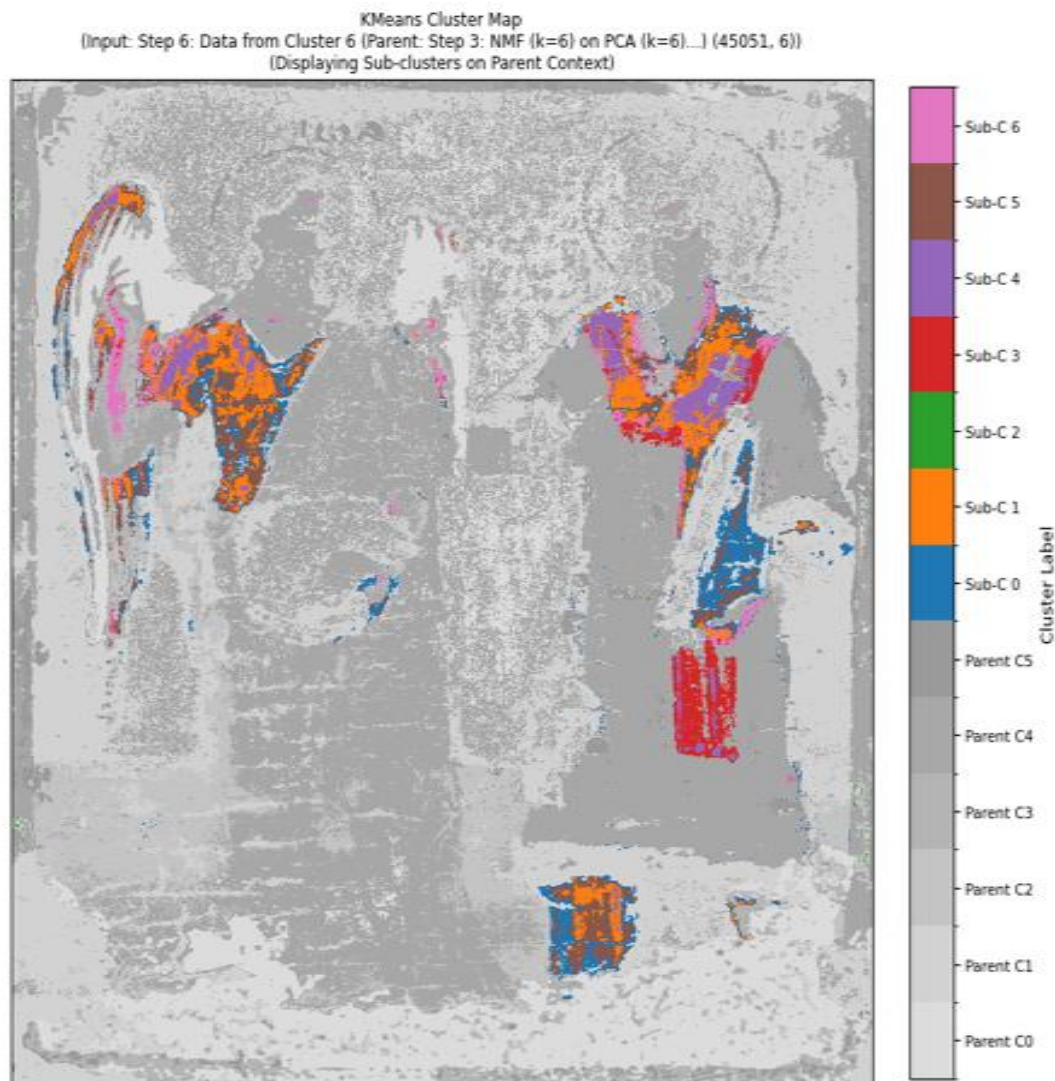
Ο χάρτης αυτός δεν είναι απλώς ένας χάρτης υλικών, αλλά ένας χάρτης της καλλιτεχνικής τεχνικής. Η αυξημένη ανάλυση καθιστά πλέον δυνατό να διακρίνουμε όχι μόνο ποια χρωστική χρησιμοποιήθηκε, αλλά και πώς ακριβώς εφαρμόστηκε. Οι φωτεινότερες περιοχές των σαρκωμάτων, που αντιστοιχούν στα τελικά φωτίσματα με παχιές στρώσεις λευκού του μολύβδου (**Cluster 6**, Ροζ), διακρίνονται πλέον με σαφήνεια από τις βασικές στρώσεις του δέρματος (**Cluster 5**, Καφέ), όπου το ίδιο υλικό πιθανότατα αναμείχθηκε με ώχρες. Ομοίως, η επιχρύσωση των φωτοστεφάνων απομονώνεται πλήρως στο **Cluster 4** (Μωβ), ενώ οι πυκνές περιοχές των ώχρων στα ενδύματα ταυτοποιούνται στο **Cluster 2** (Πράσινο). Η υψηλή ανάλυση μετατρέπει τον θεματικό χάρτη σε ένα πανίσχυρο εργαλείο για την κατανόηση της δημιουργικής διαδικασίας του αγιογράφου.

Ένα τρανταχτό παράδειγμα της αυξημένης διακριτικής ικανότητας είναι η απομόνωση του σταυρού που κρατά ο Άγιος Ιωάννης, ο οποίος αντιστοιχίζεται πλέον σε μια διακριτή συστάδα (**Cluster 3, κόκκινο**), πιθανότατα λόγω της χρήσης μιας συγκεκριμένης κόκκινης χρωστικής (π.χ. κιννάβαρι ή κόκκινη ώχρα) που δεν ήταν δυνατόν να διαχωριστεί στην ανάλυση χαμηλής ανάλυσης. Το διάγραμμα κατανομής της ενέργειας ανά συστάδα (***Energy Distribution vs KMeans Labels*** Εικόνα 102) επιβεβαιώνει τον επιτυχή διαχωρισμό, δείχνοντας διακριτές κατανομές για κάθε μία από τις επτά ομάδες.



Εικόνα 102: Διάγραμμα διασποράς της συνολικής έντασης (*Energy Distribution*) για κάθε μία από τις επτά τελικές συστάδες, επιβεβαιώνοντας τον επιτυχή και διακριτό διαχωρισμό τους.

Για την περαιτέρω διερεύνηση της τεχνικής του καλλιτέχνη, εφαρμόστηκε ιεραρχική συσταδοποίηση (***sub-clustering***) στις κύριες περιοχές ενδιαφέροντος που προέκυψαν. Η ανάλυση εστιάστηκε κυρίως στα ενδύματα και τα σαρκώματα (**Cluster 6**), με στόχο τον εντοπισμό των υλικών που χρησιμοποιήθηκαν για τις φωτοσκιάσεις και τις λεπτομέρειες Εικόνα 103.



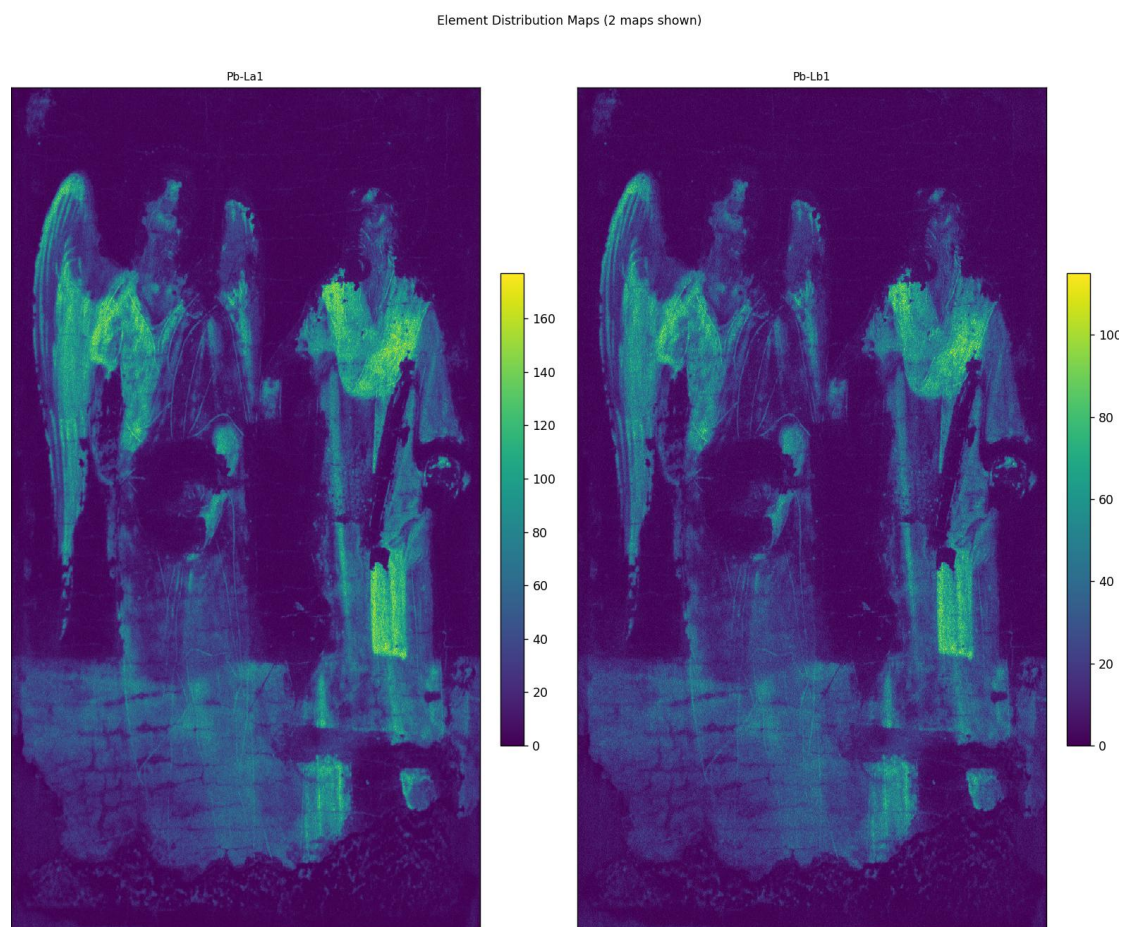
Εικόνα 103: Αποτέλεσμα της ιεραρχικής ανάλυσης (sub-clustering) στην περιοχή των σαρκωμάτων (Parent C6). Οι υπο-ομάδες που προκύπτουν αντιστοιχούν σε διακριτά στάδια της ζωγραφικής διαδικασίας, όπως η βασική στρώση, τα φωτίσματα και οι σκιές.

Ο χάρτης του **sub-clustering** αποκαλύπτει με εντυπωσιακή λεπτομέρεια τη δομή του ζωγραφικού στρώματος Εικόνα 103. Για παράδειγμα, μέσα στην ευρύτερη περιοχή των σαρκωμάτων (**Parent C6**), διακρίνονται πλέον υπο-ομάδες που αντιστοιχούν:

- Στη βασική στρώση του δέρματος (προπλασμός) και τους μεσαίους τόνους, όπου η ανάμειξη λευκού του μολύβδου και ώχρας είναι πιο ομοιογενής (**Sub-C 1**, Πορτοκαλί και **Sub-C 5**, Καφέ).
- Στα έντονα φωτίσματα (ψιμυθιές), τις τελικές δηλαδή πινελιές στα πιο προεξέχοντα σημεία του προσώπου (μύτη, μέτωπο, ζυγωματικά) με σχεδόν καθαρό λευκό του μολύβδου (**Sub-C 4**, Μωβ).
- Στις πιο σκούρες περιοχές (σκιές), όπου πιθανόν χρησιμοποιήθηκε μεγαλύτερη αναλογία ώχρας ή κάποια άλλη σκούρα χρωστική για τη δημιουργία αντίθεσης και βάθους (**Sub-C 3**, Κόκκινο).

Αυτή η λεπτομερής ανάλυση μας επιτρέπει να ανασυνθέσουμε ψηφιακά τη διαδικασία με την οποία ο καλλιτέχνης "έπλαθε" τις μορφές του, ξεκινώντας από μια βασική στρώση και προσθέτοντας διαδοχικά στρώματα φωτός και σκιάς, αποκαλύπτοντας την

τεχνοτροπία και τη δεξιοτεχνία του σε ένα επίπεδο που θα ήταν αόρατο με το γυμνό μάτι.

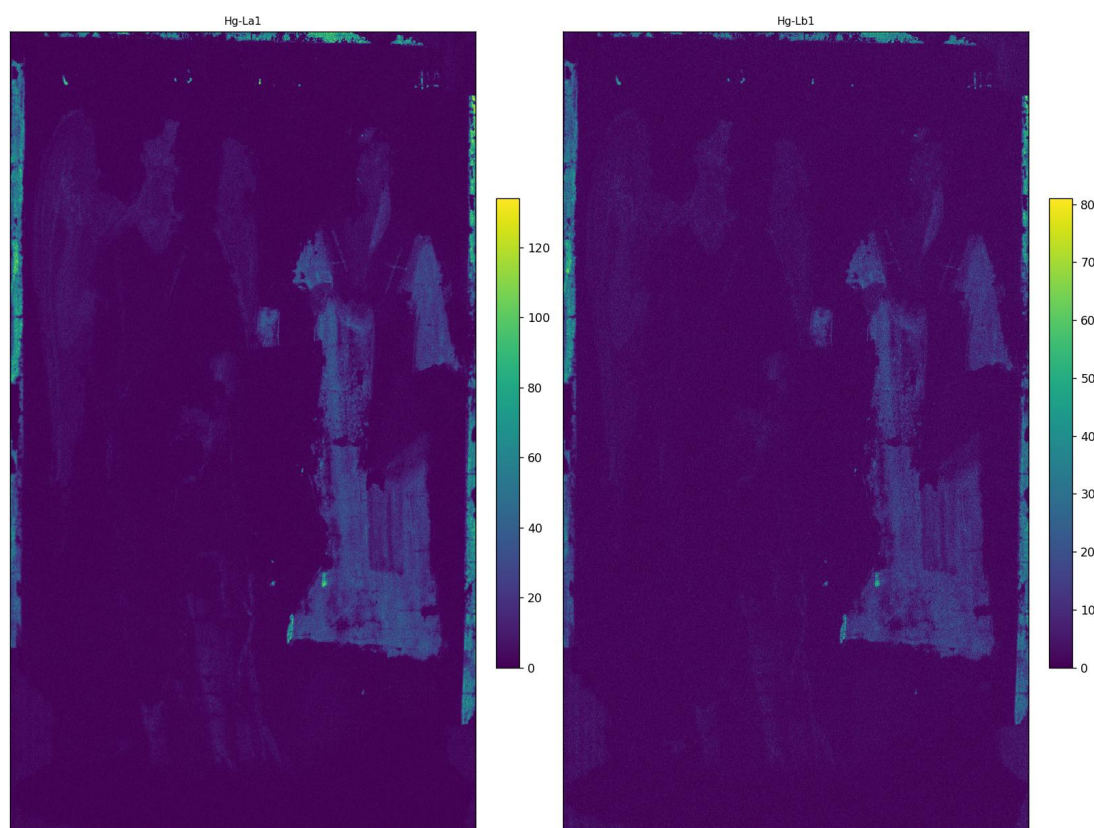


Εικόνα 104: Στοιχειακοί χάρτες κατανομής για τον Μόλυβδο (**Pb**), για τις γραμμές εκπομπής **Pb-La1**, **Pb-Lb1**. Η υψηλή συγκέντρωση του μολύβδου στα σαρκώματα και στα "φωτίσματα" ταυτοποιεί τη χρήση του Λευκού του Μολύβδου, αποκαλύπτοντας τις λεπτές πινελιές (ψιμυθιές) του καλλιτέχνη.

Μόλυβδος (**Pb**) - Το Λευκό των Σαρκωμάτων και των Φωτισμάτων

Οι χάρτες του Μολύβδου (**Pb**), και ειδικά οι εξαιρετικά ευκρινείς χάρτες που προέρχονται από τις γραμμές εκπομπής **L-shell** (**Pb-La1** και **Pb-Lb1**), παρέχουν την πιο λεπτομερή εικόνα για την τεχνική απόδοσης του όγκου και του φωτός. Αντιθέτως, οι χάρτες των γραμμών **K-shell** (**Pb-Kb**) εμφανίζουν μόνο θόρυβο, λόγω της πολύ υψηλής ενέργειας που απαιτείται για τη διέγερσή τους. Ο μολύβδος εντοπίζεται με πολύ υψηλή συγκέντρωση στις περιοχές των σαρκωμάτων (πρόσωπα και χέρια), επιβεβαιώνοντας τη χρήση του Λευκού του Μολύβδου (**2PbCO₃·Pb(OH)₂**) ως βάση για την απόδοση του δέρματος. Η υψηλή ανάλυση καθιστά πλέον ορατές τις μεμονωμένες, λεπτές πινελιές (ψιμυθιές) που χρησιμοποίησε ο καλλιτέχνης για να δημιουργήσει τα φωτίσματα (**highlights**) στις πιο προεξέχουσες περιοχές, όπως η μύτη, τα ζυγωματικά και οι πτυχώσεις των ενδυμάτων. Η καθαρότητα αυτών των χαρτών **L-shell**

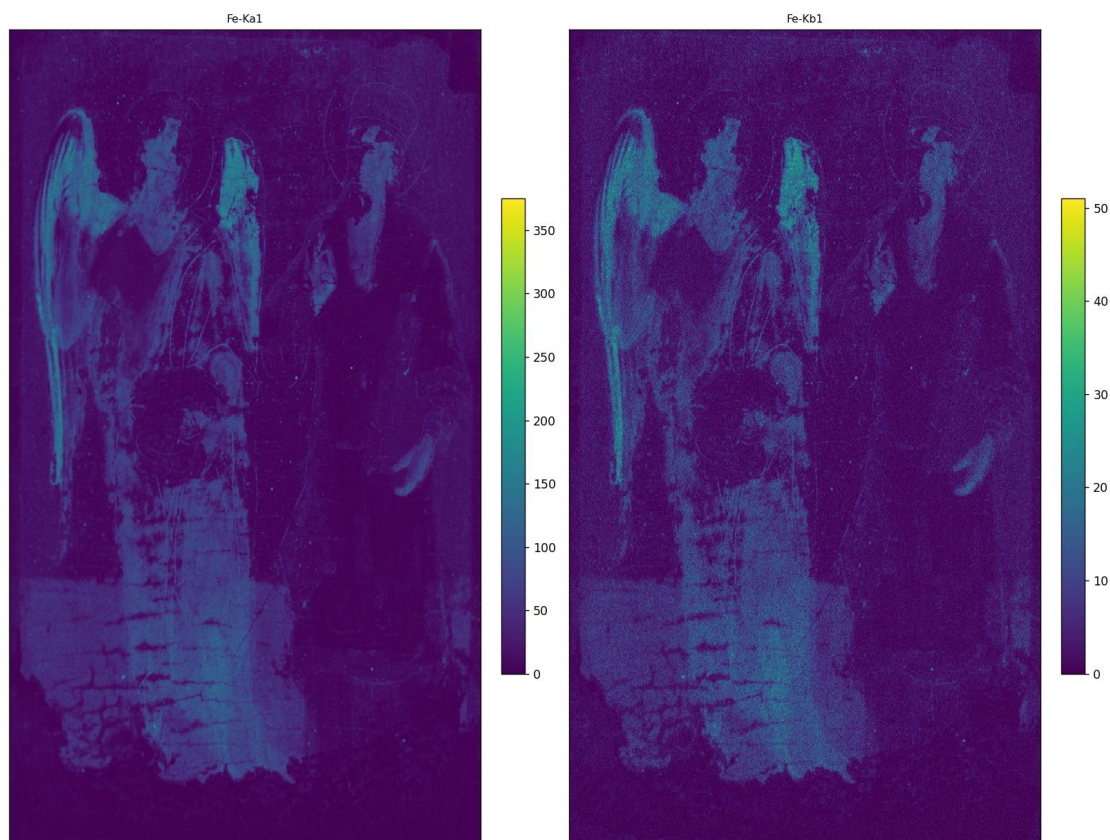
επιβεβαιώνει οπτικά και με απόλυτη ακρίβεια τα ευρήματα της ανάλυσης *sub-clustering*.



Εικόνα 105: Στοιχειακοί χάρτες κατανομής για τον Υδράργυρο (*Hg*). Η στοχευμένη παρουσία του (κυρίως στις γραμμές *La1* και *Lb1*) σε συγκεκριμένες κόκκινες περιοχές, όπως ο σταυρός που εντοπίζεται στο στήθος του Ιεράρχη, ταυτοποιεί τη χρήση της πολύτιμης χρωστικής κιννάβαρι (*HgS*).

Υδράργυρος (*Hg*) - Η Έκπληξη της Κόκκινης Χρωστικής

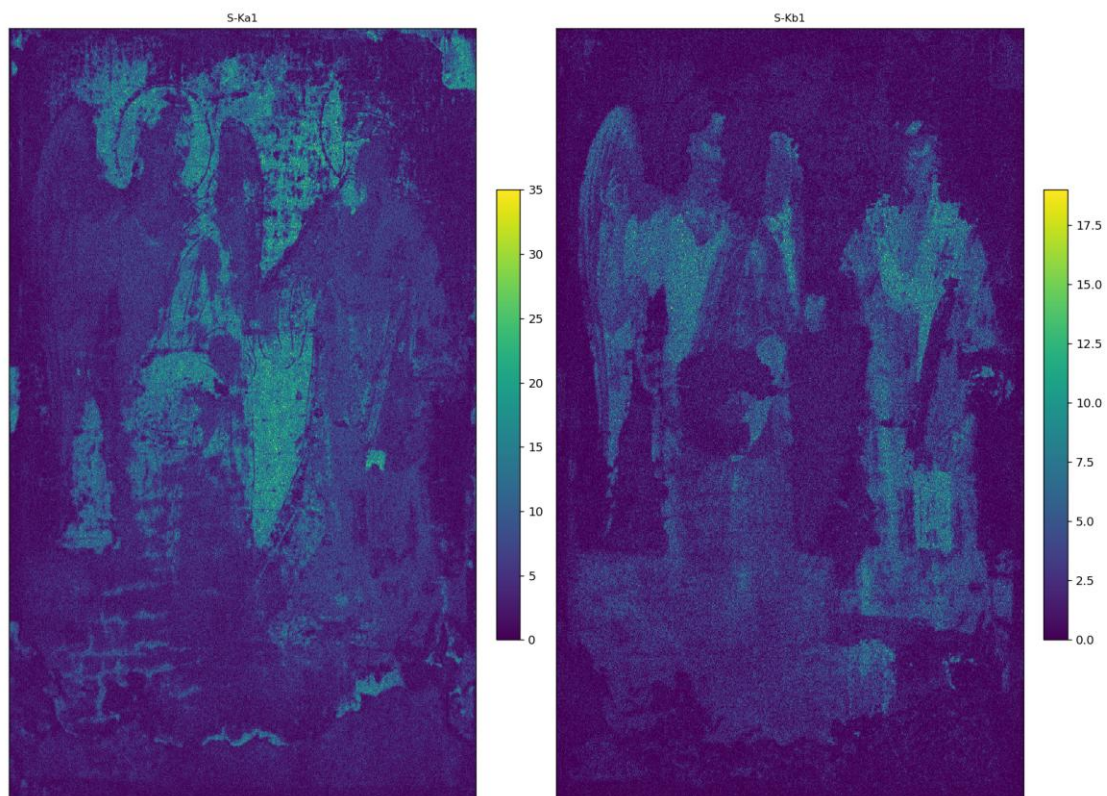
Μια από τις σημαντικότερες ανακαλύψεις της ανάλυσης υψηλής ανάλυσης είναι η σαφής ανίχνευση Υδραργύρου (*Hg*). Ενώ οι γραμμές *K-shell* ου *Hg* δεν μπορούν να διεγερθούν με λυχνία *Rh*, οι χάρτες *Hg-La1* και *Hg-Lb1* αποκαλύπτουν την παρουσία του σε πολύ συγκεκριμένες περιοχές, κυρίως στον κόκκινο σταυρό που κρατά ο Άγιος και σε ορισμένες κόκκινες λεπτομέρειες στα ενδύματα. Η παρουσία του υδραργύρου ταυτοποιεί αδιαμφισβήτητα τη χρήση του κιννάβαρι (*vermilion*, *HgS*), μιας πολύτιμης και λαμπερής κόκκινης χρωστικής που χρησιμοποιούνταν για σημεία ιδιαίτερης θεολογικής ή καλλιτεχνικής έμφασης. Αυτό το εύρημα εξηγεί πλήρως γιατί ο σταυρός απομονώθηκε ως μια διακριτή συστάδα (*Cluster 3*) στην ανάλυση *K-Means*, καθώς η χημική του υπογραφή (*Hg*) είναι μοναδική και διαφέρει από τις ώχρες σιδήρου (*Fe*).



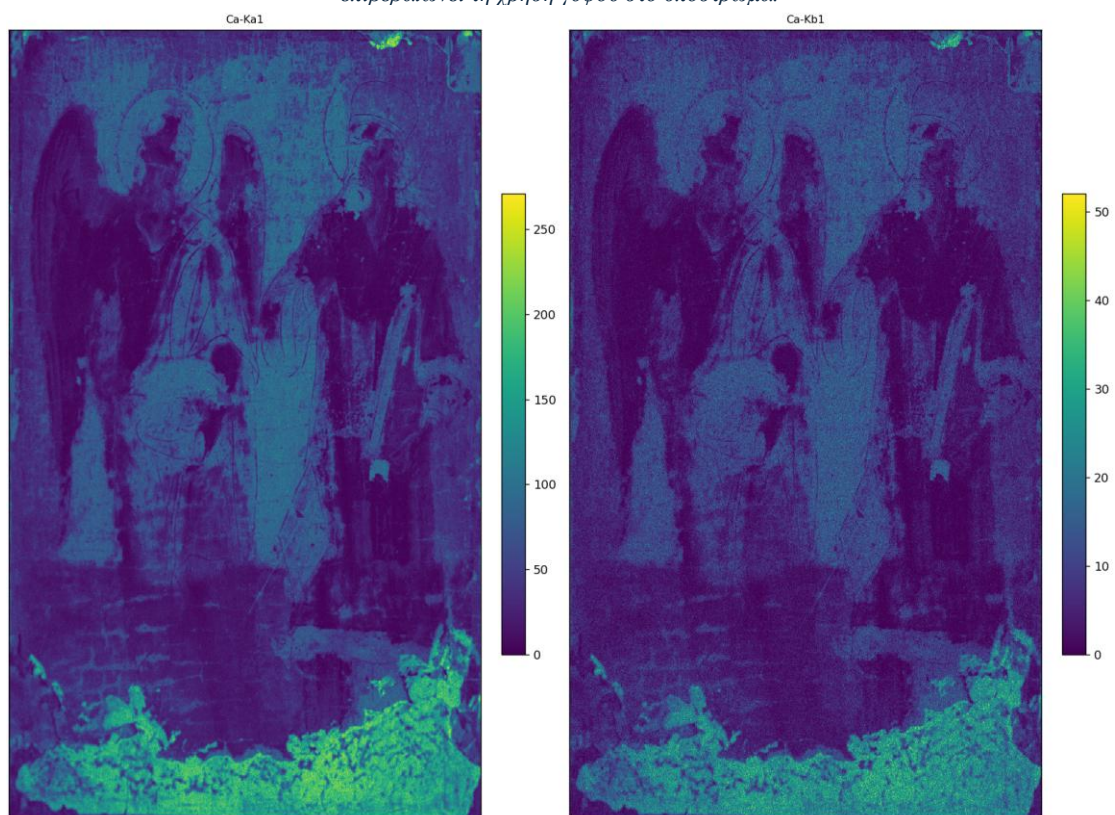
Εικόνα 106: Στοιχειακοί χάρτες κατανομής για τον Σίδηρο (*Fe*). Η κατανομή του, κυρίως μέσω των γραμμών *Kα*, συνδέεται άμεσα με τις καφε-κόκκινες και ωχρές περιοχές, επιβεβαιώνοντας τη χρήση γήινων χρωστικών (ώχρες) με βάση τα οξείδια του σιδήρου

Σίδηρος (*Fe*) - Οι Γήινες Χρωστικές των Ενδυμάτων

Η κατανομή του Σιδήρου (*Fe*), όπως αποτυπώνεται με εξαιρετική ευκρίνεια στους χάρτες *K-shell* (*Fe-Kα1*, *Kβ1*), συνδέεται άμεσα με τις καφε-κόκκινες και ωχρές περιοχές των ενδυμάτων. Αυτό επιβεβαιώνει την εκτεταμένη χρήση γήινων χρωστικών με βάση τα οξείδια του σιδήρου (ώχρες, *sienna*). Η υψηλή ανάλυση επιτρέπει πλέον να διακρίνουμε περιοχές με διαφορετική πυκνότητα εφαρμογής της χρωστικής: παχύτερες, πιο αδιαφανείς στρώσεις στις βασικές επιφάνειες και λεπτότερες, πιο αραιές στρώσεις στις φωτοσκιάσεις. Η παρατήρηση μιας μικρότερης, αλλά σαφούς, συγκέντρωσης σιδήρου και στις περιοχές των σαρκωμάτων ενισχύει την υπόθεση ότι ο καλλιτέχνης ανέμιξε μικρή ποσότητα ώχρας με το λευκό του μολύβδου για να επιτύχει τη θερμή, φυσική απόχρωση του δέρματος.



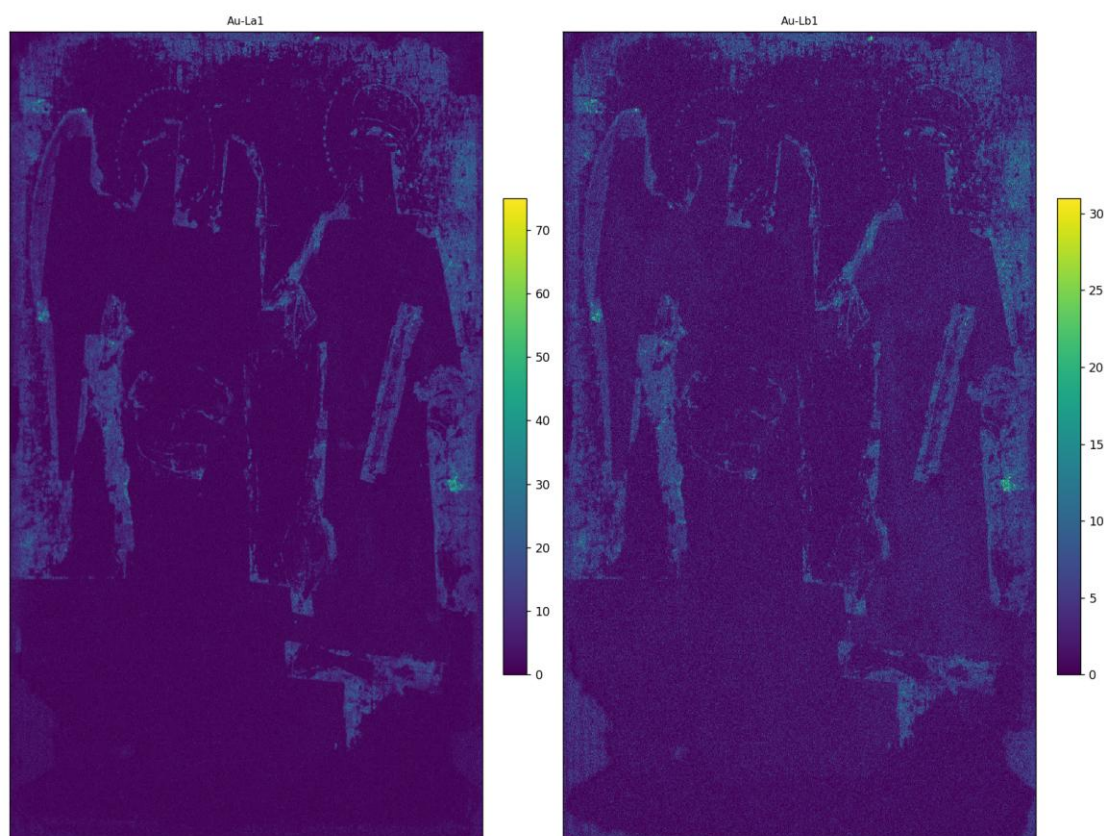
Εικόνα 107: Στοιχειακοί χάρτες κατανομής για το **Θείο (S)** (γραμμές K-edge, Kα1, Kβ1). Η διάχυτη παρουσία του επιβεβαιώνει τη χρήση γύψου στο υπόστρωμα.



Εικόνα 108: Στοιχειακοί χάρτες κατανομής για το **Ασβέστιο (Ca)** (γραμμές Kα1, Kβ1). Η ισχυρή χωρική συσχέτιση με το θείο επιβεβαιώνει την παρουσία γύψου (CaSO_4).

Ασβέστιο (Ca) και Θείο (S) - Το Υπόστρωμα Προετοιμασίας

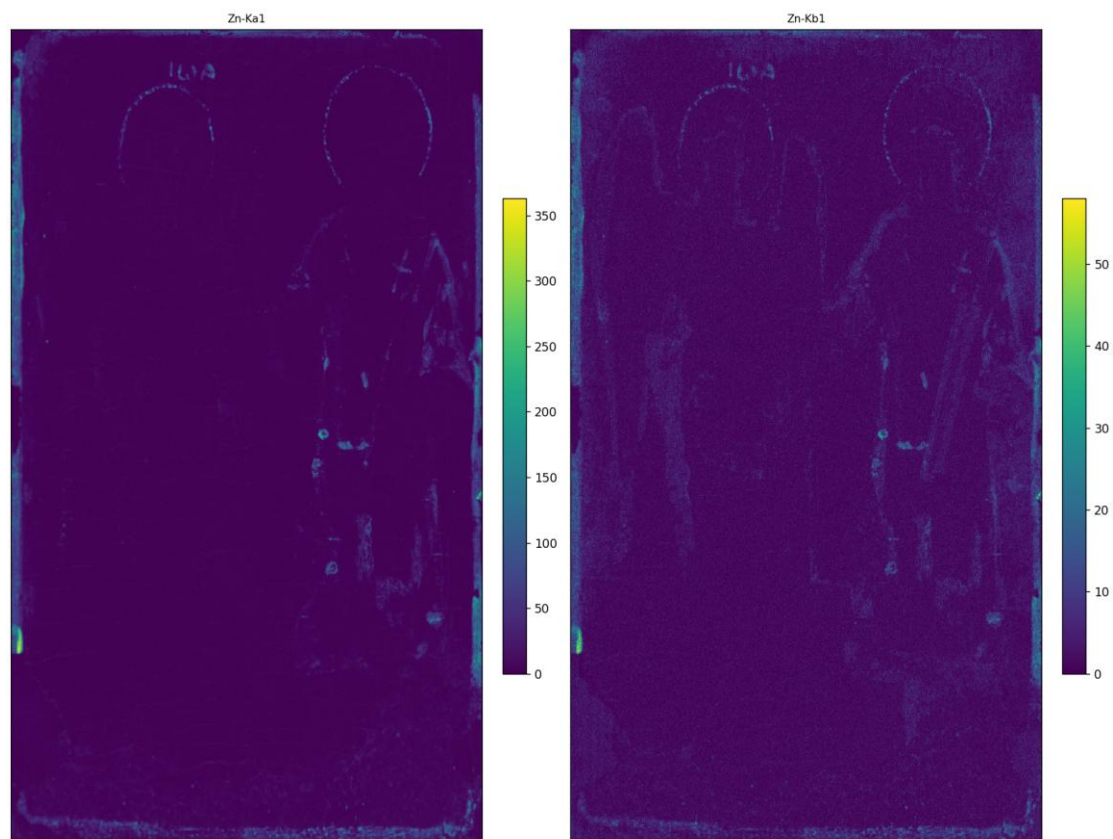
Οι χάρτες κατανομής του Ασβεστίου (**Ca**) και του Θείου (**S**) παρουσιάζουν μια σχεδόν καθολική, διάχυτη παρουσία σε ολόκληρη την επιφάνεια της εικόνας. Η ισχυρή χωρική συσχέτιση μεταξύ των δύο αυτών στοιχείων υποδεικνύει αδιαμφισβήτητη τη χρήση του γύψου (Θεικό Ασβέστιο, **CaSO₄**) ως το κύριο υλικό για το στρώμα προετοιμασίας (**gesso**). Η ελαφρά μείωση της έντασης των σημάτων τους στις περιοχές των μορφών οφείλεται στο φαινόμενο της εξασθένησης (**attenuation**) από τα πυκνά, υπερκείμενα στρώματα χρωστικών βαρέων στοιχείων (**Pb, Fe, Au**). Η ποιότητα των δεδομένων υψηλής ανάλυσης είναι τέτοια που καθιστά ορατό ακόμα και το μοτίβο του **craquelure** (δικτυωτές ρωγμές) στο υπόστρωμα, παρέχοντας πολύτιμες πληροφορίες για την κατάσταση διατήρησης του έργου.



Εικόνα 109: Στοιχειακοί χάρτες κατανομής για τον Χρυσό (**Au**) (γραμμές **L**). Η αποκλειστική παρουσία του στα φωτοστέφανα και στις χρυσοκονδυλίες ταυτοποιεί την τεχνική της επιχρύσωσης.

Χρυσός (**Au**) - Η Τέχνη της Επιχρύσωσης

Οι χάρτες του Χρυσού (**Au**), προερχόμενοι από τις έντονες γραμμές **L-shell** (**Au-La1, Au-Lb1**), χαρτογραφούν με εκπληκτική λεπτομέρεια τα επιχρυσωμένα τμήματα. Το σήμα του χρυσού είναι αποκλειστικά εντοπισμένο στα φωτοστέφανα και στις εξαιρετικά λεπτές, καλλιγραφικές γραμμές της χρυσοκονδυλιάς στα ενδύματα. Η υψηλή ανάλυση επιτρέπει πλέον την παρατήρηση της υφής και της συνέχειας κάθε πινελιάς, αποκαλύπτοντας την δεξιοτεχνία και τη σταθερότητα του χεριού του καλλιτέχνη. Σε ορισμένα σημεία, είναι ορατές μικρές φθορές ή απώλειες του φύλλου χρυσού, παρέχοντας πολύτιμες πληροφορίες και για την κατάσταση διατήρησης των πιο ευαίσθητων περιοχών του έργου.

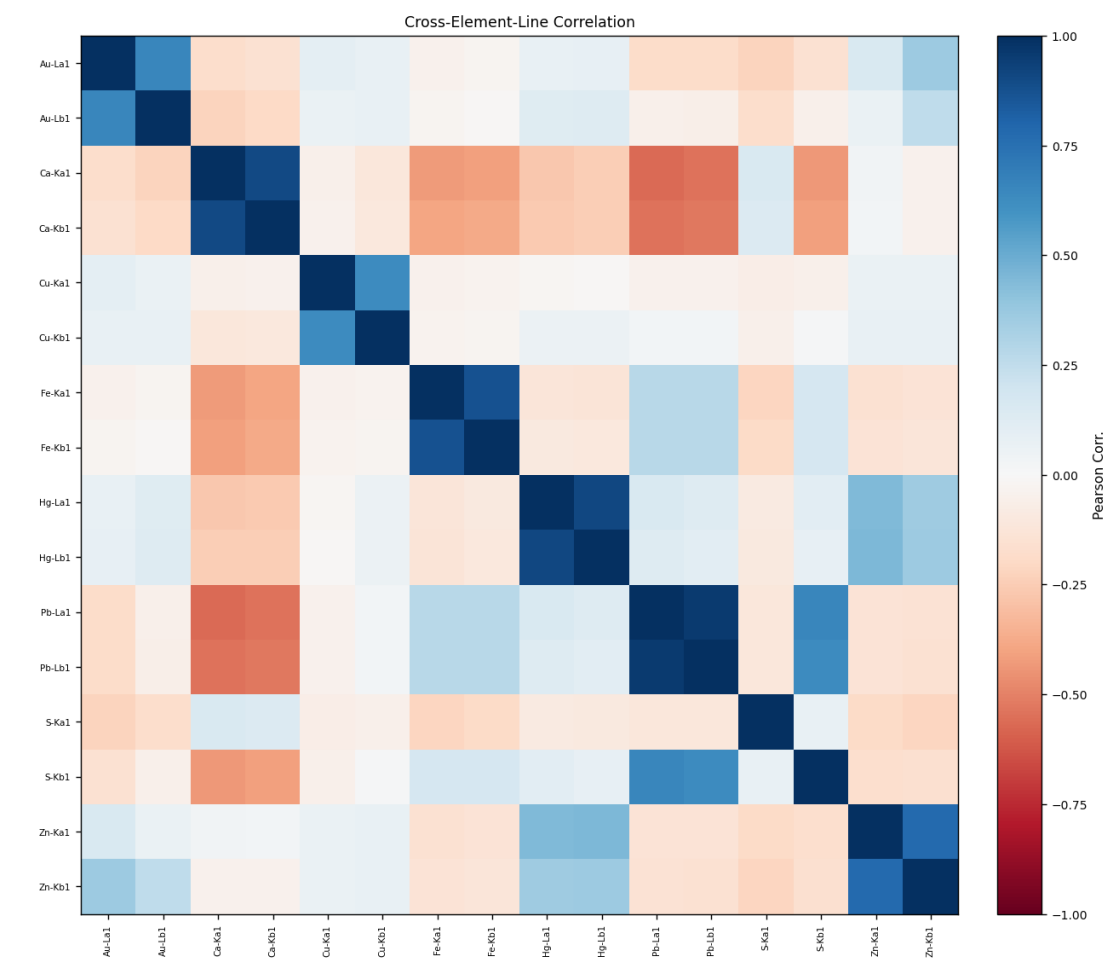


Εικόνα 110: Στοιχειακοί χάρτες κατανομής για τον **Ψευδάργυρο (Zn)**. Η σποραδική παρουσία του αποτελεί ένδειξη μεταγενέστερων επεμβάσεων συντήρησης (*retouching*).

Ψευδάργυρος (Zn) - Ιχνη Μεταγενέστερων Επεμβάσεων

Τέλος, οι χάρτες του Ψευδαργύρου (**Zn**), και κυρίως ο χάρτης **Zn-Ka**, επιβεβαιώνουν τα ευρήματα της αρχικής ανάλυσης. Η παρουσία του είναι σποραδική, εντοπίζεται σε μικρές, ασύνδετες κηλίδες και δεν συσχετίζεται με καμία από τις περιοχές της αρχικής ζωγραφικής παλέτας. Αυτή η κατανομή είναι χαρακτηριστική της χρήσης του Λευκού του Ψευδαργύρου

(ZnO) και αποτελεί σαφή ένδειξη τοπικών επεμβάσεων συντήρησης (*retouching*) που πραγματοποιήθηκαν σε μεταγενέστερη εποχή για την κάλυψη φθορών.



Εικόνα 111: Πίνακας συσχέτισης (heatmap) μεταξύ των κυριότερων στοιχειακών γραμμών που ανιχνεύθηκαν στην ανάλυση υψηλής ανάλυσης. Οι θετικές συσχετίσεις (κόκκινο) υποδεικνύουν χωρική συνύπαρξη υλικών, ενώ οι αρνητικές (μπλε) αμοιβαίο αποκλεισμό.

Η ανάλυση υψηλής ανάλυσης της εικόνας του Αγίου Ιωάννη όχι μόνο επιβεβαίωσε τα ευρήματα της αρχικής μελέτης αλλά τα επέκτεινε σημαντικά, παρέχοντας μια άνευ προηγουμένου ματιά στην τεχνική του καλλιτέχνη. Η εφαρμογή της βελτιστοποιημένης, υβριδικής μεθοδολογίας *NMF-PCA* αποδείχθηκε εξαιρετικά αποτελεσματική, οδηγώντας στην εξαγωγή θεματικών χαρτών με πρωτοφανή λεπτομέρεια.

Τα κύρια συμπεράσματα της μελέτης αυτής είναι:

1. **Πλήρης Χαρακτηρισμός της Παλέτας:** Η παλέτα του καλλιτέχνη ταυτοποιήθηκε με απόλυτη βεβαιότητα και περιλαμβάνει:
 - **Υπόστρωμα:** Γύψος (*CaSO₄*), όπως επιβεβαιώνεται από την καθολική συν-εντόπιση Ασβεστίου (*Ca*) και Θείου (*S*).
 - **Λευκά:** Λευκό του Μολύβδου (*Pb-based*), ως το κυρίαρχο υλικό για τα σαρκώματα και τα φωτίσματα.
 - **Καφέ/Κόκκινα:** Γήινες ώχρες (*Fe-based*) για τα ενδύματα και **Κιννάβαρι** (*HgS*) για σημεία ιδιαίτερης έμφασης, όπως ο σταυρός.

- ο **Επιχρύσωση:** Φύλλο Χρυσού (**Au**) για τα φωτοστέφανα και τις χρυσοκονδυλιές.
- ο **Μεταγενέστερες Επεμβάσεις:** Λευκό του Ψευδαργύρου (**Zn-based**) για τοπικές επισκευές (**retouching**).

2. **Αποκάλυψη της Καλλιτεχνικής Τεχνικής:** Η υψηλή ανάλυση επέτρεψε τη μετάβαση από την απλή ταυτοποίηση υλικών στον χαρακτηρισμό της τεχνικής. Έγινε δυνατή η απεικόνιση μεμονωμένων πινελιών, η διάκριση μεταξύ διαφορετικών τρόπων εφαρμογής της ίδιας χρωστικής (π.χ. παχιές στρώσεις έναντι αναμειγξέων για φωτοσκιάσεις), και η λεπτομερής χαρτογράφηση της χρυσοκονδυλιάς.

3. **Ανίχνευση Σπάνιων Υλικών:** Η αυξημένη ποιότητα των δεδομένων ήταν καθοριστική για την ανίχνευση του Υδραργύρου (**Hg**), και συνεπώς την ταυτοποίηση της πολύτιμης χρωστικής κιννάβαρι, ένα εύρημα που δεν ήταν εφικτό στην αρχική ανάλυση χαμηλής ανάλυσης.

Η ποσοτική ανάλυση και η μελέτη των συσχετίσεων μεταξύ των στοιχείων παρέχουν την τελική επιβεβαίωση των υποθέσεων. Από την ανάλυση των φασμάτων, ταυτοποιήθηκαν τα παρακάτω στοιχεία, με τις αποκλίσεις μεταξύ των θεωρητικών και των πειραματικών κορυφών να βρίσκονται εντός αποδεκτών ορίων.

Προχωρούμε και στην εύρεση των αντίστοιχων συγκεντρώσεων κάθε προσμίξεως σ όλη την εικόνα.

Τα αποτελέσματα δείχνουν πως το Θειούχο ασβέστιο βρίσκεται σ' όλη την εικόνα και σε μεγαλύτερη συγκέντρωση σε σχέση με τις άλλες προσμίξεις. Αποτέλεσμα που βρίσκεται σε συμφωνία με τις τεχνικές κατασκευής των εικόνων της εκάστοτε εποχής.

Η ανάλυση των φασμάτων επιβεβαίωσε την παρουσία των βασικών στοιχείων. Οι αποκλίσεις μεταξύ των θεωρητικών (βιβλιογραφικών) και των πειραματικών ενεργειακών κορυφών βρίσκονται εντός αποδεκτών ορίων, λαμβάνοντας υπόψη τη διακριτική ικανότητα του ανιχνευτή και τις πιθανές φασματικές επικαλύψεις.

Στοιχείο	Γραμμή Ακτίνων- X (keV)	Εντοπισμένη Κορυφή (keV)	Παρατηρήσεις
S	2.307 (Kα)	~2.3	Επιβεβαίωση παρουσίας CaSO ₄ (Γύψος)
Ca	3.690 (Kα)	~3.7	Κύριο συστατικό του gesso, ισχυρή συσχέτιση με S
Fe	6.404 (Kα)	~6.4	Ταυτοποίηση γήινων χρωστικών (ώχρες)
Cu	8.048 (Kα)	~8.0	Χρωστική με βάση τον Χαλκό (π.χ. αζουρίτης)
Zn	8.638 (Kα)	~8.6	Ένδειξη μεταγενέστερων επεμβάσεων (ZnO)
Au	9.713 (Kα)	~9.7	Ταυτοποίηση φύλλου χρυσού

Hg	9.989 (La)	~10.0	Ταυτοποίηση Κιννάβαρι (HgS)
Pb	10.551 (La)	~10.5	Ταυτοποίηση Λευκού του Μολύβδου

Ποσοτική Ανάλυση Φάσεων:

Η ανάλυση επιβεβαιώνει την κυριαρχία του στρώματος προετοιμασίας. Το **Θεικό Ασβέστιο (CaSO₄)** αποτελεί την κυρίαρχη φάση, καλύπτοντας το μεγαλύτερο ποσοστό της αναλυόμενης επιφάνειας. Η δεύτερη κυρίαρχη φάση είναι πλούσια σε Ασβέστιο με προσμίξεις μετάλλων, υποδεικνύοντας τις περιοχές όπου το *gesso* καλύπτεται από τα ζωγραφικά στρώματα. Οι φάσεις που περιέχουν οξείδια σιδήρου και αλουμοπυριτικά υλικά αντιστοιχούν στις χρωστικές (ώχρες, πηλώδη υλικά) και αποτελούν μικρότερο ποσοστό, όπως αναμένεται.

Φάση	Κύρια Στοιχεία	Ποσοστό Επιφάνειας (%)	Ερμηνεία
Φάση 1	Ca, S, O	~40%	Καθαρός Γύψος/Ανυδρίτης (Gesso)
Φάση 2	Ca, O, S + (Ti, Cr, Mn)	~40%	Gesso καλυμμένο από/αναμεμειγμένο με χρωστικές
Φάση 3	Al, Si, K + (Fe, Ca)	~15%	Αλουμοπυριτικά (πηλώδη υλικά, υπόστρωμα)
Φάση 4	Fe, Cr, O + (Ca)	~5%	Οξείδια Σιδήρου-Χρωμίου (Ωχρες)

Συσχετίσεις Υλικών:

Η μελέτη των συσχετίσεων **Pearson** επιβεβαιώνει τις φυσικές σχέσεις μεταξύ των υλικών. Για παράδειγμα, η ισχυρή αρνητική συσχέτιση (π.χ., $r \approx -0.7$) μεταξύ των πυριτικών υλικών (**Si-πλούσια**, πιθανόν από τον ξύλινο φορέα) και του γύψου (**Ca-πλούσια**) υποδηλώνει ότι αυτά τα δύο υλικά δεν συνυπάρχουν, αλλά αλληλοαποκλείονται χωρικά, κάτι που είναι απολύτως αναμενόμενο.

Σύγκριση Μεθόδων & Τελικό Συμπέρασμα

Η ολοκλήρωση των δύο μελετών επιτρέπει μια σαφή σύγκριση μεταξύ μιας τυπικής ανάλυσης και της προηγμένης μεθοδολογίας που αναπτύχθηκε:

- Η ανάλυση "χαμηλής ανάλυσης" ήταν ικανή να ταυτοποιήσει τα **κύρια υλικά** της παλέτας (**Pb, Fe, Au, Ca, S**) και να δώσει μια γενική εικόνα της κατανομής τους. Ωστόσο, η ικανότητά της περιοριζόταν στον απλό χαρακτηρισμό.

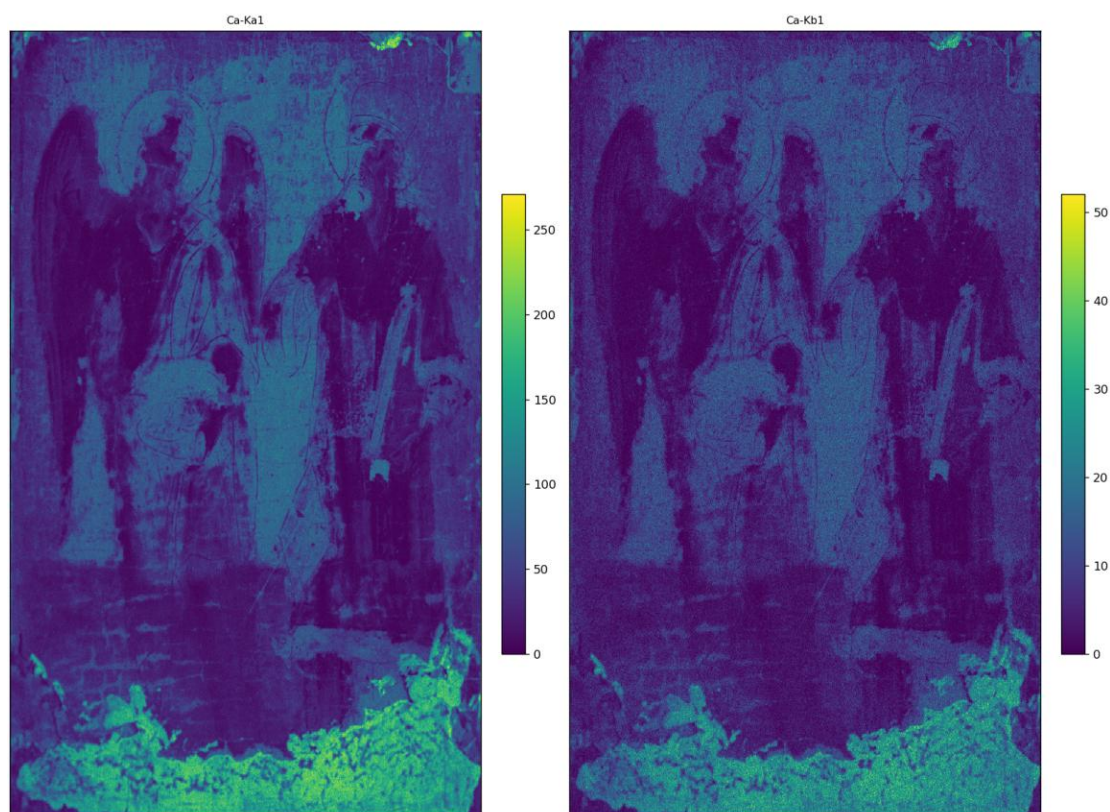
- Η ανεπτυγμένη μεθοδολογία υψηλής ανάλυσης προχώρησε πολύ πέρα από αυτό. Η αυξημένη ευαισθησία επέτρεψε την **ανίχνευση σπάνιων υλικών (HgS)**, η υψηλή χωρική ανάλυση αποκάλυψε την **καλλιτεχνική τεχνική** (πινελιές, στρώσεις), και η ισχύς των αλγορίθμων (*NMF*, *sub-clustering*) επέτρεψε την ψηφιακή ανασύνθεση της ζωγραφικής διαδικασίας.

Συνολικά, η μελέτη αυτή αποδεικνύει ότι ο συνδυασμός δεδομένων *MA-XRF* υψηλής ποιότητας με μια εξελιγμένη, πολυεπίπεδη στατιστική ανάλυση μπορεί να μετατρέψει μια τυπική εξέταση υλικών σε ένα πανίσχυρο εργαλείο για την ιστορία της τέχνης και τη συντήρηση, αποκαλύπτοντας όχι μόνο "από τι" είναι φτιαγμένο ένα έργο, αλλά και "πώς" ακριβώς δημιουργήθηκε.

Συγκριτική Ανάλυση και Επικύρωση

Σε αυτό το κρίσιμο τελικό στάδιο, πραγματοποιείται η άμεση συγκριτική ανάλυση και επικύρωση των στοιχειακών χαρτών που εξήχθησαν από το ανεπτυγμένο λογισμικό έναντι των δεδομένων αναφοράς του εργαστηριακού εξοπλισμού. Η ανάλυση αυτή έχει διττό στόχο: αφενός, την ποσοτική και ποιοτική επικύρωση της ακρίβειας των αλγορίθμων που αναπτύχθηκαν, και αφετέρου, την ανάδειξη της επιστημονικής αξίας που προσφέρει η μετάβαση από δεδομένα χαμηλής σε υψηλή χωρική ανάλυση.

Για το Ασβέστιο (*Ca*):

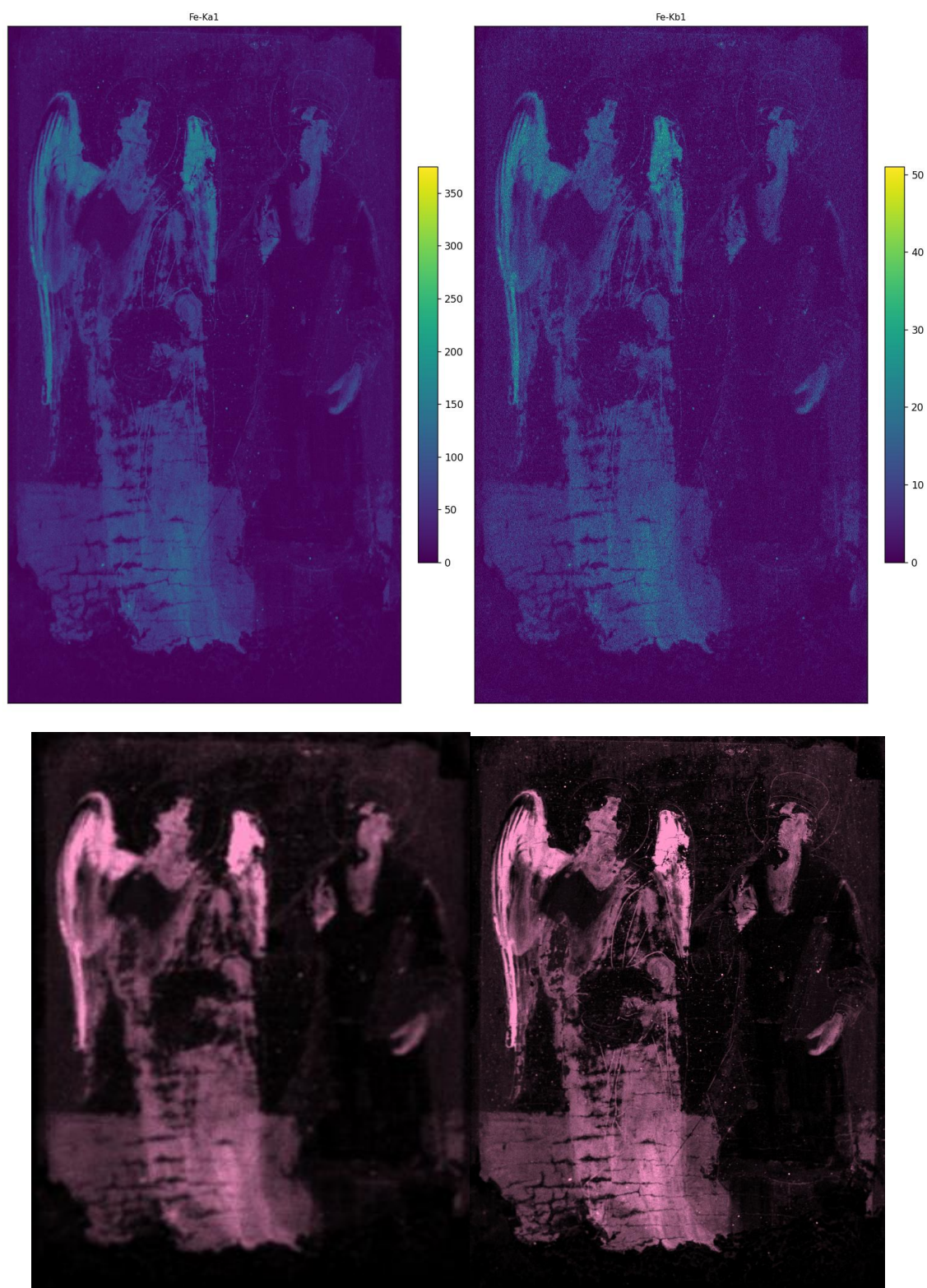




Εικόνα 112: Σύγκριση του στοιχειακού χάρτη για το Ασβέστιο (Ca-Kα) μεταξύ της ανάλυσης χαμηλής (αριστερά) και υψηλής (δεξιά) ανάλυσης. Η υψηλότερη ανάλυση αποκαλύπτει με μεγαλύτερη λεπτομέρεια τις φθορές και τη δομή του υποστρώματος.

Οι στοιχειακοί χάρτες φαίνεται πως έχουν αρκετή διαφορά μεταξύ χαμηλής και υψηλής ανάλυσης. Βεβαία η πραγματική αξία προκύπτει από τη **συνδυαστική ανάγνωση** των δύο τύπων απεικόνισης. Ενώ μεμονωμένα παρέχουν συγκεκριμένες πληροφορίες, μαζί δημιουργούν μια ολοκληρωμένη εικόνα της δομής της εικόνας. Όσον αφορά την διαφορά μεταξύ της εργαστηριακής ανάλυσης Εικόνα 112^β και του ανεπτυγμένου προγράμματος Εικόνα 112^α βλέπουμε πως έχουμε αρκετά μεγάλη συμφωνία στα αποτελέσματα τόσο στην χαμηλή όσο και στην υψηλή συχνότητα. Η ίδια διαδικασία ακολουθήθηκε και για τους υπολοίπους στοιχειακούς χάρτες με ακριβώς τα ίδια αποτελέσματα όπως αυτά εικονίζονται παρακάτω.

Για τον Σίδηρο (**Fe**):

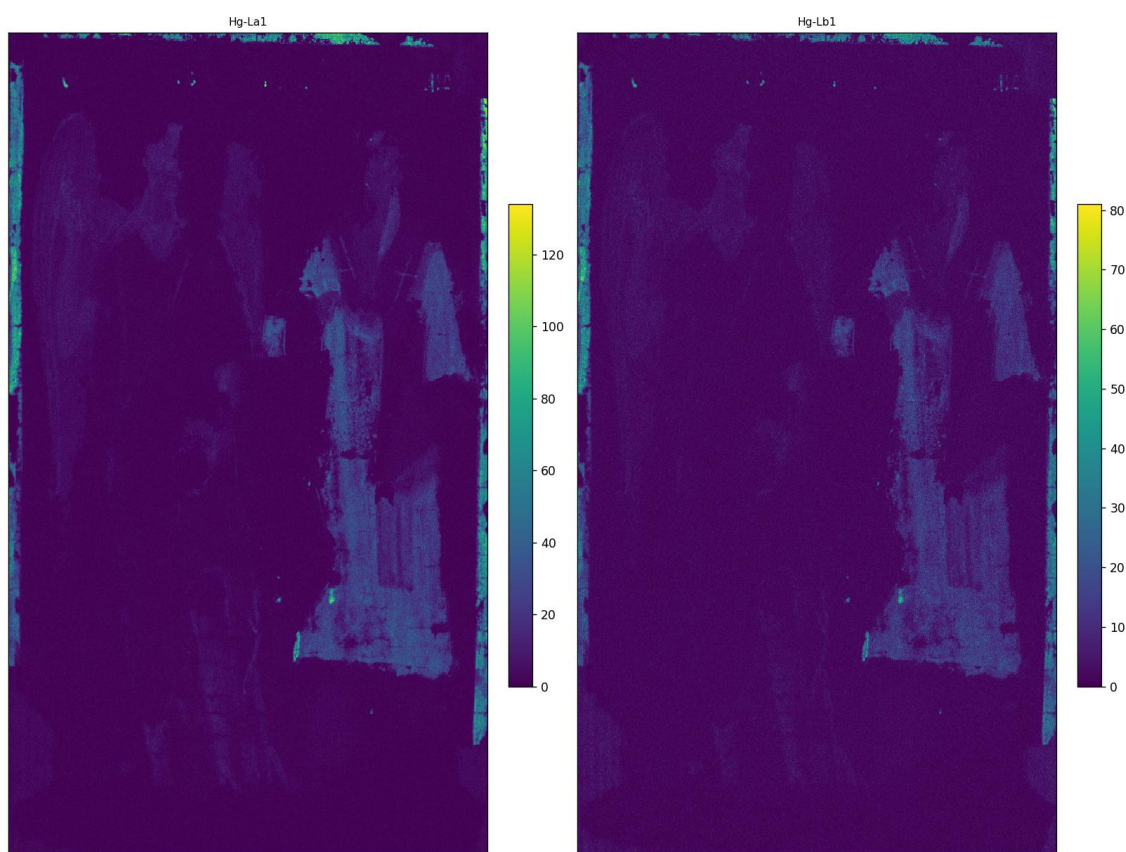


Εικόνα 113: Συγκριτικοί στοιχειακοί χάρτες για τον Σίδηρο (Fe). (Πάνω) Αποτελέσματα από το αναπτυγμένο λογισμικό. (Κάτω) Αποτελέσματα από το εργαστηριακό λογισμικό.

Η κατανομή του Σιδήρου (**Fe**) συνδέεται άμεσα και σχεδόν αποκλειστικά με τις καφε-κόκκινες και ωχρές περιοχές των ενδυμάτων των Αγίων. Αυτό το εύρημα επιβεβαιώνει την εκτεταμένη χρήση γήινων χρωστικών με βάση τα οξείδια του σιδήρου, όπως οι

ώχρες (π.χ., Fe_2O_3), η Σιένα ή η όμπρα, οι οποίες αποτελούσαν βασικό και σταθερό στοιχείο της παραδοσιακής παλέτας της αιογραφίας. Η σύγκριση των ευκρινών χαρτών **Fe-Ka** με τον χάρτη αναφοράς του εργαστηρίου αποδεικνύει την **απόλυτη ταύτιση**. Η γεωμετρία, η σχετική ένταση και η χωρική κατανομή είναι πανομοιότυπες, γεγονός που **επικυρώνει την ακρίβεια και την αξιοπιστία του αλγορίθμου** εξαγωγής στοιχειακών χαρτών. Ο **Χάρτης Χαμηλής Ανάλυσης** ταυτοποιεί σωστά τις περιοχές των ενδυμάτων, αλλά τις παρουσιάζει ως μια σχετικά ομοιογενή, θολή μάζα. Επιβεβαιώνει τι υλικό χρησιμοποιήθηκε (ώχρες), αλλά δίνει ελάχιστες πληροφορίες για το πώς χρησιμοποιήθηκε. **Χάρτης Υψηλής Ανάλυσης** (Εικόνα 119 δεξιά): Το αποτέλεσμα εδώ είναι μεταμορφωτικό. Ο χάρτης δεν δείχνει απλώς "πού υπάρχει σίδηρος". Αποκαλύπτει την υφή της πινελιάς, τις λεπτές διακυμάνσεις στην πυκνότητα του χρώματος και τη μέθοδο με την οποία ο καλλιτέχνης "έπλασε" τις πτυχώσεις των υφασμάτων για να δώσει όγκο, φως και ρεαλισμό. Γίνεται ορατή η κατεύθυνση και το πάχος της κάθε πινελιάς, μετατρέποντας τον χάρτη σε ένα παράθυρο στην ίδια τη δημιουργική διαδικασία. Επιπλέον, η παρατήρηση μιας ασθενέστερης, αλλά υπαρκτής, συγκέντρωσης σιδήρου και στις περιοχές των σαρκωμάτων επιβεβαιώνει την τεχνική της ανάμειξης μικρής ποσότητας ώχρας με το λευκό του μολύβδου για την επίτευξη της θερμής απόχρωσης του δέρματος, μια πρακτική που τεκμηριώνεται από την πολυμεταβλητή ανάλυση (sub-clustering).

Για τον υδράργυρο (**Hg**):





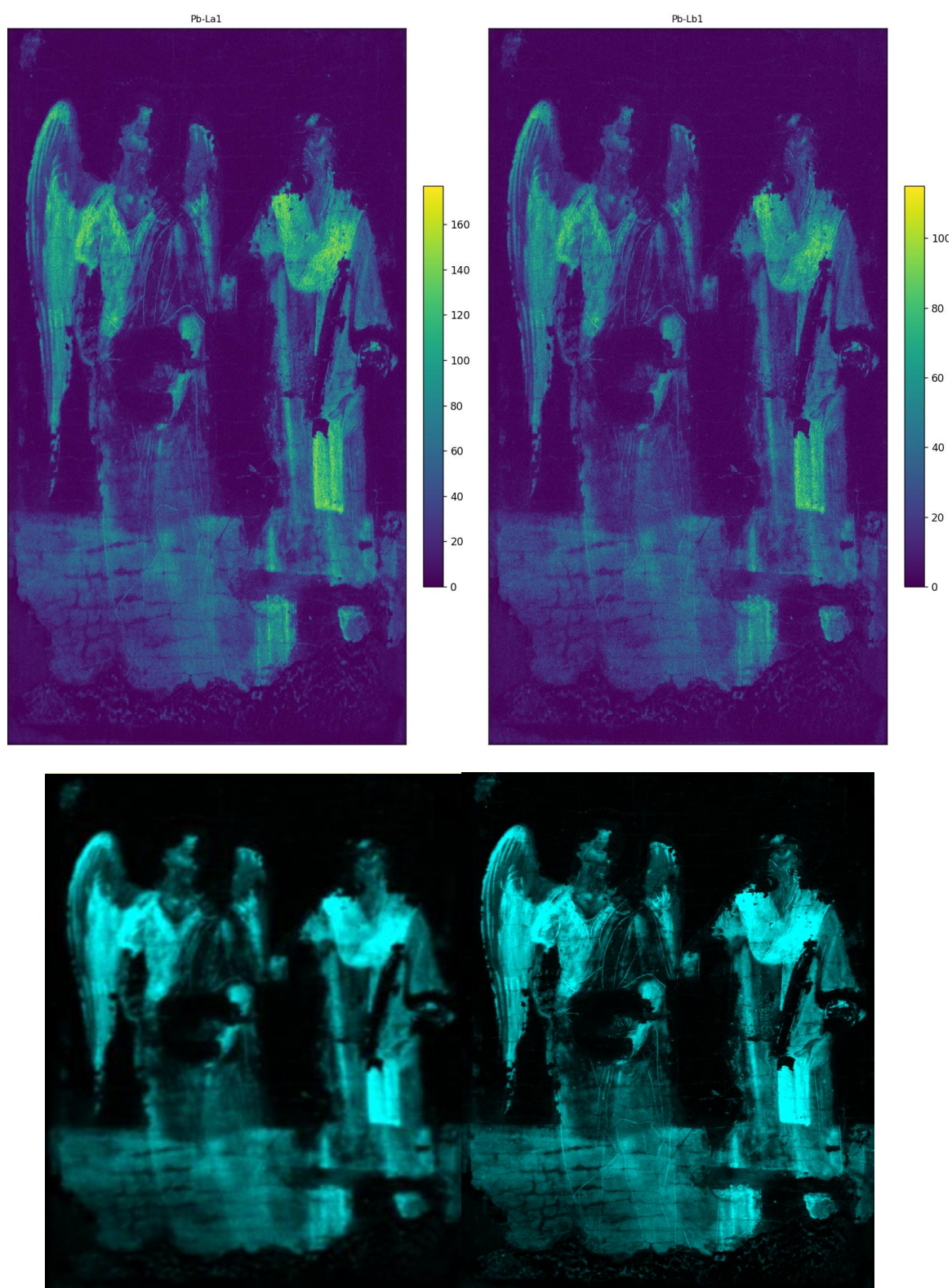
Εικόνα 114: Συγκριτικοί στοιχειακοί χάρτες για τον Υδραργύρο (**Hg**) (Πάνω) Αποτελέσματα από το αναπτυσσόμενο λογισμικό. Σύγκριση μεταξύ χαμηλής και υψηλής ανάλυσης (Κάτω) από το εργαστηριακό λογισμικό.

Η ανίχνευση και η χαρτογράφηση του Υδραργύρου (**Hg**) αποτελεί μία από τις πιο σημαντικές και αποκαλυπτικές ανακαλύψεις της παρούσας μελέτης. Η παρουσία του στοιχείου αυτού ταυτοποιεί τη χρήση του κιννάβαρι (**HgS**), μιας πολύτιμης, λαμπερής και ιστορικά σημαντικής κόκκινης χρωστικής, η οποία χρησιμοποιούνταν επιλεκτικά για σημεία ιδιαίτερης θεολογικής ή καλλιτεχνικής έμφασης. **Γραμμές L-shell (Hg-La1, Hg-La2, Hg-Lβ1):** Σε πλήρη αντίθεση, οι χάρτες αυτοί παρουσιάζουν ένα ισχυρό και σαφές σήμα. Η χωρική τους κατανομή είναι απόλυτα συνεπής, αποκαλύπτοντας την παρουσία του υδραργύρου σε πολύ συγκεκριμένες περιοχές. Η σύγκριση των ευκρινών χαρτών **L-shell** του λογισμικού με τον χάρτη αναφοράς του εργαστηρίου αποδεικνύει την **απόλυτη ταύτιση** στη χωρική κατανομή. Αυτό το γεγονός **επικυρώνει την ικανότητα του λογισμικού να ανιχνεύει και να χαρτογραφεί με ακρίβεια ακόμα και στοιχεία που υπάρχουν σε περιορισμένες ποσότητες**. Χάρτης Χαμηλής Ανάλυσης (αριστερά): Το σήμα είναι εξαιρετικά ασθενές. Ο λόγος σήματος προς θόρυβο (**SNR**) είναι πολύ χαμηλός. Αν και διακρίνονται κάποιες αχνές ενδείξεις, η ταυτοποίηση του **Hg** θα ήταν επισφαλής. Χάρτης Υψηλής Ανάλυσης (δεξιά): Το αποτέλεσμα είναι διαφορετικό. Ο χάρτης παρέχει ένα σαφές, ισχυρό και αδιαμφισβήτητο σήμα. Ο σταυρός, το περίγραμμα και οι λεπτομέρειες στο ένδυμα εμφανίζονται με απόλυτη ευκρίνεια.

Η ταυτοποίηση του κιννάβαρι προσφέρει πολύτιμες πληροφορίες για τις προθέσεις του καλλιτέχνη. Η επιλεκτική χρήση αυτής της ακριβής και λαμπερής χρωστικής, σε αντίθεση με τις φθηνότερες ώχρες σιδήρου, υποδηλώνει ότι ο αγιογράφος ήθελε να προσδώσει ιδιαίτερη σημασία και αξία στα συγκεκριμένα εικονογραφικά στοιχεία. Το εύρημα αυτό εξηγεί επίσης πλήρως γιατί οι περιοχές αυτές απομονώθηκαν ως μια μοναδική, διακριτή συστάδα (**Cluster 3**) στην ανάλυση **K-Means**: η μοναδική χημική τους υπογραφή (παρουσία **Hg**) τις κατέστησε στατιστικά διακριτές από όλες τις άλλες περιοχές της εικόνας.

Για τον Μόλυβδο (**Pb**):

Element Distribution Maps (2 maps shown)



Εικόνα 115: Συγκριτικοί στοιχειακοί χάρτες για τον **Μόλυβδο (Pb)** (Πάνω) Αποτελέσματα από το αναπτυγμένο λογισμικό. Σύγκριση μεταξύ χαμηλής και υψηλής ανάλυσης (Κάτω) από το εργαστηριακό λογισμικό.

Η χαρτογράφηση του Μολύβδου (**Pb**) παρέχει μία από τις πιο πλούσιες σε πληροφορία εικόνες για την τεχνική του καλλιτέχνη, ιδίως όσον αφορά την απόδοση του όγκου, του βάθους και του φωτός. Η παρουσία του στοιχείου αυτού ταυτοποιεί αδιαμφισβήτητα τη χρήση του **Λευκού του Μολύβδου (Lead White, $2\text{PbCO}_3 \cdot \text{Pb(OH)}_2$)**, της βασικής λευκής χρωστικής που χρησιμοποιήθηκε για αιώνες στην τέχνη λόγω της εξαιρετικής καλυπτικότητας και της λαμπρότητάς της. **Γραμμές L-shell (Pb-La1, Pb-La2, Pb-Lβ1)**: Οι χάρτες αυτοί παρουσιάζουν ένα εξαιρετικά ισχυρό, ευκρινές και λεπτομερές σήμα. Η χωρική τους κατανομή είναι απόλυτα συνεπής, χαρτογραφώντας με υψηλή πιστότητα τις περιοχές των **σαρκωμάτων (πρόσωπα, χέρια)**, τα **φωτίσματα (highlights)** στα ενδύματα και τα φτερά, καθώς και το **έδαφος** στο κάτω μέρος της εικόνας. Η σύγκριση των ευκρινών χαρτών L-shell του λογισμικού με τον χάρτη του εργαστηρίου αποδεικνύει την **απόλυτη χωρική και ποιοτική ταύτιση**. Και οι δύο απεικονίσεις αποκαλύπτουν τα ίδια μορφολογικά χαρακτηριστικά με την ίδια σχετική ένταση, γεγονός που **επικυρώνει την ακρίβεια του αλγορίθμου** στην επεξεργασία του σήματος του Μολύβδου.

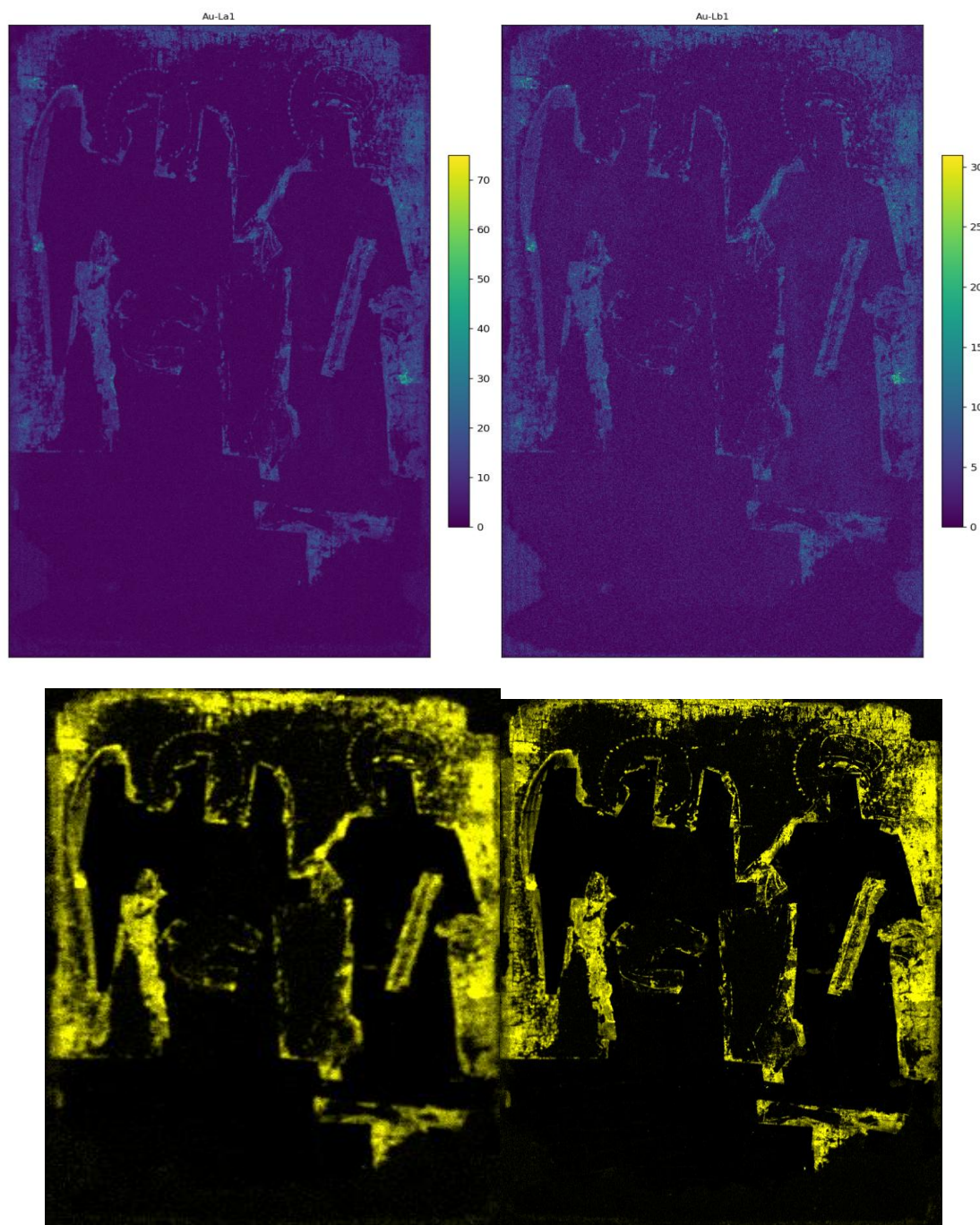
Χάρτης Χαμηλής Ανάλυσης: Ο χάρτης αυτός ταυτοποιεί σωστά τις κύριες περιοχές χρήσης του Λευκού του Μολύβδου. Τα σαρκώματα και τα φωτίσματα είναι ορατά, αλλά εμφανίζονται ως ενιαίες, θολές περιοχές, χωρίς εσωτερική δομή. **Χάρτης Υψηλής Ανάλυσης (δεξιά)**: Ο χάρτης δεν δείχνει απλώς τις περιοχές του μολύβδου, αλλά αποκαλύπτει τη μικροδομή της ζωγραφικής επιφάνειας. Γίνονται ορατές οι μεμονωμένες, λεπτές πινελιές που χρησιμοποίησε ο καλλιτέχνης για να δημιουργήσει τα τελικά φωτίσματα – οι περίφημες "ψιμυθιές" της βυζαντινής αγιογραφίας. Η τεχνική του "πλασίματος" των μορφών με διαδοχικές, όλο και πιο φωτεινές στρώσεις γίνεται ορατή σε μια κλίμακα που προσεγγίζει την παρατήρηση με μικροσκόπιο.

Η λεπτομερής χαρτογράφηση του Μολύβδου επιβεβαιώνει την εφαρμογή της κλασικής βυζαντινής τεχνικής. Ο καλλιτέχνης χρησιμοποίησε το Λευκό του Μολύβδου με δύο κύριους τρόπους:

1. **Ως βασική χρωστική**: Αναμεμειγμένο με ώχρες (**Fe**) για τη δημιουργία της βασικής απόχρωσης του δέρματος ("προπλασμός").
2. **Ως καθαρή χρωστική**: Εφαρμοσμένο σε λεπτές, τελικές πινελιές για τη δημιουργία των φωτισμάτων που δίνουν την αίσθηση του θεϊκού φωτός και του όγκου.

Η καθαρότητα αυτών των χαρτών επιβεβαιώνει οπτικά και με απόλυτη ακρίβεια τα συμπεράσματα που προέκυψαν από την πολυμεταβλητή ανάλυση **sub-clustering**, όπου οι διαφορετικές εφαρμογές του Λευκού του Μολύβδου (βασική στρώση, φωτίσματα, σκιές) ομαδοποιήθηκαν σε ξεχωριστές υπο-συστάδες.

Για τον Χρυσό (**Au**):



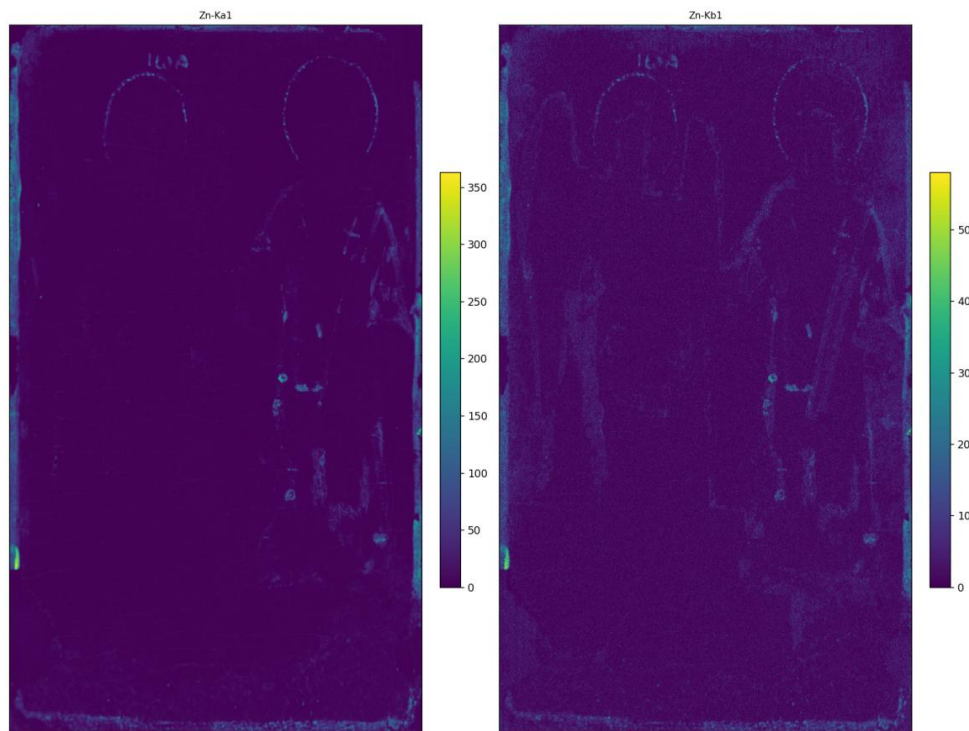
Εικόνα 116: Συγκριτικοί στοιχειακοί χάρτες για τον Χρυσό (**Au**) (Πάνω) Αποτελέσματα από το αναπτυγμένο λογισμικό. Σύγκριση μεταξύ χαμηλής και υψηλής ανάλυσης (Κάτω) από το εργαστηριακό λογισμικό.

Η χαρτογράφηση του Χρυσού (**Au**) αποκαλύπτει τα πιο πολύτιμα και συμβολικά σημαντικά τμήματα της εικόνας. Η παρουσία του στοιχείου αυτού ταυτοποιεί την τεχνική της **επιχρύσωσης**, πιθανότατα με τη χρήση λεπτών φύλλων χρυσού (**gold leaf**). Στη βυζαντινή παράδοση, ο χρυσός δεν αποτελεί απλώς διακοσμητικό στοιχείο, αλλά συμβολίζει το Θείο, άκτιστο φως, προσδίδοντας στο έργο μια υπερβατική διάσταση. Γραμμές **L-shell** (**Au-La1**, **Au-La2**, **Au-Lb1**): οι χάρτες αυτοί παρέχουν ένα ισχυρό, ευκρινές και λεπτομερές σήμα. Η χωρική τους κατανομή είναι απόλυτα συνεπής και αποκαλύπτει την παρουσία του χρυσού αποκλειστικά σε δύο τύπους

περιοχών: στα φωτοστέφανα (*halos*) των Αγίων και στις εξαιρετικά λεπτές, καλλιγραφικές διακοσμητικές γραμμές πάνω στα ενδύματα, μια τεχνική γνωστή ως χρυσοκονδυλιά (*chrysography*). Η σύγκριση των ευκρινών χαρτών L-shell του λογισμικού με τον χάρτη αναφοράς του εργαστηρίου αποδεικνύει την **απόλυτη χωρική και ποιοτική ταύτιση**. Και οι δύο απεικονίσεις απομονώνουν τις ίδιες περιοχές με την ίδια ακρίβεια, **επικυρώνοντας την ικανότητα του λογισμικού να χαρτογραφεί με επιτυχία τα ευγενή μέταλλα**. Χάρτης Χαμηλής Ανάλυσης (αριστερά): Ο χάρτης αυτός ταυτοποιεί σωστά την παρουσία του χρυσού στα φωτοστέφανα και στο γενικό "*campus*" (φόντο) της εικόνας. Ωστόσο, οι λεπτές γραμμές της χρυσοκονδυλιάς είναι είτε αόρατες είτε εμφανίζονται ως θολές, ασυνεχείς κηλίδες. Η πληροφορία είναι κυρίως ποιοτική. Χάρτης Υψηλής Ανάλυσης (δεξιά): Ο χάρτης δεν δείχνει απλώς "πού υπάρχει χρυσός", αλλά αποκαλύπτει την καλλιγραφική ποιότητα της γραμμής. Κάθε χρυσοκονδυλιά γίνεται ορατή ως μια συνεχής, σταθερή πινελιά. Είναι πλέον δυνατή η μελέτη της δεξιοτεχνίας και της σταθερότητας του χεριού του καλλιτέχνη. Επιπλέον, γίνονται ορατές οι μικρο-φθορές και οι απώλειες στο φύλλο χρυσού, παρέχοντας πολύτιμες πληροφορίες για την κατάσταση διατήρησης των πιο ευαίσθητων και πολύτιμων τμημάτων του έργου.

Η λεπτομερής χαρτογράφηση του Χρυσού επιβεβαιώνει τη χρήση δύο βασικών τεχνικών επιχρύσωσης: την εφαρμογή φύλλου χρυσού για τις μεγάλες επιφάνειες των φωτοστεφάνων και την τεχνική της χρυσοκονδυλιάς για τις διακοσμητικές λεπτομέρειες. Η καθαρότητα των χαρτών επιβεβαιώνει οπτικά τα αποτελέσματα της πολυμεταβλητής ανάλυσης, όπου οι επιχρυσωμένες περιοχές ομαδοποιήθηκαν σε μια μοναδική, στατιστικά διακριτή συστάδα (*Cluster 4*). Η ικανότητα να διακρίνουμε την ποιότητα και την κατάσταση διατήρησης αυτών των λεπτών γραμμών είναι τεράστιας σημασίας τόσο για τους ιστορικούς της τέχνης όσο και για τους συντηρητές.

Για τον Ψευδάργυρο (*Zn*):





Εικόνα 117: Συγκριτικοί στοιχειακοί χάρτες για τον Ψευδαργύρο (Zn) (Πάνω) Αποτελέσματα από το αναπτυγμένο λογισμικό. Σύγκριση μεταξύ χαμηλής και υψηλής ανάλυσης (Κάτω) από το εργαστηριακό λογισμικό.

Η ανίχνευση του Ψευδαργύρου (Zn) αποτελεί ένα εύρημα με ιδιαίτερη σημασία για την ιστορία του έργου, καθώς το στοιχείο αυτό δεν ανήκει στην παραδοσιακή παλέτα της βυζαντινής αιογραφίας. Η παρουσία του ταυτοποιεί τη χρήση του **Λευκού του Ψευδαργύρου (Zinc White, ZnO)**, μιας χρωστικής που έγινε ευρέως εμπορικά διαθέσιμη από τα μέσα του 19ου αιώνα. Συνεπώς, ο χάρτης του ψευδαργύρου δεν αποκαλύπτει την αρχική ζωγραφική, αλλά λειτουργεί ως ένας **χάρτης των μεταγενέστερων επεμβάσεων συντήρησης και επιζωγραφίσεων (retouching)**. **Γραμμές K-shell (Zn-Kα1, Zn-Kα2, Zn-Kβ1):** Οι χάρτες αυτοί παρουσιάζουν ένα ισχυρό και ευκρινές σήμα, παρά τη σποραδική φύση της κατανομής. Η χωρική τους συνέπεια επιβεβαιώνει την αξιόπιστη ανίχνευση του στοιχείου.

- **Η κατανομή είναι εξαιρετικά τοπική και περιλαμβάνει:**

1. Μια **επιγραφή ("ΙωΑ" - Ιωάννης)** πάνω από το φωτοστέφανο του αριστερού Αγίου.
2. **Λεπτές γραμμές που οριοθετούν τα φωτοστέφανα.**
3. Ένα **περίγραμμα** γύρω από την εικόνα.
4. Διάσπαρτες **κηλίδες** κυρίως πάνω στη μορφή του δεξιού Αγίου.

Η σύγκριση των χαρτών του λογισμικού με τον χάρτη αναφοράς του εργαστηρίου αποδεικνύει την **απόλυτη και εντυπωσιακή ταύτιση**. Κάθε λεπτομέρεια, από τις καλλιγραφικές γραμμές της επιγραφής μέχρι την παραμικρή κηλίδα, αποτυπώνεται με την ίδια ακρίβεια. Αυτό το γεγονός **επικυρώνει την ικανότητα του λογισμικού να εντοπίζει και να χαρτογραφεί με υψηλή πιστότητα ακόμα και σήματα χαμηλής έντασης με σποραδική κατανομή**. **Χάρτης Χαμηλής Ανάλυσης (αριστερά):** Ο χάρτης αυτός εντοπίζει την παρουσία του ψευδαργύρου, αλλά οι λεπτομέρειες είναι θολές. Η επιγραφή "ΙωΑ" είναι σχεδόν δυσανάγνωστη και θα μπορούσε να ερμηνευτεί ως τυχαίες κηλίδες. Οι λεπτές γραμμές των φωτοστεφάνων είναι ασυνεχείς. **Χάρτης Υψηλής Ανάλυσης (δεξιά):** Ο χάρτης δεν δείχνει απλώς "κηλίδες", αλλά αποκαλύπτει

την **πρόθεση πίσω από την επέμβαση**. Η επιγραφή "**ΙωΑ**" είναι πλέον απόλυτα ευανάγνωστη, αποκαλύπτοντας μια συνειδητή προσπάθεια ταυτοποίησης του Αγίου Ιωάννη. Οι γραμμές των φωτοστεφάνων είναι συνεχείς και καλλιγραφικές. Η υψηλή ανάλυση μας επιτρέπει να διακρίνουμε τη φύση της επέμβασης: δεν πρόκειται για άτεχνες επιδιορθώσεις, αλλά για μια λεπτομερή και σκόπιμη επιζωγράφιση.

Συμπερασματικά θα έχουμε πως:

- **Low vs. High Resolution:** Η χαμηλή ανάλυση ήταν επαρκής για τον γενικό χαρακτηρισμό της παλέτας. Η υψηλή ανάλυση, ωστόσο, ήταν απαραίτητη για την κατανόηση της τεχνικής, την ανίχνευση σπάνιων υλικών (κιννάβαρι) και τη μελέτη της κατάστασης διατήρησης σε μικρο-κλίμακα.
- **PCA vs. NMF vs. NMF-PCA:** Η *PCA* αποδείχθηκε ένα ισχυρό εργαλείο για τη μείωση διαστάσεων. Η *NMF* παρείχε πιο ερμηνεύσιμα αποτελέσματα. Η **υβριδική προσέγγιση NMF-PCA** αναδείχθηκε ως η βέλτιστη, συνδυάζοντας τα πλεονεκτήματα και των δύο για την επίτευξη του μέγιστου διαχωρισμού και ευκρίνειας.

ΠΑΝΑΓΙΑ Η ΚΑΘΟΔΗΓΗΤΡΙΑ

Η παρούσα μελέτη αποσκοπεί στον ποσοτικό χαρακτηρισμό και τη χωρική ανάλυση



Εικόνα 118: Η φορητή εικόνα της 'Παναγίας της Οδηγήτριας' που αποτέλεσε το αντικείμενο της ανάλυσης MA-XRF. Η εικόνα παρουσιάζει χαρακτηριστικές φθορές και αποτελεί τυπικό δείγμα μεταβυζαντινής αγιογραφίας.

της χημικής σύστασης της φορητής εικόνας 'Παναγία η Καθοδηγήτρια' Εικόνα 118. Για τον σκοπό αυτό, αξιοποιήθηκαν προηγμένες στατιστικές μέθοδοι επί υπερφασματικών δεδομένων που συλλέχθηκαν με τη τεχνική της Φασματοσκοπίας Απεικόνισης Φθορισμού Ακτίνων-X (*Macro X-Ray Fluorescence, MA-XRF*). Τα δεδομένα, με αρχικές διαστάσεις 963×720 εικονοστοιχεία (*pixels*) και 4096 ενεργειακά κανάλια ανά *pixel*, υποβλήθηκαν σε πολυεπίπεδη ανάλυση με στόχο την εξαγωγή κρίσιμων πληροφοριών σχετικά με τη σύνθεση και τη δομή των υλικών κατασκευής του έργου τέχνης. Επιπρόσθετα στο τέλος της ανάλυσης παρέχεται σύγκριση με μελέτες που ήδη έχουν πραγματοποιηθεί με βάση την εικόνα αυτή.

Στην αρχική φάση της ανάλυσης, τα φασματικά δεδομένα MA-XRF υποβλήθηκαν σε προεπεξεργασία με την εφαρμογή φίλτρου

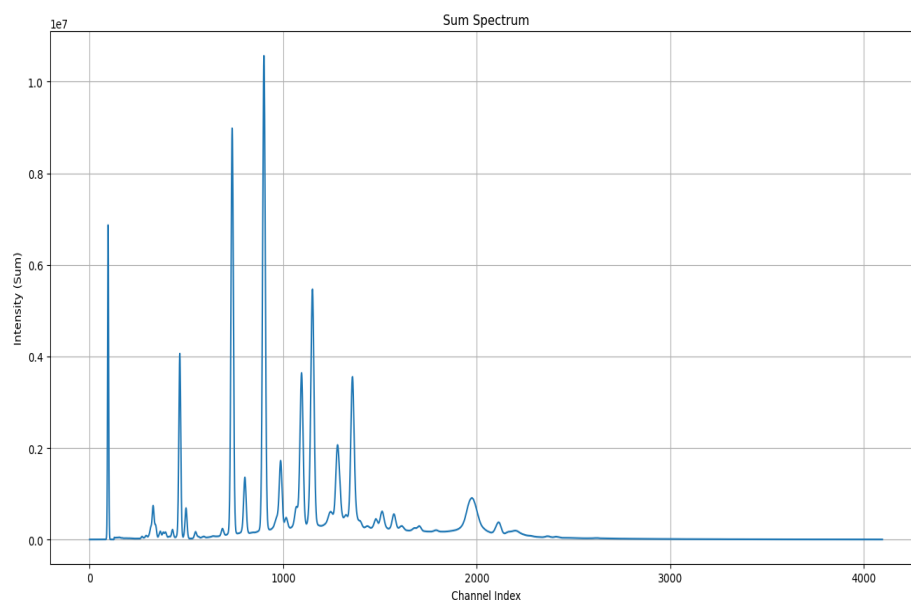
Savitzky-Golay. Η χρήση παραθύρου 11 σημείων και πολωνύμου τρίτης τάξης συνέβαλε στην εξομάλυνση των φασμάτων και τη μείωση του φασματικού θορύβου, βελτιστοποιώντας την ποιότητα των δεδομένων χωρίς σημαντική απώλεια πληροφορίας. Στη συνέχεια, τα προεπεξεργασμένα δεδομένα αναλύθηκαν διαδοχικά με τις χημειομετρικές μεθόδους της Ανάλυσης Κύριων Συνιστωσών (*Principal Component Analysis, PCA*) και της Μη Αρνητικής Παραγοντοποίησης Πινάκων (*Non-Negative Matrix Factorization, NMF*). Οι τεχνικές αυτές αποσκοπούν στον εντοπισμό και τον ποσοτικό χαρακτηρισμό των διακριτών χημικών φάσεων που απαρτίζουν το υλικό της υπό εξέταση εικόνας.

Η υπό εξέταση εικόνα είναι η 'Παναγία η Καθοδηγήτρια', για την οποία ιστορικές και τεχνοτροπικές πληροφορίες παρατέθηκαν στην εισαγωγή του πειραματικού κεφαλαίου. Οι μετρήσεις MA-XRF πραγματοποιήθηκαν με μέγεθος *pixel* $300 \mu\text{m}$ και χρόνο συλλογής 10 ms ανά *pixel*. Η πηγή ακτίνων-X λειτούργησε με τάση 50 kV και ρεύμα $600 \mu\text{A}$, χρησιμοποιώντας σωλήνα με άνοδο Ροδίου (*Rh*). Λεπτομερείς πληροφορίες σχετικά με τις παραμέτρους της πειραματικής διάταξης παρατίθενται παρακάτω Εικόνα 119.

Η αρχική επισκόπηση των φασματικών δεδομένων πραγματοποιήθηκε μέσω της δημιουργίας του φάσματος αθροίσματος (**Sum Spectrum**) Εικόνα 120, το οποίο υπολογίστηκε και απεικονίστηκε με χρήση του αναπτυχθέντος λογισμικού.

MAP INFORMATION		
Mapping parameters		
Width:	720	pixel
	216.073	mm
Height:	963	pixel
	289.471	mm
Pixel Size:	300	μm
Total number of pixel:	693360	pixel
Acquisition parameters		
Frame count:	1	
Pixel time:	10	ms/pixel
Measure time:	1:15	h
Overall time:	3:20	h
Stage speed:	30.0	mm/s
Stage position (X,Y,Z):	4.614;375.825;61.755	mm
Tube parameter		
High voltage:	50	kV
Anode current:	600	μA
Filter:	Empty	
Optic:	Lens	
Collimator diameter:	0	
SpotSize:	100	
Chamber at:	Air 1000	mbar
Flow rate:	---	l/min
Anode:	Rh	
Detector parameters		
Selected detectors:	1	
Max. pulse throughput:	275000	cps

Εικόνα 119: Πίνακας με τις πειραματικές παραμέτρους της σάρωσης MA-XRF για την εικόνα της Παναγίας. Περιλαμβάνει πληροφορίες για τις διαστάσεις της σάρωσης (*Mapping parameters*), τις συνθήκες απόκτησης δεδομένων (*Acquisition parameters*), τις ρυθμίσεις της λυχνίας ακτίνων-Χ (*Tube parameter*) και τις παραμέτρους του ανιχνευτή (*Detector parameters*).



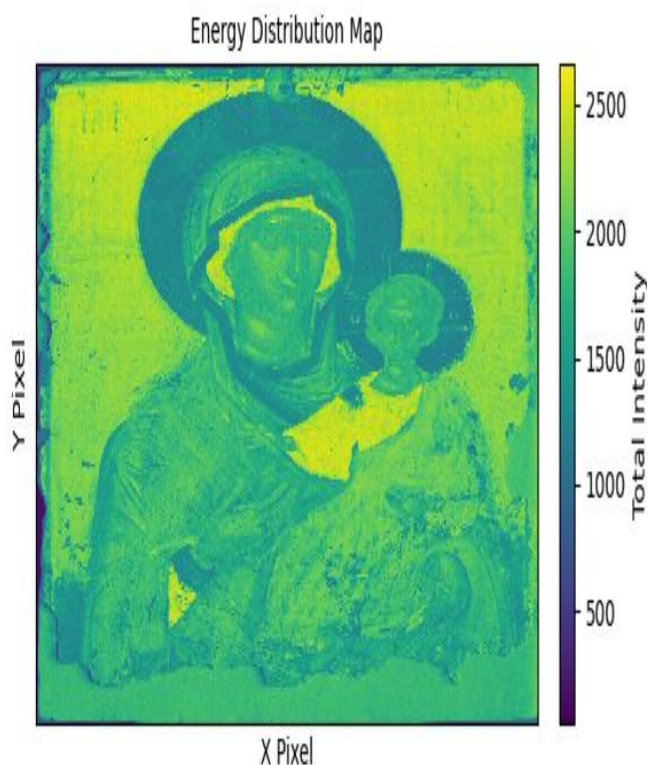
Εικόνα 120: Το φάσμα αθροίσματος (*Sum Spectrum*) των δεδομένων MA-XRF πριν από τη βαθμονόμηση. Ο οριζόντιος άξονας αντιστοιχεί στα κανάλια (*Channel Index*) του ανιχνευτή. Το φάσμα παρέχει μια αρχική, συνολική εικόνα της στοιχειακής σύστασης της εικόνας.

Το φάσμα αυτό, που αντιστοιχεί στο άθροισμα όλων των καταγεγραμμένων φασμάτων από κάθε **pixel** της εικόνας, παρουσιάζει σημαντικά μειωμένο θόρυβο λόγω της προηγηθείσας προεπεξεργασίας **Savitzky-Golay** Εικόνα 120.

Για την περαιτέρω ανάλυση των δεδομένων και την εξαγωγή ουσιαστικών πληροφοριών, εφαρμόστηκαν μέθοδοι Ανάλυσης Συνιστωσών (**Component Analysis**) και Ομαδοποίησης (**Clustering Analysis**). Οι αναλύσεις αυτές πραγματοποιήθηκαν στα δεδομένα σε γραμμική κλίμακα έντασης. Αν και η ανάλυση σε λογαριθμική κλίμακα (π.χ., **log(ένταση+1)**), όπως υποστηρίζεται από τον κώδικα) μπορεί να ενισχύσει σήματα από στοιχεία σε χαμηλές συγκεντρώσεις, για την παρούσα μελέτη προτιμήθηκε η γραμμική κλίμακα για την αμεσότερη συσχέτιση με τις σχετικές ποσότητες των υλικών. Οι κύριες στρατηγικές πολυμεταβλητής ανάλυσης που ακολουθήθηκαν ήταν:

5. Ανάλυση Κύριων Συνιστωσών (**PCA**) επί των αρχικών (δειγματοληπτημένων) δεδομένων.
6. Μη Αρνητική Παραγοντοποίηση Πινάκων (**NMF**) επί των αρχικών (δειγματοληπτημένων) δεδομένων.
7. Ανάλυση Ομαδοποίησης **K-Means** επί των αποτελεσμάτων (**scores**) της **PCA**.

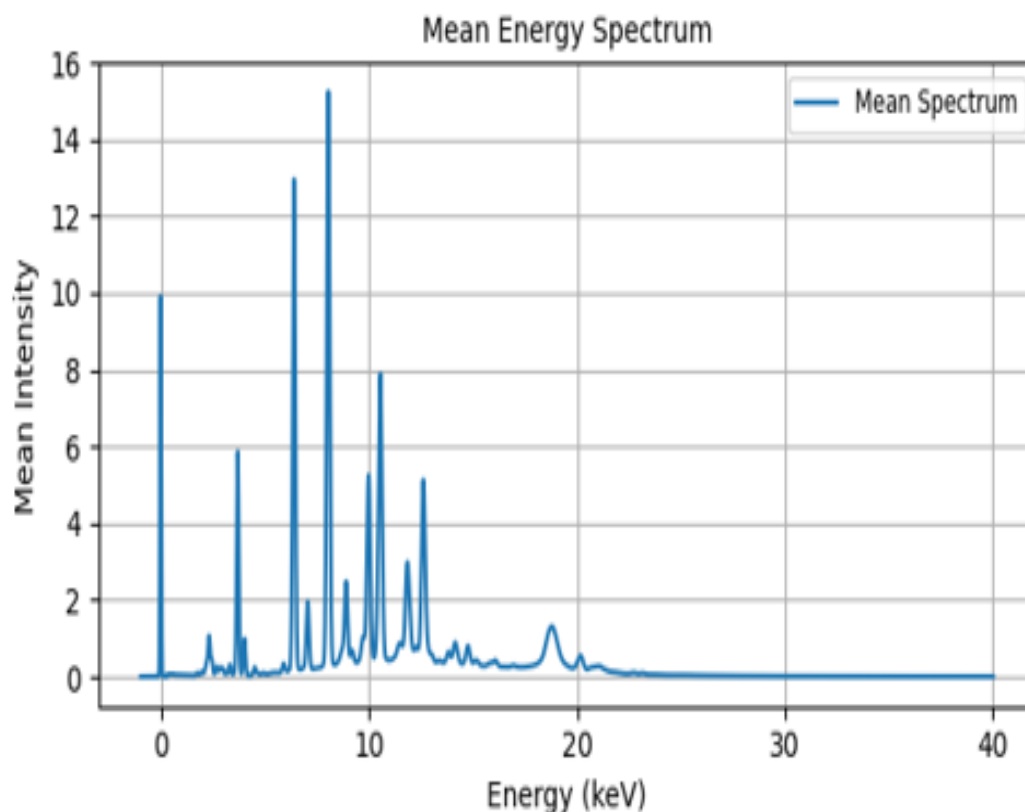
Η Εικόνα παρουσιάζει τον χάρτη της ολικής κατανομής της έντασης των υπερφασματικών δεδομένων Εικόνα 121. Ο χάρτης αυτός προκύπτει από την άθροιση της έντασης φθορισμού σε όλα τα ενεργειακά κανάλια για κάθε εικονοστοιχείο (**pixel**) της σαρωμένης περιοχής. Οι φωτεινότερες περιοχές στον χάρτη αντιστοιχούν σε υψηλότερη συνολική εκπομπή ακτίνων-X. Αυτό μπορεί να υποδηλώνει την παρουσία στοιχείων με υψηλότερο ατομικό αριθμό (**Z**), τα οποία τείνουν να φθορίζουν εντονότερα, ή/και μεγαλύτερη συγκέντρωση ή πάχος υλικού στις συγκεκριμένες περιοχές.



Εικόνα 121: Χάρτης συνολικής κατανομής έντασης (**Energy Distribution Map**). Οι φωτεινότερες περιοχές (κίτρινο) αντιστοιχούν σε υψηλότερη συνολική εκπομπή ακτίνων-X, υποδεικνύοντας την παρουσία στοιχείων με υψηλό ατομικό αριθμό (π.χ., **Pb**, **Au**) ή μεγαλύτερο πάχος υλικού.

Στην παρούσα εικόνα, παρατηρείται ότι οι περιοχές που αντιστοιχούν στα πρόσωπα της Παναγίας και του Ιησού, στα χέρια, καθώς και στα φωτοστέφανα, εμφανίζουν σημαντικά υψηλότερη συνολική ένταση σε σύγκριση με τις περιοχές των ενδυμάτων ή του φόντου. Η παρατήρηση αυτή υποδηλώνει την πιθανή χρήση πυκνότερων υλικών, στρωμάτων προετοιμασίας πλούσιων σε βαρέα στοιχεία, ή χρωστικών που περιέχουν στοιχεία υψηλού **Z** (π.χ., μόλυβδο, υδράργυρο, χρυσό) στις εν λόγω περιοχές. Η ακριβής φύση αυτών των υλικών θα διερευνηθεί περαιτέρω με τις επόμενες

φασματοσκοπικές και στατιστικές αναλύσεις. Ωστόσο σε σύγκριση με την εικόνα περιοχές όπως το φωτοστέφανο που είναι εμφανώς από χρυσό φαίνονται σκοτεινές πράγμα που υποδηλώνει πως το στοιχείο που βρίσκεται στο υπόβαθρο είναι κύριος μέτοχος του φάσματος και το στρώμα χρυσού σε εκείνη την περιοχή είναι αρκετά λεπτό.

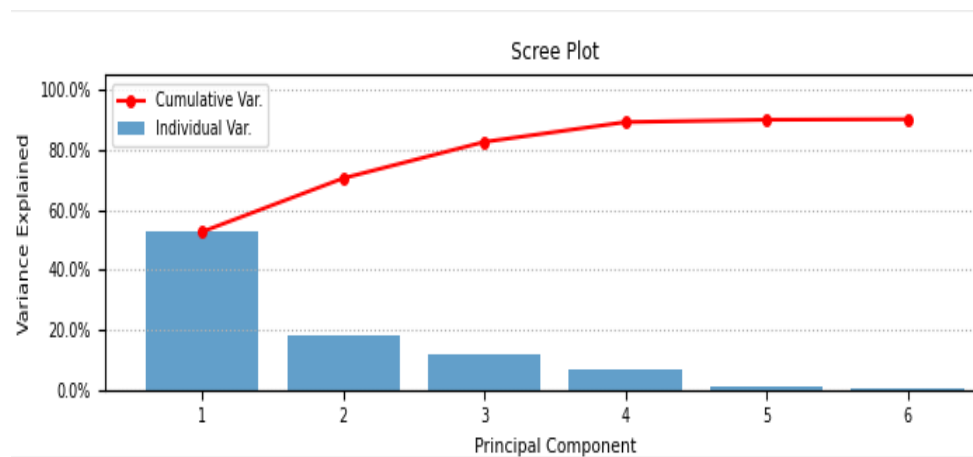


Εικόνα 122: Το μέσο φάσμα φθορισμού (Mean Energy Spectrum) μετά τη βαθμονόμηση σε ενέργεια (keV). Οι κορυφές αντιστοιχούν στις χαρακτηριστικές γραμμές εκπομπής των κυρίαρχων στοιχείων στην εικόνα, όπως Ca, S, Fe, Pb και Au.

Η Ανάλυση Κύριων Συνιστωσών (**PCA**) εφαρμόστηκε στα προεπεξεργασμένα υπερφασματικά δεδομένα με στόχο τη μείωση της διαστατικότητάς τους, τη συμπύκνωση της πληροφορίας και την ανάδειξη των κυριότερων πηγών φασματικής μεταβλητότητας. Ένα κρίσιμο βήμα στην **PCA** είναι ο προσδιορισμός του βέλτιστου αριθμού κύριων συνιστωσών (**Principal Components, PCs**) που θα διατηρηθούν. Ο αριθμός αυτός θα πρέπει να είναι επαρκής για να περιγράφει το μεγαλύτερο μέρος της διακύμανσης των αρχικών δεδομένων, απορρίπτοντας παράλληλα τις συνιστώσες που αντιστοιχούν κυρίως σε θόρυβο.

Για τον καθορισμό του αριθμού των **PCs**, εξετάστηκε το διάγραμμα **Scree Plot** (όπως φαίνεται στην παρακάτω εικόνα Εικόνα 123), το οποίο απεικονίζει το ποσοστό της συνολικής διακύμανσης που εξηγεί κάθε διαδοχική συνιστώσα, καθώς και το αθροιστικό ποσοστό. Στην παρούσα μελέτη, χρησιμοποιήθηκε η μέθοδος Incremental **PCA (IPCA)** για την επεξεργασία του πλήρους συνόλου των δεδομένων (**693.360 pixels × 4096 κανάλια**, συνολικού μεγέθους **~11.16 GB**), λόγω του μεγάλου όγκου των δεδομένων. Η **IPCA** επιτρέπει την επεξεργασία των δεδομένων τμηματικά (σε

batches), καθιστώντας την ανάλυση εφικτή ακόμα και σε συστήματα με περιορισμένη μνήμη RAM.



Εικόνα 123: Scree Plot

Όπως φαίνεται στην Εικόνα δίπλα, οι πρώτες έξι (6) κύριες συνιστώσες εξηγούν αθροιστικά το **90.2%** της συνολικής μεταβλητότητας (**PC1: 52.7%, PC2: 17.8%, PC3:**

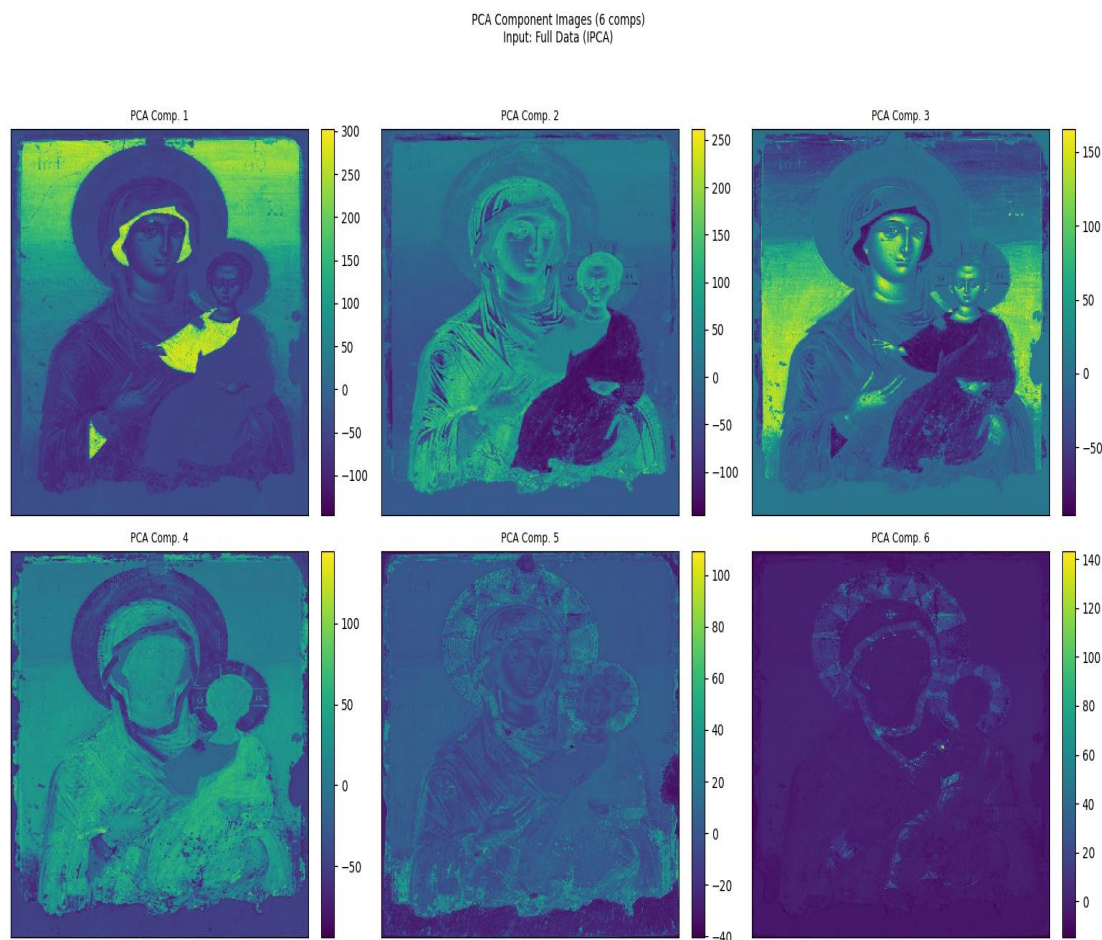
12.1%, PC4: 6.6%, PC5: 0.8%, PC6: 0.2%). Η επιλογή των έξι συνιστωσών κρίθηκε ως ένας ικανοποιητικός συμβιβασμός μεταξύ της διατήρησης της πληροφορίας και της μείωσης της διαστατικότητας, με τον χρόνο επεξεργασίας για την **IPCA** να ανέρχεται περίπου στα **15 λεπτά** σε ένα τυπικό υπολογιστικό σύστημα.

Όπως φαίνεται στην Εικόνα δίπλα, οι πρώτες έξι (6) κύριες συνιστώσες εξηγούν αθροιστικά το **90.2%** της συνολικής μεταβλητότητας (**PC1: 52.7%, PC2: 17.8%, PC3: 12.1%, PC4: 6.6%, PC5: 0.8%, PC6: 0.2%**). Η επιλογή των έξι συνιστωσών κρίθηκε ως ένας ικανοποιητικός συμβιβασμός μεταξύ της διατήρησης της πληροφορίας και της μείωσης της διαστατικότητας, με τον χρόνο επεξεργασίας για την **IPCA** να ανέρχεται περίπου στα **15 λεπτά** σε ένα τυπικό υπολογιστικό σύστημα.

Εικόνα 124: Αξιολόγηση της Ανάλυσης Κύριων Συνιστωσών (PCA). **(Πάνω)** Το διάγραμμα Scree Plot δείχνει την ατομική (μπλε μπάρες) και σωρευτική (κόκκινη γραμμή) εξηγούμενη διακύμανση. **(Κάτω)** Τα αριθμητικά αποτελέσματα της Incremental PCA (IPCA), που επιβεβαιώνουν ότι οι πρώτες 6 συνιστώσες διατηρούν περίπου το 90.2% της συνολικής πληροφορίας.

Data size after filtering: 11.16 GB
Fitting PCA: 0% | 0/70
[01:13<39:18, 38.22s/it]

Οι χωρικές κατανομές (**scores**) των έξι αυτών κύριων συνιστωσών απεικονίζονται στην παρακάτω εικόνα.



Εικόνα 125: Χωρικοί χάρτες (score maps) των έξι πρώτων κύριων συνιστωσών (PCA Components 1-6). Κάθε χάρτης απεικονίζει τη χωρική κατανομή μιας ορθογόνιας πηγής φασματικής διακύμανσης, διαχωρίζοντας διαφορετικές περιοχές της εικόνας (π.χ., υπόβαθρο, φωτοστέφανα, ενδύματα).

PC1 (52.7% της μεταβλητότητας): Αυτή η συνιστώσα, ως η κυρίαρχη πηγή διακύμανσης, επιτυγχάνει έναν σαφή και ισχυρό διαχωρισμό μεταξύ του ζωγραφικού στρώματος και του υποστρώματος. Οι έντονες θετικές τιμές (κίτρινο) αντιστοιχούν στα φωτοστέφανα και στα πρόσωπα, όπου κυριαρχούν χρωστικές με στοιχεία υψηλού ατομικού αριθμού (*Au*, *Pb*), ενώ οι αρνητικές τιμές (μωβ) εντοπίζονται κυρίως στο φόντο και στο εκτεθειμένο gesso. Η **PC1** ουσιαστικά αντιπροσωπεύει την ισχυρότερη αντισυσχέτιση στο σύνολο δεδομένων: αυτή μεταξύ των χρωστικών και του υποστρώματος (*CaSO₄*).

PC2 (17.8%): Η συνιστώσα αυτή αποκαλύπτει την ισχυρότερη χημική αντίθεση εντός του ζωγραφικού στρώματος. Οι θετικές τιμές χαρτογραφούν με ακρίβεια τα πρόσωπα και τα χέρια (σαρκώματα), περιοχές που κυριαρχούνται από το Λευκό του Μολύβδου (*Pb*). Αντίθετα, οι αρνητικές τιμές εντοπίζονται στα ενδύματα και το φόντο.

PC3 (12.1%): Η **PC3** απομονώνει μια διαφορετική χημική ομάδα, με τις θετικές τιμές να αντιστοιχούν κυρίως στα κόκκινα και καφέ ενδύματα. Αυτό υποδεικνύει ότι η συνιστώσα αυτή αντιπροσωπεύει τη φασματική υπογραφή των γήινων χρωστικών με βάση τον Σίδηρο (*Fe*).

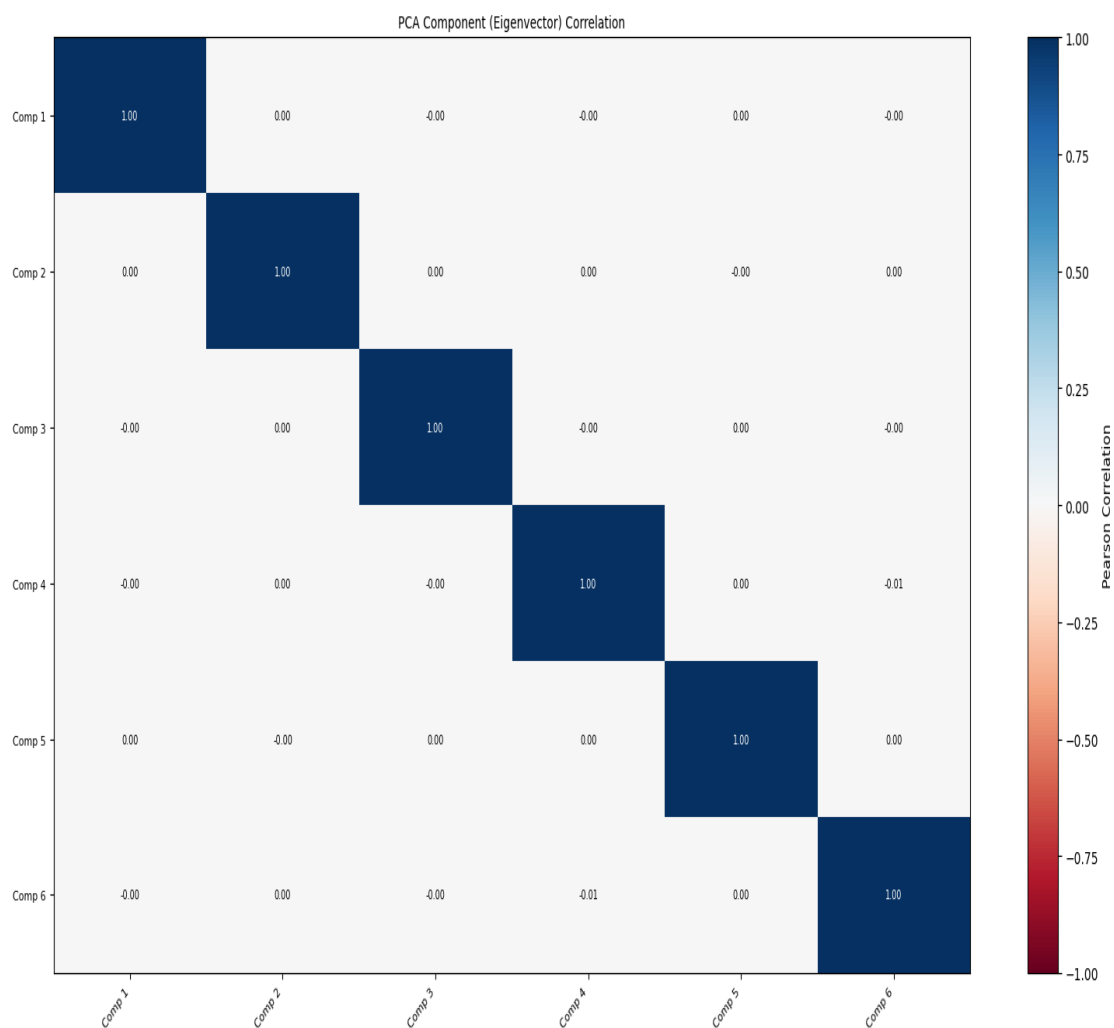
PC4, PC5, PC6 (δευτερεύουσες πηγές & θόρυβος): Οι συνιστώσες υψηλότερης τάξης καταγράφουν λεπτότερες διαφοροποιήσεις. Η **PC4** φαίνεται να απομονώνει το

μπλε της Παναγίας. Οι **PC5** και **PC6**, εξηγώντας πολύ μικρό ποσοστό της διακύμανσης (<1%), αποτελούνται κυρίως από τυχαίες κηλίδες και στατιστικό θόρυβο, υποδεικνύοντας ότι η ουσιαστική χημική πληροφορία έχει ήδη εξαχθεί.

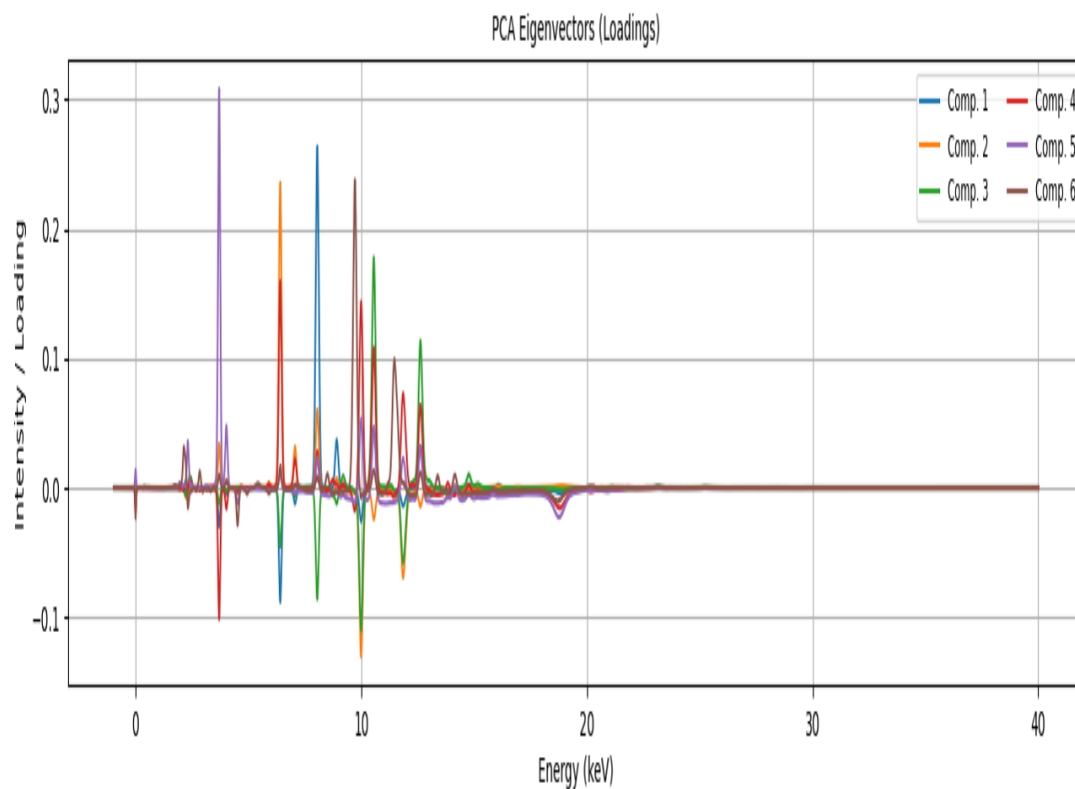
Ένα θεμελιώδες χαρακτηριστικό της **PCA** είναι ότι οι παραγόμενες κύριες συνιστώσες είναι εξ ορισμού ορθογώνιες, δηλαδή γραμμικά μη συσχετισμένες. Αυτό επιβεβαιώνεται από τον πίνακα συσχέτισης των ιδιοδιανυσμάτων (eigenvectors ή loadings vectors) των έξι **PCs**, όπου οι τιμές εκτός της κύριας διαγωνίου είναι πρακτικά μηδενικές.

Τα ιδιοδιανύσματα (**loadings**) κάθε **PC**, τα οποία απεικονίζουν τη συνεισφορά κάθε αρχικού ενεργειακού καναλιού στη διαμόρφωση της εκάστοτε συνιστώσας. Η ανάλυση των loadings επιτρέπει την ερμηνεία της χημικής σημασίας κάθε **PC**.

Εικόνα 126: Φασματική ερμηνεία των κύριων συνιστωσών. Ο πίνακας συσχέτισης (correlation matrix) των

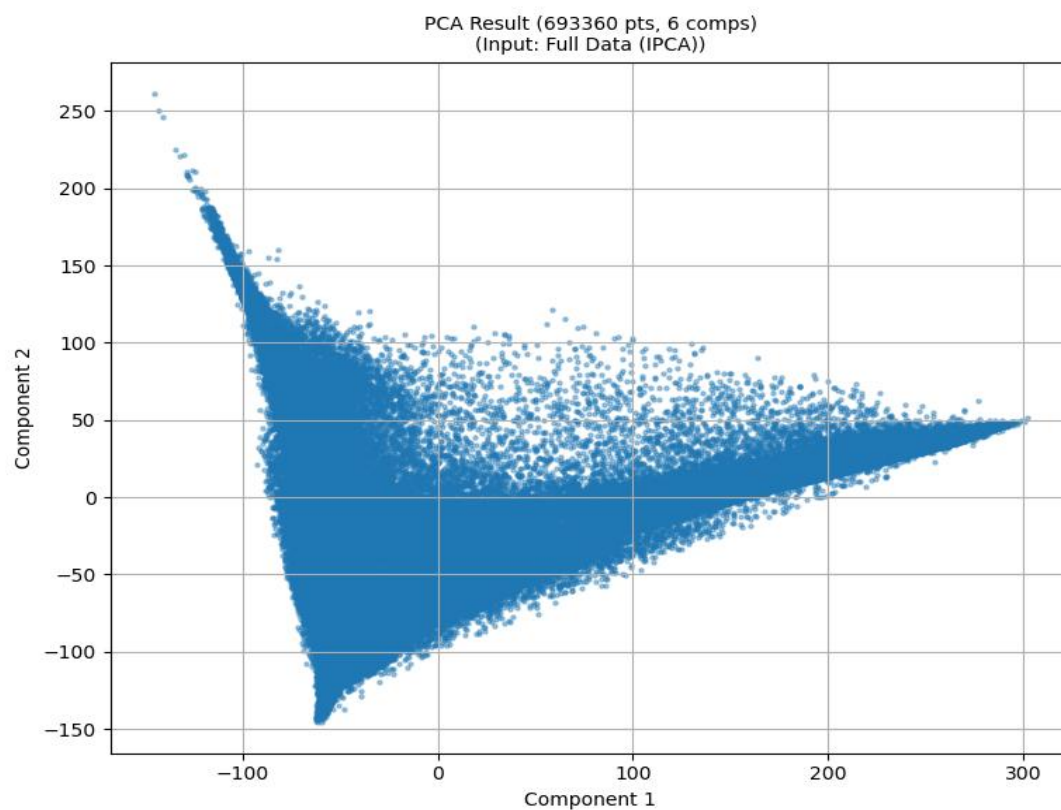


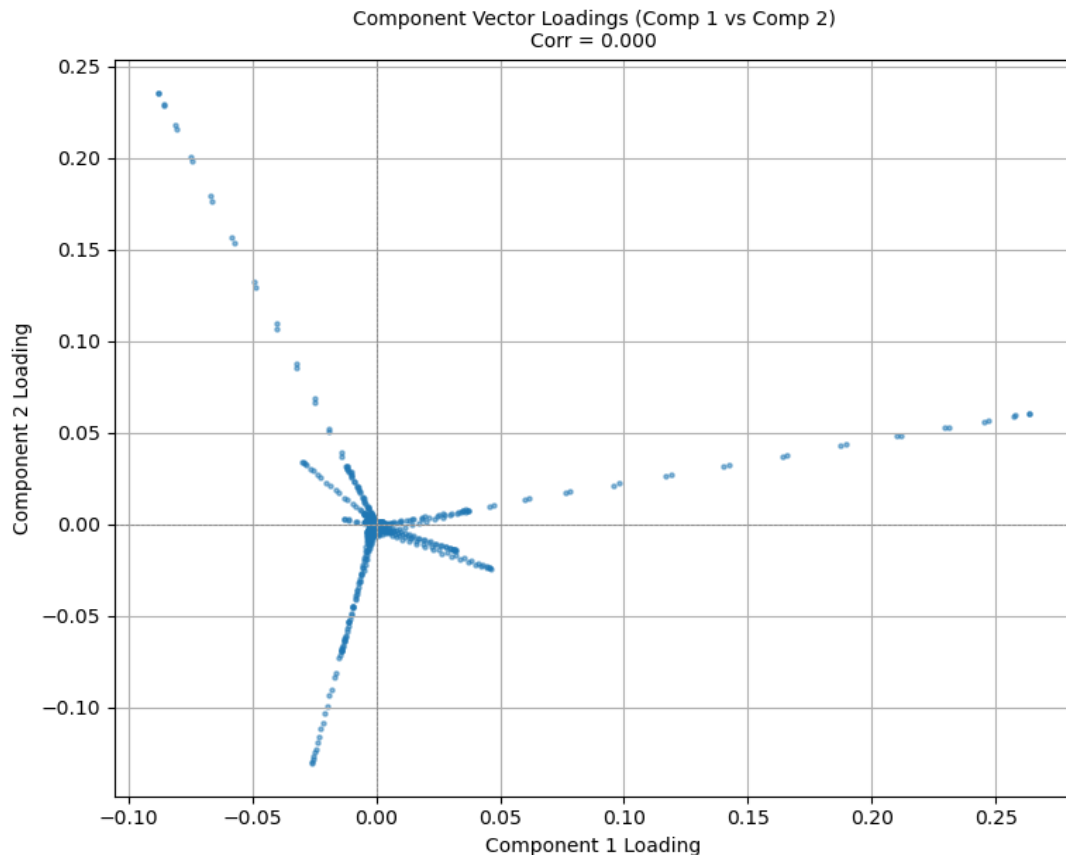
ιδιοδιανυσμάτων επιβεβαιώνει την ορθογωνιότητά τους.



Εικόνα 127: Τα φάσματα των ιδιοδιανυσμάτων (PCA Eigenvectors/Loadings), που αποκαλύπτουν τη φασματική υπογραφή κάθε συνιστώσας.

Για μια ορθή ανάλυση όμως προχωρούμε και στην απεικόνιση κάθε άξονα με βάση της άλλους ξεχωριστά αλλά και κάθε διανύσματος άξονα σε σχέση με τα άλλα Εικόνα 128.





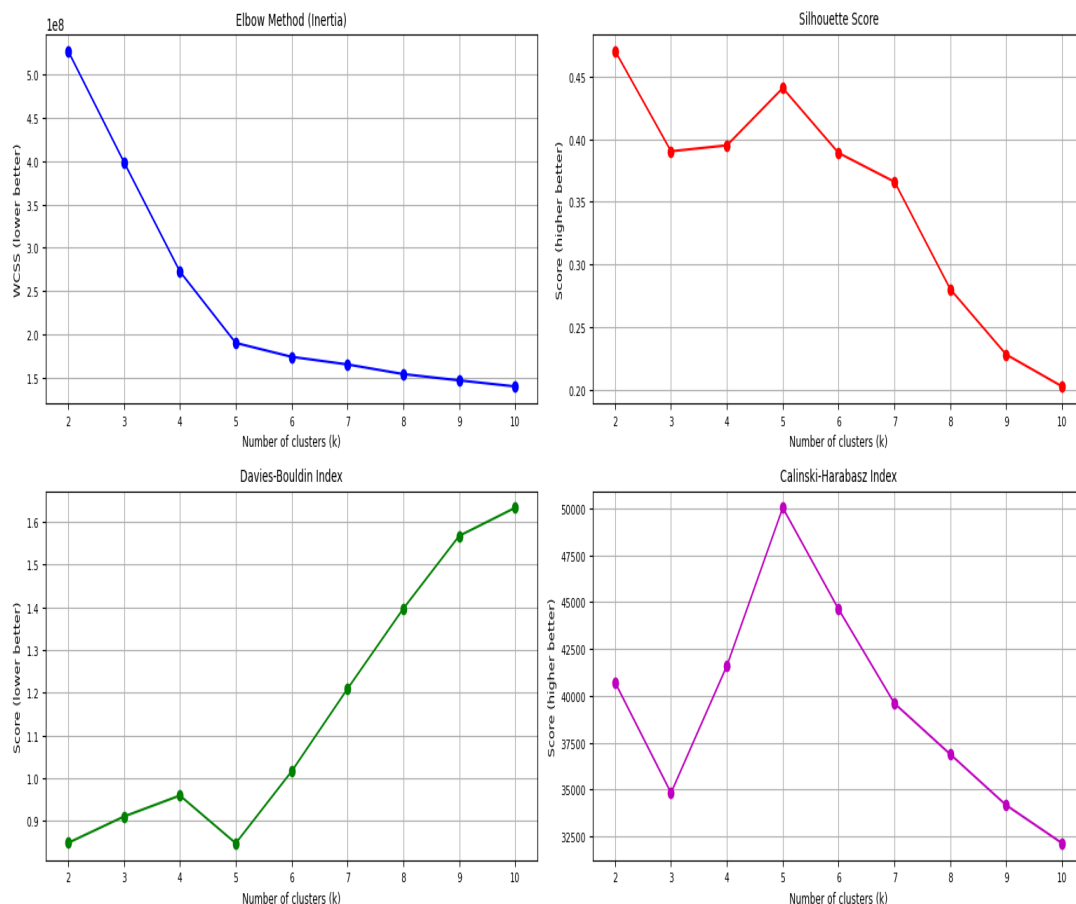
Εικόνα 128: Δισδιάστατη απεικόνιση της ανάλυσης PCA. **(Πάνω)** Διάγραμμα διασποράς των συντελεστών (scores) της Συνιστώσας 2 έναντι της Συνιστώσας 1. **(Κάτω)** Διάγραμμα των αντίστοιχων loadings, που δείχνει τη φασματική σχέση μεταξύ των δύο συνιστωσών.

Έπειτα την απεικόνιση των αξόνων και των διανυσμάτων στον δυσδιάστατο χώρο συναρτήσει των υπολοίπων αξόνων μπορούμε να προχωρήσουμε στην εύρεση του κατάλληλου αριθμού συστάδων μέσω του προγράμματος. Τοποθετούμε μέγιστο αριθμό **δέκα (10) συστάδες** για μείωση χρόνου υπολογισμού. Τα αποτελέσματα των μετρικών αξιολόγησης παρουσιάζονται στην Εικόνα παρακάτω Εικόνα 129.

- **Μέθοδος Elbow (WCSS):** Η καμπύλη του αθροίσματος των τετραγωνικών αποστάσεων εντός των συστάδων (*Within-Cluster Sum of Squares, WCSS*) ως προς τον αριθμό των συστάδων εμφανίζει μια χαρακτηριστική κάμψη ('αγκώνα') στην περιοχή $k=5$ έως $k=6$. Πέραν αυτού του σημείου, η προσθήκη επιπλέον συστάδων οδηγεί σε οριακή μόνο μείωση του *WCSS*.
- **Silhouette Score:** Ο δείκτης *Silhouette*, ο οποίος μετρά πόσο παρόμοιο είναι ένα αντικείμενο με τη δική του συστάδα σε σύγκριση με άλλες συστάδες, παρουσιάζει τη μέγιστη τιμή του για $k=2$, ενώ για $k=5$ και $k=6$ οι τιμές παραμένουν σε σχετικά υψηλά επίπεδα, υποδηλώνοντας καλό διαχωρισμό.
- **Davies-Bouldin Index:** Ο δείκτης *Davies-Bouldin*, όπου χαμηλότερες τιμές υποδηλώνουν καλύτερη ομαδοποίηση, εμφανίζει ελάχιστη τιμή για $k=5$.
- **Calinski-Harabasz Index:** Ο δείκτης *Calinski-Harabasz*, όπου υψηλότερες τιμές αντιστοιχούν σε καλύτερα ορισμένες συστάδες, παρουσιάζει τοπικό μέγιστο για $k=5$.

Συνεκτιμώντας τις παραπάνω μετρικές, και με γνώμονα την επίτευξη ενός ουσιαστικού διαχωρισμού των κύριων υλικών και χρωματικών περιοχών της εικόνας, επιλέχθηκε ο αριθμός των **έξι (6) συστάδων** για την περαιτέρω ανάλυση Εικόνα 131. Παρόλο που το πρόγραμμα παρέχει τη δυνατότητα εφαρμογής διαφόρων αλγορίθμων ομαδοποίησης, όπως *K-Means*, *DBScan*, *Gaussian Mixture Models (GMM)* και *Spectral Clustering*, για την παρούσα εργασία επιλέχθηκε ο αλγόριθμος **K-Means**. Η επιλογή αυτή βασίστηκε στην υπολογιστική του απόδοση, την ευρεία του χρήση και την ικανότητά του να παρέχει ικανοποιητικά και ερμηνεύσιμα αποτελέσματα σε σύνολα δεδομένων φασματοσκοπίας απεικόνισης, όπως διαπιστώθηκε και σε προκαταρκτικές αναλύσεις.

Cluster Evaluation Metrics (on Sample Data)



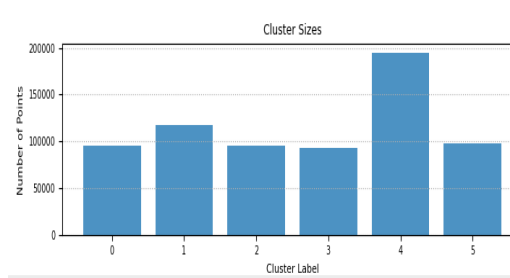
Εικόνα 129: Αποτελέσματα από την εκτέλεση του πλαισίου αξιολόγησης συσταδοποίησης για τον προσδιορισμό του βέλτιστου αριθμού συστάδων (k). Τα διαγράμματα απεικονίζουν τις μετρικές *Elbow Method*, *Silhouette Score*, *Davies-Bouldin Index* και *Calinski-Harabasz Index* για k από 2 έως 10, υποδεικνύοντας ως βέλτιστο εύρος τις 5-6 συστάδες.

Στην δίπλα εικόνα Εικόνα 130 παρατηρούμε το μέγεθος κάθε συστάδας και τον αριθμό των δεδομένων που περιλαμβάνει τόσο με απεικονιστικό τρόπο όπως και σαν λίστα.

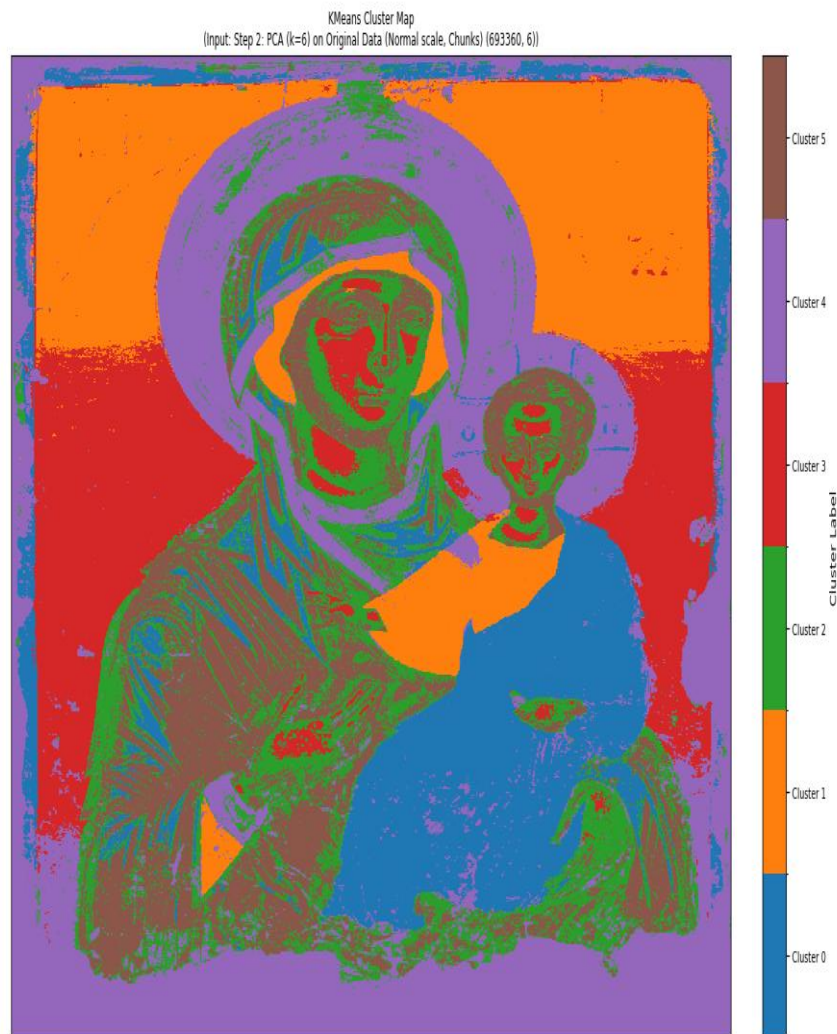

```

KMeans Results (on Step 2: PCA (k=6) on Original Data (Normal scale, Chunks) (693360, 6))
=====
Clusters found: 6
Distribution (Top 15):
Cluster 4: 194365
Cluster 1: 117925
Cluster 5: 97923
Cluster 0: 95760
Cluster 2: 94901
Cluster 3: 92486

```

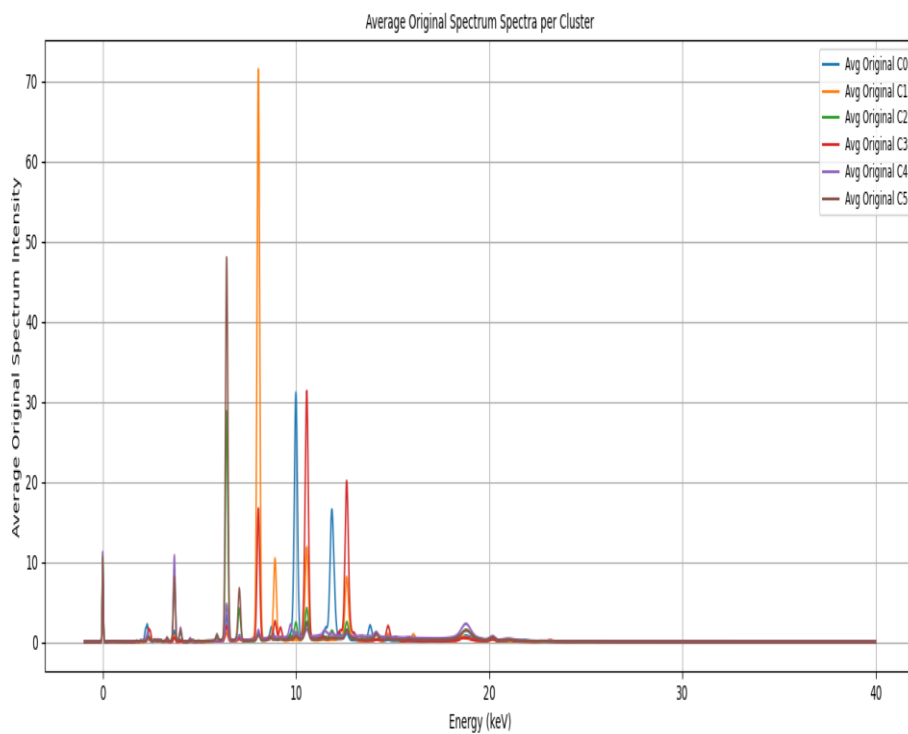
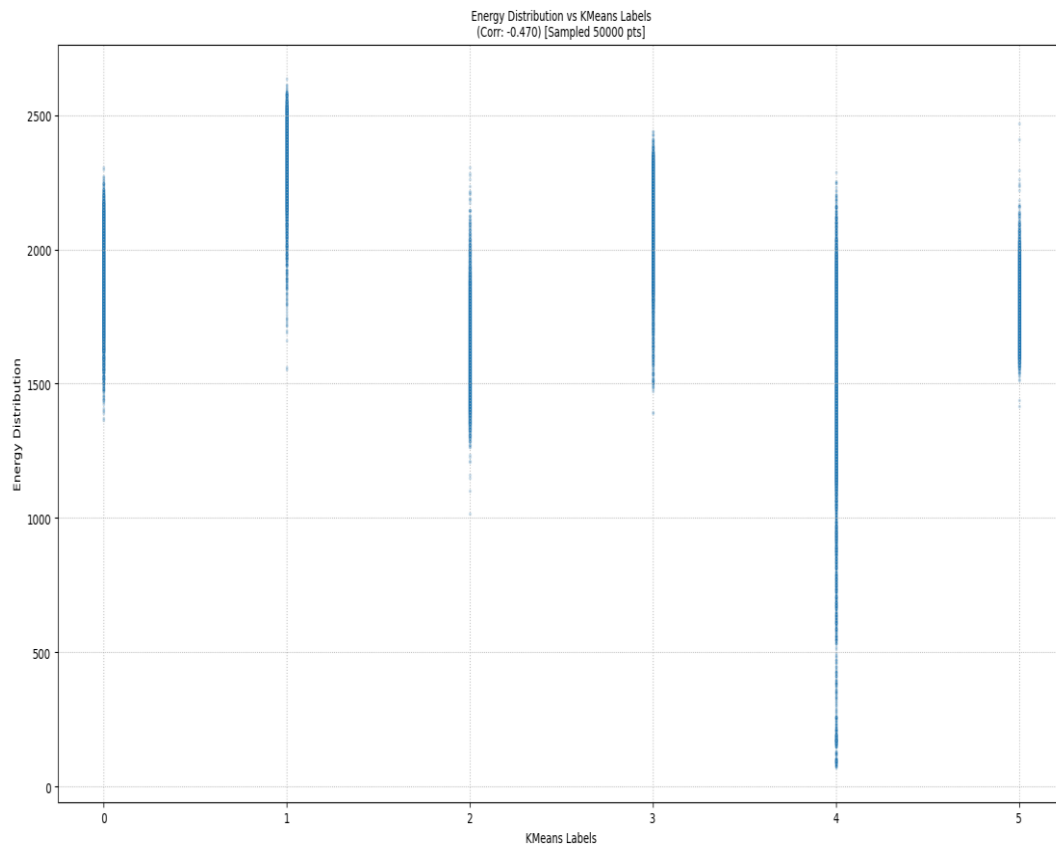


Εικόνα 130: (Αριστερά) Αριθμητική κατανομή των pixels ανά συστάδα. (Δεξιά) Γραφική απεικόνιση του μεγέθους κάθε συστάδας.



Εικόνα 131: Θεματικός χάρτης (cluster map) που προέκυψε από την εφαρμογή του αλγορίθμου K-Means ($k=6$) στα αποτελέσματα της PCA. Κάθε χρώμα αντιστοιχεί σε μια διακριτή συστάδα, αποκαλύπτοντας τη χωρική κατανομή των κυρίαρχων υλικών ομάδων.

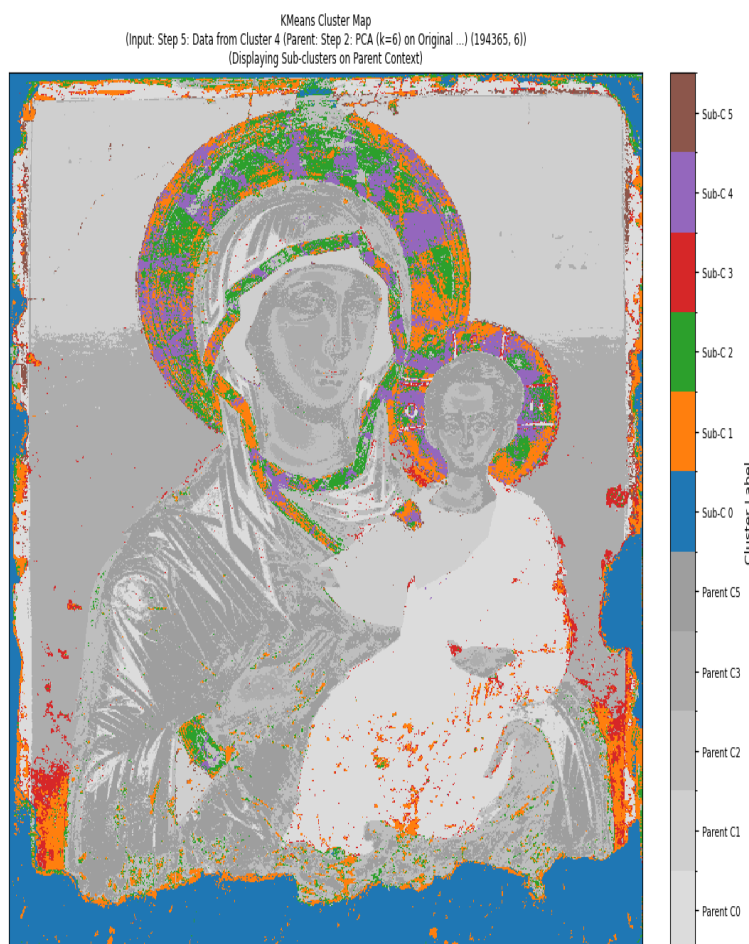
Με την Ενέργεια κάθε συστάδας αλλά και το φάσμα να απεικονίζεται παρακάτω Εικόνα 132.



Εικόνα 132: Χαρακτηρισμός των συστάδων K-Means. **(Πάνω)** Διάγραμμα κατανομής της έντασης για κάθε συστάδα. **(Κάτω)** Τα μέσα φάσματα φθορισμού για κάθε μία από τις έξι συστάδες, επιτρέποντας την ακριβή χημική τους ερμηνεία.

Από την οπτική και φασματική ανάλυση, ιδιαίτερο ενδιαφέρον παρουσιάζει η περιοχή των στεφάνων (φωτοστέφανων) της Παναγίας και του Ιησού. Οι περιοχές αυτές, που οπτικά φαίνονται χρυσές, αντιστοιχούν κυρίως στη συστάδα 4 Εικόνα 131. Το μέσο φάσμα αυτής της συστάδας εμφανίζει τις χαρακτηριστικές γραμμές L του χρυσού (*Au*), αλλά η έντασή τους είναι σχετικά χαμηλή, υποδηλώνοντας πιθανόν τη χρήση πολύ λεπτών φύλλων χρυσού (τεχνική χρυσώματος). Επιπλέον, στο ίδιο φάσμα παρατηρείται και σημαντική παρουσία ασβεστίου (*Ca*), πιθανόν από το υπόστρωμα (*gesso*) ή το συνδετικό υλικό που χρησιμοποιήθηκε για την επικόλληση του χρυσού.

Λόγω της σύνθετης φύσης αυτής της περιοχής και της σημασίας της για την κατανόηση της τεχνικής κατασκευής, κρίθηκε σκόπιμη η περαιτέρω διερεύνηση μέσω υπο-ομαδοποίησης (*sub-clustering*). Συγκεκριμένα, επιλέχθηκαν τα pixels που ανήκαν στην προαναφερθείσα συστάδα της αρχικής ομαδοποίησης, και εφαρμόστηκε εκ νέου ο αλγόριθμος *K-Means* με στόχο τον διαχωρισμό τους σε έξι (6) νέες υπο-ομάδες. Ο χάρτης που προέκυψε από αυτή τη διαδικασία, όπου οι υπο-ομάδες απεικονίζονται με διακριτά χρώματα εντός της γκριζαρισμένης περιοχής των υπόλοιπων αρχικών συστάδων, παρουσιάζεται στην Εικόνα παρακάτω Εικόνα 133.



Εικόνα 133: Αποτελέσματα της ιεραρχικής ανάλυσης (*sub-clustering*) που εφαρμόστηκε στην 'Cluster 4' (φωτοστέφανα). Ο χάρτης αποκαλύπτει την εσωτερική δομή της συστάδας με έξι νέες υπο-ομάδες (*Sub-C 0-5*), επιτρέποντας τη διερεύνηση λεπτομερών υλικών διαφοροποιήσεων (π.χ., διαφορετικά πάχη φύλλου χρυσού,

παρουσία κόλλας ή Μπόλου Αρμενίας που είναι λεπτότατο, καθαρό, κόκκινο άργιλο που χρησιμοποιείται ως ενδιάμεσο στρώμα κάτω από τις επιχρυσώσεις και βοηθά σε μεταγενέστερη στίλβωση του φύλλου Au).

Όπως παρατηρούμε από πάνω βλέπουμε πως η κατανομή των δεδομένων σε συστάδες γίνεται με καλύτερο τρόπο και παρατηρούμε ακριβώς που είναι η επικάλυψη με τα φύλλα χρυσού αλλά και η ποσότητα με την οποία έχει πραγματοποιηθεί.

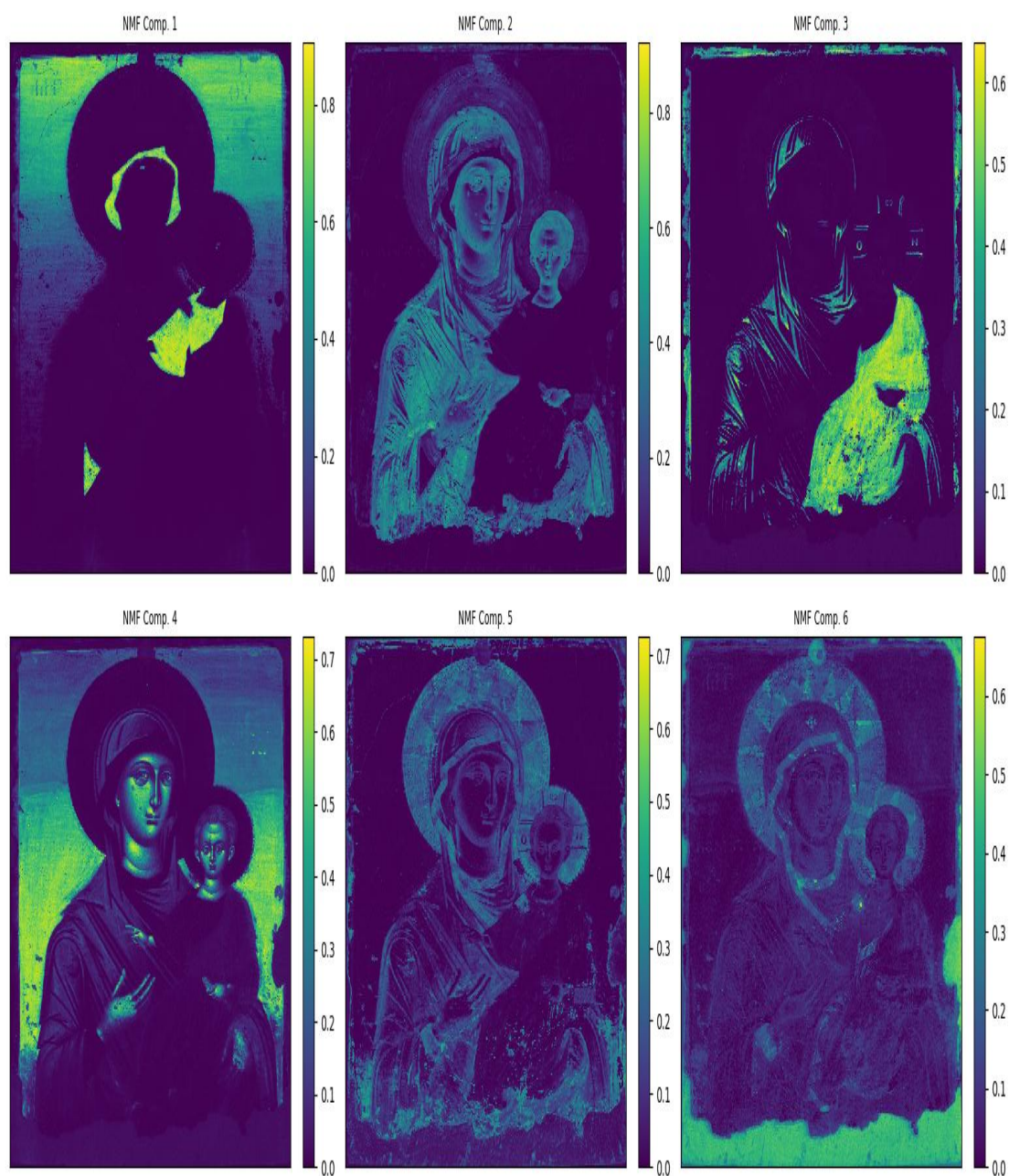
Για την περαιτέρω διερεύνηση της χημικής σύστασης της εικόνας, εφαρμόστηκε επίσης η μέθοδος της Μη Αρνητικής Παραγοντοποίησης Πινάκων (*Non-Negative Matrix Factorization, NMF*). Η *NMF*, παρόμοια με την *PCA*, είναι μια τεχνική μείωσης διαστατικότητας, αλλά με τον επιπλέον περιορισμό ότι τόσο οι παραγόμενες συνιστώσες (που αντιστοιχούν σε 'φασματικές υπογραφές') όσο και οι συντελεστές ανάμειξής τους (που αντιστοιχούν στις χωρικές τους 'αφθονίες') πρέπει να είναι μη αρνητικές. Αυτός ο περιορισμός συχνά οδηγεί σε αποτελέσματα που έχουν πιο άμεση φυσικοχημική ερμηνεία.

Στην παρούσα μελέτη, η *NMF* εφαρμόστηκε:

1. Απευθείας στα αρχικά προεπεξεργασμένα υπερφασματικά δεδομένα (μετά από δειγματοληψία, όπως υποδεικνύεται από την περιγραφή της Εικόνας: *'Input: Original Data (Sample)'*).
2. Στα αποτελέσματα (scores) της προηγούμενης ανάλυσης *PCA* (με τις **6 κύριες συνιστώσες**, ("*NMF Result (Input: Step 2: PCA (k=6)...*")). Αυτή η προσέγγιση διερευνά αν ο συνδυασμός των δύο μεθόδων μπορεί να προσφέρει επιπλέον διαχωρισμό ή βελτιωμένη ερμηνεία.

Για την πρώτη περίπτωση (*NMF* στα αρχικά δεδομένα), επιλέχθηκε η εξαγωγή **έξι (6)** συνιστωσών *NMF*. Η πληροφορία που συλλαμβάνεται από αυτές τις συνιστώσες εκτιμήθηκε ότι αντιστοιχεί περίπου στο **80%** της συνολικής μεταβλητότητας που εξηγούσαν οι αντίστοιχες έξι *PCs*. Στην Εικόνα από κάτω Εικόνα 134 παρουσιάζονται οι χωρικές κατανομές των έξι συνιστωσών *NMF* που προέκυψαν από την ανάλυση του αρχικού δείγματος δεδομένων.

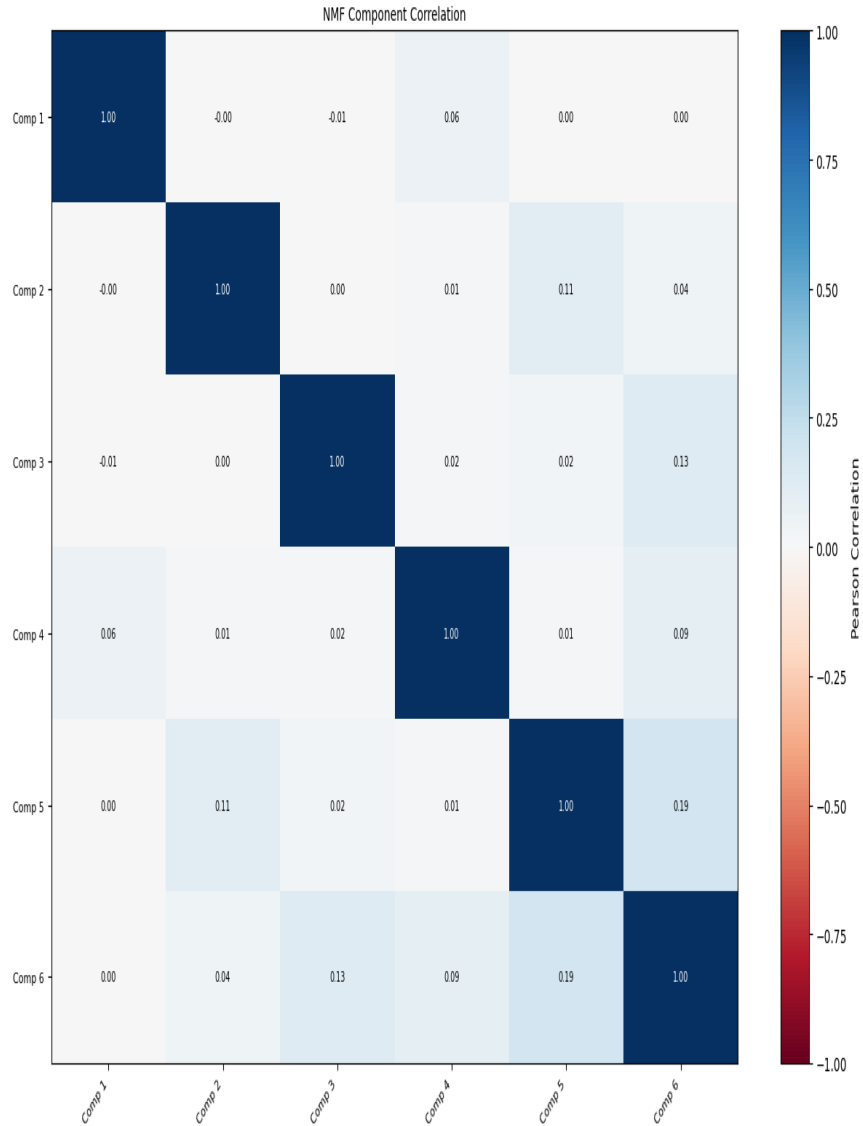
NMF Component Images (6 comps)
Input: Original Data (Sample)



Εικόνα 134: Χωρικοί χάρτες των έξι συνιστωσών που προέκυψαν από την ανάλυση NMF. Λόγω της μη-αρνητικότητας, οι χάρτες αυτοί μπορούν να ερμηνευθούν ως χάρτες αφθονίας, αναδεικνύοντας τη χωρική κατανομή διακριτών χημικών συστατικών (π.χ., gesso, χρυσός, ώχρες).

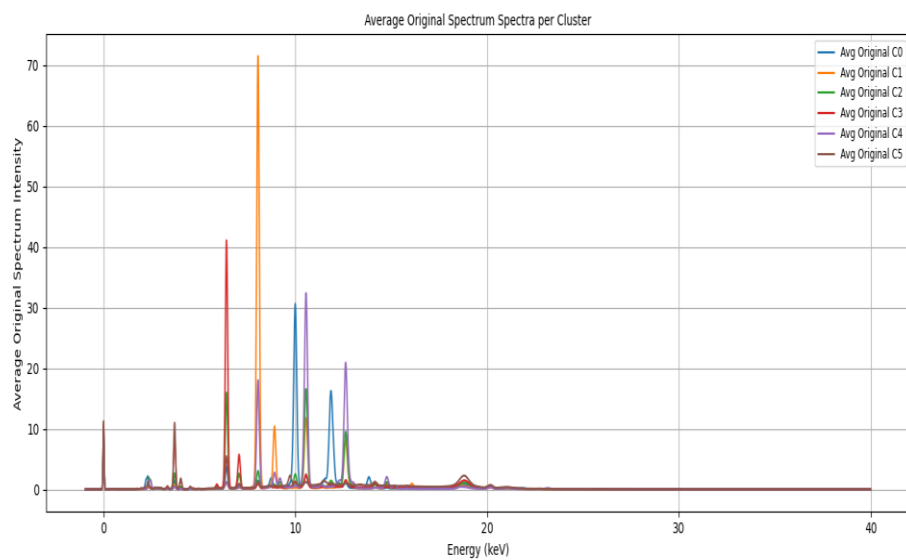
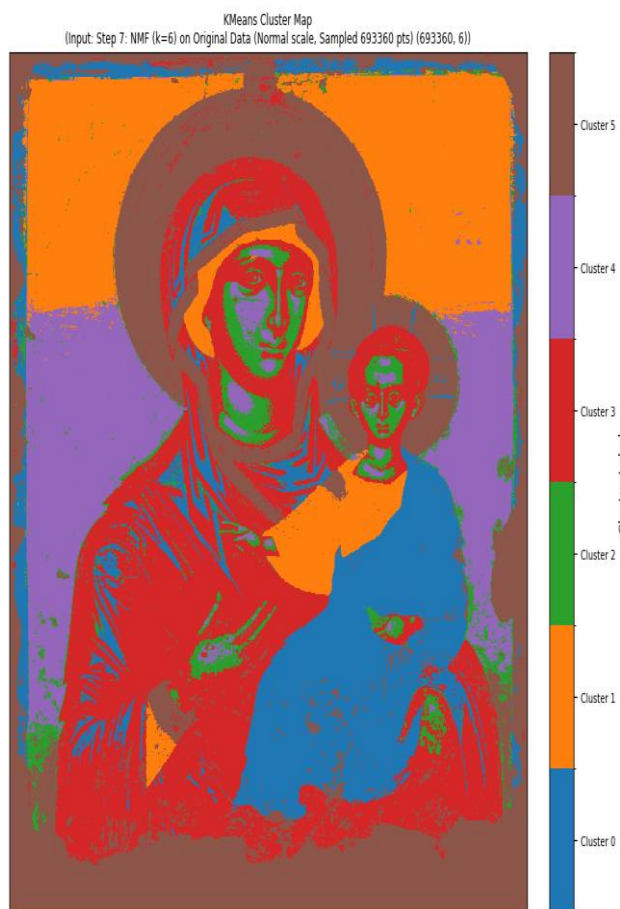
Η εφαρμογή της **NMF** στα αρχικά δεδομένα παρέχει συνιστώσες που, λόγω του περιορισμού της μη-αρνητικότητας, μπορούν να ερμηνευθούν άμεσα ως χάρτες αφθονίας για συγκεκριμένες χημικές φάσεις. Οι έξι χάρτες (Εικόνα 134) αποκαλύπτουν έναν εξαιρετικά σαφή διαχωρισμό των υλικών, ο οποίος μπορεί να ερμηνευθεί ως εξής:

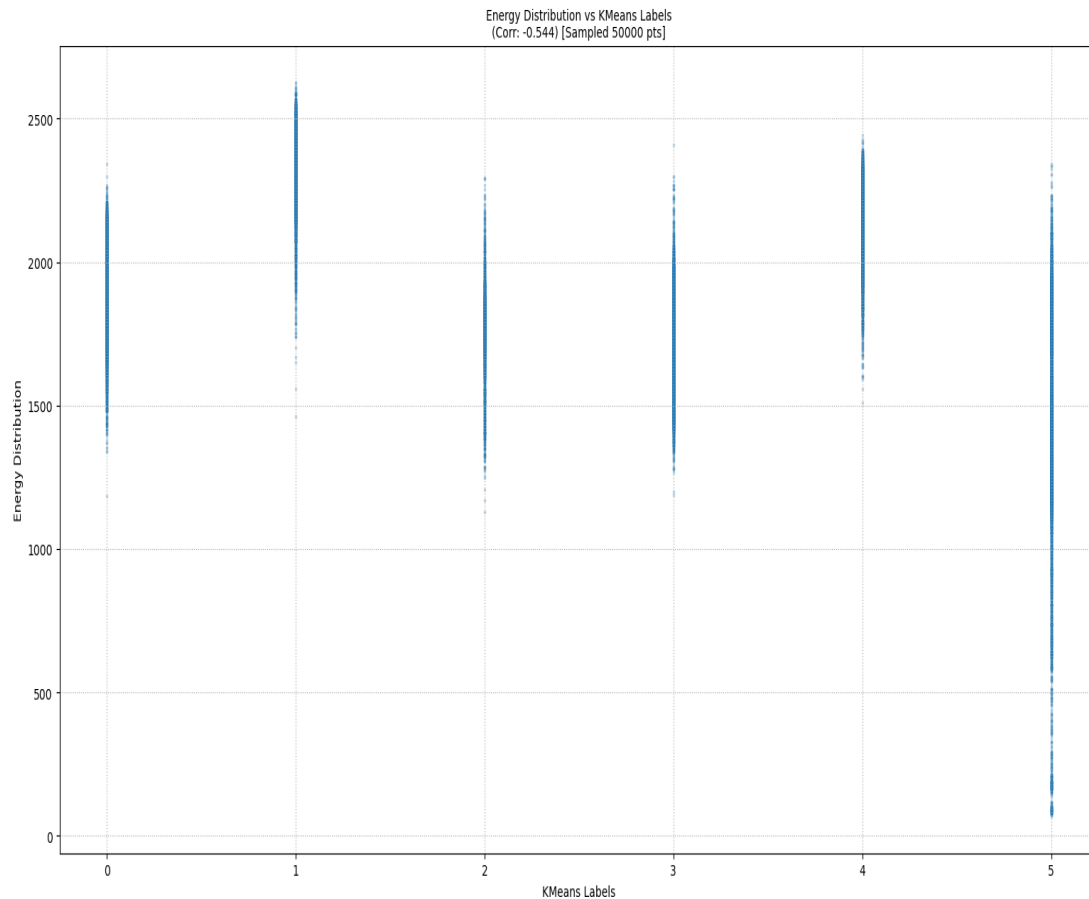
- **NMF Comp. 1** (Χρυσός - *Au* & Άργυρος - *Ag*): Αυτή η συνιστώσα απομονώνει με εξαιρετική ευκρίνεια τις επιμεταλλωμένες περιοχές. Η χωρική της κατανομή ταυτίζεται απόλυτα με στοιχειακούς χάρτες του Χρυσού (*Au*) στα φωτοστέφανα και του Αργύρου (*Ag*) στις επιγραφές, αποδεικνύοντας την ικανότητα της **NMF** να ομαδοποιεί υλικά με παρόμοια φασματικά χαρακτηριστικά (ευγενή μέταλλα).
- **NMF Comp. 2** (Λευκό του Μολύβδου - *Pb*): Η συνιστώσα αυτή χαρτογραφεί με απόλυτη ακρίβεια τα σαρκώματα (πρόσωπα, χέρια) και τα φωτεινότερα σημεία (**highlights**) στα ενδύματα. Η κατανομή της είναι σχεδόν πανομοιότυπη με τον στοιχειακό χάρτη του Μολύβδου (*Pb*), ταυτοποιώντας έτσι το Λευκό του Μολύβδου, τη βασική λευκή χρωστική της παλέτας.
- **NMF Comp. 3** (Κιννάβαρι - *Hg*): Αυτή η συνιστώσα εντοπίζεται σχεδόν αποκλειστικά στο έντονο κόκκινο ένδυμα της Παναγίας. Η χωρική της θέση αντιστοιχεί απόλυτα στον στοιχειακό χάρτη του Υδραργύρου (*Hg*), επιβεβαιώνοντας τη χρήση της πολύτιμης χρωστικής κιννάβαρι (**HgS**).
- **NMF Comp. 4** (Γαιώδεις/Μπλε Χρωστικές - *Fe, Cu*): Αυτή η συνιστώσα παρουσιάζει υψηλή ένταση στα ενδύματα του Χριστού και της Παναγίας (εκτός της κόκκινης περιοχής). Η κατανομή της δείχνει ισχυρή συσχέτιση με τους στοιχειακούς χάρτες του Σιδήρου (*Fe*) και του Χαλκού (*Cu*), υποδεικνύοντας ότι αντιπροσωπεύει ένα μείγμα ή μια συνύπαρξη γαιωδών χρωστικών (ώχρες) και χρωστικών με βάση τον χαλκό (αζουρίτης/μαλαχίτης).
- **NMF Comp. 5 & 6** (Υπόστρωμα - *Ca, S*): Οι συνιστώσες αυτές κυριαρχούν στο φόντο, στο περίγραμμα και στις περιοχές με εκτεταμένες φθορές. Η κατανομή τους ταυτίζεται με τους στοιχειακούς χάρτες του Ασβεστίου (*Ca*) και του Θείου (*S*). Απομονώνουν συνεπώς το υπόστρωμα προετοιμασίας (**gesso**). Η ύπαρξη δύο συνιστωσών για το ίδιο υλικό υποδηλώνει πιθανώς διαφορετικές καταστάσεις διατήρησης ή ελαφρώς διαφορετική σύσταση (π.χ., διαφορετική αναλογία συνδετικού μέσου).



Εικόνα 135: Πίνακας συσχέτισης (heatmap) των συνιστωσών *NMF*. Σε αντίθεση με την *PCA*, οι συνιστώσες δεν είναι κατ' ανάγκη ορθογώνιες. Οι θετικές συσχετίσεις υποδεικνύουν κοινά φασματικά χαρακτηριστικά ή χωρική συνύπαρξη υλικών μεταξύ των συνιστωσών.

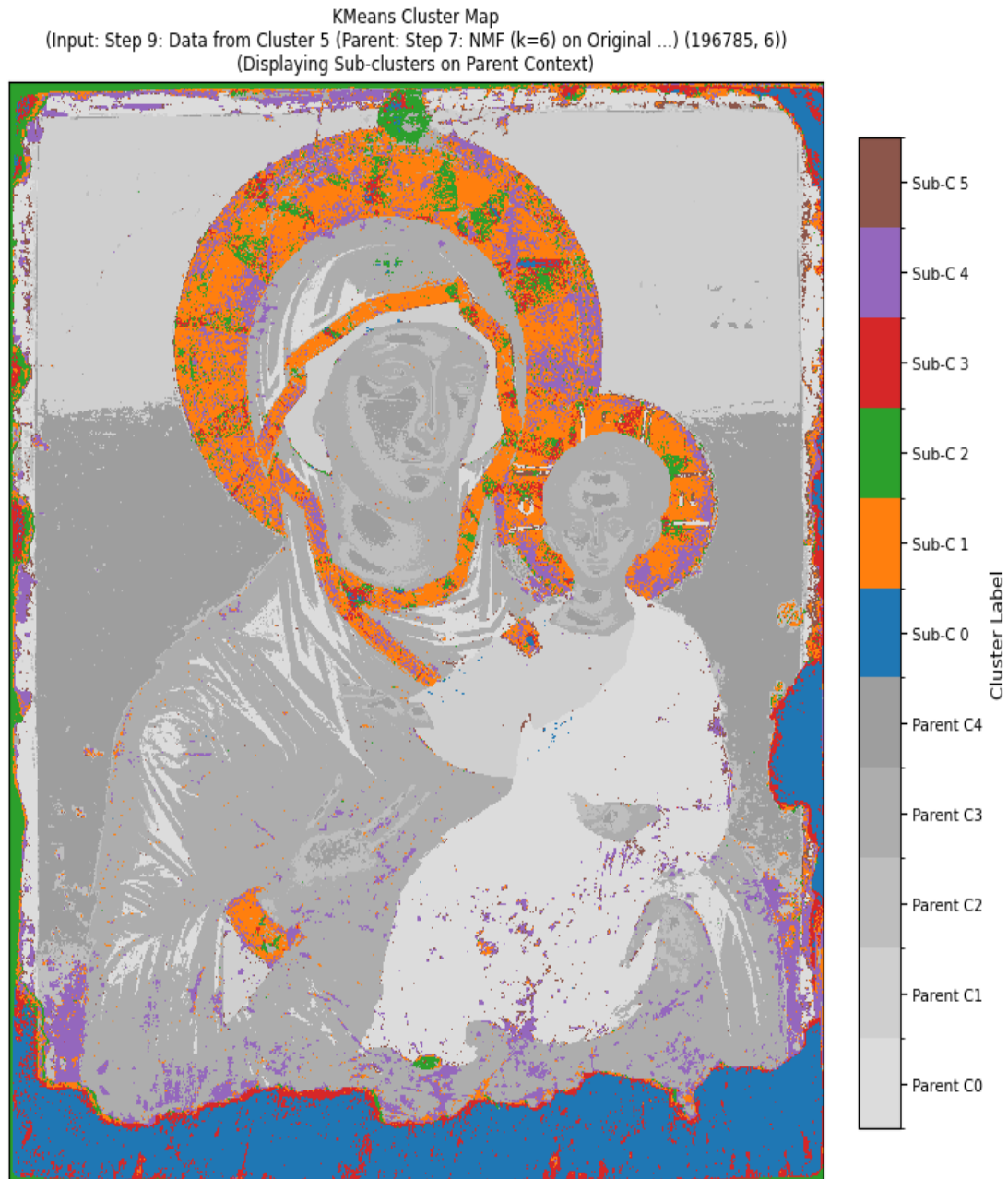
"Ο πίνακας συσχέτισης των συνιστωσών *NMF* (Εικόνα 135) δείχνει ότι, σε αντίθεση με την *PCA*, οι συνιστώσες *NMF* δεν είναι κατ' ανάγκη ορθογώνιες. Παρατηρούνται ορισμένες μέτριες έως ισχυρές θετικές συσχετίσεις, για παράδειγμα μεταξύ της **Comp 3** και **Comp 2** (τιμή ~ 0.0) και της **Comp 5** και **Comp 6** (τιμή ~ 0.19). Αυτό υποδηλώνει ότι αυτές οι συνιστώσες μπορεί να μοιράζονται κοινά φασματικά χαρακτηριστικά ή να συνυπάρχουν χωρικά. Για να ολοκληρώσουμε την ανάλυση με βάση την διαδικασία για την *PCA* προχωρούμε στην συσταδοποίηση μέσω της μεθόδου *Kmeans* πάλι με τον ίδιο αριθμό συστάδων. Τα αποτελέσματα των εικόνων παρέχονται παρακάτω Εικόνα 136.





Εικόνα 136: Αποτελέσματα της συσταδοποίησης *K-Means* ($k=6$) επί των δεδομένων NMF. **(Πάνω)** Ο θεματικός χάρτης. **(Μέση)** Τα μέσα φάσματα για κάθε συστάδα. **(Κάτω)** Το διάγραμμα κατανομής της έντασης ανά συστάδα.

Σε σύγκριση με την συσταδοποίηση της μεθόδου της **PCA** παρατηρούμε πως βρίσκονται σε μεγάλη συμφωνία μεταξύ ακόμα και στην κατανομή των εντάσεων με βάση τις συστάδες Εικόνα 136. Πάλι κατανοούμε πως το στέμμα της Παναγίας και του Ιησού περιέχουν την σημαντική πληροφορία οπότε και προχωρούμε σε **subclustering** της **συστάδας 5**, υπο μεθόδου **Kmeans** με έξι (**6**) συστάδες (Εικόνα 137).



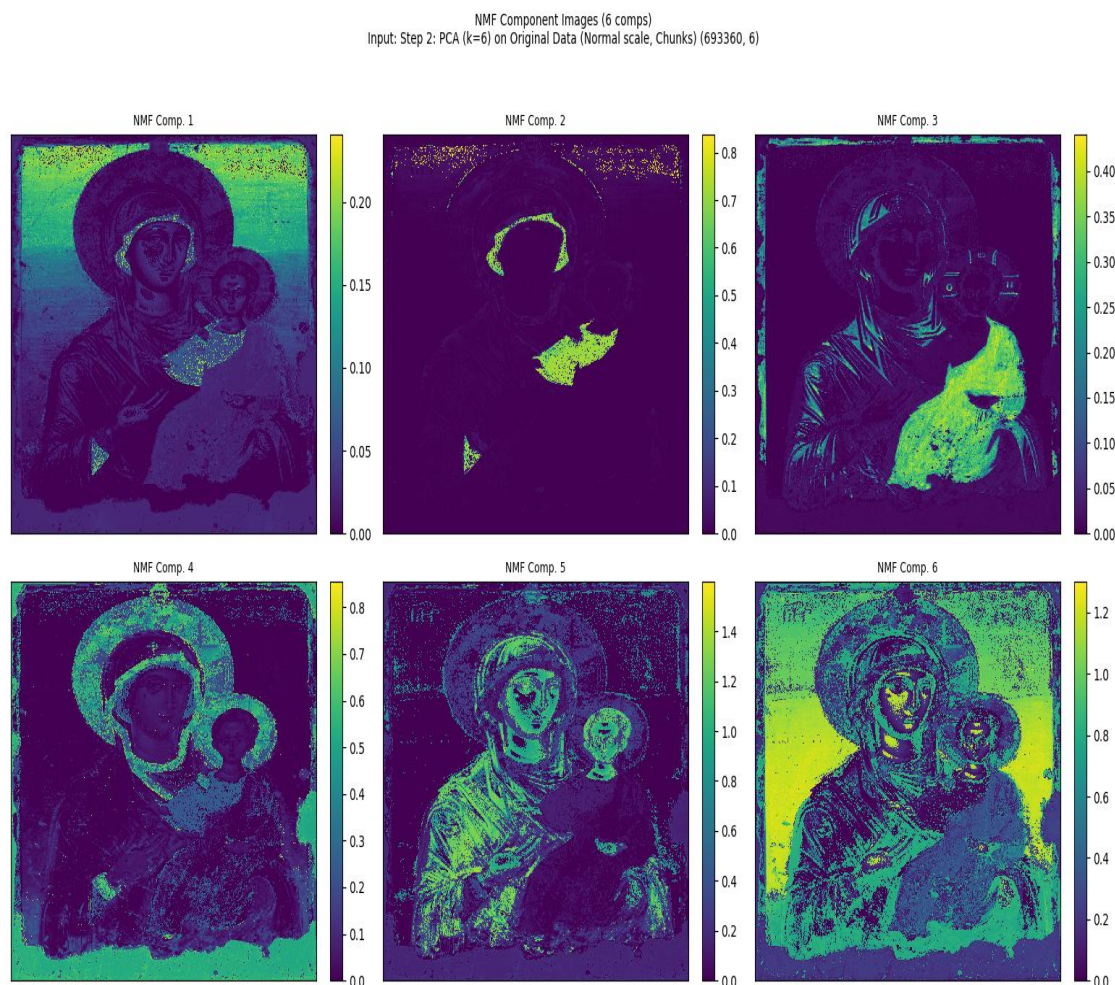
Εικόνα 137: Μέθοδος συσταδοποίησης Kmeans επάνω στα δεδομένα της NMF με subclustering επάνω στην συστάδα 5

Εδώ παρατηρούμε πως η συσταδοποίηση έχει κάνει καλύτερη διαφοροποίηση στην εικόνα του υποβάθρου και των περιοχών που βρίσκονται στα όρια της εικόνας και του υποβάθρου όπως παρατηρούμε που βρίσκονται στην συστάδες **3** και **1** Εικόνα 137. Επιπρόσθετα στα στέμματα παρατηρούμε μια πιο λεπτομερή σε σχέση με την βασική εικόνα συσταδοποίηση μεταξύ των περιοχών **2** και **1**. Βεβαία είναι μια πιο ομοιογενοποιημένη συσταδοποίηση σε σχέση με αυτή της **PCA**.

Υβριδική Ανάλυση NMF-PCA: Από τους Χάρτες Υλικών στους Χάρτες Τεχνικής

Τέλος για να προχωρήσουμε και στην ολοκληρωμένη ανάλυση προχωρούμε στα δεδομένα που λάβαμε από την **PCA** και τα μετασχηματίζουμε με βάση την **NMF** Εικόνα 138. Υπο αριθμό αξόνων **6** και επομένως λαμβάνουμε σχεδόν το **99%** της πληροφορίας

της εικόνας, σε χρόνο περίπου **10 λεπτών**. Προχωρούμε λοιπόν και στην απεικόνιση των **Component** πάνω στην εικόνα μας.

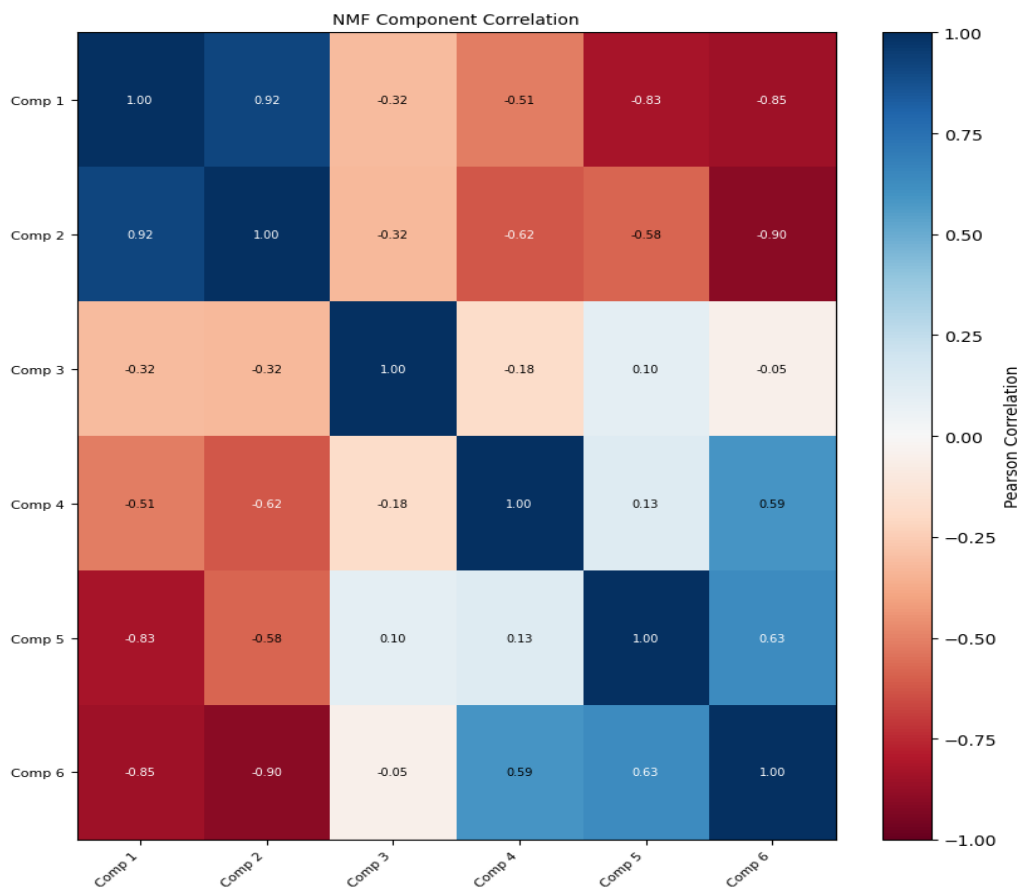


Εικόνα 138: Απεικόνιση αξόνων NMF με επάνω στην εικόνα, χρωματίζοντας μόνο τις περιοχές εμφάνισης.

Εκ πρώτης όψεως είναι ξεκάθαρο πως η πληροφορία που λαμβάνουμε αλλά και οι άξονες πλέον έχουν αποδώσει την μέγιστη πληροφορία και έχουν διαφοροποιηθεί, με πολλούς από τους άξονες να μη βρίσκονται σε συσχέτιση μεταξύ τους Εικόνα 139. Η εφαρμογή της υβριδικής μεθόδου **NMF-PCA** απέδωσε χάρτες συνιστωσών με εξαιρετική ευκρίνεια και άμεση φυσικοχημική ερμηνεία, συνδυάζοντας τα πλεονεκτήματα και των δύο τεχνικών. Οι έξι χάρτες αφθονίας που προέκυψαν (Εικόνα 138) αποκαλύπτουν με σαφήνεια την παλέτα και την τεχνική του καλλιτέχνη:

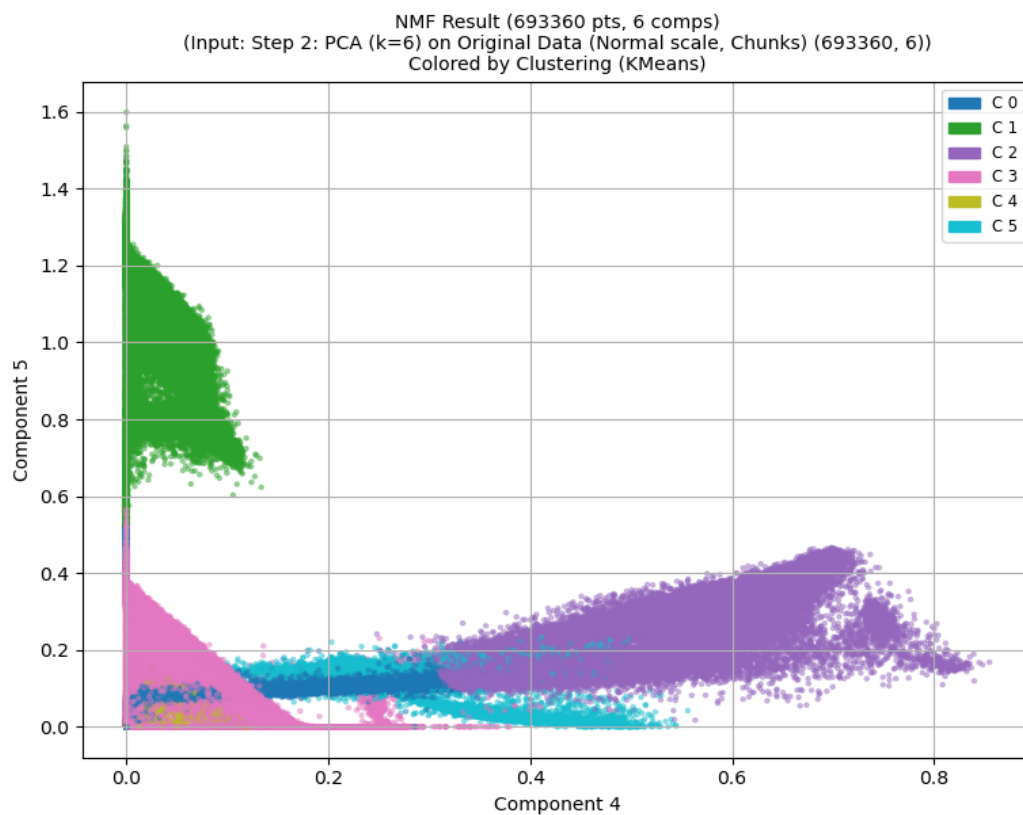
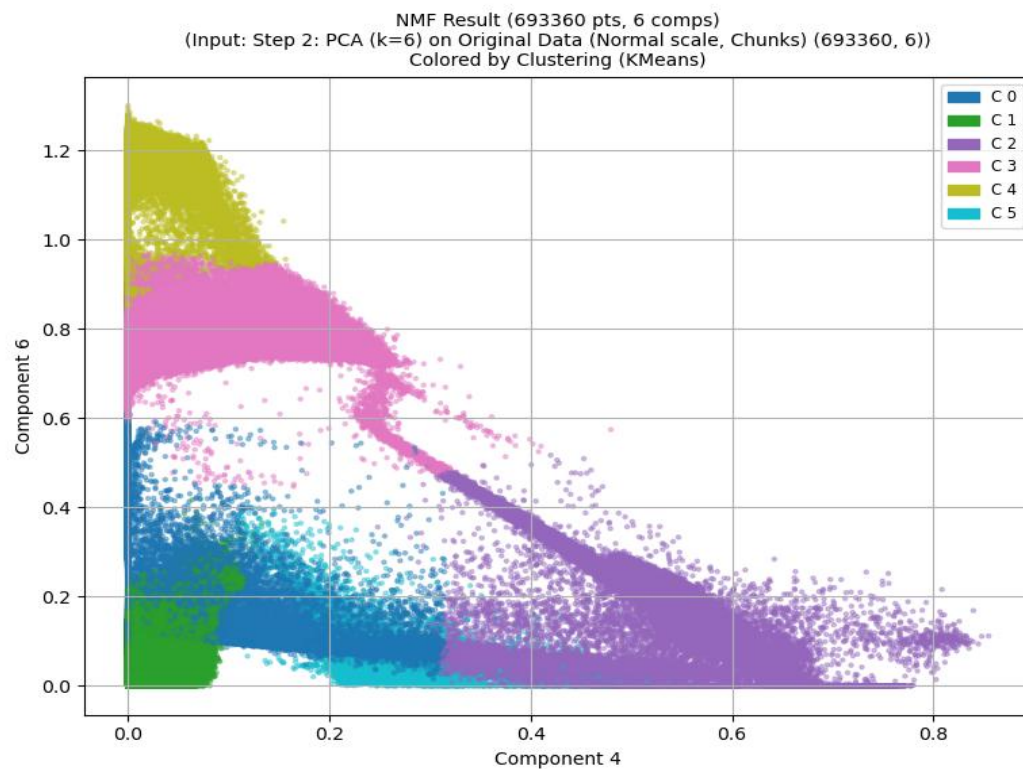
- **NMF-PCA Comp. 5** (Γαιώδεις Χρωστικές - **Fe**): Αυτή η συνιστώσα απομονώνει τις καφε-κόκκινες περιοχές, κυρίως στα ενδύματα. Η χωρική της κατανομή και το αντίστοιχο φάσμα της την ταυτοποιούν με τις γήινες χρωστικές με βάση τον Σίδηρο (Fe), όπως οι ώχρες και οι σιένες.
- **NMF-PCA Comp. 4,2** (Χρυσός - **Au** & Άργυρος - **Ag**): Η συνιστώσα αυτή χαρτογραφεί με απόλυτη ακρίβεια τις επιμεταλλωμένες περιοχές. Απομονώνει τα φωτοστέφανα που έχουν φιλοτεχνηθεί με Χρυσό (**Au**), καθώς και τις επιγραφές που περιέχουν Άργυρο (**Ag**).

- **NMF-PCA Comp. 3,5,6** (Λευκό του Μολύβδου - **Pb**): Πρόκειται για την πιο ευδιάκριτη συνιστώσα, η οποία αναδεικνύει αποκλειστικά τα σαρκώματα (πρόσωπα και χέρια). Η κατανομή της αντιστοιχεί απόλυτα στη χρήση του Λευκού του Μολύβδου (**Pb**), του βασικού υλικού για την απόδοση του δέρματος.
- **NMF-PCA Comp. 6** (Υπόστρωμα - **Ca, S**): Παρουσιάζει μια διάχυτη, ομοιογενή κατανομή που αντιστοιχεί στο στρώμα προετοιμασίας (*gesso*), με το σήμα της να είναι ισχυρότερο στις εκτεθειμένες περιοχές φθοράς. Το φάσμα της κυριαρχείται από το Ασβέστιο (**Ca**) και το Θείο (**S**) του γύψου.
- **NMF-PCA Comp. 4** (Σπάνιες Χρωστικές - **Hg, Cu**): Αυτή η συνιστώσα είναι ιδιαίτερα αποκαλυπτική, καθώς ομαδοποιεί υλικά που χρησιμοποιήθηκαν επιλεκτικά. Εντοπίζει το έντονο κόκκινο του ενδύματος της Παναγίας, που αντιστοιχεί στο κιννάβαρι (**HgS**), και το μπλε μαφόριο, που αντιστοιχεί σε χρωστική χαλκού (**Cu**).
- **NMF-PCA Comp. 1** (Φωτίσματα & Τεχνική): Αυτή η συνιστώσα δεν απομονώνει ένα μόνο υλικό, αλλά μια τεχνική. Χαρτογραφεί τις πιο φωτεινές περιοχές σε όλη την εικόνα, δηλαδή τα **highlights** στα σαρκώματα, τις πτυχωσεις των ενδυμάτων και τις λεπτομέρειες των φωτοστεφάνων. Φασματικά, κυριαρχείται από το **Pb** (καθαρό λευκό) και το **Au** (έντονη αντανάκλαση), αποκαλύπτοντας πώς ο καλλιτέχνης εφάρμοσε τα τελικά φωτίσματα για να δώσει όγκο και πνευματικότητα στο έργο.



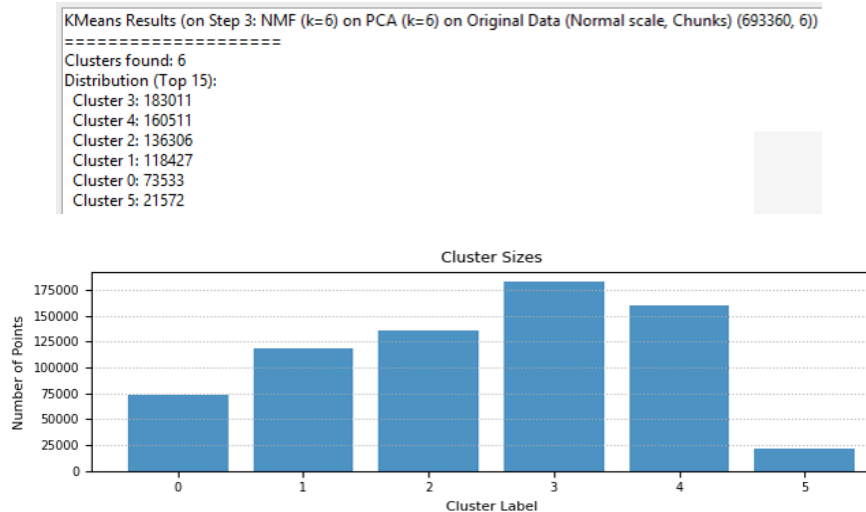
Εικόνα 139: Πίνακας συσχέτισης αξόνων μεθόδου NMF

Προχωρούμε επομένως και στην συσταδοποίησης με βάση της μεθόδου **Kmeans** όπως και στις προηγούμενες ανάλυσης με βάση τον αριθμό των **6** συστάδων.

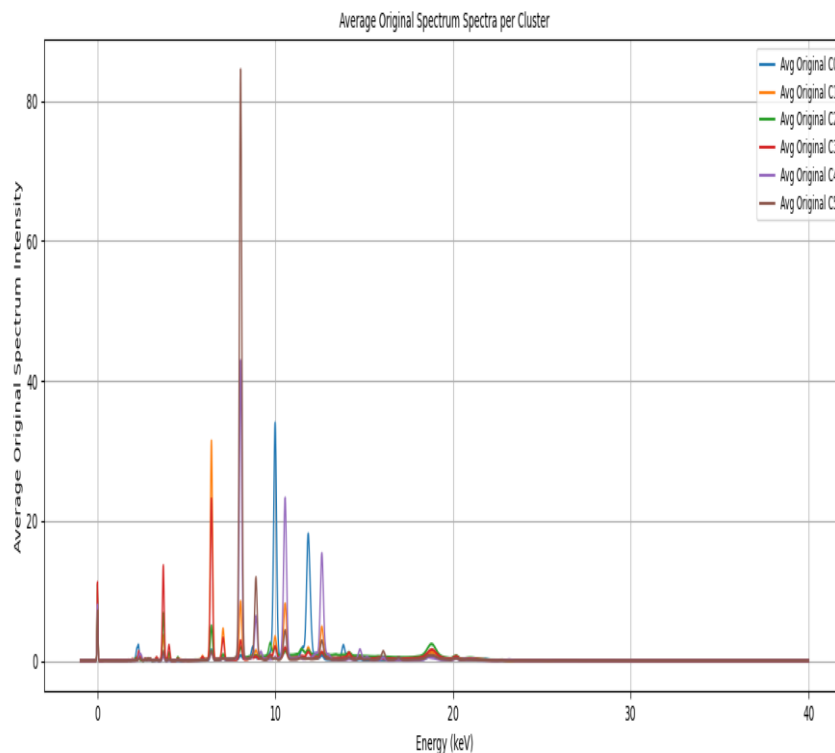


Εικόνα 140: Ενδεικτική απεικόνιση των Αξόνων από την μέθοδο NMF με βάση κάθε συστάδα διαφορετικού χρώματος

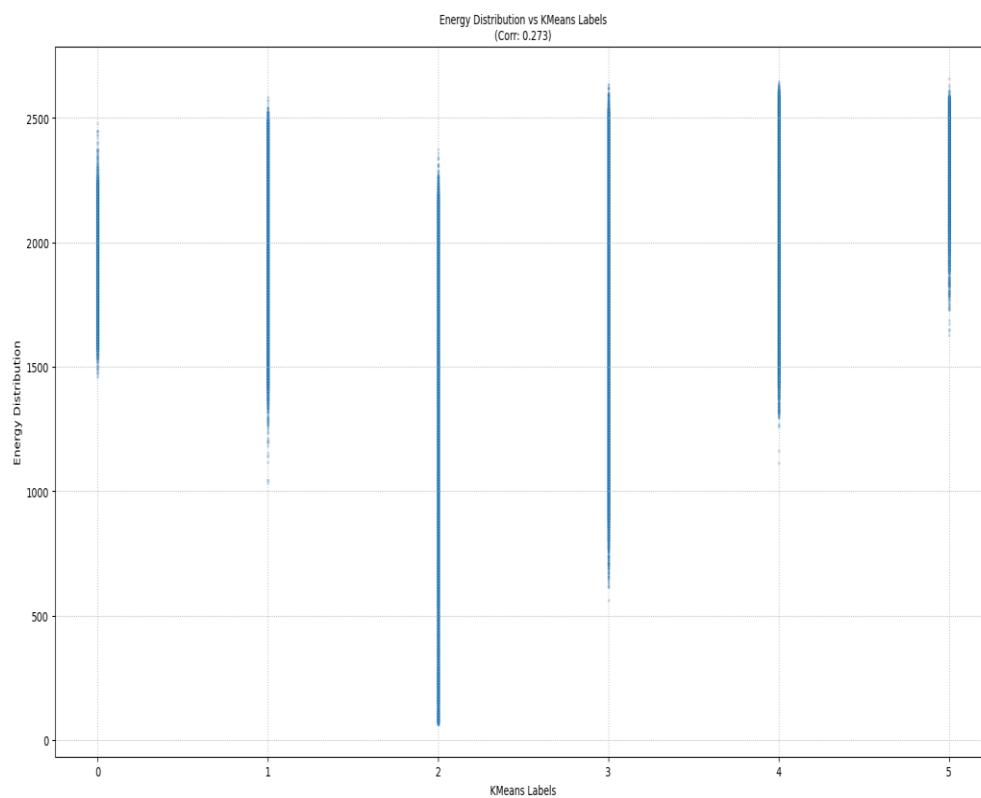
Οι παραπάνω εικόνες αποτελούν την απεικόνιση της συσταδοποίησης πάνω στην απεικόνιση της σύγκρισης των **Component** της **NMF** μεταξύ τους Εικόνα 140. Παρακάτω παραθέτουμε και την τα αποτελέσματα της μεθόδου Kmeans Εικόνα 141 και της εικόνας των συστάδων επάνω στην αρχική εικόνα, αλλά όπως επίσης και το φάσμα της κάθε συστάδας επάνω στο ολικό φάσμα Εικόνα 142.



Εικόνα 141: Γραφική Απεικόνιση συσταδοποίησης, με πλήθος δεδομένων ανά συστάδα

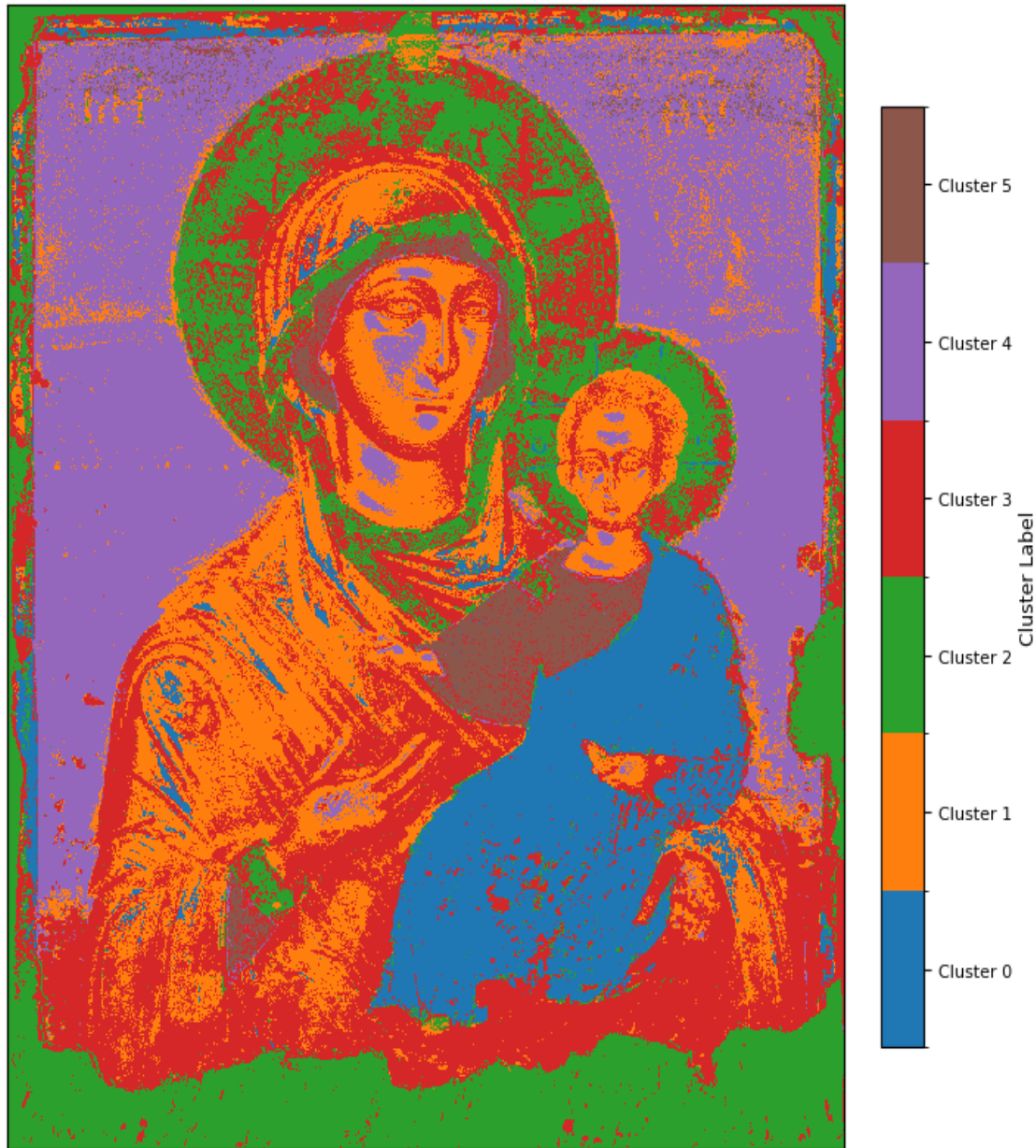


Εικόνα 142: Εικόνα απεικόνισης των Συστάδων με βάση το Φάσμα συναρτήσει της ενέργειας.



Εικόνα 143: Απεικόνιση του εύρους ενέργειας με βάση τα κανάλια κάθε συστάδας

KMeans Cluster Map
(Input: Step 3: NMF (k=6) on PCA (k=6) on Original Data (Normal scale, Chunks) (693360, 6))



Εικόνα 144: Αποτέλεσμα συσταδοποίησης μέσω μεθόδου Kmeans

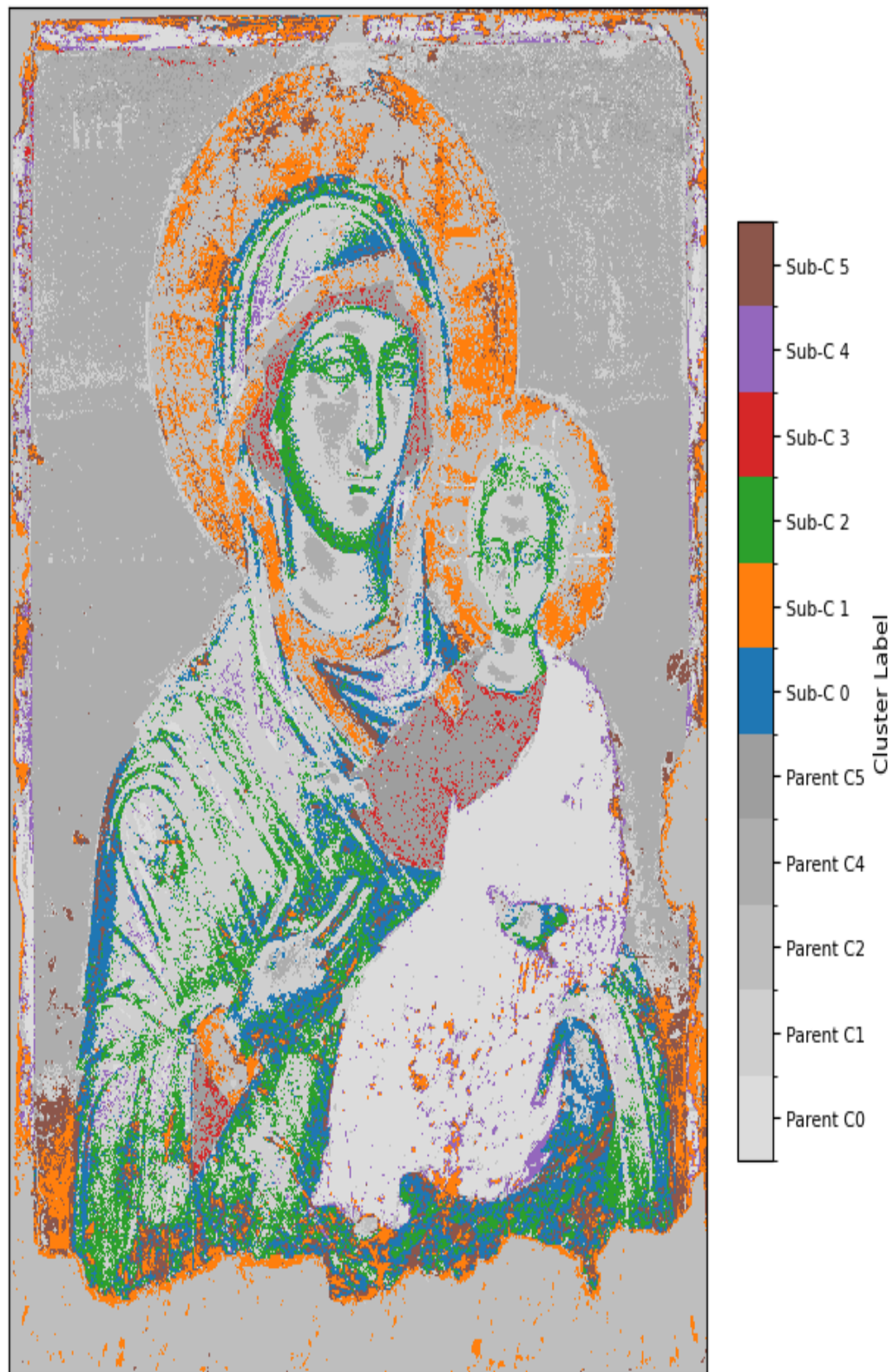
Τώρα παρατηρούμε Εικόνα 144 ότι η πληροφορία που μπορούμε να λάβουμε βάση τα στοιχεία αλλά και τη συσταδοποίηση είναι η μέγιστη δυνατή. Παρατηρούμε πως πλέον διακρίνουμε τα γράμματα στην εικόνα τα οποία είναι απόρροια δυο συστάδων, το στέμμα επίσης, αλλά πλέον το υπόβαθρο ορθά κατανεμήθηκε σε μια συστάδα. Πάλι όμως περιοχή ενδιαφέροντος αποτελεί το στέμμα, οπότε και προχωρούμε στις εξής απεικονίσεις. Θα λάβουμε τις συστάδες 2 και 3 για **subclustering** με την ίδια μεθοδολογία όπως και πριν Εικόνα 146 Εικόνα 147.

KMeans Cluster Map
(Input: Step 3: NMF (k=6) on PCA (k=6) on Original Data (Normal scale, Chunks) (693360, 6))



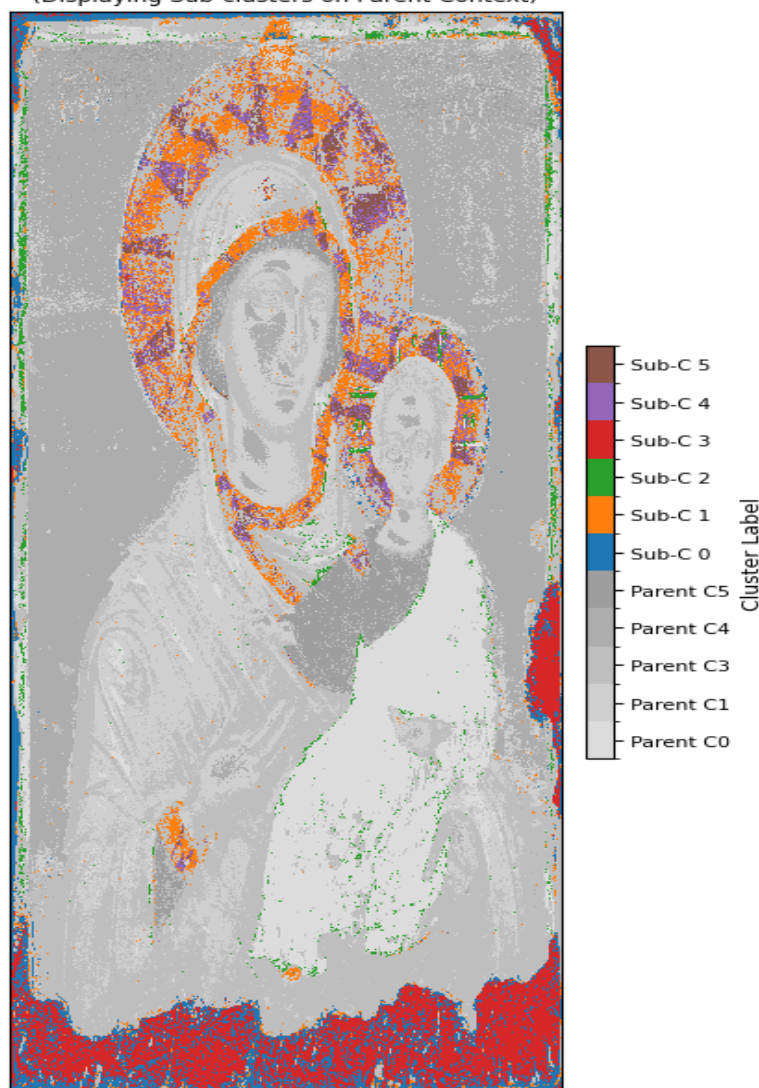
Εικόνα 145: Συστάδα 3 απομονωμένη

KMeans Cluster Map
 (Input: Step 5: Data from Cluster 3 (Parent: Step 3: NMF (k=6) on PCA (k=6)...) (183011, 6))
 (Displaying Sub-clusters on Parent Context)



Εικόνα 146: Αποτέλεσμα SubClustering συστάδας (3) με έξι συστάδες

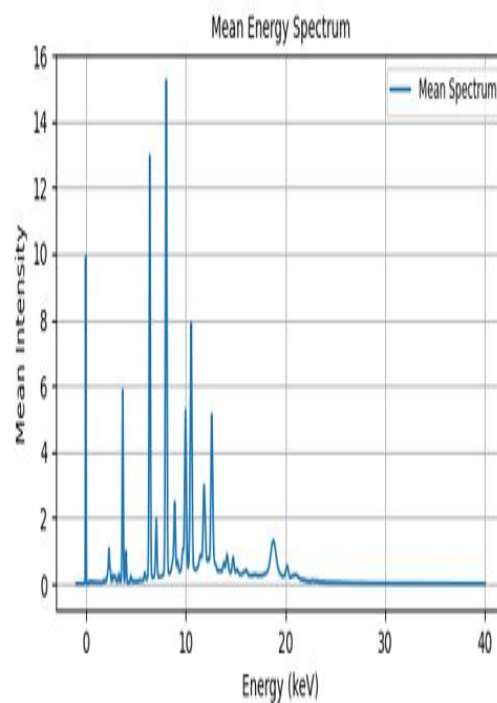
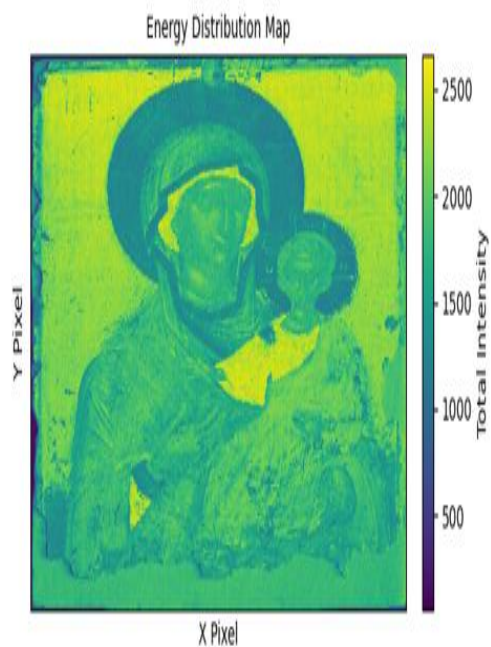
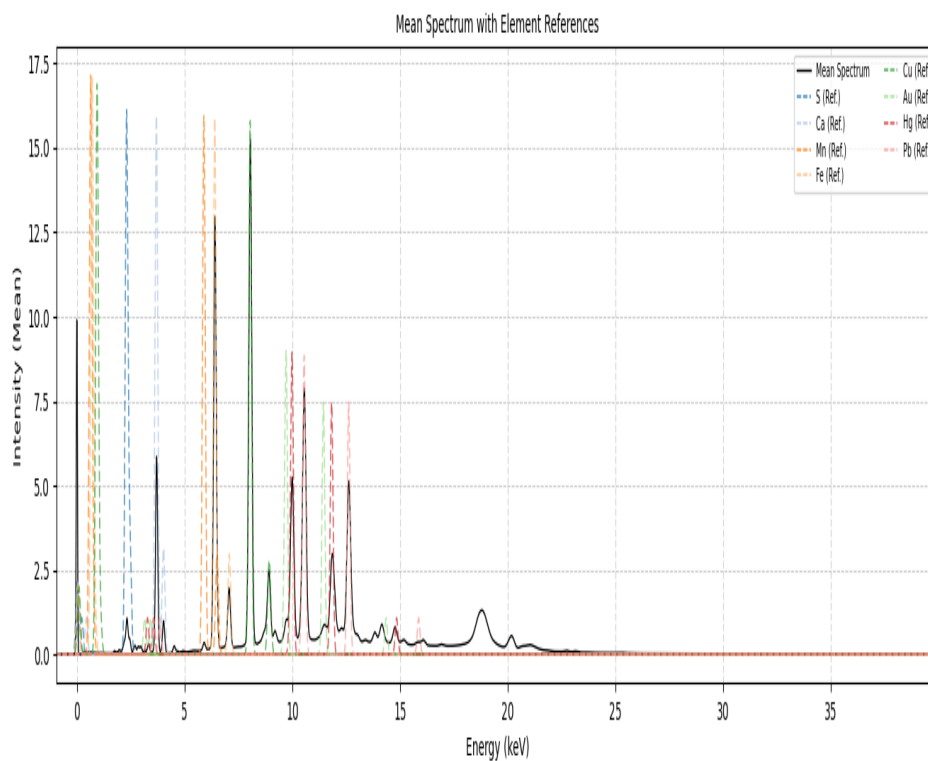
KMeans Cluster Map
(Input: Step 9: Data from Cluster 2 (Parent: Step 3: NMF (k=6) on PCA (k=6)...) (136306, 6))
(Displaying Sub-clusters on Parent Context)



Εικόνα 147: Αποτέλεσμα SubClustering συστάδας (2) με έξι συστάδες

Διακρίνουμε λοιπόν πως οι αρχικές **συστάδες 2** (Εικόνα 147) και **3** (Εικόνα 146) αποτελούν τον κύριο παράγοντα τους στέμματος των Αγίων Προσώπων αλλά και τις περιοχές που είναι έντονες και είναι οριακές, δηλαδή περιοχές μετάβασης των υλικών. Για την άμεση ταυτοποίηση και τη χωρική απεικόνιση των επιμέρους χημικών στοιχείων που συνθέτουν την εικόνα, πραγματοποιήθηκε στοιχειακή χαρτογράφηση. Αυτό επιτεύχθηκε μέσω του καθορισμού ενεργειακών παραθύρων που αντιστοιχούν στις χαρακτηριστικές γραμμές φθορισμού κάθε ενδιαφέροντος στοιχείου και της ολοκλήρωσης της έντασης εντός αυτών των παραθύρων για κάθε pixel.

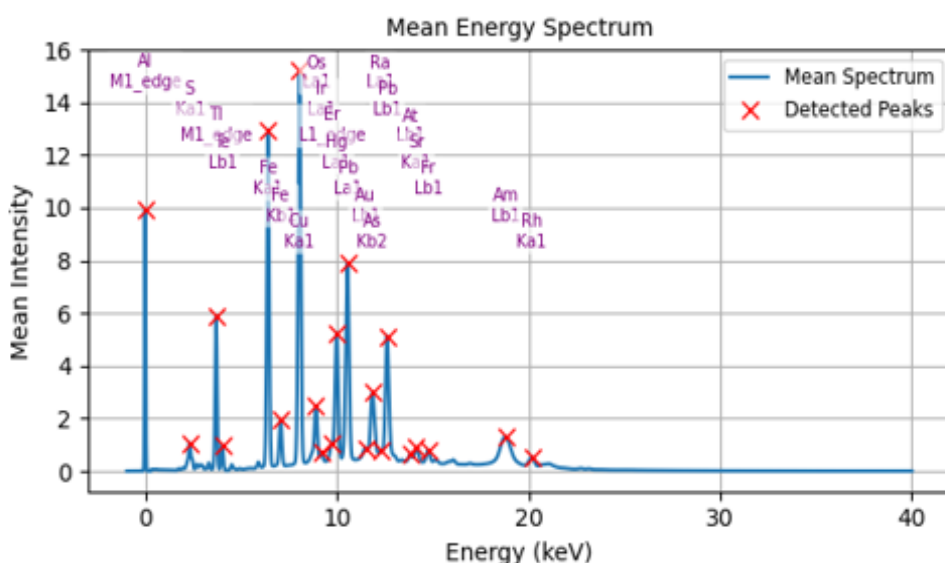
Προκειμένου να γίνει ακριβής αντιστοίχιση των ενεργειακών καναλιών με τις θεωρητικές ενέργειες των γραμμών φθορισμού, εφαρμόστηκε βαθμονόμηση του ενεργειακού άξονα. Μέσω της λειτουργικότητας του αναπτυχθέντος λογισμικού, ορίστηκαν οι παράμετροι βαθμονόμησης **offset (A) = -0.95529552 keV** και **gain (B) = 0.010001 keV/κανάλι**. Το αποτέλεσμα της βαθμονόμησης φαίνεται στο μέσο φάσμα ενέργειας που παρουσιάζεται στην Εικόνα 148 (αντίστοιχο της Εικόνα 122 πάνω), όπου οι κύριες κορυφές έχουν πλέον αντιστοιχιστεί σε συγκεκριμένες ενεργειακές τιμές (keV).



Εικόνα 148: Απεικόνιση Επιλεγμένων στοιχείων στο φάσμα (επάνω), και ολική κατανομή έντασης πάνω στην εικόνα (κάτω αριστερά) και φάσμα ενέργειας (κάτω δεξιά)

Το λογισμικό παρέχει τη δυνατότητα αυτόματης ανίχνευσης κορυφών (**peak finding**) στο μέσο φάσμα έντασης Εικόνα 149. Στη συνέχεια, επιχειρείται η αυτόματη συσχέτιση

(*peak matching*) αυτών των ανιχνευμένων κορυφών με τις θεωρητικές ενέργειες γραμμών φθορισμού από μια ενσωματωμένη βιβλιοθήκη *XRF (XRFLib)*. Τα αποτελέσματα αυτής της αυτόματης διαδικασίας φαίνονται στο αναδυόμενο παράθυρο της Εικόνα 149. Όπως παρατηρείται, ορισμένες ανιχνευμένες κορυφές αντιστοιχίζονται σε συγκεκριμένες γραμμές στοιχείων (π.χ., *Al M1_edge* στα **-0.005 keV**, *S Ka1* στα **2.325 keV**, *Fe Ka1* στα **6.405 keV**, κ.λπ.). Ωστόσο, η αυτόματη αυτή αντιστοίχιση μπορεί να οδηγήσει σε εσφαλμένες ή μη ρεαλιστικές ταυτοποιήσεις, ειδικά για ασθενείς κορυφές ή σε περιοχές με επικαλύψεις, καθώς ο αλγόριθμος αξιολόγησης βασίζεται κυρίως στην εγγύτητα της ενέργειας χωρίς να λαμβάνει υπόψη το συνολικό φασματικό 'αποτύπωμα' ή τη σχετική ένταση των γραμμών μιας σειράς (π.χ., *Ka/Kβ*). Για τον λόγο αυτό, η αυτόματη ανίχνευση χρησιμοποιείται κυρίως ως καθοδηγητικό Εργαλείο.



Peak Matching Results

Matched Peaks:

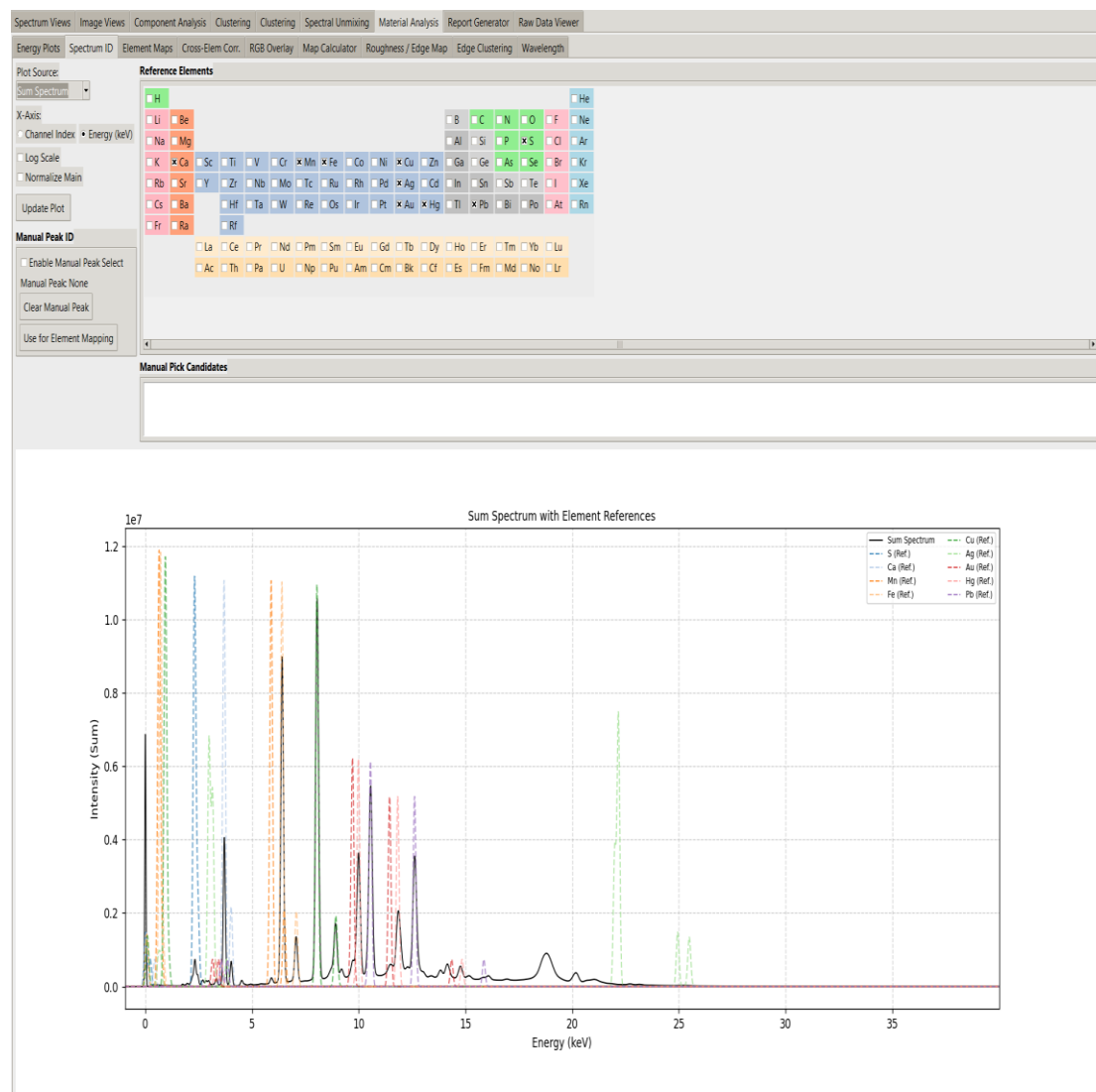
Element	Line	Lib.(keV)	Det.(keV)	Diff(keV)
Al	M1_edge	0.001	-0.005	0.006
S	Ka1	2.308	2.325	0.017
Ti	M1_edge	3.704	3.705	0.001
Te	Lb1	4.029	4.025	0.004
Fe	Ka1	6.404	6.405	0.002
Fe	Kb1	7.058	7.066	0.008
Cu	Ka1	8.048	8.046	0.002
Os	La1	8.911	8.916	0.005
Ir	La1	9.175	9.196	0.021
Er	L1_edge	9.751	9.736	0.016
Hg	La1	9.989	9.996	0.007
Pb	La1	10.551	10.556	0.005
Au	Lb1	11.443	11.496	0.053
As	Kb2	11.864	11.856	0.008
Ra	La1	12.339	12.286	0.053
Pb	Lb1	12.614	12.626	0.012
At	Lb1	13.875	13.836	0.039
Sr	Ka1	14.165	14.156	0.009
Fr	Lb1	14.770	14.766	0.004
Am	Lb1	18.856	18.797	0.059
Rh	Ka1	20.216	20.177	0.039

OK

Εικόνα 149: Αποτέλεσμα Κατανομής Κορυφών με βάση τον αλγόριθμο απόλυτης ταύτισης. Χωρίς όμως την συνολική το βαθμονομημένο αλγόριθμο συνολικής ταύτισης που παρέχεται στο επόμενο παράθυρο

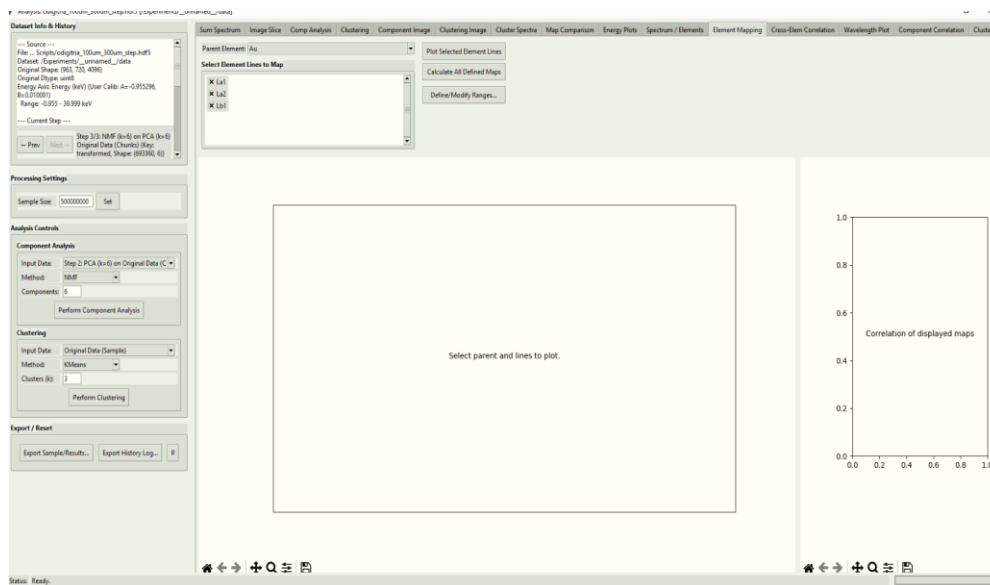
Για μεγαλύτερη ακρίβεια και έλεγχο, το λογισμικό επιτρέπει τη χειροκίνητη επιλογή των στοιχείων ενδιαφέροντος μέσω ενός διαδραστικού περιοδικού πίνακα Εικόνα 150.

Ο χρήστης μπορεί να επιλέξει τα στοιχεία των οποίων οι χαρακτηριστικές γραμμές είναι ορατές στο φάσμα ή αναμένονται με βάση την τεχντροπία και τα υλικά της εποχής. Στην προκειμένη περίπτωση, επιλέχθηκαν τα στοιχεία [**Au, Pb, Fe, Mn, Ca, S, Cu**, όπως φαίνονται στην κάτω]. Για κάθε επιλεγμένο στοιχείο και τις κύριες γραμμές φθορισμού του (π.χ., **K α** , **K β** , **L α** , L β), ορίζονται αυτόματα ή μπορούν να τροποποιηθούν χειροκίνητα τα ενεργειακά παράθυρα ολοκλήρωσης (**Regions of Interest, ROIs**). Η οπτικοποίηση των επιλεγμένων στοιχείων και των αντίστοιχων γραμμών τους πάνω στο μέσο φάσμα φαίνεται στην Εικόνα 150.



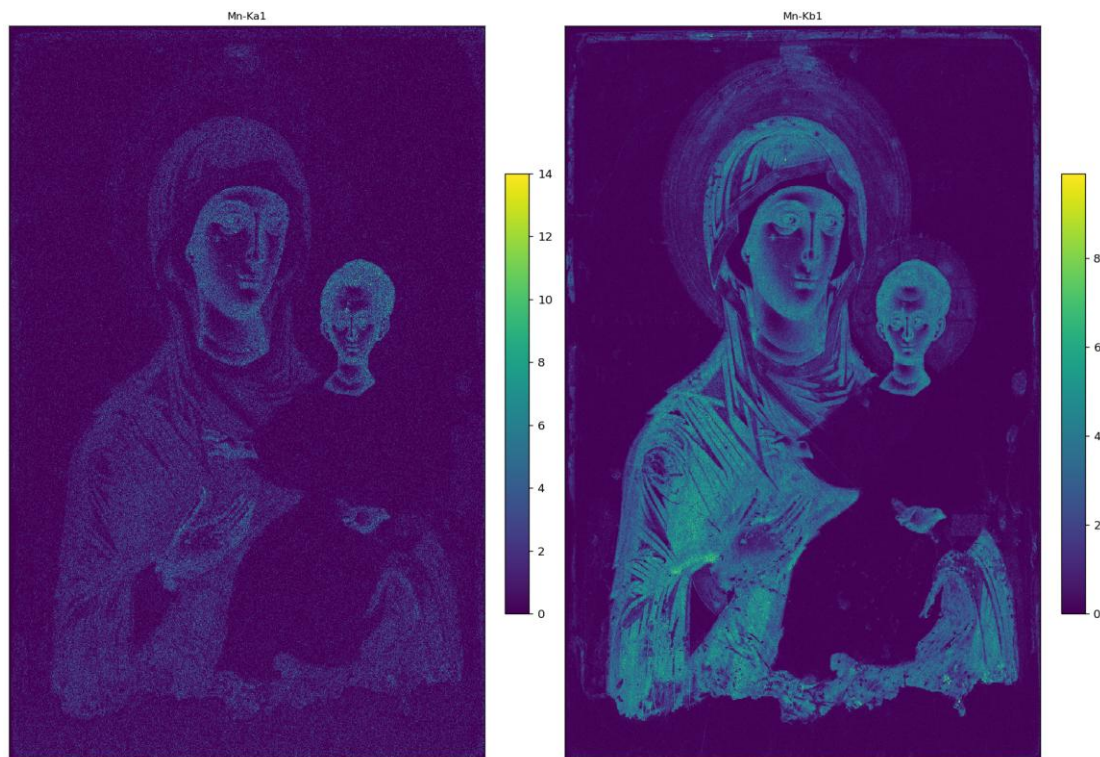
Εικόνα 150: Παράθυρο επιλογής στοιχείων που ταιριάζουν στο φάσμα με βάση την ενέργεια ή τα κανάλια. Όπως επίσης και επιλογή Κορυφών για ανάγνωση πιθανών Στοιχείων.

Μετά τον καθορισμό των στοιχείων και των αντίστοιχων ενεργειακών παραθύρων τους, ενεργοποιείται η διαδικασία δημιουργίας των στοιχειακών χαρτών μέσω της επιλογής '**Calculate All Defined Maps**' Εικόνα 151.



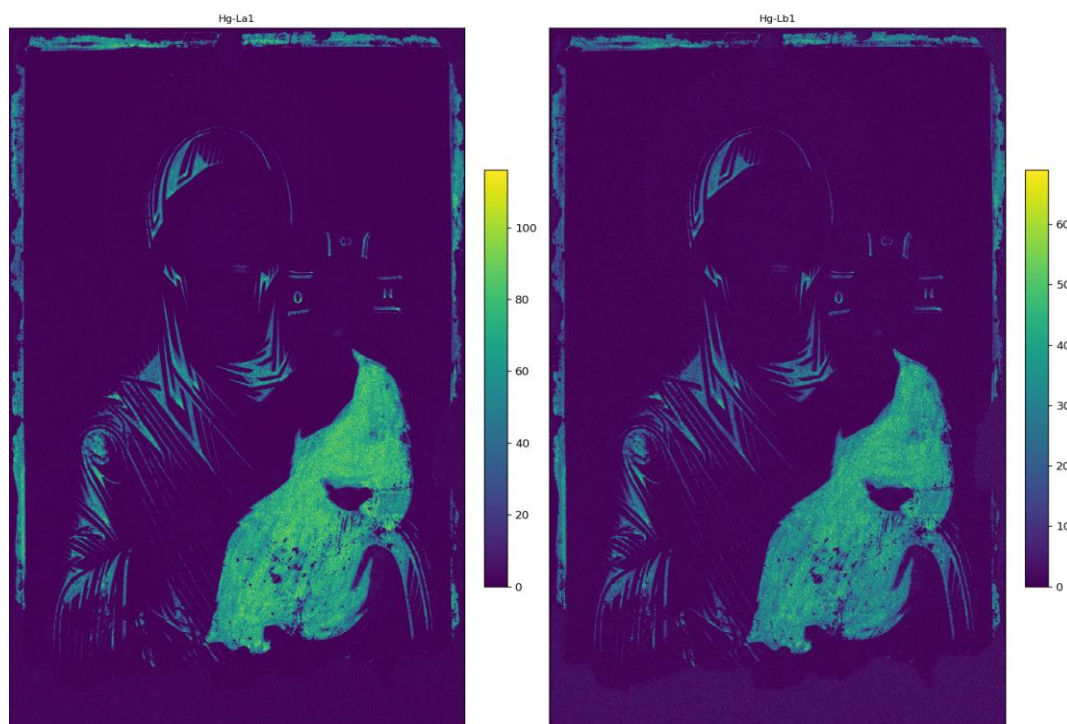
Εικόνα 151: Παράθυρο Υπολογισμού Στοιχειομετρικών Χαρτών μέσω του αναπτυσσόμενου προγράμματος

Για κάθε **pixel** της εικόνας, το λογισμικό υπολογίζει τη συνολική ένταση φθορισμού εντός του καθορισμένου ενεργειακού παραθύρου για κάθε επιλεγμένη γραμμή στοιχείου. Οι τιμές αυτές της έντασης αποτελούν τις τιμές των **pixels** στους αντίστοιχους στοιχειακούς χάρτες. Οι χάρτες που προκύπτουν απεικονίζουν τη χωρική κατανομή της έντασης κάθε επιλεγμένης γραμμής στοιχείου στην επιφάνεια της εικόνας. Οι φωτεινότερες περιοχές υποδηλώνουν υψηλότερη σχετική συγκέντρωση του αντίστοιχου στοιχείου.



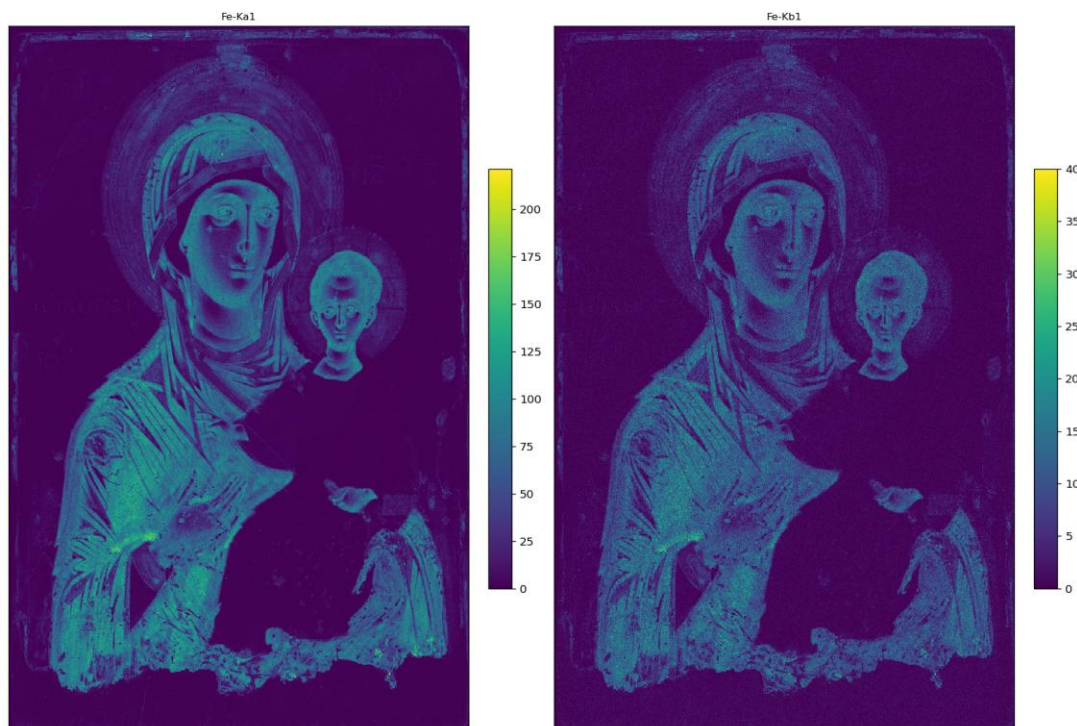
Εικόνα 152: Στοιχειομετρικός Χάρτης Μαγγανίου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K

Μαγγάνιο (Mn): Το μαγγάνιο μπορεί να υπάρχει ως συστατικό σε ορισμένες γαιώδεις χρωστικές (π.χ., *umber*) ή να έχει χρησιμοποιηθεί για τη δημιουργία σκούρων αποχρώσεων ή περιγραμμάτων Εικόνα 152.



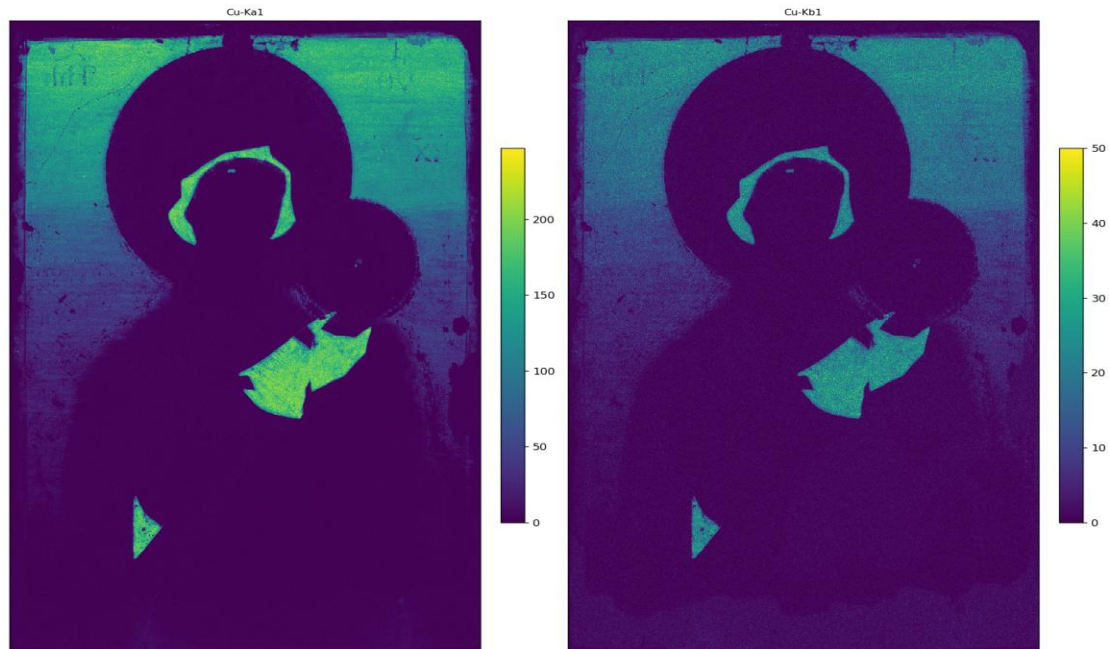
Εικόνα 153: Στοιχειομετρικός Χάρτης Υδραργύρου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές L

Υδράργυρος (Hg): Ο υδράργυρος Εικόνα 153 εντοπίζεται σε έντονα κόκκινες περιοχές, κυρίως στο ένδυμα του Ιησού αλλά και σε γραμμές στο ένδυμα της Παναγίας, και σχετίζεται με τη χρήση κινναβάρεως (*HgS*), μιας συνηθισμένης κόκκινης χρωστικής.



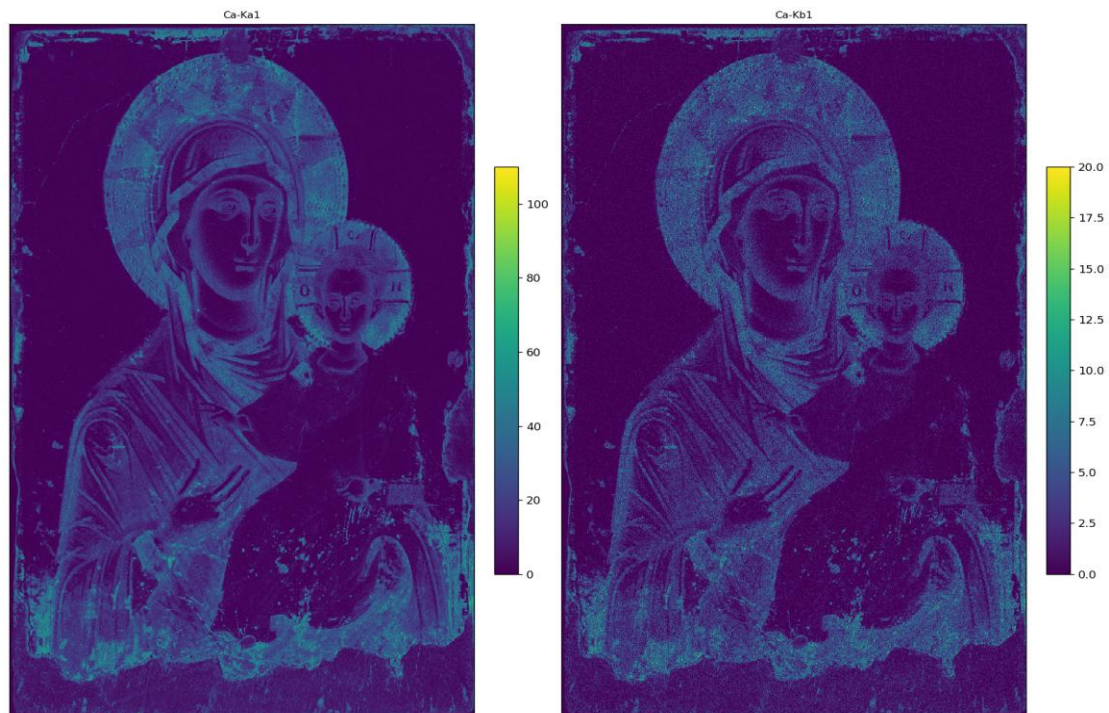
Εικόνα 154: Στοιχειομετρικός Χάρτης Σιδήρου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K

Σίδηρος (Fe): Ο σίδηρος ανιχνεύεται κυρίως στις καφέ και κοκκινωπές αποχρώσεις των μαλλιών, σε ορισμένες σκιές των ενδυμάτων αλλά και στον χιτώνα της Παναγίας και πιθανόν στο προπλάσμο των προσώπων. Αυτό υποδηλώνει τη χρήση γαιωδών χρωστικών όπως οι ώχρες (κόκκινη ώχρα, κίτρινη ώχρα, σιένα). Εμφανής στα κατώτερα τμήματα της μορφής της Παναγίας Εικόνα 154.



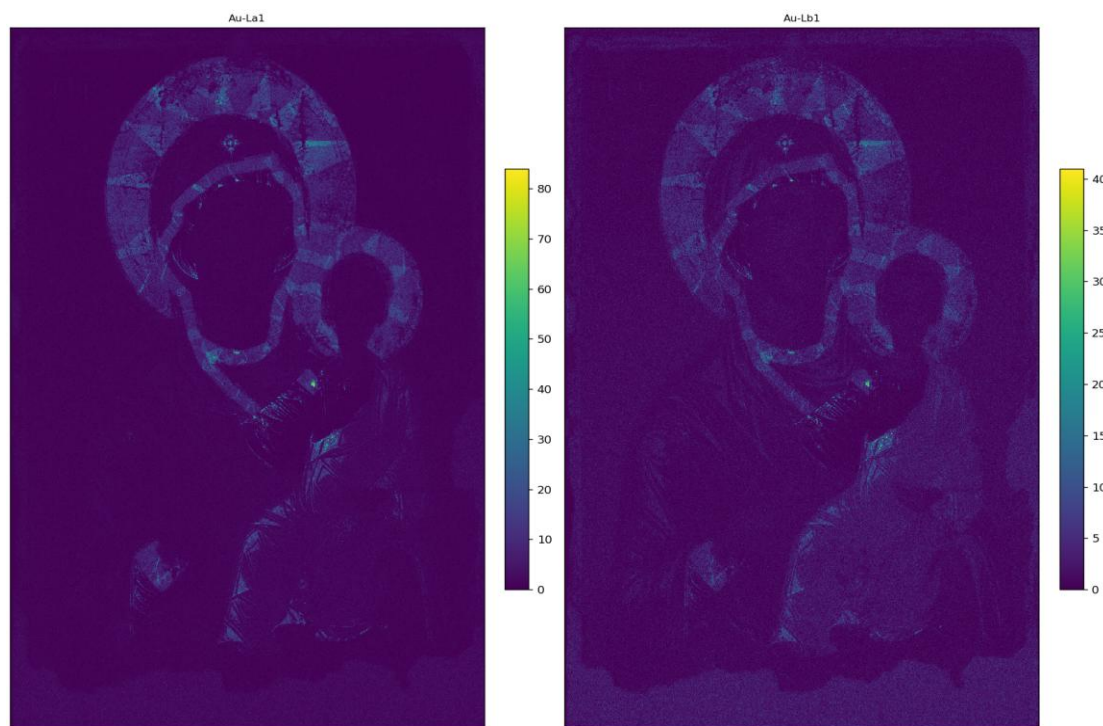
Εικόνα 155: Στοιχειομετρικός Χάρτης Χαλκού από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K

Χαλκός (Cu): Ο χαλκός Εικόνα 155 εντοπίζεται κυρίως στον χιτώνα του Χριστού και σε τμήματα του ενδύματος της Παναγίας), υποδεικνύοντας τη χρήση χρωστικών με βάση τον χαλκό, όπως ο μαλαχίτης (πράσινο).



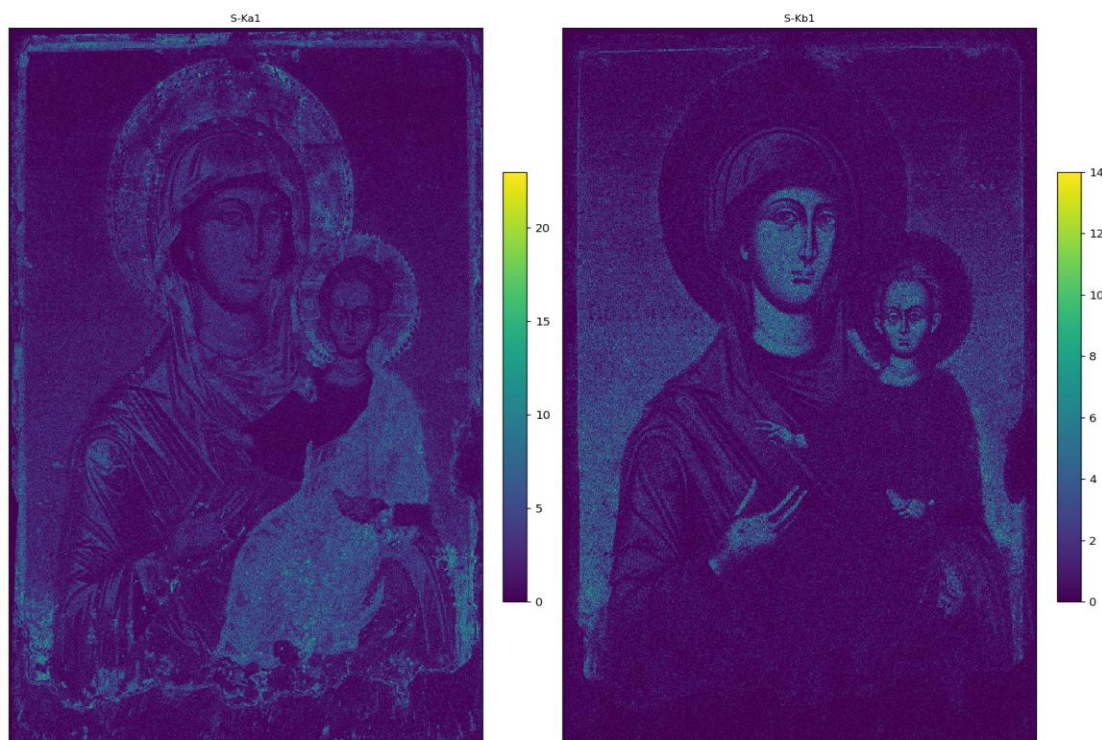
Εικόνα 156: Στοιχειομετρικός Χάρτης Ασβεστίου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K

Ασβέστιο (Ca): Το ασβέστιο Εικόνα 156 παρουσιάζει ευρεία κατανομή, με ιδιαίτερα υψηλές συγκεντρώσεις στις περιοχές προετοιμασίας (*gesso ground*), αλλά και σε ανάμειξη με χρωστικές, πιθανόν ως πληρωτικό υλικό ή ως ανθρακικό ασβέστιο (κιμωλία). Η συνύπαρξή του με θείο (*S*) σε ορισμένες περιοχές μπορεί να υποδηλώνει την παρουσία γύψου ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$)



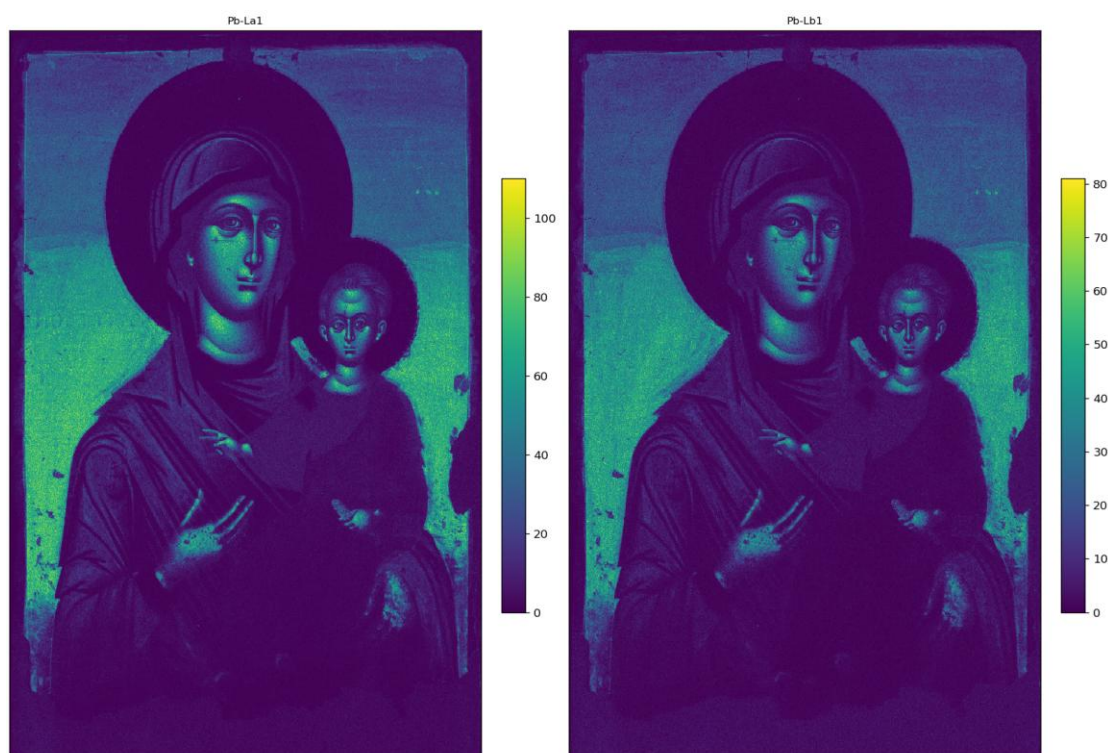
Εικόνα 157: Στοιχειομετρικός Χάρτης Χρυσού από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές *L*

Χρυσός (Au): Οι στοιχειακοί χάρτες για τις γραμμές *L* του χρυσού (*Au-La1*, *Au-La2*, *Au-Lb1* - βλ. Εικόνα 157) αποκαλύπτουν την παρουσία του στοιχείου αυτού κυρίως στα φωτοστέφανα της Παναγίας και του Ιησού, καθώς και σε διακοσμητικές λεπτομέρειες των ενδυμάτων και των επιγραφών (π.χ., "*ΜΡ ΘΥ*", "*ΙC ΧC*", "*Η ΟΔΗΓΗΤΡΙΑ*"). Η κατανομή είναι αρκετά σαφής και οριοθετημένη, υποδηλώνοντας την τεχνική του χρυσώματος (*gilding*). Η σχετικά χαμηλή ένταση του σήματος του χρυσού, παρά την οπτική εντύπωση, είναι συμβατή με τη χρήση πολύ λεπτών φύλλων χρυσού, μια συνήθης πρακτική στη βυζαντινή και μεταβυζαντινή αγιογραφία. Παρατηρούνται επίσης ασθενέστερες ενδείξεις χρυσού σε άλλες περιοχές, πιθανόν λόγω φθοράς, μεταφοράς υλικού ή ιχνών από προγενέστερες επεμβάσεις. Η σύγκριση των χαρτών των διαφορετικών γραμμών *L* του χρυσού (*La1*, *La2*, *Lb1*) δείχνει παρόμοια χωρική κατανομή, επιβεβαιώνοντας την ταυτοποίηση του στοιχείου.



Εικόνα 158: Στοιχειομετρικός Χάρτης Θείου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K

Θείο (S): Το θείο Εικόνα 158, εκτός από την πιθανή του παρουσία ως συστατικό του γύψου (σε συνδυασμό με **Ca**), εντοπίζεται και σε μπλε περιοχές, γεγονός που θα μπορούσε να υποδηλώνει τη χρήση κυανού (που περιέχει θειούχες ενώσεις του λαζουρίτη)."

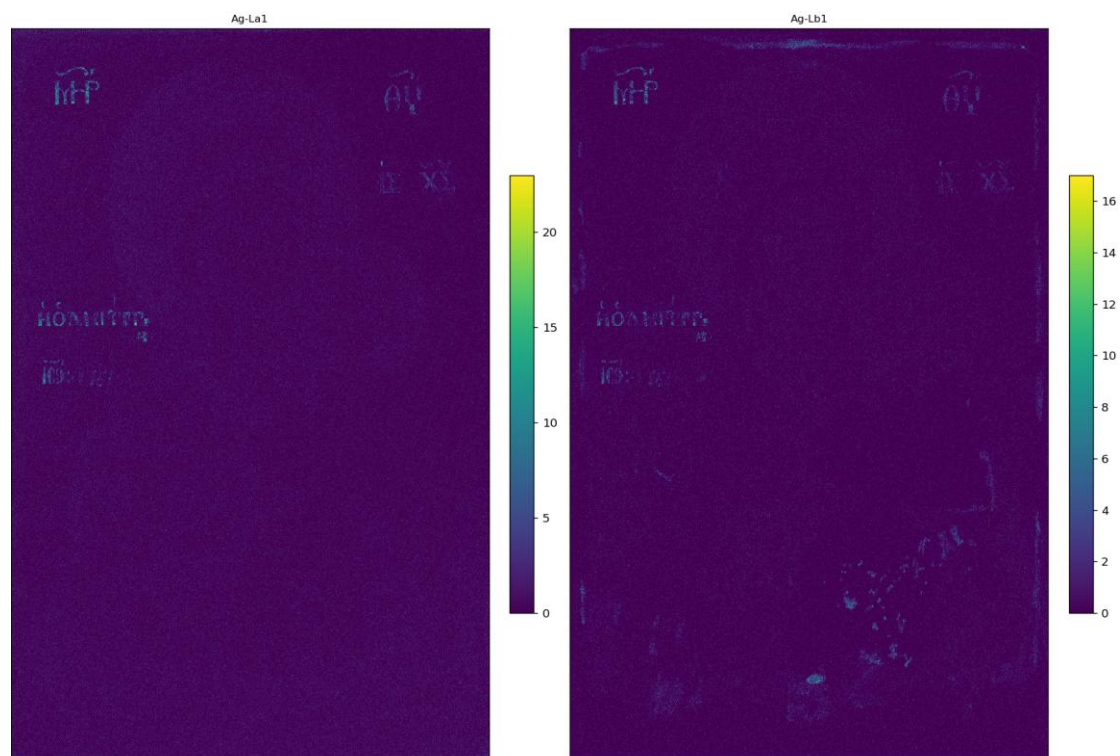


Εικόνα 159: Στοιχειομετρικός Χάρτης Μολύβδου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές L

Μόλυβδος (Pb): Εντοπίζεται εκτενώς στις λευκές και ανοιχτόχρωμες περιοχές των προσώπων (*incarnates*), των χεριών, καθώς και σε φωτεινά σημεία των ενδυμάτων, υποδηλώνοντας τη χρήση λευκού του μολύβδου ($2PbCO_3 \cdot Pb(OH)_2$) ως βασικής λευκής χρωστικής Εικόνα 159. Η παρουσία του σε ορισμένες κόκκινες περιοχές μπορεί να υποδηλώνει τη μινίου (Pb_3O_4).

Οι χάρτες κατανομής των στοιχείων αποκαλύπτουν σαφή μοτίβα τοποθετοποίησης, υποδεικνύοντας χημική ετερογένεια στο δείγμα:

- Ο Σίδηρος (**Fe**) και το Ασβέστιο (**Ca**) παρουσιάζουν μερική συνύπαρξη σε ορισμένες περιοχές, γεγονός που θα μπορούσε να ερμηνευθεί ως ανάμειξη γαιωδών χρωστικών (πλούσιων σε **Fe**) με υλικά προετοιμασίας ή πληρωτικά υλικά πλούσια σε **Ca**.
- Η ασθενής χωρική συσχέτιση του Ασβεστίου (**Ca**) με το Θείο (**S**) σε ορισμένες περιοχές μπορεί να υποδηλώνει τον περιορισμένο σχηματισμό θεικού ασβεστίου (γύψου).



Εικόνα 160: : Στοιχειομετρικός Χάρτης Αργύρου Ag από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές L

Άργυρος (Ag): Οι στοιχειακοί χάρτες για τις γραμμές **L** του Αργύρου (**Ag-La1**, **Ag-Lb1**) αποκαλύπτουν μια ενδιαφέρουσα και απροσδόκητη πληροφορία. Η παρουσία του αργύρου είναι εξαιρετικά τοπική και εντοπίζεται αποκλειστικά στις επιγραφές που υπάρχουν στην εικόνα (π.χ., "**MP ΘΥ**", "**IC XC**", "Η ΟΔΗΓΗΤΡΙΑ"). Σε αντίθεση με τον χρυσό που χρησιμοποιήθηκε για τα φωτοστέφανα, ο καλλιτέχνης επέλεξε ένα διαφορετικό, λιγότερο πολύτιμο αλλά εξίσου λαμπερό μέταλλο για την απόδοση των γραμμάτων. Αυτή η επιλεκτική χρήση υποδηλώνει μια συνειδητή καλλιτεχνική απόφαση, πιθανώς για λόγους κόστους ή για τη δημιουργία μιας διακριτής οπτικής αντίθεσης. Η ανίχνευση του αργύρου, που συχνά μπορεί να συγχέεται με άλλα υλικά, επιβεβαιώνει την υψηλή ευαισθησία της τεχνικής **MA-XRF** στην ταυτοποίηση ευγενών μετάλλων.

Σύγκριση με την Βιβλιογραφία (*Gerodimos et al. (2022)* και *Gerodimos et al. (2024)*)

Η παρούσα ενότητα αποσκοπεί στην τοποθέτηση των ευρημάτων της ανάλυσης της εικόνας 'Παναγία η Καθοδηγήτρια' (εφεξής "υπό μελέτη εικόνα") στο ευρύτερο πλαίσιο της επιστημονικής έρευνας για τη μελέτη έργων τέχνης, και συγκεκριμένα βυζαντινών/μεταβυζαντινών εικόνων, μέσω *MA-XRF* και προηγμένων τεχνικών ανάλυσης δεδομένων. Για τον σκοπό αυτό, τα αποτελέσματα θα συγκριθούν με δύο συναφείς δημοσιευμένες μελέτες: τη μελέτη των *Gerodimos et al. (2022)* ["Scanning X-ray Fluorescence Data Analysis for the Identification of Byzantine Icons' Materials, Techniques, and State of Preservation: A Case Study"], η οποία εξετάζει μια εικόνα της Παναγίας "Οδηγήτριας" του 18ου αιώνα, και τη μελέτη των *Gerodimos et al. (2024)* ["MA-XRF datasets analysis based on convolutional neural network: A case study on religious panel paintings"], η οποία παρουσιάζει μια προσέγγιση βασισμένη σε Συνελκτικά Νευρωνικά Δίκτυα (*CNN*) για την ανάλυση δεδομένων *MA-XRF* από θρησκευτικές εικόνες, συμπεριλαμβανομένης και της εικόνας "Παναγία η Οδηγήτρια" και άλλων εικόνων.

Αρχικά θα παρατεθεί ο πίνακας για τις εν λόγω καταστάσεις και πειραματικές διατάξεις των ερευνών.

Πίνακας 2: Πίνακας πειραματικών διατάξεων

Παράμετρος	Υπό Μελέτη Εικόνα	[Gerodimos 2022]	[Gerodimos 2024] - Dataset C
Πηγή (Άνοδος)	Ρόδιο (Rh)	Βολφράμιο (W)	Ρόδιο (Rh)
Τάση / Ένταση	50 kV / 600 μ A	50 kV / 800 μ A	50 kV / 600 μ A
Μέγεθος Pixel	300 μ m	1000 μ m (1 mm)	300 μ m
Χρόνος / Pixel	10 ms	2000 ms (2 s)	10 ms
Σύνολο Pixels	~693k	~2.7k	~674k
Κανάλια	4096	2048	4096

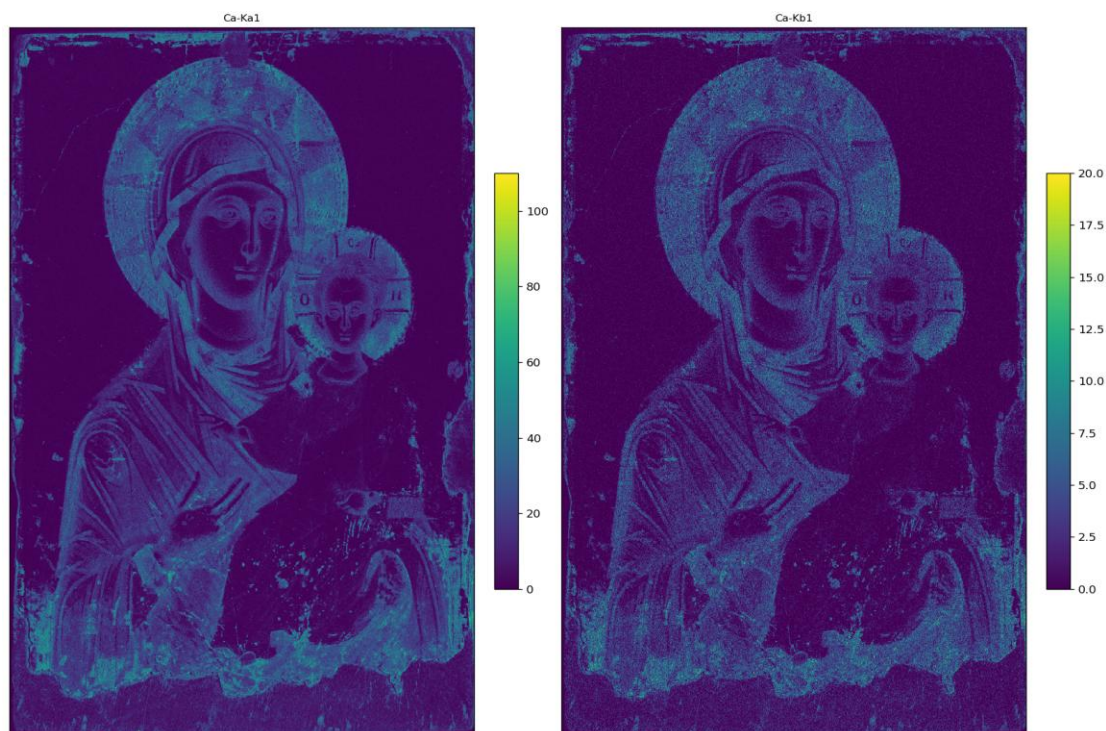
Όλες οι μελέτες χρησιμοποιούν *MA-XRF*, αλλά με κάποιες διαφορές. Η υπό μελέτη εικόνα και το *Dataset "C"* του έχουν τις ίδιες ακριβώς πειραματικές παραμέτρους. Αυτό καθιστά τη σύγκριση των αποτελεσμάτων και των μεθόδων επεξεργασίας (πολυμεταβλητή ανάλυση έναντι *CNN*) ιδιαίτερα άμεση και πολύτιμη.

Στο [Gerodimos2022] χρησιμοποιείται διαφορετική άνοδος (*W αντί Rh*), σημαντικά μεγαλύτερο μέγεθος pixel (*1000 μ m vs 300 μ m*) και πολύ μεγαλύτερο χρόνο συλλογής ανά pixel (*2000 ms vs 10 ms*). Αυτό σημαίνει ότι είχε λιγότερα pixels αλλά πιθανόν καλύτερη στατιστική (λόγω χρόνου) ανά pixel, αν και με χαμηλότερη χωρική ανάλυση. Η διαφορετική άνοδος επηρεάζει το φάσμα διέγερσης και κατ' επέκταση την ευαισθησία σε συγκεκριμένα στοιχεία.

Η βαθμονόμηση του ενεργειακού άξονα είναι κρίσιμη. Στην έρευνα μας χρησιμοποιείται *offset* (A) = -0.95529552 keV και *gain* (B) = $0.010001\text{ keV/κανάλι}$. Στο [Gerodimos2022] αναφέρει *gain* $\sim 0.01999\text{ keV/κανάλι}$ και *offset* $\sim -0.9647\text{ keV}$. Οι τιμές *gain* είναι διαφορετικές, που σημαίνει ότι η αντιστοίχιση καναλιών σε keV θα διαφέρει, κάτι που πρέπει να ληφθεί υπόψη κατά τη σύγκριση φασμάτων. Στο [Gerodimos2022] δίνεται έμφαση στη λεπτομερή φασματοσκοπική ανάλυση με *PyMCA* για τους στοιχειακούς χάρτες και χρησιμοποιείται *clustering* για την ομαδοποίηση των φασμάτων. Στο [Gerodimos2024] εισάγεται μια νέα προσέγγιση με *CNNs* για την επιτάχυνση και αυτοματοποίηση της εξαγωγής των εντάσεων των γραμμών και της δημιουργίας στοιχειακών χαρτών. Για τους στοιχειακούς χάρτες έχουμε πως:

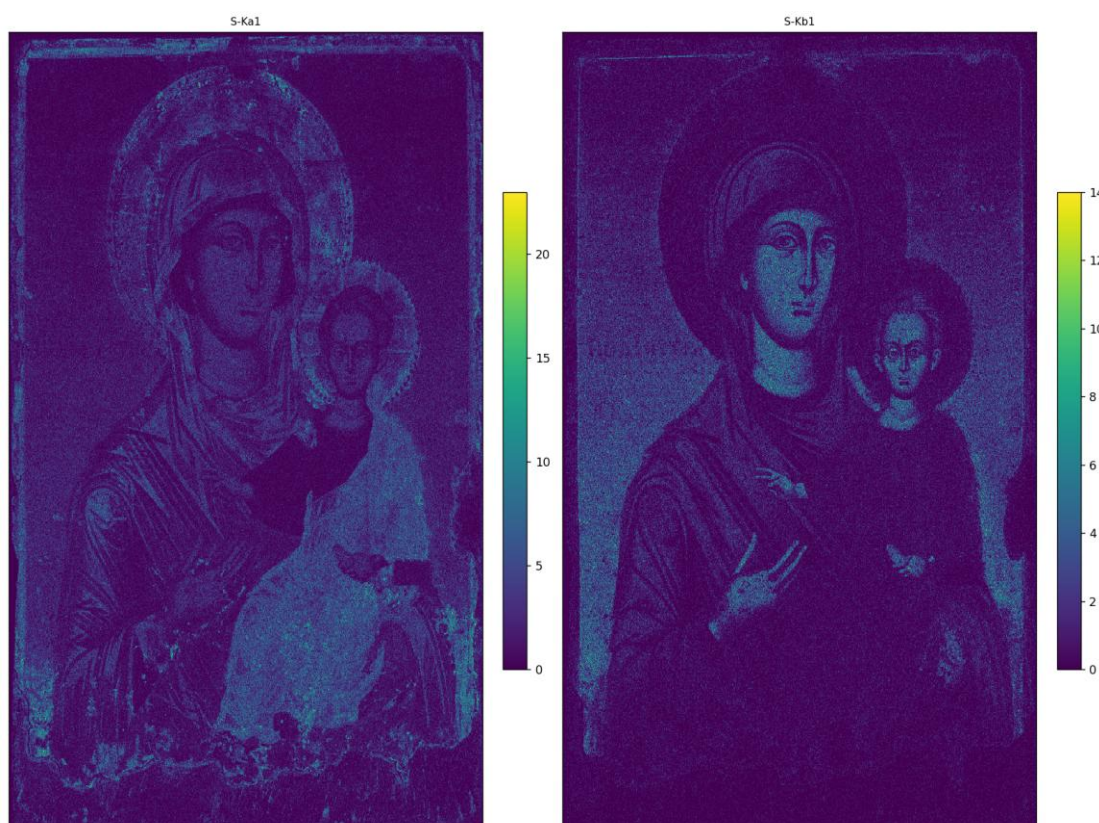
- **Υπόστρωμα (Gesso Ground):**

Υπό Μελέτη Εικόνα: κυρίως *Ca* Εικόνα 161 και *S* Εικόνα 162, υποδηλώνοντας γύψο ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) ή/και ανθρακικό ασβέστιο (CaCO_3) με θειούχες προσμίξεις. Στο [Gerodimos2022] αναφέρεται ότι τα υποστρώματα στις ελληνικές μεταβυζαντινές εικόνες ήταν σχεδόν αποκλειστικά από ενώσεις θεικού ασβεστίου (γύψος, ανυδρίτης) που συχνά περιέχουν προσμίξεις θεικού στροντίου (σελεστίνη). Αυτό επιβεβαιώνεται από τους χάρτες *Ca* και *S* (**Figure**). Η ασθενέστερη ένταση του *S* στην παρούσα μελέτη είναι συνεπής με τις φυσικές αρχές της φασματοσκοπίας *XRF*.





Εικόνα 161: Στοιχειομετρικός Χάρτης Ασβεστίου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K(επάνω) και εικόνα εργαστήριου κάτω

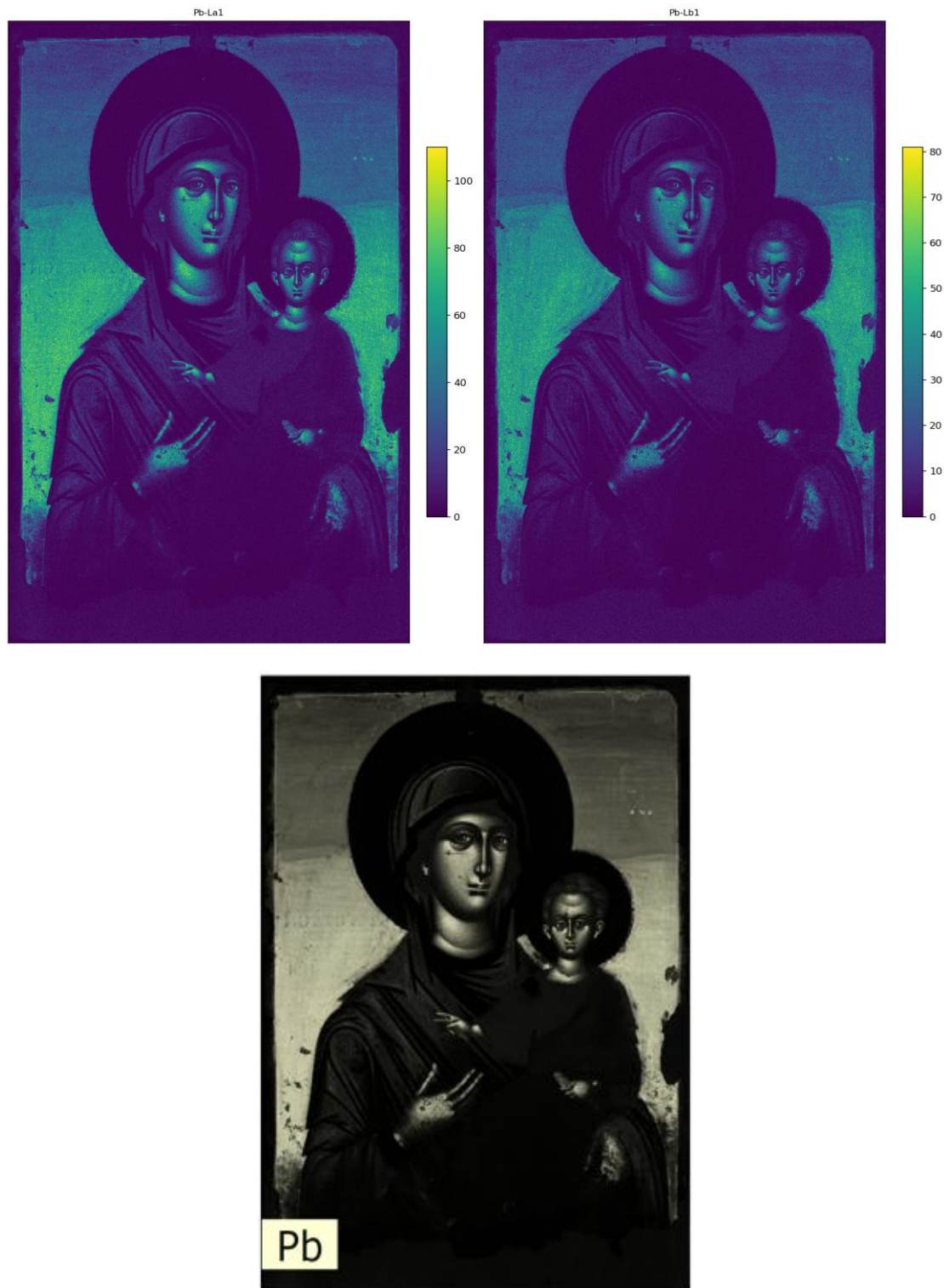


Εικόνα 162: Στοιχειομετρικός Χάρτης Θείου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K

- **Χρωστικές Προσώπων (Incarnates):**

Υπό Μελέτη Εικόνα: Ο χάρτης **Pb** υποδηλώνει χρήση του μολύβδου Εικόνα 163. Ο χάρτης **Fe** δείχνει παρουσία σιδήρου (ώχρες) στον προπλασμό και τις σκιές. Το **Mn** μπορεί να υποδηλώνει **umber**. Αντίστοιχα στο [Gerodimos2022] αναφέρεται εκτεταμένη χρήση λευκού του μολύβδου αναμεμιγμένου με άλλες χρωστικές για την

απόδοση φωτεινότητας και όγκου. Το **Fe** ανιχνεύεται στις σάρκες και σχετίζεται με ώχρες. Το **Mn** Εικόνα 164 σε σκοτεινά περιγράμματα και σκιές υποδηλώνει **umber**. **Σύγκριση:** Υπάρχει ισχυρή συμφωνία. Και οι δύο μελέτες υποδεικνύουν λευκό του μολύβδου και γαιώδεις χρωστικές (ώχρες, **umber**) για τα πρόσωπα.



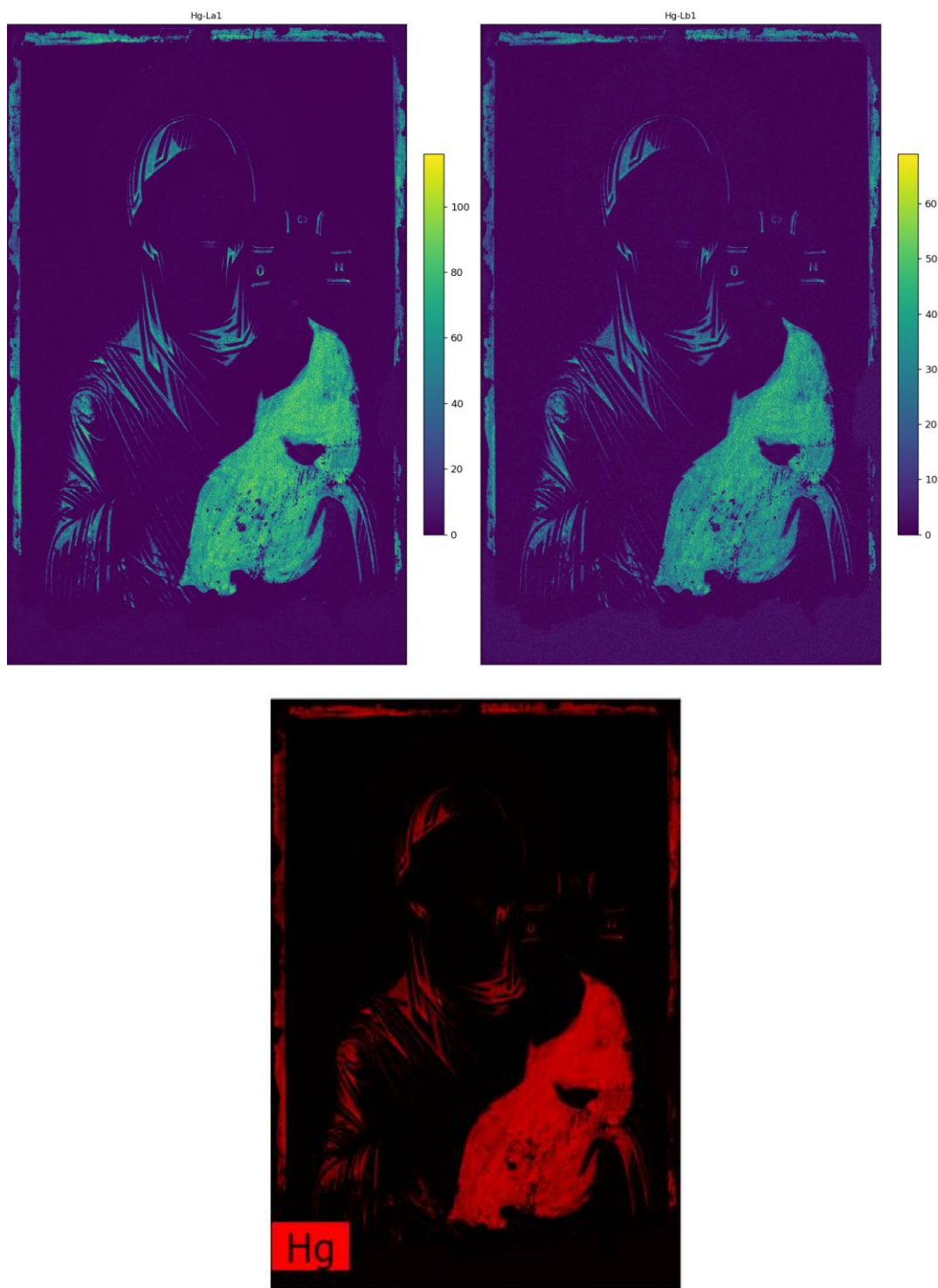
Εικόνα 163: Στοιχειομετρικός Χάρτης Μολύβδου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές L (επάνω) και εικόνα εργαστήριου κάτω



Εικόνα 164: Στοιχειομετρικός Χάρτης Μαγγανίου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K (επάνω) και εικόνα εργαστήριου κάτω

- **Κόκκινες Χρωστικές (π.χ., Ένδυμα Παναγίας):**

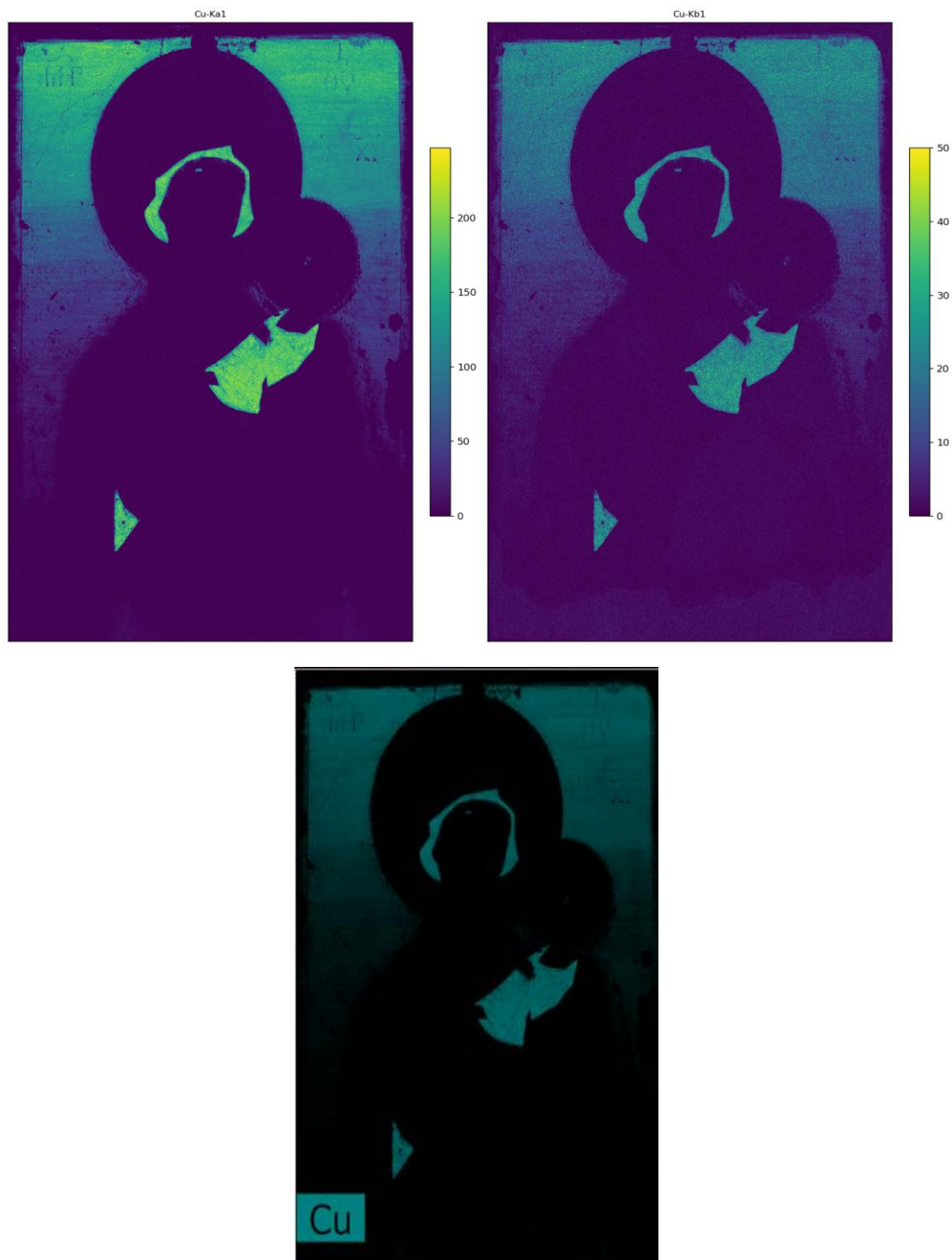
Υπό Μελέτη Εικόνα: Ο χάρτης **Hg** Εικόνα 165 υποδηλώνει χρήση κινναβάρως (**HgS**). Πιθανή παρουσία **Pb** σε κόκκινες περιοχές μπορεί να υποδηλώνει μίνιο. Στο [Gerodimos2022] εντοπίζεται **HgS** (κιννάβαρι/**vermilion**) για τις κόκκινες περιοχές του ενδύματος, συχνά αναμεμειγμένο με **Pb₃O₄** (μίνιο) ή πάνω σε στρώση ώχρας. Πράγμα που δηλώνει συμφωνία όσον αφορά το κιννάβαρι.



Εικόνα 165: Στοιχειομετρικός Χάρτης Υδραργύρου από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές L (επάνω) και εικόνα εργαστηρίου κάτω

- **Μπλε/Πράσινες Χρωστικές:**

Υπό Μελέτη Εικόνα: Ο χάρτης **Cu** Εικόνα 166 υποδηλώνει χρήση χρωστικών με βάση τον χαλκό (αζουρίτης/μπλε, μαλαχίτης/πράσινο). Η παρουσία **S** σε μπλε περιοχές μπορεί να υποδηλώνει ουλτραμαρίνη. Στο [Gerodimos2022] αναφέρεται ότι το μπλε του κεφαλόδεσμου της Παναγίας πιθανόν αποδόθηκε με χρωστική με βάση τον χαλκό, αναμεμιγμένη με λευκό του μολύβδου. Συμφωνία για τις χρωστικές χαλκού.



Εικόνα 166: Στοιχειομετρικός Χάρτης Χαλκού από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές K (επάνω) και εικόνα εργαστηρίου κάτω

- **Χρύσωμα (Gilding):**

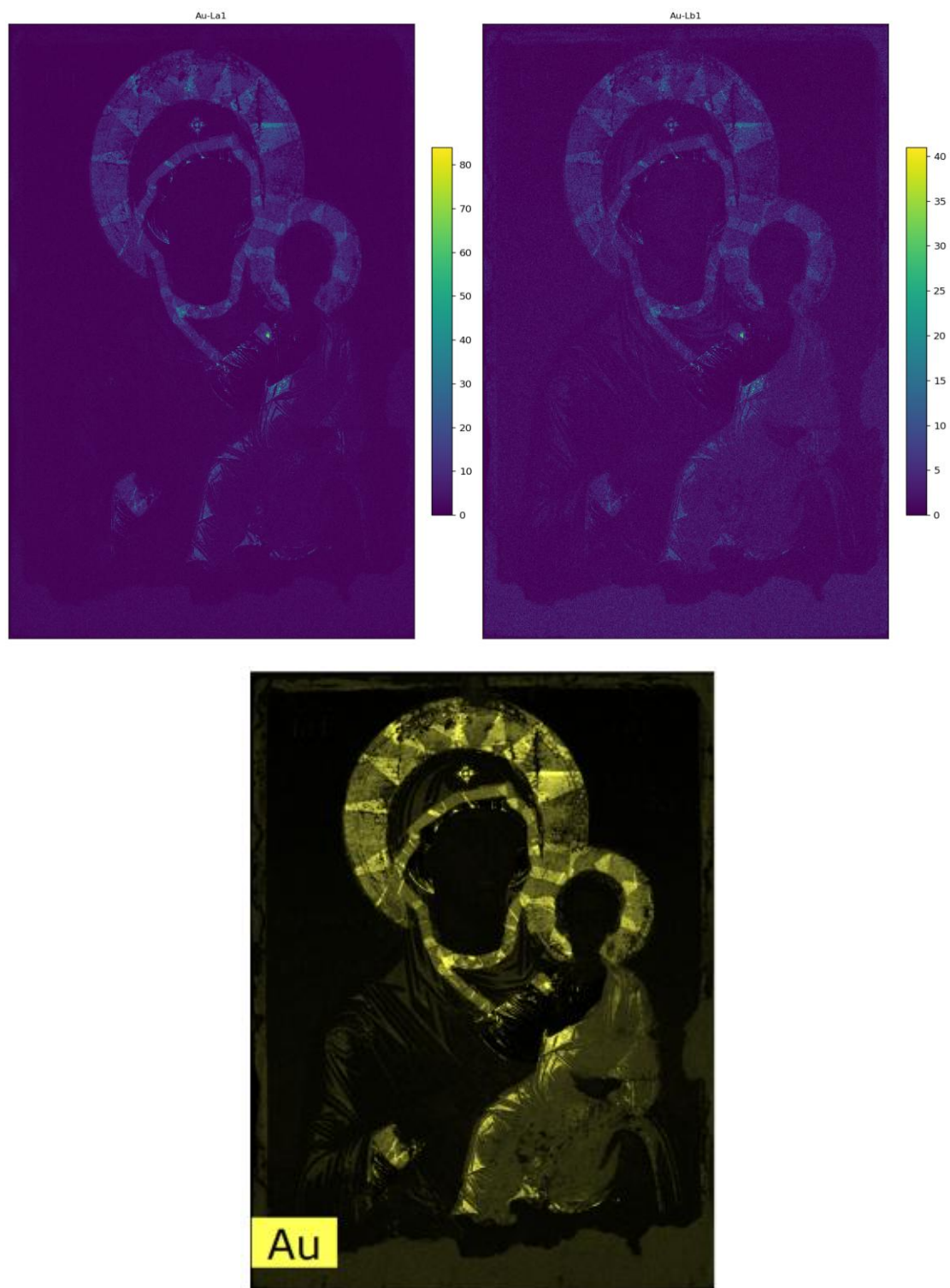
Υπό Μελέτη Εικόνα: Εντοπίζονται Au στα φωτοστέφανα και τις επιγραφές, με σχετικά χαμηλή ένταση, υποδηλώνοντας λεπτά φύλλα χρυσού Εικόνα 168. [Gerodimos2022]: Αναφέρει χρήση φύλλων χρυσού για το φωτοστέφανο και τα διακοσμητικά στοιχεία. Επίσης, ανιχνεύει κηλίδες χρυσού (*shell gold*) στον μπλε

κεφαλόδεσμο. Η χαμηλή ένταση του Au αποδίδεται στο μικρό πάχος των φύλλων ($<1\ \mu\text{m}$). Σημειώνεται η παρουσία Fe κάτω από το χρυσό, υποδηλώνοντας χρήση μπόλου (*bole*). Συμφωνία για τη χρήση των λεπτών φύλλων χρυσού, βέβαια παρατηρείται εμφάνιση **Ca** από κάτω από τα στρώματα χρυσού (πιθανόν *gesso* ή κόλλα). σαφής συνύπαρξη του σιδήρου κάτω από τον χρυσό στα φωτοστέφανα αποτελεί αδιάψευστη απόδειξη της παραδοσιακής τεχνικής χρυσώματος με χρήση μπόλου, επιβεβαιώνοντας τα ευρήματα της βιβλιογραφίας [Gerodimos2022].

RGB Overlay: R: Element Fe-Ka1 | G: Element Au-La1 | B: Element Ca-Ka1



Εικόνα 167: Σύνθετος χάρτης στοιχειακής κατανομής (RGB Overlay) της εικόνας, που αποκαλύπτει τη στρωματογραφία των υλικών. Αντιστοίχιση χρωμάτων: Κόκκινο (R) = Σίδηρος (Fe-Ka1), Πράσινο (G) = Χρυσός (Au-La1), Μπλε (B) = Ασβέστιο (Ca-Ka1). Η επικάλυψη του σήματος του σιδήρου (κόκκινο) και του χρυσού (πράσινο) στα φωτοστέφανα επιβεβαιώνει τη χρήση μπόλου ως υπόστρωμα για τη χρύσωση.



Εικόνα 168: Στοιχειομετρικός Χάρτης Χρυσού από το πρόγραμμα ανάλυσης. Χάρτες για τις Γραμμές L (επάνω) και εικόνα εργαστηρίου κάτω

ΣΥΜΠΕΡΑΣΜΑΤΑ

Η παρούσα διπλωματική εργασία επικεντρώθηκε στη μη-καταστροφική ανάλυση βυζαντινών και μεταβυζαντινών εικόνων μέσω της φασματοσκοπίας απεικόνισης φθορισμού ακτίνων-X (*MA-XRF*), με κύριο στόχο την ανάπτυξη και εφαρμογή μιας ολοκληρωμένης μεθοδολογίας για την "αποκωδικοποίηση" των πολύπλοκων υπερφασματικών δεδομένων. Μέσα από τη διεξαγωγή μιας σειράς πειραμάτων, συμπεριλαμβανομένων αναλύσεων σε χαμηλή και υψηλή χωρική ανάλυση, και την εφαρμογή ενός εύρους χημειομετρικών τεχνικών, εξήχθησαν κρίσιμα συμπεράσματα τόσο για την αποτελεσματικότητα των μεθόδων όσο και για την υλική σύσταση και την καλλιτεχνική τεχνική των υπό μελέτη έργων.

Αξιολόγηση και Βελτιστοποίηση της Αναλυτικής Μεθοδολογίας

Η μελέτη κατέδειξε ότι η επιλογή της κατάλληλης αναλυτικής στρατηγικής είναι καθοριστικής σημασίας για την εξαγωγή ουσιαστικής πληροφορίας από δεδομένα MA-XRF.

- **Αξιολόγηση Μεθόδων Μείωσης Διαστάσεων: Η Ανάλυση Κύριων Συνιστωσών (PCA)** αποδείχθηκε ένα εξαιρετικά ισχυρό εργαλείο για την αρχική μείωση της διαστατικότητας και την απομάκρυνση του θορύβου, συμπυκνώνοντας πάνω από το **80%** της πληροφορίας σε μόλις 6 συνιστώσες. Ωστόσο, η ερμηνεία της παραμένει έμμεση λόγω της φύσης των συνιστωσών. Αντίθετα, η **Μη Αρνητική Παραγοντοποίηση Πινάκων (NMF)**, επιβάλλοντας τον περιορισμό της μη-αρνητικότητας, παρήγαγε συνιστώσες με άμεση φυσικοχημική ερμηνεία, λειτουργώντας ως "χάρτες αφθονίας" υλικών. Η **υβριδική προσέγγιση NMF-PCA**, όπου η NMF εφαρμόστηκε στα αποθορυβοποιημένα δεδομένα της **PCA**, αναδείχθηκε ως η **βέλτιστη μεθοδολογία**, συνδυάζοντας την υπολογιστική ισχύ της **PCA** με την ανώτερη ερμηνευσιμότητα της **NMF** και οδηγώντας στον ακριβέστερο διαχωρισμό των υλικών.
- **Σημασία της Κλίμακας Ανάλυσης:** Η ανάλυση σε γραμμική κλίμακα ήταν αποτελεσματική στην ανάδειξη των κυρίαρχων υλικών. Ωστόσο, η εφαρμογή ενός **λογαριθμικού μετασχηματισμού (log-scale)** αποδείχθηκε κρίσιμη για την ενίσχυση και τον εντοπισμό σημάτων από υλικά σε χαμηλές συγκεντρώσεις ή σε πολύ λεπτά στρώματα (π.χ. χρυσός, υδράργυρος), τα οποία διαφορετικά χάνονταν μέσα στο ισχυρό σήμα του υποστρώματος.

Αποτελέσματα από την Ανάλυση των Έργων Τέχνης

Η εφαρμογή της βελτιστοποιημένης μεθοδολογίας στα εξεταζόμενα έργα οδήγησε σε μια ολοκληρωμένη κατανόηση της υλικής τους ταυτότητας.

- **Πλήρης Χαρακτηρισμός της Παλέτας:** Ταυτοποιήθηκε με απόλυτη βεβαιότητα η παραδοσιακή παλέτα της μεταβυζαντινής αγιογραφίας. Επιβεβαιώθηκε η χρήση γύψου (**CaSO₄**) για το υπόστρωμα, λευκού του μολύβδου (**Pb-based**) ως βασικής λευκής χρωστικής, γήινων χρωστικών (**Fe-based** όπως και **Mn-based umber**) για τα σαρκώματα και τις σκιάσεις, φύλλου χρυσού (**Au**) για την επιχρύσωση, και χρωστικών χαλκού (**Cu-based**). Ιδιαίτερα σημαντική ήταν η ταυτοποίηση της πολύτιμης κόκκινης χρωστικής **κιννάβαρι (HgS)** σε σημεία έμφασης, καθώς και η ανίχνευση

ιχνών **ψευδαργύρου (Zn)** που υποδεικνύουν μεταγενέστερες επεμβάσεις συντήρησης.

- **Αποκάλυψη της Καλλιτεχνικής Τεχνικής:** Η ανάλυση υψηλής ανάλυσης επέτρεψε τη μετάβαση από τον απλό χαρακτηρισμό υλικών στην **εις βάθος διερεύνηση της τεχνικής του καλλιτέχνη**. Έγινε δυνατή η απεικόνιση μεμονωμένων πινελιών (ψιμυθιές), η διάκριση μεταξύ παχιών και λεπτών στρώσεων της ίδιας χρωστικής, και η λεπτομερής χαρτογράφηση της καλλιγραφικής δεξιοτεχνίας στη χρυσοκονδυλιά. Επιπλέον, η διαδικασία του **sub-clustering** αποκάλυψε με εντυπωσιακή λεπτομέρεια τη σύνθετη στρωματογραφία των υλικών, όπως τη διάκριση της βασικής στρώσης, των φωτισμάτων και των σκιάσεων στα σαρκώματα, επιτρέποντας την ψηφιακή ανασύνθεση της δημιουργικής διαδικασίας.
- **Σημασία της Χωρικής Ανάλυσης:** Η σύγκριση των πειραμάτων χαμηλής και υψηλής ανάλυσης απέδειξε ότι η επιλογή της ανάλυσης είναι κρίσιμη και εξαρτάται από τον ερευνητικό στόχο. Η χαμηλή ανάλυση (**500 $\mu\text{m}/\text{pixel}$**) ήταν επαρκής για τον γενικό χαρακτηρισμό της παλέτας, ενώ η υψηλή ανάλυση (**200-300 $\mu\text{m}/\text{pixel}$**) ήταν απαραίτητη για την κατανόηση της καλλιτεχνικής τεχνικής, την ανίχνευση σπάνιων υλικών και τη μελέτη της κατάστασης διατήρησης σε μικρο-κλίμακα.

Κεντρικό ρόλο στην επιτυχή εφαρμογή και αξιολόγηση αυτών των σύνθετων μεθοδολογιών διαδραμάτισε το **εξειδικευμένο λογισμικό που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας**. Η αρθρωτή του αρχιτεκτονική και το διαδραστικό του περιβάλλον επέτρεψαν την αποδοτική διαχείριση τεράστιων συνόλων δεδομένων (της τάξης των **gigabytes**) μέσω της τμηματικής φόρτωσης (**chunking**) και της δυνατότητας επιλογής αριθμητικής ακρίβειας. Το λογισμικό δεν λειτούργησε απλώς ως ένα εργαλείο εκτέλεσης, αλλά ως ένα ολοκληρωμένο ερευνητικό περιβάλλον που διευκόλυνε τη γρήγορη σύγκριση μεθόδων, την οπτική επιθεώρηση των αποτελεσμάτων και την ιεραρχική, εις βάθος διερεύνηση περιοχών ενδιαφέροντος (**sub-clustering**). Η ανάπτυξή του αποδεικνύει ότι η δημιουργία προσαρμοσμένων υπολογιστικών εργαλείων είναι απαραίτητη προϋπόθεση για την πλήρη αξιοποίηση των σύγχρονων αναλυτικών τεχνικών στην επιστήμη της πολιτιστικής κληρονομιάς.

Συνοψίζοντας, η παρούσα διπλωματική εργασία ανέπτυξε, εφάρμοσε και αξιολόγησε με επιτυχία μια ισχυρή, πολυεπίπεδη μεθοδολογία για την ανάλυση δεδομένων **MA-XRF**. Η προσέγγιση αυτή, υποστηριζόμενη από ένα προσαρμοσμένο υπολογιστικό εργαλείο, υπερβαίνει τους περιορισμούς της απλής στοιχειακής απεικόνισης και αποτελεί ένα ολοκληρωμένο πλαίσιο για την βαθύτερη κατανόηση της τεχνολογίας, της στρωματογραφίας και της υλικής ιστορίας των πολιτιστικών αγαθών.

Επικύρωση Αποτελεσμάτων και Όρια Μεθοδολογίας

Η αξιοπιστία των συμπερασμάτων που προκύπτουν από τη χημειομετρική ανάλυση δεδομένων MA-XRF εξαρτάται από την ορθότητα της μεθοδολογίας. Για τον λόγο αυτό, πραγματοποιήθηκε μια διαδικασία επικύρωσης, ενώ ταυτόχρονα αναγνωρίζονται τα εγγενή όρια της προσέγγισης.

X.1 Επικύρωση των Αποτελεσμάτων

Η επικύρωση των ταυτοποιήσεων των υλικών που προέκυψαν από το λογισμικό **NICE**

πραγματοποιήθηκε μέσω μιας συνδυαστικής προσέγγισης: Διασταύρωση με το Φάσμα Συνολικής Έντασης (**Sum Spectrum**): Τα αντιπροσωπευτικά φάσματα που εξήχθησαν για κάθε συστάδα (**cluster**) από το **AnalysisWindow** συγκρίθηκαν οπτικά με το συνολικό φάσμα της εικόνας. Η παρουσία των χαρακτηριστικών κορυφών μιας χρωστικής (π.χ., **Hg La** και **Lβ** για τον κιννάβαρι) στο φάσμα της συστάδας της επιβεβαιώνει την ταυτοποίηση.

Χωρική Συσχέτιση (Spatial Correlation): Η χωρική κατανομή (χάρτης) μιας συστάδας που ταυτοποιήθηκε με μια συγκεκριμένη χρωστική (π.χ., κόκκινη) συγκρίθηκε με την οπτική εικόνα του έργου τέχνης. Η υψηλή χωρική συσχέτιση μεταξύ της συστάδας και των περιοχών του αντίστοιχου χρώματος παρείχε ισχυρή οπτική επικύρωση. Η λειτουργία "**Plot Linking**" του λογισμικού ήταν καθοριστική σε αυτή τη διαδικασία, επιτρέποντας τον άμεσο συσχετισμό μεταξύ των σημείων στο διάγραμμα **PCA/t-SNE** και της θέσης τους στην εικόνα.

Σύγκριση με Συμπληρωματικές Τεχνικές (όπου είναι εφικτό): Αν και η παρούσα μελέτη βασίστηκε σε μη-καταστροφικές τεχνικές, τα αποτελέσματα είναι συμβατά με τη γνωστή βιβλιογραφία για τις μεταβυζαντινές χρωστικές. Σε μελλοντικές εργασίες, τα αποτελέσματα της **MA-XRF** θα μπορούσαν να διασταυρωθούν με σημειακές αναλύσεις μέσω άλλων φασματοσκοπικών τεχνικών (π.χ., **Raman**, **FTIR**) σε επιλεγμένες περιοχές για την απόλυτη επιβεβαίωση της μοριακής σύστασης.

Χ.2 Όρια και Περιορισμοί της Μεθόδου

Παρά την ισχύ της, η εφαρμοζόμενη μεθοδολογία υπόκειται σε ορισμένους περιορισμούς:

Περιορισμοί της XRF: Η τεχνική **XRF** είναι ευαίσθητη στη στοιχειακή σύσταση, αλλά όχι στη χημική ένωση ή την κρυσταλλική δομή. Για παράδειγμα, δεν μπορεί να διακρίνει μεταξύ διαφορετικών ορυκτών του χαλκού (π.χ., μαλαχίτης vs. αζουρίτης) εάν δεν υπάρχουν χαρακτηριστικά ιχνοστοιχεία. Επιπλέον, η **XRF** έχει περιορισμένη ευαισθησία σε πολύ ελαφρά στοιχεία ($Z < 13$, π.χ., **Na**, **Mg**, **Al**) και δεν μπορεί να ανιχνεύσει οργανικές χρωστικές που βασίζονται στον άνθρακα.

Επιλογή Παραμέτρων Αλγορίθμων: Η απόδοση των αλγορίθμων, ιδίως του **K-Means**, εξαρτάται από την επιλογή του αριθμού των συστάδων (**k**). Αν και το λογισμικό **NICE** παρέχει πολλαπλούς δείκτες (**Silhouette**, κ.λπ.) για να καθοδηγήσει αυτή την επιλογή, η τελική απόφαση εξακολουθεί να απαιτεί την κρίση του ερευνητή. Μια εσφαλμένη επιλογή του **k** μπορεί να οδηγήσει είτε σε συγχώνευση διακριτών υλικών σε μία συστάδα, είτε σε διαχωρισμό ενός υλικού σε πολλαπλές.

Ερμηνεία των Συνιστωσών PCA/NMF: Οι συνιστώσες που προκύπτουν από την **PCA** και την **NMF** δεν αντιστοιχούν πάντα σε μεμονωμένες, καθαρές χρωστικές. Συχνά, μια συνιστώσα μπορεί να αντιπροσωπεύει ένα μείγμα υλικών ή μια χωρική συσχέτιση (π.χ., μια συνιστώσα που αναδεικνύει τις άκρες ή τις σκιές). Η ερμηνεία τους απαιτεί εμπειρία και προσεκτική σύγκριση με τα φάσματα και τις εικόνες.

Φαινόμενα Επικάλυψης Στρωμάτων: Η ακτινοβολία **XRF** διεισδύει σε κάποιο βάθος στο έργο τέχνης, με αποτέλεσμα το τελικό φάσμα να αποτελεί το άθροισμα του σήματος από πολλαπλά στρώματα (π.χ., βερνίκι, χρωστική, υπόστρωμα). Οι αλγόριθμοι μηχανικής μάθησης προσπαθούν να αποσυνθέσουν αυτά τα μεικτά σήματα, αλλά η πλήρης αποσύνθεση είναι μια εξαιρετικά δύσκολη πρόκληση και μπορεί να οδηγήσει σε λανθασμένες ταυτοποιήσεις εάν δεν ληφθεί υπόψη η συγκεκριμένη στρωματογραφία υποστρώματος και χρωστικών.

BIBΛΙΟΓΡΑΦΙΑ

Case Studies – Important Papers for Analysis and Results

[**Gerodimos2022**] Gerodimos, T., Asvestas, A., Mastrotheodoros, G. P., Chantas, G., Liougos, I., Likas, A., & Anagnostopoulos, D. F. (2022). **Scanning X-ray Fluorescence data analysis for the identification of Byzantine icons' materials, techniques, and state of preservation: A case study.** *Journal of Imaging*, 8(5), 147. <https://doi.org/10.3390/jimaging8050147>

[**Gerodimos2024**] Gerodimos, T., Georvasilis, I., Asvestas, A., Mastrotheodoros, G. P., Likas, A., & Anagnostopoulos, D. F. (2024). **MA-XRF datasets analysis based on convolutional neural network: A case study on religious panel paintings.** *Manuscript submitted for publication.*

[**Mastrotheodoros**] Mastrotheodoros, G. P., Asvestas, A., Gerodimos, T., & Anagnostopoulos, D. F. (2022). **Revealing the materials, painting techniques, and state of preservation of a heavily altered early 19th Century Greek icon through MA-XRF.** *Heritage*, 5(3), 1637–1654. <https://doi.org/10.3390/heritage5030085>

Bibliography – Electromagnetic Waves

- [1] D'Agostino, S. (1975). Hertz's researches on electromagnetic waves. *Historical Studies in the Physical Sciences*, 6, 261–323.
- [2] Kumar, V., Kalra, J., Shah, M., & Pant, B. (2019). Analytical study on the effects of electromagnetic waves on human beings. *International Journal of Innovative Technology and Exploring Engineering*, 8(6), 2278–3075. <https://doi.org/10.35940/ijitee.D1017.0384S319>
- [3] Amineh, R. K., & Roodaki-Lavasani, M. (2020). Applications of electromagnetic waves: Present and future. *Electronics*, 9(5), 808. <https://doi.org/10.3390/electronics9050808>
- [4] Schwartz, M. D. (n.d.). *Lecture 9: Reflection, Transmission and Impedance - 1. Boundary conditions at a junction* [Lecture notes]. Harvard University.
- [5] Schwartz, M. D. (n.d.). *Lecture 10: Energy and Power in Waves - 1. Energy in a string* [Lecture notes]. Harvard University.
- [6] Heisenberg, W. (1925). On the quantum reinterpretation of kinematical and mechanical relationships. *Zeitschrift für Physik*, 33, 879–893.
- [7] Morin, D. (n.d.). *Electromagnetic waves* [Lecture notes]. Harvard University.
- [8] Schwartz, M. D. (n.d.). *Lecture 11: Wavepackets and dispersion* [Lecture notes]. Harvard University.
- [9] Schwartz, M. D. (n.d.). *Lecture 14: Polarization - 1. Polarization vectors* [Lecture notes]. Harvard University.

[10] Schwartz, M. D. (n.d.). *Lecture 16: Accelerating charges and prisms - 1. Larmor formula* [Lecture notes]. Harvard University.

Bibliography – Auger

[1] Burhop, E. H. S. (1934). The Auger effect. *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, 148(864), 272–

284. <https://doi.org/10.1098/rspa.1935.0018>

[2] Barthés-Labrousse, M.-G. (1995). The Auger effect. *Microscopy Microanalysis Microstructures*, 6(3), 253–262. <https://doi.org/10.1051/mmm:1995122>

[3] Nikjoo, H., Emfietzoglou, D., & Charlton, D. E. (2008). The Auger effect in physical and biological research. *International Journal of Radiation Biology*, 84(12), 1011–1026. <https://doi.org/10.1080/09553000802460172>

Bibliography – Machine Learning – Supervised – Unsupervised

[1] Li, X., et al. (2021). Unsupervised content-preserving transformation for optical microscopy. *Light: Science & Applications*, 10(1), 44. <https://doi.org/10.1038/s41377-021-00484-y>

[2] Jafarpour, S., Polatkan, G., Brevdo, E., Hughes, S., Brasoveanu, A., & Daubechies, I. (2009). *Stylistic analysis of paintings using wavelets and machine learning* (Master's thesis). Princeton University.

[3] Kontolati, K., Loukrezis, D., Giovanis, D. G., Vandanapu, L., & Shields, M. D. (2022). A survey of unsupervised learning methods for high-dimensional uncertainty quantification in black-box-type problems. *Journal of Computational Physics*, 464, 111313. <https://doi.org/10.1016/j.jcp.2022.111313>

[4] Subasi, A. (2020). Clustering examples. In *Practical Machine Learning for Data Analysis Using Python* (pp. 465–511). Academic Press. <https://doi.org/10.1016/b978-0-12-821379-7.00007-2>

[5] Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: A brief primer. *Behavior Therapy*, 51(5), 675–687. <https://doi.org/10.1016/j.beth.2020.05.002>

[6] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

[7] Li, N., Shepperd, M., & Guo, Y. (2020). A systematic review of unsupervised learning techniques for software defect prediction. *Information and Software*

- Technology*, 122, 106287. <https://doi.org/10.1016/j.infsof.2020.106287>
- [8] Roberts, M., Garg, M., & Lipton, Z. (2022). Unsupervised learning under latent label shift. *arXiv preprint arXiv:2212.00000*.
- [9] Balestrieri, R., et al. (2023). A cookbook of self-supervised learning. *arXiv preprint arXiv:2304.12210*.
- [10] Abiodun, O. S., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology*, 48(3), 128–138. <https://doi.org/10.14445/22312803/ijctt-v48p126>
- [11] El Bouchefry, K., & de Souza, R. S. (2020). Learning in Big Data: Introduction to machine learning. In *Knowledge Discovery in Big Data from Astronomy and Earth Observation* (pp. 225–249). Elsevier. <https://doi.org/10.1016/b978-0-12-819154-5.00023-0>
- [12] Ghahramani, Z. (2004). Unsupervised learning. In *Advanced Lectures on Machine Learning* (pp. 72–112). Springer.
- [13] Greene, D., Cunningham, P., Mayer, R., & Wien, T. (2017). Unsupervised learning and clustering. In *Machine Learning and Knowledge Discovery in Databases* (pp. 51–89). Springer. <https://doi.org/10.1007/978-3-540-75171-7-3>
- [14] Wang, X. (2015). *Machine Learning Basics* [Lecture notes]. The Chinese University of Hong Kong.
- [15] Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26(3), 159–190. <https://doi.org/10.1007/s10462-007-9052-3>
- [16] Mirzal, A. (2020). Statistical analysis of microarray data clustering using NMF, spectral clustering, Kmeans, and GMM. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6), 1–1. <https://doi.org/10.1109/tcbb.2020.3025486>
- [17] Usmani, U. A., Happonen, A., & Watada, J. (2022). A review of unsupervised machine learning frameworks for anomaly detection in industrial applications. In *Lecture Notes in Networks and Systems* (pp. 158–189). Springer. https://doi.org/10.1007/978-3-031-10464-0_11

Bibliography – K Means

- [1] Nurhayati, T., Kania, L., Wardhani, N., & Maarif, H. (2018). Big Data technology for comparative study of K-Means and Fuzzy C-Means algorithms performance.

In *2018 International Conference on Information and Communications Technology (ICOIACT)*.

- [2] Esteves, R. M., Hacker, T., & Rong, C. (2013). Competitive K-Means, a new accurate and distributed K-Means algorithm for large datasets. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science* (pp. 17-24). IEEE. <https://doi.org/10.1109/cloudcom.2013.89>
- [3] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm. *Applied Statistics*, 28(1), 100. <https://doi.org/10.2307/2346830>
- [4] Huang, Z. (1998). Extensions to the k-Means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2, 283–304.
- [5] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881–892. <https://doi.org/10.1109/tpami.2002.1017616>
- [6] Xu, L., Li, J., Wong, A., & Peng, J. (2014). K-P-Means: A clustering algorithm of K ‘purified’ means for hyperspectral endmember estimation. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1787–1791. <https://doi.org/10.1109/lgrs.2014.2309340>
- [7] Olukanmi, P., & Twala, B. (2017). K-means-sharp: modified centroid update for outlier-robust k-means clustering. In *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*.
- [8] Shah, S., & Singh, M. (2012). Comparison of a time efficient modified K-mean algorithm with K-mean and K-medoid algorithm. In *2012 International Conference on Communication Systems and Network Technologies* (pp. 1-5). IEEE. <https://doi.org/10.1109/csnt.2012.100>
- [9] Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451–461. [https://doi.org/10.1016/s0031-3203\(02\)00060-2](https://doi.org/10.1016/s0031-3203(02)00060-2)

Bibliography – DBScan

- [1] Bushra, A. A., & Yi, G. (2021). Comparative analysis review of pioneering DBSCAN and successive density-based clustering algorithms. *IEEE Access*, 9, 87918–87935. <https://doi.org/10.1109/access.2021.3089036>
- [2] Chen, Y., et al. (2021). KNN-BLOCK DBSCAN: Fast clustering for large-scale

data. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(6), 3939–3953. <https://doi.org/10.1109/tsmc.2019.2956527>

[3] Chen, Y., Zhou, L., Bouguila, N., Wang, C., Chen, Y., & Du, J. (2021). BLOCK-DBSCAN: Fast clustering for large scale data. *Pattern Recognition*, 109,

107624. <https://doi.org/10.1016/j.patcog.2020.107624>

[4] Han, X., Armenakis, C., & Jadidi, M. (2021). Modeling vessel behaviours by clustering AIS data using optimized DBSCAN. *Sustainability*, 13(15),

8162. <https://doi.org/10.3390/su13158162>

Bibliography – NMF

[1] Hafshejani, S., & Moaberfard, Z. (2021). *Initialization for nonnegative matrix factorization: A comprehensive review*. arXiv preprint arXiv:2108.08272.

[2] Feng, X.-R., et al. (2022). Hyperspectral unmixing based on nonnegative matrix factorization: A comprehensive review. *IEEE Geoscience and Remote Sensing Magazine*, 10(4), 8–35.

[3] Alfeld, M., Wahabzada, M., Bauckhage, C., Kersting, K., Wellenreuther, G., & Falkenberg, G. (2014). Non-negative factor analysis supporting the interpretation of elemental distribution images acquired by XRF. *Journal of Physics: Conference Series*, 499, 012013. <https://doi.org/10.1088/1742-6596/499/1/012013>

[4] Dong, L., Yuan, Y., & Lu, X. (2021). Spectral–spatial joint sparse NMF for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3), 2391–2402. <https://doi.org/10.1109/tgrs.2020.3006109>

[5] Wang, Y.-X., & Zhang, Y.-J. (2013). Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6), 1336–1353. <https://doi.org/10.1109/tkde.2012.51>

[6] Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology*, 7, 886498. <https://doi.org/10.3389/fsoc.2022.886498>

[7] Shen, X., Zhang, X., Lan, L., Liao, Q., & Luo, Z. (2019). Another robust NMF: Rethinking the hyperbolic tangent function and locality constraint. *IEEE Access*, 7, 31089–31102. <https://doi.org/10.1109/access.2019.2903309>

[8] Wang, S., Deng, C., Lin, W., Huang, G.-B., & Zhao, B. (2017). NMF-based image quality assessment using extreme learning machine. *IEEE Transactions on Cybernetics*, 47(1), 232–243. <https://doi.org/10.1109/tcyb.2015.2512852>

- [9] Mirzal, A. (2020). Statistical analysis of microarray data clustering using NMF, spectral clustering, Kmeans, and GMM. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6), 1–1. <https://doi.org/10.1109/tcbb.2020.3025486>
- [10] Łach, B., et al. (2021). Application of factorisation methods to analysis of elemental distribution maps acquired with a full-field XRF imaging spectrometer. *Sensors*, 21(23), 7965. <https://doi.org/10.3390/s21237965>
- [11] Lee, D. D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 13, 556–562.
- [12] Gillis, N. (2014). The why and how of nonnegative matrix factorization. *SIAM Review*, 56(2), 257–291.
- [13] Lin, C. J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Computation*, 19(10), 2756–2779.

Bibliography – PCA

- [1] Huang, X., Wu, L., & Ye, Y. (2019). A review on dimensionality reduction techniques. *International Journal of Pattern Recognition and Artificial Intelligence*, 33(10), 1950017. <https://doi.org/10.1142/s0218001419500174>
- [2] Shlens, J. (2014). *A tutorial on principal component analysis*. arXiv preprint arXiv:1404.1100.
- [3] Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- [4] Allegretta, I., et al. (2020). Macro-classification of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal component analysis (PCA) and machine learning algorithms. *Talanta*, 212, 120785. <https://doi.org/10.1016/j.talanta.2020.120785>
- [5] Cao, L. J., Chua, K. S., Chong, W. K., Lee, H. P., & Gu, Q. M. (2003). A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, 55(1–2), 321–336. [https://doi.org/10.1016/s0925-2312\(03\)00433-8](https://doi.org/10.1016/s0925-2312(03)00433-8)
- [6] Donais, M. K., Wojtas, S., Desmond, A., Duncan, B., & George, D. B. (2012). Differentiation of hypocaust and floor tiles at Coriglia, Castel Viscardo (Umbria, Italy) using principal component analysis (PCA) and portable X-ray fluorescence (XRF) spectrometry. *Applied Spectroscopy*, 66(9), 1005–1012. <https://doi.org/10.1366/12-06616>

- [7] Donais, M. K., Alrais, M., Konomi, K., George, D., Ramundt, W. H., & Smith, E. (2020). Energy dispersive X-ray fluorescence spectrometry characterization of wall mortars with principal component analysis: Phasing and ex situ versus in situ sampling. *Journal of Cultural Heritage*, 43, 90–97. <https://doi.org/10.1016/j.culher.2019.12.007>
- [8] Moggia, C., et al. (2022). Improving bitter pit prediction by the use of X-ray fluorescence (XRF): A new approach by multivariate classification. *Frontiers in Plant Science*, 13, 1033308. <https://doi.org/10.3389/fpls.2022.1033308>
- [9] Gawali, P. B., Sangode, S. J., & Herlekar, M. (2021). An application of Principal Component Analysis (PCA) using XRF data to delineate the continentality factor over beach sediments from Vengurla, West Coast, Sindhudurg district, Maharashtra, India. *International Advanced Research Journal in Science, Engineering and Technology*, 8(9). <https://doi.org/10.17148/iarjset.2021.8902>
- [10] Rehman, A., Khan, A., Ali, M. A., Khan, M. U., Khan, S. U., & Ali, L. (2020). Performance analysis of PCA, Sparse PCA, Kernel PCA and Incremental PCA algorithms for heart failure prediction. In *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)* (pp. 1-5). IEEE. <https://doi.org/10.1109/ICECCE49384.2020.9179199>

Bibliography – GMM

- [1] Ward, J., Marvin, R., O'Halloran, T., Jacobsen, C., & Vogt, S. (2013). Rapid and accurate analysis of an X-ray fluorescence microscopy data set through Gaussian mixture-based soft clustering methods. *Microscopy and Microanalysis*, 19(5), 1281–1289. <https://doi.org/10.1017/s1431927613012737>

Bibliography – XRF

- [1] Zhang, Z., He, X., Cheng, H., & Li, F. (2020). Background deduction approach based on iterative adaptive window-based wavelet transform and Gaussian convolution filtering for XRF. In *2020 2nd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM)*. <https://doi.org/10.1109/aiam50918.2020.00055>
- [2] Reiche, I., Müller, K., Mysak, E., Eveno, M., & Mottin, B. (2015). Toward a three-dimensional vision of the different compositions and the stratigraphy of the painting L'Homme blessé by G. Courbet: coupling SEM–EDX and confocal micro-XRF. *Applied Physics A*, 121(3), 903–913. <https://doi.org/10.1007/s00339-015-9428->

- [3] Romano, F. P., Pappalardo, G., Pappalardo, L., Garraffo, S., Gigli, R., & Pautasso, A. (2005). Quantitative non-destructive determination of trace elements in archaeological pottery using a portable beam stability-controlled XRF spectrometer. *X-Ray Spectrometry*, 35(1), 1–7. <https://doi.org/10.1002/xrs.880>
- [4] Longoni, A., et al. (2006). XRF spectrometers based on monolithic arrays of silicon drift detectors: elemental mapping analyses and advanced detector structures. *IEEE Transactions on Nuclear Science*, 53(2), 641–647. <https://doi.org/10.1109/tns.2006.872640>
- [5] Grau-Luque, E., Atlan, F., Becerril-Romero, I., Perez-Rodriguez, A., Guc, M., & Izquierdo-Roca, V. (2021). Spectrapepper: A Python toolbox for advanced analysis of spectroscopic data for materials and devices. *Journal of Open Source Software*, 6(67), 3781. <https://doi.org/10.21105/joss.03781>
- [6] Gholizadeh, A., et al. (2021). Vis–NIR and XRF data fusion and feature selection to estimate potentially toxic elements in soil. *Sensors*, 21(7), 2386. <https://doi.org/10.3390/s21072386>
- [7] Desnica, V., & Schreiner, M. (2006). A LabVIEW-controlled portable x-ray fluorescence spectrometer for the analysis of art objects. *X-Ray Spectrometry*, 35(5), 280–286. <https://doi.org/10.1002/xrs.906>
- [8] Campanella, L., Ferretti, M., Polese, C., García, C. R., & Sabatini, I. (2013). Non-destructive approach to multilayer objects: XRF analysis of gilt and enamelled metals of the medieval Cross of Rosciolo. *Procedia Chemistry*, 8, 284–291. <https://doi.org/10.1016/j.proche.2013.03.035>
- [9] Marguá, E., Queralt, I., & de Almeida, E. (2022). X-ray fluorescence spectrometry for environmental analysis: Basic principles, instrumentation, applications and recent trends. *Chemosphere*, 303, 135006. <https://doi.org/10.1016/j.chemosphere.2022.135006>
- [10] Mazzinghi, A., et al. (2021). MA-XRF for the characterisation of the painting materials and technique of the Entombment of Christ by Rogier van der Weyden. *Applied Sciences*, 11(13), 6151. <https://doi.org/10.3390/app11136151>
- [11] Gherase, M. R., & Fleming, D. E. B. (2011). Calculation of depth-dependent elemental concentration with X-ray fluorescence using a layered calibration method. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 269(10), 1150–

1156. <https://doi.org/10.1016/j.nimb.2011.02.086>
- [12] Van Grieken, R., & Markowicz, A. (2001). *Handbook of X-Ray Spectrometry* (2nd ed.). CRC Press.
- [13] Haschke, M., Flock, J., & Haller, M. (2021). *X-Ray Fluorescence Spectroscopy for Laboratory Applications*. John Wiley & Sons.
- [14] Marguá, E., & van Grieken, R. (2013). *X-ray fluorescence spectrometry and related techniques: An introduction*. Momentum Press.
- [15] Gerodimos, T., Asvestas, A., Mastrotheodoros, G. P., Chantas, G., Liougos, I., Likas, A., & Anagnostopoulos, D. F. (2022). Scanning X-ray Fluorescence data analysis for the identification of Byzantine icons' materials, techniques, and state of preservation: A case study. *Journal of Imaging*, 8(5), 147. <https://doi.org/10.3390/jimaging8050147>
- [16] Kadachi, A. N., & Al-Eshaikh, M. A. (2012). Limits of detection in XRF spectroscopy. *X-Ray Spectrometry*, 41(5), 350–354. <https://doi.org/10.1002/xrs.2412>
- [17] Kawai, J., Yamasaki, K., & Tanaka, R. (2019). Fundamental parameter method in X-ray fluorescence analysis. In *Encyclopedia of Analytical Chemistry* (pp. 1–14). <https://doi.org/10.1002/9780470027318.a9666>
- [18] Kramar, U. (1999). X-Ray fluorescence spectrometers. In *Encyclopedia of Spectroscopy and Spectrometry* (pp. 2469–2477). Academic Press. <https://doi.org/10.1016/b978-0-12-374413-5.00314-6>
- [19] Laclavetine, K., Ager, F. J., Arquillo, J., Respaldiza, M. Á., & Scrivano, S. (2016). Characterization of the new mobile confocal micro X-ray fluorescence (CXRF) system for in situ non-destructive cultural heritage analysis at the CNA: μ XRF-CONCHA. *Microchemical Journal*, 125, 62–68. <https://doi.org/10.1016/j.microc.2015.11.013>
- [20] Mantler, M., & Schreiner, M. (2011). *X-Ray Fluorescence Spectrometry (XRF) in Geoarchaeology*. New York, NY: Springer. <https://doi.org/10.1007/978-1-4419-6886-9>
- [21] Mantouvalou, I., et al. (2010). A compact 3D micro-X-ray fluorescence spectrometer with X-ray tube excitation for archaeometric applications. *Journal of Analytical Atomic Spectrometry*, 25(4), 554. <https://doi.org/10.1039/b915912f>
- [22] Oyedotun, T. D. T. (2018). X-ray fluorescence (XRF) in the investigation of the composition of earth materials: a review and an overview. *Geology, Ecology, and Landscapes*, 2(2), 148–154. <https://doi.org/10.1080/24749508.2018.1452459>

- [23] Patterson, B., & Havrilla, G. (2008). Confocal micro X-ray fluorescence: A new paradigm in 3D elemental imaging for materials characterization. *Microscopy and Microanalysis*, 14(S2), 1088–1089. <https://doi.org/10.1017/s1431927608083487>
- [24] Perez, R. D., Sánchez, H. J., Rubio, M., & Perez, C. A. (2010). Analysis of thin intermediate layers by confocal μ -XRF. *X-Ray Spectrometry*, 40(1), 19–23. <https://doi.org/10.1002/xrs.1277>
- [25] Potts, P. J., & Tsuji, K. (2018). X-Ray fluorescence and emission | Energy dispersive X-Ray fluorescence. In *Encyclopedia of Analytical Science* (3rd ed., pp. 1–12). Elsevier. <https://doi.org/10.1016/b978-0-12-409547-2.14519-6>
- [26] Pimenta, A., Felix, V., Oliveira, M., Andrade, M., Oliveira, M., & Freitas, R. (2023). Investigating Brazilian paintings from the 19th Century by MA-XRF. *Quantum Beam Science*, 7(1), 9. <https://doi.org/10.3390/qubs7010009>
- [27] Shackley, M. S. (2018). X-Ray Fluorescence Spectrometry (XRF). In *The Encyclopedia of Archaeological Sciences* (pp. 1–5). <https://doi.org/10.1002/9781119188230.saseas0620>
- [28] Streli, C., Wobrauschek, P., & Kregsamer, P. (1999). X-ray fluorescence spectroscopy, applications. In *Encyclopedia of Spectroscopy and Spectrometry* (pp. 2478–2487). <https://doi.org/10.1006/rwsp.2000.0337>
- [29] Tao, G. Y., Zhang, Z. Y., & Ji, A. (1990). XRF procedures for analysis of standard reference materials. *X-Ray Spectrometry*, 19(2), 85–88. <https://doi.org/10.1002/xrs.1300190211>
- [30] Brouwer, P. (2003). *Theory of XRF: Getting acquainted with the principles*. Almelo: Panalytical.
- [31] Tighe, M., Rogan, G., Wilson, S. C., Grave, P., Kealhofer, L., & Yukongdi, P. (2018). The potential for portable X-ray fluorescence determination of soil copper at ancient metallurgy sites, and considerations beyond measurements of total concentrations. *Journal of Environmental Management*, 206, 373–382. <https://doi.org/10.1016/j.jenvman.2017.10.052>
- [32] Uhler, K., et al. (2008). Applications of a new portable (micro) XRF instrument having low-Z elements determination capability in the field of works of art. *X-Ray Spectrometry*, 37(4), 450–457. <https://doi.org/10.1002/xrs.1074>
- [33] Wang, S., Ward, J., Leyffer, S., Wild, S. M., Jacobsen, C., & Vogt, S. (2014). Unsupervised cell identification on multidimensional X-ray fluorescence datasets. *Journal of Synchrotron Radiation*, 21(3), 568–

579. <https://doi.org/10.1107/s1600577514001416>

[34] Zwicky, C. N., & Lienemann, P. (2004). Quantitative or semi-quantitative?—laboratory-based WD-XRF versus portable ED-XRF spectrometer: results obtained from measurements on nickel-base alloys. *X-Ray Spectrometry*, 33(4), 294–

300. <https://doi.org/10.1002/xrs.730>

[35] Janssens, K., et al. (2000). Use of microscopic XRF for non-destructive analysis in art and archaeometry. *X-Ray Spectrometry*, 29, 73–

91. [https://doi.org/10.1002/\(SICI\)1097-4539\(200001/02\)29:1<73::AID-](https://doi.org/10.1002/(SICI)1097-4539(200001/02)29:1<73::AID-XRS416>3.0.CO;2-M)

XRS416>3.0.CO;2-M

[36] Rousseau, R. M., Willis, J. P., & Duncan, A. R. (1996). Practical XRF calibration procedures for major and trace elements. *X-Ray Spectrometry*, 25(4), 179–189.

Bibliography – Ma – XRF

[1] Adler, L. A., Axelrod, J., & Branco, J. J. R. (1960). *Advances in X-ray Analysis*, 2, 167.

[2] Asher, R. K., & Ruiz, E. R. (1986). *Plating and Surface Finishing*, 73, 30.

[3] Attaelmanan, A. (1995). *On resolution enhancement of capillary based scanning x-ray microfluorescence spectroscopy* (Doctoral dissertation). Chalmers University of Technology, Göteborg, Sweden.

[4] Bethe, A. (1930). *Annalen der Physik*, 5, 325.

[5] Boehme, D. R. (1987). *Advances in X-ray Analysis*, 30, 39.

[6] Carpenter, D. A. (2000). In K. Janssens, F. Adams, & A. Rindby (Eds.), *Microscopic X-ray Fluorescence Analysis* (p. 347). Chichester: Wiley.

[7] Carpenter, D. A., & Taylor, M. A. (1991). *Advances in X-ray Analysis*, 34, 217.

[8] Carpenter, D. A., Gorin, A., & Shor, J. T. (1995a). *Advances in X-Ray Analysis*, 38, 557.

[9] Carpenter, D. A., Taylor, M. A., & Lawson, R. L. (1995b). *Journal of Trace and Microprobe Techniques*, 13, 141.

[10] Czechanski, M. (1986). Measurement of precious metal coating thickness using x-ray fluorescence instrumentation. In *Proceedings of the 9th International Precious Metals Conference* (p. 457).

[10] Engstrom, P. (1991). *Development of capillary optics for x-ray focussing* (Doctoral dissertation). Chalmers University of Technology, Göteborg, Sweden.

[11] Fukumoto, N., Kobayashi, Y., Uchiumi, A., Kuruhashi, M., Kawase, A., & Watanabe, H. (1990). In K. Shinohara (Ed.), *X-ray Microscopy in Biology and Medicine* (p. 227). Tokyo: Japan Science Society Press.

[12] Gordon, B., & Jones, K. (1993). In M. D. Morris (Ed.), *Microscopic and Spectroscopic Imaging of the Chemical State* (p. 303). New York: Marcel Dekker.

- [13] Herman, G. T. (1980). *Image Reconstructions from Projections*. London: Academic Press.
- [14] Hoffman, S., Bilderback, D., & Thiel, D. (1994). *Nuclear Instruments and Methods in Physics Research Section A*, 347, 384.
- [15] Holynska, B., Ostachowicz, B., Ostachowicz, J., Ostrowski, A., Ptasinski, J., & Wegrzynek, D. (1995). *Journal of Trace and Microprobe Techniques*, 13, 163.
- [16] Iida, A. (2000). In K. Janssens, F. Adams, & A. Rindby (Eds.), *Microscopic X-ray Fluorescence Analysis* (p. 117). Chichester: Wiley.
- [17] Janssens, K., Van Langevelde, F., Adams, F. C., Vis, R. D., Sutton, S. R., Rivers, M. L., Jones, K. W., & Bowen, D. K. (1992). *Advances in X-ray Analysis*, 35, 1265.
- [18] Janssens, K., Adams, F., & Rindby, A. (Eds.). (2000a). *Microscopic X-ray Fluorescence Analysis*. Chichester: Wiley.
- [19] Janssens, K., Vincze, L., Vekemans, B., Aerts, A., Adams, F., Jones, K. W., & Knöchel, A. (1996c). *Microchimica Acta*, 13, 87.
- [20] Janssens, K., Vincze, L., Wei, F., Proost, K., Vekemans, B., Vittiglio, G., He, Y., Yan, Y., & Falkenberg, G. (2000c). Feasibility of polycapillary-based trace-level m-XANES by means of bending magnet radiation. In *Book of Abstracts of European Conference on Energy Dispersive X-ray Spectrometry* (p. 18). Krakow, Poland.
- [21] Jones, P. D., Briffa, K. R., & Schweingruber, F. H. (1995). *Geophysical Research Letters*, 22, 1333.
- [22] Kardiawarman, K., York, B. R., Qian, X. W., Xiao, Q. F., MacDonald, C. A., & Gibson, W. M. (1995). In R. Hoover & M. B. Williams (Eds.), *X-ray and Ultraviolet Sensors and Applications, Proc SPIE 2519* (p. 197). Society of Photo-Optical Instrumentation and Engineering.
- [23] Lankosz, M., Holynska, B., & Pella, P. A. (1993). *X-Ray Spectrometry*, 22, 54.
- [24] Larsson, S. (1991). *X-ray microbeam spectroscopy* (Doctoral dissertation). Chalmers University of Technology, Göteborg, Sweden.
- [25] Mott, N. F., & Massey, H. S. W. (1949). *The Theory of Atomic Collisions*. Oxford: Oxford University Press.
- [26] Nellessen, J. (1996). *Erweiterung der micro-röntgenfluoreszenzanalyse auf dreidimensionale proben* (Diplomarbeit). Fachbereich Chemie, Universität Dortmund.
- [27] Osan, J., Torok, S., & Rindby, A. (2000). In K. Janssens, F. Adams, & A. Rindby (Eds.), *Microscopic X-ray Fluorescence Analysis* (p. 315). Chichester: Wiley.
- [28] Paque, J. M., Browning, R., King, P. L., & Pianetta, P. (1990). In *Proceedings 19th Microbeam Analysis Society Meeting* (p. 202). San Francisco: San Francisco Press.
- [29] Reed, S. J. B. (1975). *Electron Microprobe Analysis*. Cambridge: Cambridge University Press.
- [30] Robinson, J. W. (1974). *Handbook of Spectroscopy* (Vol. 1). Boca Raton, FL: CRC Press.
- [31] Simionovici, A., Chukalina, M., Schroer, C., Drakopoulos, M., Snigireva, I., Snigirev, A., ... & Adams, F. (2000). *IEEE Transactions on Nuclear Science*, 47, 2736.

- [32] Stebel, M. D., & Silverman, W. M. (1984). XRF programmable plating thickness measurement instrumentation for automated testing of small parts and printed wiring boards. In *Proceedings of the ICWA 1984 Coil Winding Conference* (p. 75).
- [33] Stocklein, W., Gunaratnam, L., Hicks, T., Kidfelt, G., Rindby, A., Warman, G., & Widstedt, I. (1997). The analysis of float glass using elemental analysis: Comparison of techniques. In *Proceedings of the EAFS (European Association of Forensic Societies) 1997 Conference*. Lyon.
- [34] Tissot, R. G. Jr., & Boehme, D. R. (1994). The novel use of XRMF spectroscopy for environmental toxicity evaluation and recyclable potential in electronic components. In *Proceedings Denver X-ray Conference*.
- [35] Trebbia, P., Wulveryk, J. M., & Bonnet, M. (1995). *Microbeam Analysis*, 4, 85.
- [36] Vekemans, B., Vincze, L., Szaloki, I., Feng, H., Riekel, C., Van Grieken, R., Adams, F., & Jones, K. W. (2000). Micro X-ray fluorescence tomography on single sediment particles from New York/New Jersey Harbor. In *Book of Abstracts of European Conference on Energy Dispersive X-ray Spectrometry* (p. 106). Krakow.
- [37] Wegrzynek, D. (2000). A feasibility study: Computer micro tomography using a laboratory x-ray fluorescence micro-beam spectrometer. In *Book of Abstracts of European Conference on Energy Dispersive X-ray Spectrometry* (p. 25). Krakow.
- [38] Wuttman, M. (1996). The beginnings of copper and copper alloys metallurgy in Egypt. The need for a better characterisation of the artefacts. In *Proceedings First Int. Conference on Ancient Egyptian Mining and Metallurgy and Conservation of Metallic Artefacts*. Cairo.
- [39] Zietz, L., & Baez, A. V. (1957). In V. E. Cosslett, A. Engstrom, & H. H. Pattee (Eds.), *X-ray Microscopy and Microradiography*. New York: Academic Press.
- [40] Bichlmeier, S., Janssens, K., Heckel, J., Gibson, D., Hoffmann, P., & Ortner, H. (2001). Component analysis of paint layers by micro X-ray fluorescence. *X-Ray Spectrometry*, 30(1), 8.
- [41] Bonnet, N. (1995). *Microchimica Acta*, 120, 195.