

Σχολή Μηχανικών Ορυκτών Πόρων
Πολυτεχνείο Κρήτης



Διπλωματική Εργασία

Χημειομετρική πρόβλεψη περιεκτικότητας αρωματικών
συστατικών σε μεσαία πετρελαϊκά κλάσματα από
δεδομένα υπέρυθρης φασματοσκοπίας

ΝΟΔΑΡΑΣ ΕΥΣΤΑΘΙΟΣ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ:

ΠΑΣΑΔΑΚΗΣ ΝΙΚΟΛΑΟΣ (Επιβλέπων), Αν. Καθηγητής Π.Κ.

ΖΕΡΒΑΚΗΣ ΜΙΧΑΗΛ, Καθηγητής Π.Κ.

ΜΑΘΙΟΥΔΑΚΗΣ ΕΜΜΑΝΟΥΗΛ, Επ. Καθηγητής Π.Κ.

*Χανιά
Δεκέμβριος 2014*

ΕΥΧΑΡΙΣΤΙΕΣ

Η παρούσα διπλωματική δεν θα μπορούσε να έχει πραγματοποιηθεί χωρίς την καθοδήγηση και την βοήθεια διαφόρων ατόμων, που με τον ένα ή τον άλλο τρόπο, προσέφεραν την πολύτιμη στήριξη στην προετοιμασία και την ολοκλήρωση της εργασίας αυτής.

Πρώτα από όλα θα ήθελα να ευχαριστήσω ειλικρινά και θερμά τον επιβλέποντα της παρούσας εργασίας Αν. Καθηγητή Νικόλαο Πασαδάκη ο οποίος αρχικά μου έδωσε τη δυνατότητα να πραγματοποιήσω την εργασία αυτή και στην συνέχεια με καθοδήγησε με τις πολύτιμες συμβουλές αλλά και με τις καίριες παρατηρήσεις του σε όλη την διάρκεια διεκπεραίωσης της.

Επιπλέον θα ήθελα να εκφράσω τις ευχαρηστίες μου στον Επ.Καθηγητή Μαθιουδάκη Εμμανουήλ και στον Καθηγητή Ζερβάκη Μιχαήλ για την συμμετοχή τους στην εξεταστική επιτροπή.

Επίσης ευχαριστώ την κ.Χαμηλάκη Ελίνα για την καθοδήγηση της κατά τις εργαστηριακές αναλύσεις, τον κ.Παπαγεωργίου Κωσταντίνο για την βοήθεια στην κατανόηση επιμέρους θεμάτων της εργασίας και την κ.Κουκούνια Άννα για την συμβολή της στην επεξεργασία του κειμένου.

Τέλος οφείλω το μεγαλύτερο ευχαριστώ στην οικογενειά μου που με στήριξε στην πορεία μου στο τμήμα Μηχανικών Ορυκτών Πόρων του Πολυτεχνείου Κρήτης.

Σε αυτούς, που με την καθημερινή τους συμπαράσταση, την υπομονή τους και την θετική τους σκέψη, συνέβαλαν στην εκπλήρωση του στόχου μου αφιερώνεται η εργασία αυτή.

ΠΕΡΙΛΗΨΗ.....	4
ΕΙΣΑΓΩΓΗ.....	5
1. Η ΣΥΣΤΑΣΗ ΤΩΝ ΠΕΤΡΕΛΟΕΙΔΩΝ.....	7
1.1 Παραφινικοί Υδρογονάνθρακες.....	8
1.2 Ναφθενικοί Υδρογονάνθρακες.....	8
1.3 Αρωματικοί Υδρογονάνθρακες.....	8
1.3.1 Γεωχημική Σημασία Αρωματικών Υδρογονανθράκων.....	12
1.3.2 Παράμετροι και Ιδιότητες που σχετίζονται με τη συγκέντρωση Αρωματικών Υδρογονανθράκων.....	12
2. ΧΡΩΜΑΤΟΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ.....	15
2.1 Εισαγωγή.....	15
2.2 Ταξινόμηση των Βασικών Χρωματογραφικών Τεχνικών.....	16
2.3 Βασικοί Χρωματογραφικοί Όροι.....	18
2.4 Ποιοτική και Ποσοτική Χρωματογραφική Ανάλυση.....	21
2.5 Υγρή Χρωματογραφία Υψηλής Πίεσης (High Pressure Liquid Chromatography – HPLC).....	21
2.6 Χρωματογραφική Ανάλυση δειγμάτων.....	24
2.6.1 Πρότυπες Μέθοδοι Χρωματογραφικής Ανάλυσης (ASTM).....	24
2.6.2 Αποτελέσματα Χρωματογραφικής Ανάλυσης των δειγμάτων.....	32
3. ΦΑΣΜΑΤΟΣΚΟΠΙΑ ΥΠΕΡΥΘΡΟΥ.....	38
3.1 Εισαγωγή.....	38
3.2 Φασματοσκοπία Υπερύθρου.....	39
3.3 Οργανολογία Υπερύθρου.....	44
3.4 Αρχές Λειτουργίας Φασματοσκοπίας FTIR.....	46
3.5 Περιγραφή Συλλογής Φασματοσκοπικών Δεδομένων των δειγμάτων.....	48
3.6 Οι αρωματικοί υδρογονάνθρακες στο φάσμα του υπερύθρου.....	51
4 .ΠΟΛΥΜΕΤΑΒΛΗΤΕΣ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ.....	55

4.1 Εισαγωγή.....	55
4.2 Βασικές Έννοιες και Μεγέθη της Στατιστικής.....	55
4.3 Προεπεξεργασία Φασματικών Δεδομένων.....	61
4.4 Πολυμεταβλητές Μέθοδοι Επεξεργασίας Δεδομένων.....	65
4.4.1 Εισαγωγή.....	65
4.4.2 Σημασία της Πολυμεταβλητής Ανάλυσης και Βασικές Έννοιες.....	66
4.4.3 Ανάλυση Κύριων Συνιστωσών PCA (Principle Component Analysis)...	69
4.4.4 Πολλαπλή Γραμμική Παλινδρόμηση.....	76
4.4.5 Παλινδρόμηση Κυρίων Συνιστωσών PCR (Principle Component Regression).....	77
4.4.6 Μέθοδος των Μερικών Ελαχίστων Τετραγώνων PLS (Partial Least Squares).....	79
5 .ΑΠΟΤΕΛΕΣΜΑΤΑ.....	82
5.1 Μοντελα Βασισμένα στο σύνολο των δειγμάτων.....	83
5.2 Μοντέλα Πρότυπης Αναλυτικής Μεθόδου ASTM D1319.....	91
5.3 Μοντέλα που έχουν αναλυθεί με Πρότυπες Αναλυτικές Μεθόδους HPLC.....	99
5.4 Μοντέλα δειγμάτων που έχουν αναλυθεί με τις Πρότυπες Αναλυτικές Μεθόδους ASTM D7419, D2549.....	107
5.5 Μοντέλα Πρότυπων Αναλυτικών Μεθόδων ASTM D1319, D2549 (Ανοιχτής Στήλης).....	115
5.6 Μοντέλα Βασισμένα σε δείγματα με Σημείο Βρασμού Μεγαλύτερο των 400°C..	123
5.7 Μοντέλα Βασισμένα σε δείγματα με Σημείο Βρασμού Μικρότερο των 400°	131
6.ΣΥΜΠΕΡΑΣΜΑΤΑ.....	141
BIBΛΙΟΓΡΑΦΙΑ.....	142

ΠΕΡΙΛΗΨΗ

Η μέτρηση επιμέρους ιδιοτήτων σε πετρέλαια και πετρελαϊκά κλάσματα είναι συχνά χρονοβόρα και υψηλού κόστους διαδικασία. Για το λόγο αυτό έχουν χρησιμοποιηθεί έμμεσοι τρόποι υπολογισμού τους, οι οποίοι στηρίζονται στη χρήση μαθηματικών μοντέλων πρόβλεψης με βάση δεδομένα χημικής σύστασης που προέρχονται από φασματοσκοπικές αναλύσεις. Οι αναλύσεις αυτές απαιτούν μικρό χρόνο, έχουν περιορισμένο κόστος ενώ υπάρχει και η δυνατότητα της on-line εφαρμογής τους.

Στην παρούσα εργασία αναπτύχθηκαν μοντέλα πρόβλεψης της συγκέντρωσης των αρωματικών συστατικών σε κλάσματα πετρελαίου, με βάση τα φάσματα τους στο μέσο υπέρυθρο (mid-IR). Χρησιμοποιήθηκαν 130 πετρελαϊκά δείγματα, που καλύπτουν ένα ευρύ φάσμα περιοχών σημείων βρασμού, από ελαφρές νάφθες έως και βαριά λιπαντικά κλάσματα. Η περιεκτικότητα των αρωματικών συστατικών σε αυτά προσδιορίστηκε εργαστηριακά, χρησιμοποιώντας χρωματογραφικές τεχνικές, όπως υγρή χρωματογραφία υψηλής πίεσης (HPLC), χρωματογραφία ανοικτής στήλης κ.ά. Η συγκέντρωση των αρωματικών επιλέχθηκε ως η ιδιότητα-στόχος πρόβλεψης των μοντέλων, λόγω της επίδρασης της στη διαμόρφωση των ιδιοτήτων των πετρελαϊκών προϊόντων, όσο και για την περιβαλλοντική σημασία της.

Στην ανάπτυξη των μοντέλων πρόβλεψης χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS). Η εκπαίδευση και ο έλεγχος των μοντέλων υλοποιήθηκε σε περιβάλλον Matlab, με προσαρμογή ήδη υφιστάμενου στο εργαστήριο κώδικα.

Συγκεκριμένα αναπτύχθηκαν μοντέλα για επιμέρους ομάδες πετρελαϊκών κλασμάτων, αλλά και για ομάδες δειγμάτων στις οποίες η περιεκτικότητα των αρωματικών είχε προσδιοριστεί με την ίδια αναλυτική μέθοδο. Η ακρίβεια της πρόβλεψης της συγκέντρωσης των αρωματικών, υπολογισμένη με βάση τα δείγματα ελέγχου, στα περισσότερα από τα μοντέλα που αναπτύχθηκαν ήταν συγκρίσιμη εκείνης των πρότυπων εργαστηριακών μεθόδων.

ΕΙΣΑΓΩΓΗ

Η μέτρηση των ιδιοτήτων πετρελαιοειδών είναι μια απαιτητική αναλυτική μέθοδος καθώς θεωρείται χρονοβόρα και δύσκολη και αυτά την καθιστούν ακατάλληλη για χρήση σε on-line εφαρμογές. Τις τελευταίες δυο δεκαετίες η υπέρυθη φασματοσκοπία χρησιμοποιείται σε διάφορους τομείς λόγω των πλεονεκτημάτων της σε σχέση με άλλες μεθόδους ανάλυσης. Το κυριότερο από αυτά είναι η ικανότητα της να καταγράφει φάσματα, χωρίς την ανάγκη προεπεξεργασίας για οποιοδήποτε στερεό ή υγρό δείγμα, τα οποία είναι χαρακτηριστικά της σύστασης των δειγμάτων. Η εξοικονόμηση κόστους των μετρήσεων IR έγκειται στο ότι είναι η ταχύτερη μέθοδος και αυτό εξυπηρετεί μετρήσεις ελέγχου και βελτίωσης ποιότητας προϊόντων. Επιπλέον είναι μια μη καταστρεπτική τεχνική μέτρησης για πολλές χημικές ενώσεις που παρέχει μεγάλη ακρίβεια και ευκολία στη χρήση. Λόγω αυτών των χαρακτηριστικών της, έχει αποδειχθεί ιδιαίτερος αποτελεσματική σε διάφορους τομείς όπως η γεωργία, η βιομηχανία τροφίμων, οι ιατρικές διαγνώσεις, η φαρμακευτική βιομηχανία και βέβαια η βιομηχανία πετροχημικών.

Όσον αφορά την βιομηχανία πετρελαίου είναι γνωστό πως τα προϊόντα της διύλισης του πετρελαίου πρέπει να πληρούν κάποιες συγκεκριμένες προδιαγραφές ποιότητας. Εξαιτίας της σημαντικότητας των προϊόντων αυτών στην καθημερινή ανθρώπινη δραστηριότητα ο συνεχής έλεγχός τους είναι επιτακτικός. Ωστόσο οι πρότυπες μέθοδοι μέτρησης για τα προϊόντα πετρελαίου είναι δαπανηρές, χρονοβόρες αλλά και περίπλοκες. Για παράδειγμα για την μέτρηση του αριθμού οκτανίων ενός δείγματος, που είναι η πιο γνωστή ιδιότητα των βενζινών στο ευρύ κοινό, είναι απαραίτητο ένα τυπικό όργανο δοκιμής ASTM-CFR. Όμως η λειτουργία αυτού του οργάνου έχει μεγάλο κόστος ενώ ταυτόχρονα η μέτρηση ενός μόνο δείγματος διαρκεί 20 λεπτά [12]. Αυτό σημαίνει πως μετρήσεις σαν και αυτή δεν είναι κατάλληλες για on-line εφαρμογές, όπως οι λειτουργίες ανάμιξης διυλιστηρίου.

Από την άλλη μεριά οι κανόνες και η νομοθεσία που αφορούν την ποιότητα των προϊόντων πετρελαίου συνεχίζουν να γίνονται όλο και πιο αυστηρά λόγω νέων περιβαλλοντικών ζητημάτων που προκύπτουν συνεχώς. Η περιεκτικότητα των προϊόντων αυτών σε αρωματικούς υδρογονάνθρακες επηρεάζει καθοριστικά την ποιότητα τους και γι' αυτό είναι απαραίτητη η συνεχής μέτρηση της κατά την παραγωγή αλλά και διακίνηση τους. Κατά την διακίνηση υπάρχει και η περίπτωση παράνομης ανάμειξης τελικών προϊόντων, σε βάρος του καταναλωτή και του περιβάλλοντος.

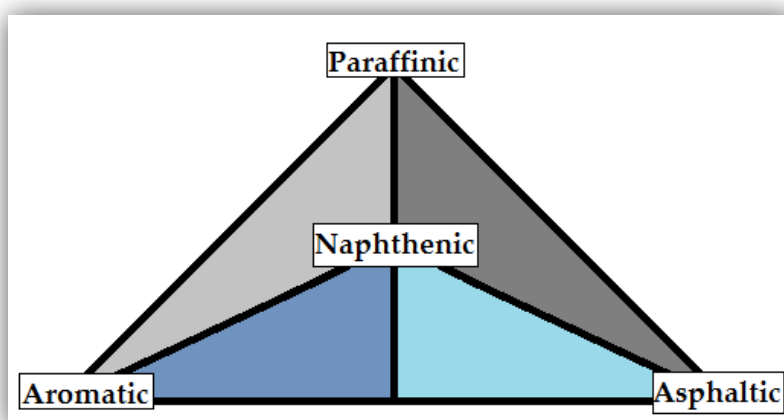
Η φασματική ανάλυση σε συνδιασμό με την χημειομετρία έχει αποδεδειγμένη αποτελεσματικότητα στους τομείς πρόβλεψης ιδιοτήτων πετρελαιοειδών. Η χημειομετρία είναι ένας όρος που μπορεί να οριστεί ως η υποστηριζόμενη χημική ανάλυση από μαθηματικές, στατιστικές αλλά και άλλες μεθόδους που βασίζονται στην τυπική λογική ώστε να σχεδιαστούν και να επιλεγθούν οι βέλτιστες διαδικασίες μετρήσεων και πειραμάτων και να παραχθεί η μέγιστη σχετική πληροφορία αναλύοντας χημικά δεδομένα. Τα δεδομένα που προκύπτουν από την φασματομετρία είναι μεγάλα σύνολα αριθμών και αυτό συνεπάγεται την ύπαρξη μεγάλου αριθμού μεταβλητών. Η πολυμεταβλητή ανάλυση των δεδομένων θεωρείται η πιο ισχυρή μέθοδος για την ερμηνεία φασματικών δεδομένων. Τα πλεονεκτήματα της πολυμεταβλητής ανάλυσης σε σχέση με την ανάλυση μιας μεταβλητής είναι τα εξής: η μείωση του θορύβου των δεδομένων, η χρήση της πληροφορίας που προκύπτει από αλληλεπιδράσεις των επιμέρους τμημάτων και ο εντοπισμός των δεδομένων που βρίσκονται εκτός του εύρους που διαμορφώνεται από τη θέση της πλειοψηφείας των υπολοίπων.

Είναι σημαντικό να σημειωθεί σε αυτό το σημείο πως οι συνθήκες λήψης φασματικών δεδομένων για τα δείγματα (πίεση, θερμοκρασία) παίζουν σημαντικό ρόλο στην αξιοπιστία του μοντέλου πρόβλεψης και γιαυτό πρέπει να είναι όμοιες. Επίσης τα χαρακτηριστικά του φασματομέτρου πρέπει να είναι ίδια σε μετρήσεις νέων δειγμάτων.

1. Η ΣΥΣΤΑΣΗ ΤΩΝ ΠΕΤΡΕΛΑΙΟΕΙΔΩΝ

Το πετρέλαιο δεν είναι μόνο η πρώτη ύλη πρωτογενούς ενέργειας που χρησιμοποιείται, αλλά είναι επίσης η βάση για τη βιομηχανική ανάπτυξη. Παρά την ταχεία ανάπτυξη και την πρόοδο των ανανεώσιμων και άλλων πηγών ενέργειας, το πετρέλαιο είναι και θα είναι στο εγγύς μέλλον η κύρια πηγή πρωτογενούς ενέργειας. Επιπλέον, το πετρέλαιο χρησιμοποιείται σε πολλές εφαρμογές της καθημερινής ζωής, που εξυπηρετούν τις ανθρώπινες ανάγκες: τις μεταφορές (ως καύσιμο για κινητήρες ντίζελ και βενζίνης), θέρμανση (ως καύσιμο για το σπίτι και βιομηχανικούς καυστήρες), της ενέργειας (ως καύσιμα σε σταθμούς παραγωγής ηλεκτρικής ενέργειας) και πρώτων υλών (για τη παραγωγή λιπαντικών και υλικών κατασκευής, όπως άσφαλτο και την παραγωγή των υλικών από τη βιομηχανία πετροχημικών, όπως τα πολυμερή, τα λιπάσματα, τα φάρμακα και άλλα). Το 90% του παραγόμενου πετρελαίου χρησιμοποιείται ως καύσιμο, το οποίο καλύπτει το ήμισυ των συνολικών ενεργειακών αναγκών, ενώ το 10% ως πρώτη ύλη.

Το πετρέλαιο είναι ένα εξαιρετικά πολύπλοκο μείγμα υδρογονανθράκων και άλλων ενώσεων του άνθρακα και υδρογόνου που περιέχει επιπλέον άζωτο, οξυγόνο, θείο και μέταλλα [7]. Η χημική σύνθεσή του ποικίλλει έντονα ανάλογα με την προέλευση και την ηλικία του. Το στοιχείο με την υψηλότερη περιεκτικότητα είναι ο άνθρακας και το επόμενο στοιχείο το οποίο βρίσκεται σε αφθονία έναντι των άλλων είναι το υδρογόνο. Αυτό δείχνει ότι το πετρέλαιο αποτελείται κυρίως από υδρογονάνθρακες. Οι υδρογονάνθρακες που βρίσκονται στο πετρέλαιο είναι κυρίως αλκάνια, κυκλοαλκάνια και αρωματικά. Τα αλκένια (ολεφίνες) είναι σπάνια και αλκίνια (υδρογονάνθρακες της σειράς ασετυλίνης) πρακτικά δεν υπάρχουν.



Σχήμα 1.1: Το θεωρητικό τρίγωνο που εισάγει διακρίσεις στις τάξεις του πετρελαίου.

Ο παλαιότερος και πιο διαδεδομένος τρόπος για ταξινόμηση των πετρελαίων ανάλογα με το περιεχόμενο των παραφινών, αρωματικών και ασφαλικών συστατικών. Κάθε πετρέλαιο αντιστοιχεί σε ένα σημείο στο εσωτερικό του τριγώνου και ονομάζεται ανάλογα με τη σχετική θέση του, όπως φαίνεται στο Σχήμα 1.1. Συνήθως αυτό το σχήμα χρησιμοποιείται σε συνδυασμό με το διαχωρισμό του πετρελαίου σε ελαφρύ και βαρύ και έτσι οι πιθανές υποκατηγορίες διπλασιάζονται[4].

1.1 Παραφινικοί υδρογονάνθρακες

Πρόκειται για κορεσμένους υδρογονάνθρακες ευθείας ή διακλαδισμένης δομής χωρίς κλειστούς δακτυλίους. Οι παραφίνες με 1-4 άτομα άνθρακα είναι αέρια, με 5-16 άτομα άνθρακα είναι υγρά και με 17 και άνω είναι στερεά σε κανονικές συνθήκες. Είναι δυνατή η ύπαρξη ισομερών δηλαδή μορίων με τον ίδιο μοριακό αλλά διαφορετικό συντακτικό τύπο. Ισομερή εμφανίζονται για αλκάνια με 4 ή περισσότερα άτομα άνθρακα. Παρουσιάζουν τα χαμηλότερα σημεία ζέσεως συγκριτικά με υδρογονάνθρακες άλλων κατηγοριών με ίδιο αριθμό ατόμων άνθρακα. Η πυκνότητα των αλκανίων είναι η μικρότερη σε σχέση με όλες τις ομάδες οργανικών συστατικών. Είναι αδιάλυτα στο νερό. Έχουν ιδιαίτερα υψηλό ενεργειακό περιεχόμενο, λόγω όμως της χαμηλής πυκνότητάς τους, το περιεχόμενο αυτό είναι το χαμηλότερο από κάθε άλλη ομάδα υδρογονανθρακών [5].

1.2 Ναφθενικοί υδρογονάνθρακες

Πρόκειται για κορεσμένους υδρογονάνθρακες με έναν ή περισσότερους δακτυλίους ατόμων και με ή χωρίς πλευρικές διακλαδώσεις παραφινικών αλυσίδων. Οι δακτύλιοι περιλαμβάνουν 3-7 άτομα άνθρακα. Μαζί με τα παραφινικά αναφέρονται ως το κορεσμένο κλάσμα του πετρελαίου. Η συγκέντρωσή τους φθάνει μέχρι και το 50%, ενώ αυξάνει από τα ελαφρύτερα στα βαρύτερα κλάσματα.

1.3 Αρωματικοί υδρογονάνθρακες

Ένας αρωματικός υδρογονάνθρακας είναι μια ένωση άνθρακα και υδρογόνου που παρουσιάζει εναλλασσόμενους διπλούς και απλούς δεσμούς μεταξύ των ατόμων άνθρακα που σχηματίζουν δακτυλίους. Ο όρος αρωματικός οφείλεται στο γεγονός ότι οι ενώσεις με αρωματικό δακτύλιο έχουν ευχάριστο άρωμα. Ο δακτύλιος που εμφανίζεται στους αρωματικούς υδρογονάνθρακες είναι ο βενζολικός που διαμορφώνεται από έξι άτομα άνθρακα όπως φαίνεται παρακάτω.



Σχήμα 1.2: Δακτύλιος Βενζολίου

Στα περισσότερα πετρέλαια το ποσοστό τους δεν ξεπερνά το 15% κ.β.. Σε κλάσματα πετρελαίου, όπως η βενζίνη, η υψηλή περιεκτικότητα αρωματικών είναι επιθυμητή λόγω του υψηλού αριθμού οκτανίων που εξασφαλίζουν. Απο την άλλη μεριά, λόγω της μεγάλης μεταβολής που παρουσιάζει το ιξώδες τους με τη θερμοκρασία, σε κλάσματα που χρησιμοποιούνται σε λίπανση είναι ανεπιθύμητοι [4]. Οι αρωματικοί υδρογονάνθρακες έχουν μεγαλύτερη πυκνότητα συγκριτικά με τα αλκάνια και τα ναφθένια και διαλύονται κατά κύριο λόγο σε πολικούς διαλύτες. Οι αρωματικοί δακτύλιοι εμφανίζονται μεμονωμένοι (μονοκυκλικοί αρωματικοί υδρογονάνθρακες) ή συζυγείς (πολυκυκλικοί αρωματικοί υδρογονάνθρακες). Οι πολυκυκλικοί αρωματικοί υδρογονάνθρακες αποτελούν αντικείμενο έρευνας τα τελευταία χρόνια σαν επικίνδυνοι ρυπαντές με καρκινογόνο δράση[26]. Μπορεί να έχουν ευθύγραμμη ή γωνιακή συμπίκνωση, όπως φαίνεται παρακάτω.


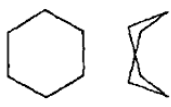
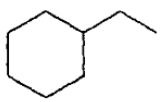
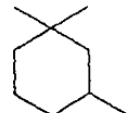
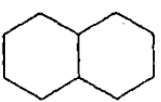

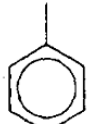
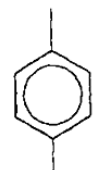

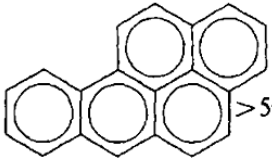


Βενζοπυρένιο

Κορονένιο

Σχήμα 1.3: Συμπύκνωση αρωματικών δακτυλίων.

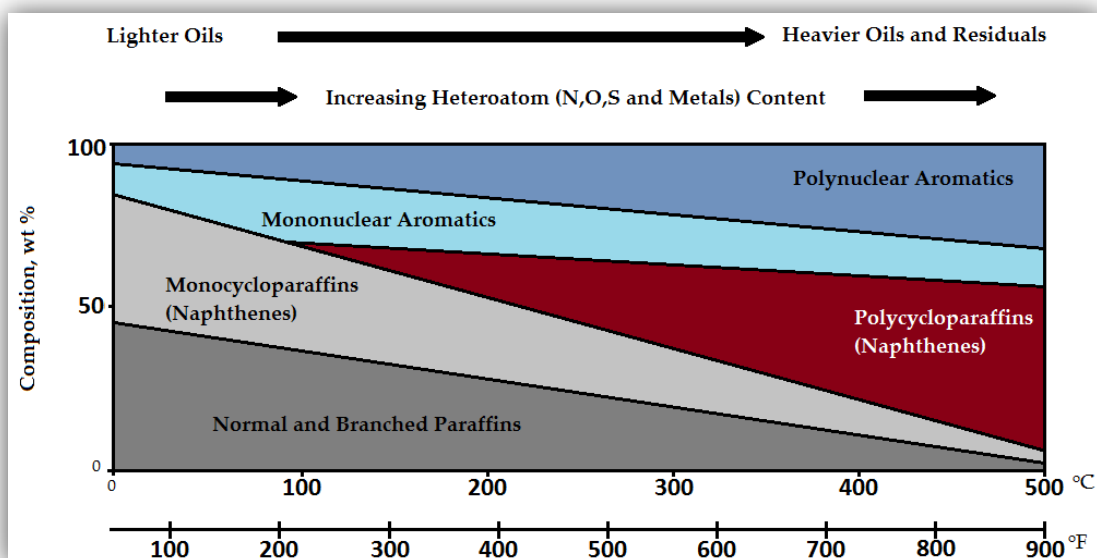
Στο Σχήμα 1.4 παρουσιάζονται ναφθενικοί και αρωματικοί υδρογονάνθρακες του πετρελαίου.

Naphthenes (Cycloparaffins)		Boiling point,
C_6H_{12}	 Methylcyclopentane	72°C
C_6H_{12}	 Cyclohexane (Side View)	81°C
C_8H_{16}	 Ethylcyclohexane	132°C
C_9H_{18}	 1,1,3-Trimethylcyclohexane	137°C
$C_{10}H_{18}$	 Decalin Trans form,	187°C
Aromatics		Boiling point,
C_6H_6	 Benzene	80°C
C_7H_8	 Toluene	111°C
C_8H_{10}	 Paraxylene	138°C
C_9H_{12}	 Isopropylbenzene	152°C
$C_{20}H_{12}$	 3,4-Benzpyrene	>500°C

Σχήμα 1.4 :Ναφθενικοί και αρωματικοί υδρογονάνθρακες του πετρελαίου

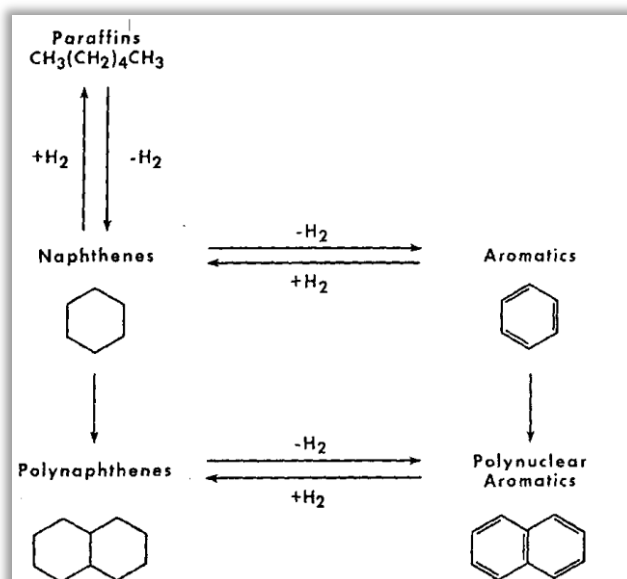
Η συμμετοχή των διαφόρων ομάδων στην σύσταση των κλασμάτων του πετρελαίου συναρτήσει των σημείων ζέσεως τους παρουσιάζεται στο Σχήμα 1.5. Με την αύξηση του σημείου ζέσεως μειώνεται η περιεκτικότητα του πετρελαίου σε αλκάνια, μονοκυκλικά ναφθένια και μονοαρωματικούς υδρογονάνθρακες, ενώ παράλληλα αυξάνονται οι

συγκεντρώσεις των πολυαρωματικών και πολυκυκλικών ναφθενικών συστατικών [4] [5]. Παρατηρείται αύξηση των πολυαρωματικών και πολυκυκλικών ναφθενίων με την αύξηση του σημείου ζέσεως, με ταυτόχρονη μείωση των παραφινών και των μονοκυκλικών ναφθενίων.



Σχήμα 1.5 : Κατανομή των ομάδων συστατικών συναρτήσει του σημείου ζέσεως στο πετρέλαιο

Οι πιθανές χημικές μετατροπές μεταξύ των ομάδων υδρογονανθράκων του πετρελαίου παρουσιάζονται στο Σχήμα 1.6 :

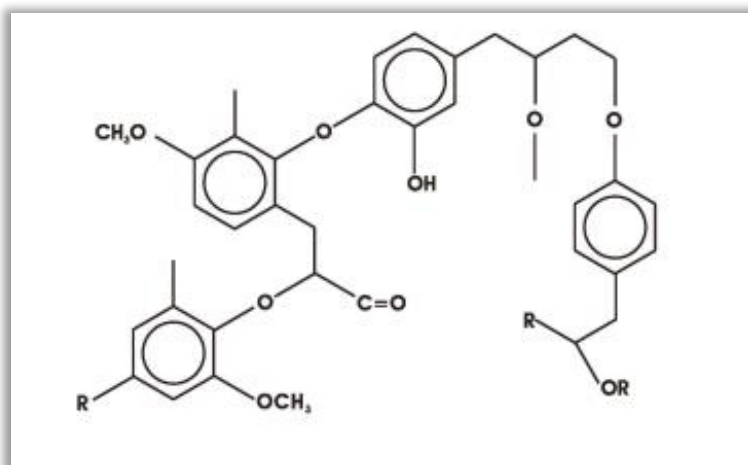


Σχήμα 1.6 : Αντιδράσεις ομάδων υδρογονανθράκων

Όπως φαίνεται στο σχήμα οι αρωματικοί υδρογονάνθρακες μπορούν να δημιουργηθούν με αφυδρογόνωση ναφθενικών και παραφινικών συστατικών.

1.3.1 Γεωχημική Σημασία Αρωματικών Υδρογονανθράκων

Κατά κύριο λόγο οι αρωματικοί υδρογονάνθρακες προέρχονται από την αποδόμηση της λιγνίνης. Η λιγνίνη και οι συγγενείς προς αυτή ενώσεις (ταννίνες) είναι ευρέως διαδεδομένα συστατικά στους φυτικούς οργανισμούς. Είναι ένα πολυμερές που εξασφαλίζει στα φυτά την ακεραιότητα της δομής τους και περιλαμβάνει αρωματικές δομές, όπως φαίνεται στο Σχήμα 1.7. Με το θάνατο των οργανισμών αποδομείται και δημιουργεί τη βασική δομή των χουμικών ανθράκων [6].



Σχήμα 1.7 : Χαρακτηριστική δομή λιγνίνης

1.3.2 Παράμετροι και Ιδιότητες που σχετίζονται με την συγκέντρωση Αρωματικών υδρογονανθράκων.

➤ Smoke Point (Σημείο Καπνού).

Το σημείο καπνού (S.P.) είναι ένα χαρακτηριστικό των καυσίμων και δείχνει την τάση αυτών να καίγονται με φλόγα που συνοδεύεται με καπνό. Η ύπαρξη υψηλού ποσοστού αρωματικών προκαλεί αύξηση του καπνού κατά την καύση και απώλειες ενέργειας λόγω της θερμικής ακτινοβολίας[24]. Το σημείο καπνού είναι το μέγιστο ύψος της φλόγας στο οποίο μπορεί να πραγματοποιηθεί καύση του καυσίμου σε ένα πρότυπο φυτίλι-λαμπτήρα χωρίς να παραχθεί καπνός. Εκφράζεται σε χιλιοστά και υψηλό σημείο καπνού υποδεικνύει ένα καύσιμο με

χαμηλή καπνογόνο τάση. Μετρηση του S.P. μπορεί να πραγματοποιηθεί με τις μεθόδους ASTM D1322 ή IP57.

➤ **Aniline Point (Σημείο Ανιλίνης).**

Το σημείο ανιλίνης (A.P.) για έναν υδρογονάνθρακα ή ένα κλάσμα πετρελαίου ορίζεται η ελάχιστη θερμοκρασία στην οποία ίσοι όγκοι του υγρού υδρογονάνθρακα και ανιλίνης είναι πλήρως αναμιξιμοί. Η ανιλίνη είναι μια αρωματική ένωση με δομή ενός μορίου βενζολίου, όπου ένα άτομο υδρογόνου έχει αντικατασταθεί από την ομάδα NH_2 . Το σημείο ανιλίνης είναι σημαντικό στον χαρακτηρισμό των κλασμάτων πετρελαίου και στην ανάλυση του μοριακού τους τύπου. Χρησιμοποιείται επίσης ως παράμετρος χαρακτηρισμού για την ποιότητα ανάφλεξης καυσίμων Diesel. Μετράται με την μέθοδο δοκιμής ASTM D611. Το σημείο ανιλίνης για διαφορετικά δείγματα παρουσιάζει αύξηση σε αυτά με αυξημένο μοριακό βάρος. Για τον ίδιο αριθμό ατόμων άνθρακα παρουσιάζει αύξηση από τους αρωματικούς προς τους παραφινικούς υδρογονάνθρακες. Τα αρωματικά έχουν πολύ χαμηλά σημεία ανιλίνης δεδομένου ότι η ίδια η ανιλίνη είναι μια αρωματική ένωση και έχει καλύτερη αναμιξιμότητα με αρωματικούς υδρογονάνθρακες. Γενικότερα έλαια με αυξημένο σημείο ανιλίνης έχουν χαμηλότερη περιεκτικότητα σε αρωματικά [12].

➤ **Octane number (αριθμός Οκτανίων).**

Ο αριθμός οκτανίων αποτελεί μια από τις πιο ενδεικτικές, για την ποιότητα της, ιδιότητα της βενζίνης και χαρακτηρίζει το καύσιμο ως προς την ικανότητα του να αποτρέπει την πρόωρη έκρηξη λόγω ανάφλεξης στις μηχανές εσωτερικής καύσης ή αλλιώς ως προς την αντικροτική συμπεριφορά του (antiknock property). Η ενεργειακή απόδοση του κινητήρα εξαρτάται άμεσα από τον αριθμό οκτανίων του καυσίμου, αφού η χρήση βενζινών υψηλού αριθμού οκτανίων επιτρέπει την αύξηση της επιτρεπόμενης αναλογίας συμπίεσης. Οι αρωματικές ενώσεις έχουν πολύ υψηλούς αριθμούς οκτανίων και καθώς αποτελούν το 20-40% των καυσίμων και επηρεάζουν αυξητικά το συνολικό αριθμό οκτανίων. Λόγω της τοξικότητάς τους, σήμερα η αύξηση του αριθμού οκτανίων επιτυγχάνεται μέσω οξυγονούχων προσθέτων (αλκοόλες R-OH και αιθέρες R-O-R) [3].

➤ **Watson K Number**

Από τα πρώτα χρόνια της βιομηχανίας πετρελαίου ήταν επιθυμητό να καθοριστεί μια μέθοδος ταξινόμησης του πετρελαίου βάσει μετρήσιμων παραμέτρων. Ο συντελεστής Watson συμβολίζεται από K_w και είναι από τους παλαιότερους παράγοντες χαρακτηρισμού [13].

Ορίζεται ως:

$$K_w = \frac{(1,8T_b)^{1/2}}{SG}$$

Όπου T_b : το σημείο βρασμού σε K

SG : ειδικό βάρος σε 15,5 °C

Σκοπός του ορισμού του συντελεστή αυτού είναι να χαρακτηρίσει τον τύπο των υδρογονανθράκων σε μίγματα πετρελαίου. Οι ναφθενικοί υδρογονάνθρακες έχουν τιμές K_w ενδιάμεσες αυτών των παραφινικών και αρωματικών υδρογονανθράκων. Σε γενικές γραμμές οι αρωματικοί υδρογονάνθρακες έχουν χαμηλές τιμές K_w ενώ οι παραφίνες υψηλές.

➤ Πυκνότητα-Ιξώδες

Όσον αφορά την πυκνότητα τα πετρελαιοειδή με μεγάλα ποσοστά αρωματικών υδρογονανθράκων παρουσιάζουν και αυξημένη πυκνότητα.

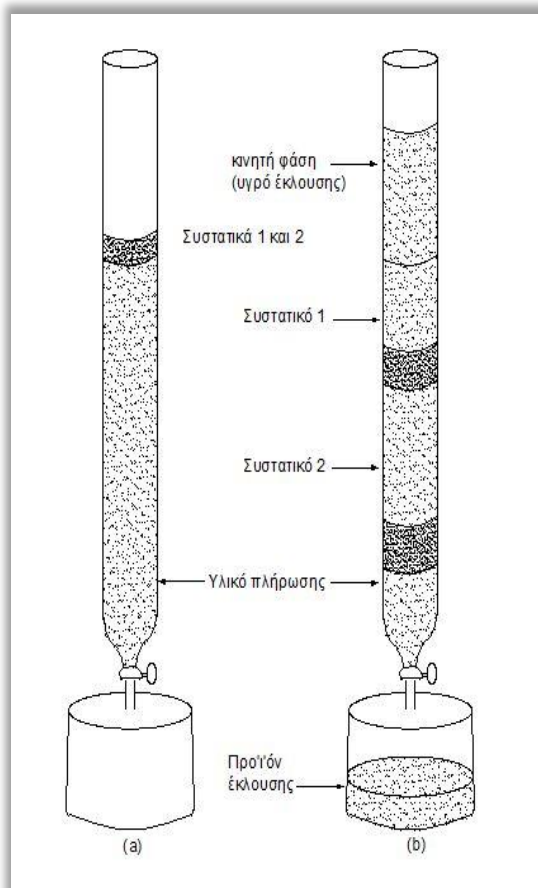
Το ιξώδες επηρεάζεται επίσης από την περιεκτικότητα σε αρωματικούς υδρογονάνθρακες και όσο μεγαλύτερη είναι αυτή τόσο μεγαλύτερο είναι και το ιξώδες. Αυτό σημαίνει ότι τα δείγματα που είναι πλούσια σε αρωματικά ρέουν πιο δύσκολα[27].

2 ΧΡΩΜΑΤΟΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

2.1 Εισαγωγή

Ο «χρωματογραφικός» διαχωρισμός των συστατικών ενός μίγματος είναι ένα σύννηδες φαινόμενο στην φύση. Η επιλεκτική κατακράτηση κάποιων συστατικών προκύπτει από τη διάχυση και μεταφορά μιγμάτων αερίων ή υγρών μέσω των πετρωμάτων. Ο Ρώσος βοτανολόγος Tswett χρησιμοποίησε για πρώτη φορά το χρωματογραφικό διαχωρισμό ως αναλυτική εφαρμογή κατά την προσπάθειά του να διαχωρίσει διαλύματα φυτικών χρωστικών ουσιών (plant pigments) διαμέσου γυάλινων στήλων, οι οποίες περιείχαν ανθρακικό ασβέστιο σε λεπτή διασπορά. Απεδείχθη ότι επιμέρους συστατικά διαχωρίζονταν κατά την κίνηση του διαλύματος μέσα στην στήλη, δημιουργώντας έγχρωμες ζώνες[27]. Έτσι δημιουργήθηκε ο όρος χρωματογραφία.

Σήμερα η χρωματογραφία στην ενόργανη μορφή της αποτελεί την πλέον διαδεδομένη τεχνική για τον διαχωρισμό μιγμάτων και την ταυτοποίηση των συστατικών τους σε αναλυτικό και παρασκευαστικό επίπεδο. Ο χρωματογραφικός διαχωρισμός παρουσιάζεται στο Σχήμα 2.1. Το μίγμα εισάγεται εντός στήλης πληρωμένης με μία στατική φάση στερεή ή υγρή (stationary phase) και κινείται κατά μήκος της διαλυμένο σε κατάλληλη υγρή ή αέρια φάση, η οποία ονομάζεται κινητή (mobile phase). Οι δύο φάσεις, κινητή και στατική, επιλέγονται έτσι ώστε τα συστατικά του μίγματος που πρόκειται να αναλυθεί να έχουν διαφορετική συνάφεια (διαλυτότητα, τάση προσρόφησης κ.λ.π.) με κάθε μία από αυτές. Ένα συστατικό που είναι ευδιάλυτο ή προσροφάται ευκολότερα στη σταθερή φάση θα χρειαστεί περισσότερο χρόνο να την διανύσει συγκριτικά με άλλο συστατικό με περιορισμένη διαλυτότητα ή τάση προσρόφησης[26]. Η διαφορετική κινητικότητα των συστατικών οδηγεί στον διαχωρισμό τους και την δημιουργία ξεχωριστών ζωνών στο χώρο. Ο διαχωρισμός δηλαδή επιτυγχάνεται στην βάση της διαφορετικής κατανομής των συστατικών ανάμεσα στις δύο αυτές φάσεις. Τα συστατικά που «παραμένουν» στην κινούμενη φάση περισσότερο διαπερνούν τη χρωματογραφική στήλη και εξέρχονται γρηγορότερα από εκείνα που παραμένουν προσροφημένα ή διαλυμένα περισσότερο χρόνο στη σταθερή φάση.



Σχήμα 2.1 : Σχηματική παράσταση χρωματογραφικού διαχωρισμού

2.2 Ταξινόμηση των βασικών χρωματογραφικών τεχνικών

Τα κριτήρια ταξινόμησης των χρωματογραφικών τεχνικών είναι η φυσική κατάσταση της κινητής φάσης, η φύση και η μορφή της στατικής φάσης και ο φυσικοχημικός μηχανισμός διαχωρισμού των συστατικών.

Ανάλογα με την φύση της κινητής φάσης η χρωματογραφία διακρίνεται σε αέρια (Gas Chromatography, GC) και υγρή (Liquid Chromatography, LC). Τελευταία, και με αρκετές εφαρμογές στην ανάλυση ορυκτών καυσίμων, έχει αναπτυχθεί η υπερκρίσιμη χρωματογραφία (Supercritical Fluid Chromatography, SFC), όπου ως κινούμενη φάση χρησιμοποιείται ένα αέριο σε υπερκρίσιμη κατάσταση, συνήθως CO_2 [22]. Σε συνδυασμό με την φύση της στατικής φάσης (υγρή ή στερεή) η χρωματογραφία διακρίνεται σε υγρή-υγρή (LLC), υγρή-στερεή (LSC), αέρια-στερεή (GSC) και αέρια-υγρή (GLC).

Ανάλογα με το είδος των φυσικοχημικών διεργασιών που λαμβάνουν χώρα κατά τον διαχωρισμό, οι χρωματογραφικές τεχνικές διακρίνονται σε:

- Χρωματογραφία προσρόφησης (adsorption)
- Χρωματογραφία κατανομής (partition)
- Χρωματογραφία ιοντοεναλλαγής (Ion Exchange)
- Χρωματογραφία μοριακού αποκλεισμού (Size exclusion)

Αναλυτικά:

Χρωματογραφία Προσρόφησης

Ο διαχωρισμός των διαφόρων ουσιών βασίζεται στο διαφορετικό βαθμό προσρόφησης στη στατική φάση. Οι κυριότερες αλληλεπιδράσεις που λαμβάνουν χώρα είναι ηλεκτροστατικής φύσης. Η χρωματογραφία προσρόφησης βρίσκει εφαρμογή στο διαχωρισμό ουσιών με παρόμοια δομή, αλλά με διαφορετική πολικότητα. Ανάλογα με τη σχέση πολικότητας μεταξύ της στατικής και της κινητής φάσης διακρίνονται δύο είδη χρωματογραφίας προσρόφησης:

Χρωματογραφία Κανονικής Φάσης

Εδώ, η στατική φάση (συνήθως SiO_2 ή Al_2O_3) είναι πολικότερη από την κινητή, η οποία αποτελείται από μη πολικούς διαλύτες, όπως εξάνιο, χλωροφόρμιο.

Χρωματογραφία Αντίστροφης Φάσης

Εδώ, η στατική φάση, η οποία είναι λιγότερο πολική της κινητής, αποτελείται από οξείδιο πυριτίου συζευγμένο με διάφορες ομάδες όπως αλκύλια (ακετύλιο, δεκαοκτύλιο), φαινύλιο, διόλες, αμινομάδες, κυανομάδες κ.ά., ενώ η κινητή φάση αποτελείται από μείγματα οργανικών διαλυτών (μεθανόλη, ακετονιτρίλιο, κ.ά.) με υδατικά ρυθμιστικά διαλύματα και νερό.

Χρωματογραφία Κατανομής

Ο διαχωρισμός στηρίζεται στη διαφορετική κατανομή των συστατικών ενός μείγματος μεταξύ της κινητής και της υγρής στατικής φάσης και εφαρμόζεται στην ανάλυση ομόλογων, μη ιονικών ενώσεων.

Χρωματογραφία Ιοντοανταλλαγής

Ο διαχωρισμός οφείλεται στις ηλεκτροστατικές αλληλεπιδράσεις μεταξύ των αναλυόμενων ιόντων και των φορτισμένων ομάδων της στατικής φάσης. Οι κυριότερες παράμετροι που

καθορίζουν τη συγκράτηση στη χρωματογραφία ιοντοανταλλαγής είναι το ιόν της δραστηκής ομάδας της στατικής φάσης, η ιονική ισχύς, το pH, ο τροποποιητής της κινητής φάσης και η θερμοκρασία.

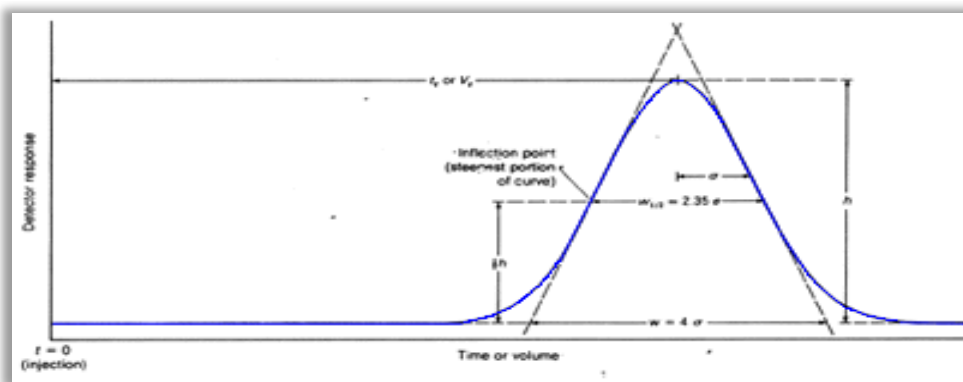
Χρωματογραφία μοριακού αποκλεισμού

Τα μόρια των συστατικών του μίγματος διαχωρίζονται ανάλογα με το μέγεθός τους καθώς με τη βοήθεια της υγρής κινούμενης φάσης διέρχονται μέσω των πόρων ενός πολυμερούς. Τα μόρια με μεγάλο μέγεθος κινούνται μόνο στους μεγάλους πόρους σε αντίθεση με τα μικρότερα που «καθυστερούν» στο δίκτυο των μικρών πόρων[27]. Στην έξοδο της στήλης τα μόρια φθάνουν διατεταγμένα σύμφωνα με το μέγεθός των, ή για μόρια συστατικών που ανήκουν στην ίδια χημική κατηγορία σύμφωνα με το Μοριακό Βάρος (MW) τους. Προϋπόθεση βέβαια αποτελεί η έλλειψη αλληλεπιδράσεων ανάμεσα στα συστατικά του μίγματος και το υλικό της χρωματογραφικής στήλης. Χρησιμοποιείται κυρίως για τον προσδιορισμό του μοριακού βάρους μιγμάτων και ως παρασκευαστική τεχνική στον διαχωρισμό μεγαλομορίων.

2.3 Βασικοί χρωματογραφικοί όροι

Το χρωματογράφημα

Στην ενόργανη χρωματογραφική ανάλυση υπάρχει μία διάταξη η οποία ονομάζεται ανιχνευτής που παρακολουθεί τη σύσταση της κινητής φάσης όπως αυτή εξέρχεται από την χρωματογραφική στήλη. Ο ανιχνευτής παρουσιάζει ευαισθησία σε κάποια ιδιότητα των συστατικών του μίγματος, η οποία πρέπει να είναι εύκολα μετρήσιμη αλλά και να έχει τιμές για τα συστατικά που αναλύονται διαφορετικές από εκείνες της κινούμενης φάσης. Όταν εκλύεται μόνο κινητή φάση το σήμα του ανιχνευτή ρυθμίζεται σε μηδενική τιμή. Με την εμφάνιση κάποιου συστατικού το σήμα αλλάζει και δημιουργείται μία κορυφή (peak). Η καμπύλη αυτή αποτελεί το προφίλ έκλουσης του συστατικού και στην ιδανική χρωματογραφική ανάλυση έχει μορφή κανονικής Gaussian κατανομής, δεδομένου ότι ανάλογης μορφής είναι και η κατανομή του πληθυσμού των μορίων στο μέτωπο κίνησης τους μέσα στην στήλη (Σχήμα 2.2).

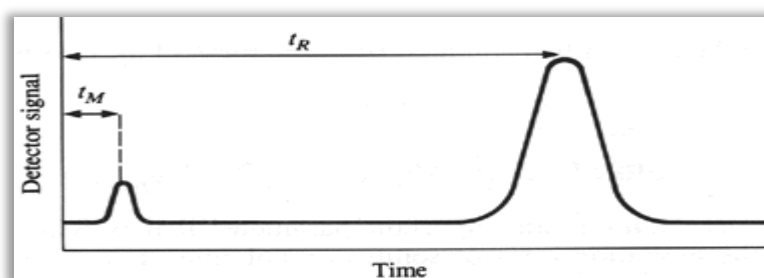


Σχήμα 2.2 :Χρωματογραφική κορυφή

Το διάγραμμα του προφίλ έκλουσης των συστατικών ενός μίγματος συναρτήσει του χρόνου ονομάζεται χρωματογράφημα.

Χρόνος συγκράτησης

Οι τρεις βασικοί παράγοντες που καθορίζουν την ταχύτητα των συστατικών ενός μίγματος μέσα σε μία χρωματογραφική στήλη είναι οι εξής: η γραμμική ταχύτητα της κινητής φάσης, ο λόγος του όγκου της στατικής φάσης προς τον όγκο της κινητής και ο συντελεστής κατανομής. Οι δύο πρώτοι παράγοντες είναι ίδιοι για όλα τα συστατικά, ενώ ο τρίτος είναι χαρακτηριστικός για κάθε συστατικό[22]. Έτσι για ένα δεδομένο χρωματογραφικό σύστημα, κάθε συστατικό κινείται με την δική του χαρακτηριστική ταχύτητα και εξέρχεται από την χρωματογραφική στήλη σε ένα χαρακτηριστικό χρόνο, ο οποίος ονομάζεται χρόνος συγκράτησης ή έκλουσης (retention time) t_R . Ο χρόνος αυτός χρησιμοποιείται για την ταυτοποίηση του συστατικού. Ο χρόνος συγκράτησης ενός συστατικού το οποίο δεν αλληλεπιδρά με την στατική φάση ονομάζεται νεκρός χρόνος (t_M) και ισούται με τον χρόνο κίνησης της κινητής φάσης μέσα στην στήλη (Σχήμα 2.3).

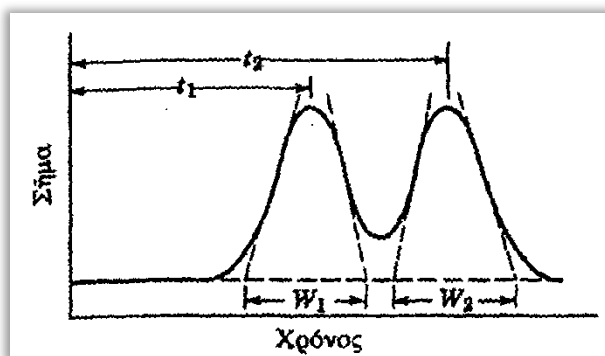


Σχήμα 2.3: Χρωματογραφικός χρόνος συγκράτησης συστατικού

Η διαχωριστική ικανότητα (resolution)

Το κυριότερο μέγεθος χαρακτηρισμού μιας χρωματογραφικής στήλης είναι η διαχωριστική ικανότητα και καθορίζεται ως η ικανότητα να διαχωρίζει συστατικά με παρόμοια χημικά χαρακτηριστικά, επομένως και ιδιότητες. Αυτό είναι ιδιαίτερα κρίσιμο στην περίπτωση των αναλύσεων δειγμάτων από οργανικά ορυκτά καύσιμα, λόγω της ύπαρξης ιδιαίτερα μεγάλου αριθμού των συστατικών σε αυτά αλλά και της χημικής των συγγένειας. Η διαχωριστική ικανότητα προσδιορίζεται συνήθως για ένα ζεύγος συστατικών του μίγματος ο διαχωρισμός του οποίου είναι ιδιαίτερα δύσκολος (critical pair).

Ποσοτικά για τις χρωματογραφικές κορυφές του Σχήματος 2.4 η διαχωριστική ικανότητα ορίζεται ως:



Σχήμα 2.4: Προσδιορισμός διαχωριστικής ικανότητας

$$R = \frac{2^*(t_2 - t_1)}{W_1 + W_2}$$

Σχήμα 2.5: Διαχωριστική ικανότητα

Το ζητούμενο σε μια ανάλυση είναι η εξασφάλιση του λεγόμενου διαχωρισμού γραμμής βάσης (baseline separation).

2.4 Ποιοτική και ποσοτική χρωματογραφική ανάλυση

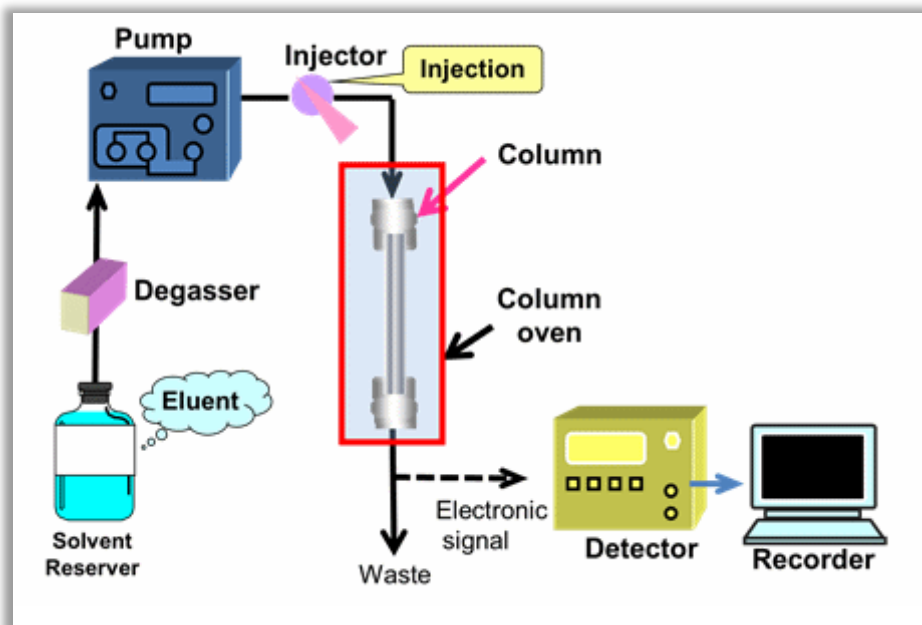
Η ποιοτική ανάλυση έχει στόχο τον προσδιορισμό των ενώσεων που υπάρχουν σε ένα μίγμα, ενώ η ποσοτική τον προσδιορισμό της περιεκτικότητας των συστατικών στο μίγμα. Ο ποιοτικός προσδιορισμός των συστατικών βασίζεται στους χρόνους συγκράτησής τους, δεδομένου ότι υπό τις ίδιες χρωματογραφικές συνθήκες ο χρόνος αυτός είναι χαρακτηριστικός για ένα συστατικό. Οι χρόνοι συγκράτησης προσδιορίζονται πειραματικά, χρησιμοποιώντας πρότυπες ενώσεις. Βέβαια ίδιους χρόνους συγκράτησης μπορούν να εμφανίσουν περισσότερα του ενός συστατικά. Πρέπει να τονιστεί ότι η εμφάνιση χρωματογραφικής κορυφής σε κάποιο χρόνο που αντιστοιχεί σε ένα συγκεκριμένο συστατικό δεν αποτελεί μονοσήμαντη απόδειξη της ύπαρξης του στο αναλυόμενο μίγμα. Για δείγματα, όπως αυτά των ορυκτών καυσίμων, λόγω του μεγάλου αριθμού των συστατικών και της χημικής συγγένειας μεταξύ των η οποία οδηγεί σε κοντινούς χρόνους συγκράτησης, η ταυτοποίηση με μόνο κριτήριο το χρόνο συγκράτησης είναι προβληματική έως αδύνατη. Λύση στο πρόβλημα αυτό δίδει η χρήση ανιχνευτών ικανών να προσδιορίζουν την ταυτότητα των εκλούμενων συστατικών. Τέτοιοι ανιχνευτές είναι ο φασματογράφος μάζας και το υπέρυθρο και υπεριώδες φασματοφωτόμετρο[6]. Ειδικά ο φασματογράφος μάζας συνδεδεμένος με αέριο χρωματογράφο έχει καθιερωθεί ως το αναλυτικό σύστημα με την πλέον ευρεία εφαρμογή στην ποιοτική ανάλυση σύνθετων μιγμάτων.

Η ποσοτική ανάλυση στηρίζεται στην αναλογία του εμβαδού της κορυφής με την συγκέντρωση ή την μάζα των αναλυόμενων συστατικών. Ο λόγος της συγκέντρωσης (ή μάζας) προς το εμβαδόν ονομάζεται συντελεστής απόκρισης (response factor) και προσδιορίζεται πειραματικά με αναλύσεις δειγμάτων γνωστής σύστασης. Η ποσοτική αυτή μεθοδολογία ονομάζεται βαθμονόμηση εξωτερικού προτύπου.

2.5 Υγρή χρωματογραφία υψηλής πίεσης (High Pressure Liquid Chromatography - HPLC)

Η υγρή χρωματογραφία υψηλής πίεσης αναπτύχθηκε στα μέσα της δεκαετίας του '70. και οφείλει την γρήγορη ανάπτυξη της στην εμφάνιση χρωματογραφικών στηλών με υψηλή δυνατότητα διαχωρισμού και στην δημιουργία on-line ανιχνευτών. Βρήκε ευρύτατη εφαρμογή στο χώρο της ανάλυσης ορυκτών καυσίμων, κυρίως λόγω της ευκολίας της να αναλύει μίγματα με ιδιαίτερα ψηλά μοριακά βάρη που ήταν

αδύνατο να μελετηθούν με αέρια χρωματογραφία. Σχηματικά ένα σύστημα HPLC παρουσιάζεται στο Σχήμα 2.6.



Σχήμα 2.6.: Διάταξη υγρής χρωματογραφίας

Ένα σύστημα HPLC περιλαμβάνει μία αντλία, ένα σύστημα εισαγωγής, χρωματογραφική στήλη, ανιχνευτή και σύστημα συλλογής και επεξεργασίας δεδομένων. Δεδομένου ότι οι στήλες περιέχουν πόρους με διάμετρο μερικά μm , η αντλία πρέπει να έχει την δυνατότητα να “κινεί” την κινητή φάση υπό υψηλή πίεση που φθάνει τα 10000 psi. Επίσης η ροή πρέπει να είναι ιδιαίτερα σταθερή ανεξάρτητα από την πίεση που αναπτύσσεται, ώστε να μην επηρεάζεται η διαχωριστική ικανότητα της στήλης και οι χρόνοι συγκράτησης των συστατικών. Το δείγμα εισάγεται μέσω βαλβίδας εισαγωγής που είναι συνδεδεμένη με βρόγχο (loop) γνωστού όγκου. Ο διαχωρισμός των συστατικών στην HPLC στηρίζεται στην αλληλεπίδραση των με το υλικό πλήρωσης της στήλης. Είναι μία δυναμική διεργασία προσρόφησης. Οι συνηθέστερες αλληλεπιδράσεις που απαντώνται είναι δίπολο-δίπολο και ιοντικές. Τα μόρια των συστατικών του μίγματος “συναγωγίζονται” με τα μόρια της κινητής φάσης και μεταξύ των για μία θέση στην επιφάνεια της στατικής φάσης. Όσο ισχυρότερη είναι η τάση προσρόφησης τόσο μεγαλύτερος θα είναι ο χρόνος παραμονής στην στήλη για το συγκεκριμένο συστατικό, με συνέπεια τον διαχωρισμό του από τα υπόλοιπα συστατικά του μίγματος. Οι συνηθέστερες κινητές φάσεις

HPLC που χρησιμοποιούνται στην ανάλυση ορυκτών καυσίμων είναι οργανικοί διαλύτες μικρής ή μεσαίας πολικότητας, όπως αλκάνια, χλωροφόρμιο, τετραυδροφουράνιο και μίγματα τους. Η επιλογή του διαλύτη εξαρτάται από το είδος του δείγματος (πολικότητα των συστατικών του) και από το είδος της στάσιμης φάσης. Η στάσιμη φάση αποτελείται από μικρά πορώδη σωματίδια με μεγάλη ειδική επιφάνεια. Τυπικά χαρακτηριστικά του υλικού πλήρωσης των στηλών είναι:

Διάμετρος πόρων 3 -10 μm , στενή κατανομή μεγέθους (10% του μέσου όρου), ειδική επιφάνεια 50 - 250 m^2/g [11]. Το υλικό που καλύπτει την στατική φάση και το οποίο εξασφαλίζει τον διαχωρισμό επιλέγεται σε σχέση με την χημεία των μιγμάτων που αναλύονται. Για μίγματα προερχόμενα από ορυκτά καύσιμα είναι συνήθως πολυμερή που περιέχουν $-\text{OH}$ και $-\text{NH}_2$ ή $-\text{CN}$.

Ανιχνευτές υγρής χρωματογραφίας

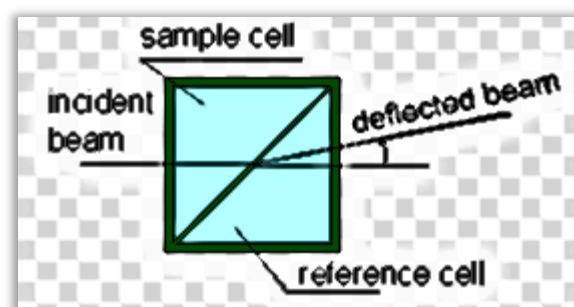
Στην ανάλυση μιγμάτων υδρογονανθράκων με HPLC, οι συνηθέστεροι ανιχνευτές είναι ο ανιχνευτής δείκτη διάθλασης και ο ανιχνευτής υπεριώδους. Χρησιμοποιούνται επίσης ο ανιχνευτής υπερύθρου, ο φασματογράφος μάζας και ο ανιχνευτής διάχυσης φωτεινής δέσμης (Light scattering).

Ανιχνευτής Δείκτη Διάθλασης (Refractive Index Detector)

Η αρχή της ανίχνευσης περιλαμβάνει την μέτρηση της αλλαγής στο δείκτη διάθλασης του εκρέοντος της στήλης διαλύματος και της κινητής φάσης. Όσο μεγαλύτερη είναι η διαφορά μεταξύ του R.I. του δείγματος και της κινητής φάσης τόσο μεγαλύτερη θα είναι η μεταβολή που θα καταγραφεί. Σε περιπτώσεις πολύπλοκων μιγμάτων, τα συστατικά του δείγματος μπορεί να καλύπτουν ένα ευρύ φάσμα τιμών δείκτη διάθλασης και μερικές από αυτές μπορεί να ταιριάζουν σε μεγάλο βαθμό με εκείνη της κινητής φάσης[11]. Τέτοιες μικρές διαφορές καθίστανται αόρατες στον ανιχνευτή. Για τους παραπάνω λόγους η χρήση του ανιχνευτή σε περιπτώσεις βαθμωτής έκλυσης καθίσταται αδύνατη.

Ο ανιχνευτής αποτελείται από ένα κελί 2 διαδρομών. Κατά την ανάλυση η μια είναι μόνιμα πληρωμένη με την κινητή φάση ενώ η άλλη έχει συνεχή ροή του δείγματος. Μια δέσμη φωτός διαπερνά αυτές τις διαδρομές και όταν υφίσταται μεταβολές στη σύνθεση του δείγματος παρατηρούνται εκτροπές της δέσμης φωτός σε σχέση με την διαδρομή αναφοράς. Η μεταβαλλόμενη γωνία της δέσμης φωτός βασίζεται στην αρχή

της εκτροπής της διαθλασιμετρίας και αποτυπώνεται σε καταγραφέα ταινίας. Ο καταγραφέας επηρεάζεται από τα προσπίπτοντα σε αυτόν φωτόνια. Τα παραπάνω αποτυπώνονται στο Σχήμα 2.7.



Σχήμα 2.7 : Σχηματικό διάγραμμα μιας εκτροπής ανιχνευτή δείκτη διάθλασης.

2.6 Χρωματογραφική ανάλυση δειγμάτων

Τα αναλυτικά δεδομένα της συγκέντρωσης των περιεχομένων αρωματικών υδρογονανθράκων που χρησιμοποιήθηκαν στην παρούσα εργασία παρήχθησαν με πρότυπες χρωματογραφικές μεθόδους ανάλυσης HPLC και ανοιχτής στήλης. HPLC μέθοδοι είναι οι: ASTM D-6591, D-7419, D-6379, D-5186, ενώ ανοιχτής στήλης είναι οι: ASTM D-2549, D-1319. Συνολικά αναλύθηκαν 131 δείγματα κλασμάτων πετρελαίου από τα διυλιστήρια της MOTOR OIL HELLAS (Κόρινθος), των ΕΛΛΗΝΙΚΩΝ ΠΕΤΡΕΛΑΙΩΝ (Ασπρόπυργος), καθώς και από πρατήρια υγρών καυσίμων στα Χανιά.

2.6.1 Πρότυπες Μέθοδοι Χρωματογραφικής Ανάλυσης (ASTM)

Παρακάτω περιγράφονται αναλυτικά οι πρότυπες μέθοδοι χρωματογραφικής ανάλυσης που χρησιμοποιήθηκαν στην εργασία αυτή:

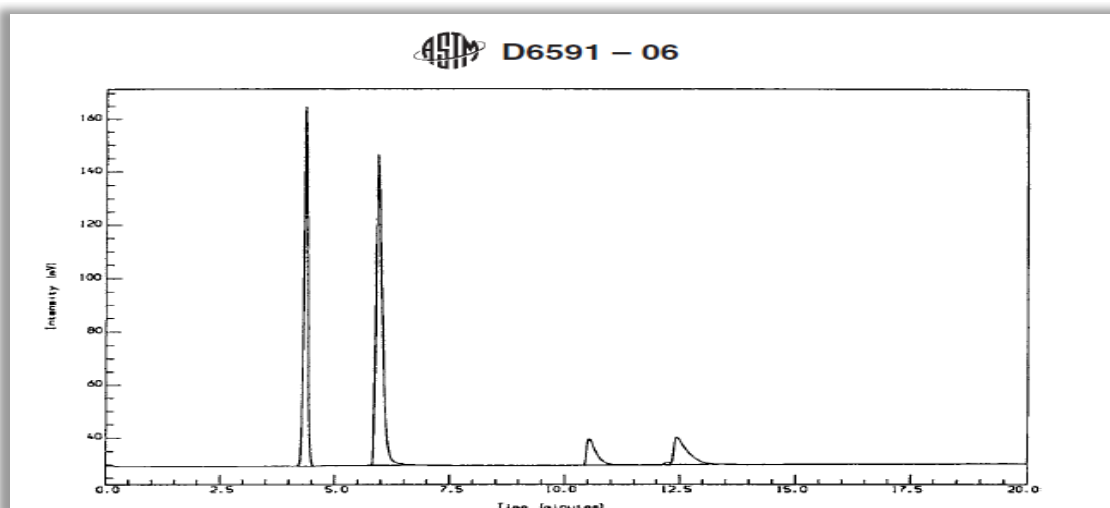
1. ASTM D 6591. Πρότυπη μέθοδος ανάλυσης για τον προσδιορισμό των αρωματικών υδρογονανθράκων σε μεσαία αποστάγματα μέσω υγρής χρωματογραφίας υψηλής απόδοσης (HPLC) με ανιχνευτή δείκτη διάθλασης.

Η μέθοδος αυτή είναι μια μέθοδος υγρής χρωματογραφίας υψηλής απόδοσης και καλύπτει τον προσδιορισμό μονοαρωματικών, διαρωματικών, τριαρωματικών καθώς και πολυκυκλικών αρωματικών υδρογονανθράκων περιεχομένων σε καύσιμα ντήζελ και σε αποστάγματα πετρελαιοειδών [32]. Το εύρος του σημείου βρασμού των παραπάνω κυμαίνεται από 150 - 400°C.

Η μέθοδος εφαρμόζεται σε δείγματα που περιέχουν 4 - 40% w/w μονοαρωματικούς, 0 - 20% διαρωματικούς, 0 - 6% τριαρωματικούς και 0 - 26% πολυκυκλικούς αρωματικούς υδρογονανθρακες και καλύπτει δείγματα με συγκεντρώσεις αρωματικών 4 - 65% w/w .

Το χρωματογραφικό σύστημα περιλαμβάνει αντλία υψηλής πίεσης , εισαγωγέα με βρόγχο 10 μ l, σύστημα δύο χρωματογραφικών στηλών σε σειρά (στήλη 1 Zorbax-NH₂ με διάμετρο 4.6 mm, μήκος 150 mm, μέγεθος σωματιδίων 5 μ m και στήλη 2 Zorbax-CN ιδίων χαρακτηριστικών), ανιχνευτή υπεριώδους ακτινοβολίας (UV DAD) που λειτουργεί σε μήκη κύματος 200 - 500 nm και ανιχνευτή δείκτη διάθλασης. Η ανάλυση γίνεται με διαλύτη επτάνιο και ροή 1 ml/min.

Το σύστημα ελέγχεται με την ανάλυση προτύπου διαλύματος (SPS) το οποίο περιέχει κυκλοεξάνιο (cyclo-C₆), οξυλένιο (o-xylene), διβενζοθειοφαίνιο (dibenzothiophene) και ανθρακένιο (anthracene). Με βάση αυτό καθορίζονται η διαχωριστική του ικανότητα (Resolution, R) και ο χρόνος αντίστροφης ροής (Backflush time, BF) σύμφωνα με τους χρόνους που προκύπτουν από το χρωματογράφημα (Σχήμα 2.8) και τους παρακάτω τύπους.



Σχήμα 2.8: Ανάλυση δείγματος γνωστών συγκεντρώσεων SPS με τη μέθοδο D6591-06.

$$BF = t_A + 0,4 (t_B - t_A)$$

Όπου t_A : χρόνος συγκράτησης διβενζοθιοφαίνιου σε δευτερόλεπτα

t_B : χρόνος συγκράτησης του ανθρακενίου σε δευτερόλεπτα.

$$R = \frac{2(t_2 - t_1)}{1,699(y_1 + y_2)}$$

Όπου t_1 : χρόνος κατακράτησης της κορυφής κυκλοεξανίου σε δευτερόλεπτα

t_2 : χρόνος κατακράτησης της κορυφής ο-ξυλόλιου σε δευτερόλεπτα

y_1 : πλάτος στο μισό ύψος κορυφής του κυκλοεξανίου σε δευτερόλεπτα

y_2 : πλάτος στο μισό ύψος κορυφής του ο-ξυλόλιου σε δευτερόλεπτα.

Στην συνέχεια το SPS αναλύεται με αντιστροφή της ροής για επαλήθευση της ικανότητας διαχωρισμού των θειοενώσεων με τα αρωματικά. Ακόλουθα παρασκευάζονται και αναλύονται διαλύματα με τα παρακάτω συστατικά: κυκλοεξάνιο (cyclo-C₆) (κυκλικό κορεσμένο), ορθο-ξυλόλιο (o-xylene, μονοαρωματικό), μεθυλναφθαλένιο (methylnaphthalene, διαρωματικό), φαινανθρένιο (phenanthrene, πολυαρωματικό). Τα διαλύματα αυτά παρασκευάζονται με 4

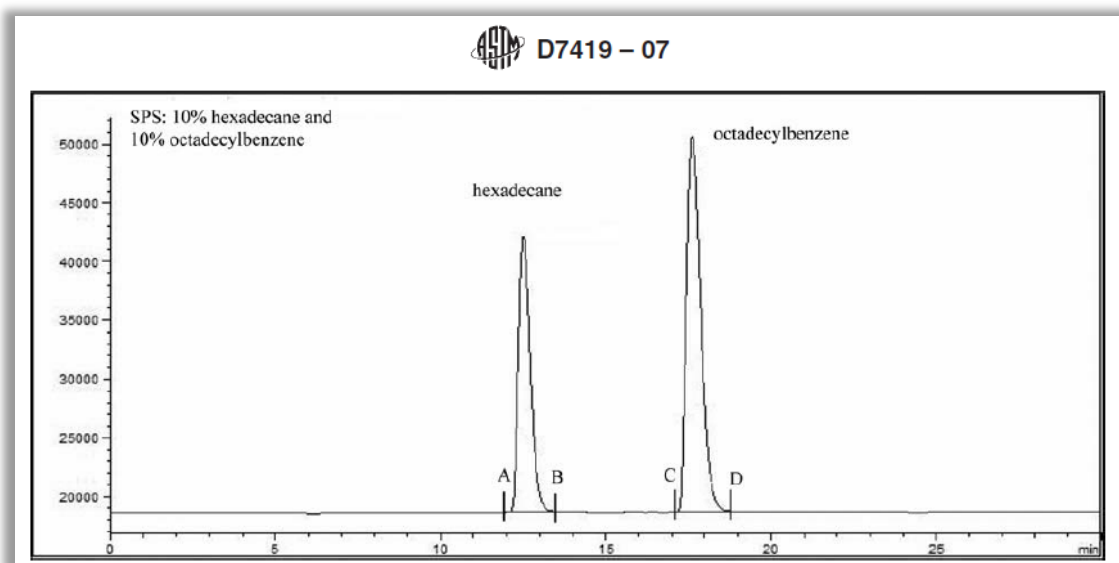
τουλάχιστον διαφορετικές συγκεντρώσεις και απο την ανάλυσή τους κατασκευάζονται οι καμπύλες βαθμονόμησης (calibration curves).

2. ASTM D 7419. Πρότυπη μέθοδος ανάλυσης για προσδιορισμό του συνόλου των αρωματικών και των κορεσμένων σε έλαια με υγρή χρωματογραφία υψηλής απόδοσης (HPLC).

Η μέθοδος αυτή καλύπτει τον προσδιορισμό της συνολικής συγκέντρωσης αρωματικών και κορεσμένων σε λιπαντικά έλαια, χρησιμοποιώντας υγρή χρωματογραφία υψηλής απόδοσης (HPLC) σε συνδιασμό με ανιχνευτή δείκτη διάθλασης (RI). Εφαρμόζεται σε δείγματα που περιέχουν σύνολο αρωματικών απο 0,2 - 46 % w/w.

Το χρωματογραφικό σύστημα περιλαμβάνει αντλία υψηλής πίεσης , εισαγωγή με βρόγχο 10 μ l, σύστημα τριών χρωματογραφικών στηλών σε σειρά (1 στήλη Zorbax-CN με διάμετρο 4.6 mm, μήκος 150 mm, μέγεθος σωματιδίων 5 μ m και 2 στήλες Zorbax-NH₂ ιδίων χαρακτηριστικών [33]), ανιχνευτή υπεριώδους ακτινοβολίας και ανιχνευτή δείκτη διάθλασης. Η ανάλυση γίνεται με διαλύτη επτάνιο και ροή 1 ml/min.

Το σύστημα ελέγχεται με την ανάλυση προτύπου διαλύματος (SPS) δυο συστατικών, δεκαοκτύλ-βενζόλιο (Octadecylbenzene) και δεκαεξάνιο (hexadecane) και καθορίζονται η διαχωριστική του ικανότητα (Resolution, R) και ο χρόνος αντίστροφης ροής (Backflush time, BF) σύμφωνα με τους χρόνους που προκύπτουν απο το χρωματογράφημα (Σχήμα 2.9) και τους παρακάτω τύπους.



Σχήμα 2.9: Ανάλυση δείγματος γνωστών συγκεντρώσεων SPS με τη μέθοδο D7419-07.

$$R = \frac{2(t_2 - t_1)}{3(y_1 + y_2)}$$

Όπου t_1 : χρόνος κατακράτησης της κορυφής δεκαεξανίου σε λεπτά,

t_2 : χρόνος κατακράτησης της κορυφής octadecylbenzene σε λεπτά

y_1 : πλάτος στο μισό ύψος της κορυφής δεκαεξανίου σε λεπτά,

y_2 : πλάτος στο μισό ύψος της κορυφής octadecylbenzene σε λεπτά

$$BFT = t_1 + 0,1(t_2 - t_1)$$

Όπου t_1 : χρόνος κατακράτησης σε λεπτά δεκαεξανίου

t_2 : χρόνος κατακράτησης της octadecylbenzene σε λεπτά.

Όταν υπολογιστεί ο χρόνος αντίστροφης ροής αναλύεται ξανά το SPS με αντιστροφή της ροής για να επαληθευτεί ο διαχωρισμός. Ακολουθεί η βαθμονόμηση του οργάνου με ανάλυση 5 προτύπων διαλυμάτων διαφορετικών συγκεντρώσεων των παραπάνω ουσιών, δηλαδή δημιουργούνται οι καμπύλες βαθμονόμησης για κάθε μία από τις ομάδες συστατικών (αρωματικά και κορεσμένα). Τα χρωματογραφήματα δημιουργούνται από τον ανιχνευτή δείκτη διάθλασης (RI detector) ενώ ο ανιχνευτής

UV, που ανιχνεύει μόνο τα αρωματικά, χρησιμοποιείται προαιρετικά για έλεγχο των αποτελεσμάτων.

3. ASTM D 6379-04. Πρότυπη μέθοδος ανάλυσης με υγρή χρωματογραφία υψηλής απόδοσης για προσδιορισμό των αρωματικών υδρογονανθράκων σε καύσιμα αεροσκαφών και αποστάγματα πετρελαίου.

Η μέθοδος αυτή καλύπτει τον προσδιορισμό των περιεχομένων μονοαρωματικών και διαρωματικών υδρογονανθράκων σε καύσιμα αεροσκαφών καθώς και σε αποστάγματα πετρελαίου με περιοχή βρασμού από 50 - 300°C. Η συνολική περιεκτικότητα σε αρωματικά υπολογίζεται από το άθροισμα των επιμέρους ομάδων. Η μέθοδος αυτή ενδείκνυται για αποστάγματα που περιέχουν 10-25% μονοαρωματικούς και 0 - 7% διαρωματικούς υδρογονάνθρακες. Δείγματα που περιέχουν θείο, άζωτο ή οξυγόνο μπορεί να παρουσιάσουν προβληματικά αποτελέσματα.

Σύμφωνα με τη διαδικασία το δείγμα αραιώνεται με αναλογία 1:1 με την κινητή φάση. Η κινητή φάση είναι συνήθως επτάνιο. Ένας σταθερός όγκος αυτού του διαλύματος εισάγεται στον υγρό χρωματογράφο υψηλής απόδοσης, ο οποίος είναι εφοδιασμένος με μια πολική στήλη. Η στήλη αυτή έχει πολύ μικρή επιλεκτικότητα στους μη αρωματικούς και έντονη επιλεκτικότητα στους αρωματικούς υδρογονάνθρακες. Ως αποτέλεσμα της επιλεκτικότητας αυτής προκύπτει ο διαχωρισμός των παραπάνω καθώς και ο διαχωρισμός των αρωματικών σύμφωνα με την δομή των δακτυλίων τους σε μονοαρωματικούς και πολυαρωματικούς [31].

Η στήλη συνδέεται με ανιχνευτή δείκτη διάθλασης που ανιχνεύει τα συστατικά καθώς εκλούνται από την στήλη. Το ηλεκτρονικό σήμα του ανιχνευτή παρακολουθείται συνεχώς από έναν επεξεργαστή δεδομένων. Τα χαρακτηριστικά των σημάτων (εμβადόν κορυφής) από τα αρωματικά του δείγματος συγκρίνονται με εκείνα που ελήφθησαν από τα πρότυπα βαθμονόμησης προκειμένου να υπολογιστεί το ποσοστό μονοαρωματικών και πολυαρωματικών, το άθροισμα των οποίων αναφέρεται ως η συνολική περιεκτικότητα αρωματικών.

4. ASTM D-2549. Πρότυπη μέθοδος ανάλυσης για διαχωρισμό αρωματικών και μη-αρωματικών σε κλάσματα ελαίων υψηλού σημείου βρασμού με έκλυση.

Η μέθοδος αυτή καλύπτει τον διαχωρισμό και ποσοτικό προσδιορισμό των αρωματικών σε κλάσματα απο μίγματα υδρογονανθράκων που έχουν σημείο βρασμού μεταξύ των 232 – 538°C.

Αρχικά διαλύονται 2 gr του δείγματος σε 10 ml πεντανίου και θερμαίνονται, αν χρειαστεί, για να επιτευχθεί η διάλυση. Το δείγμα τοποθετείται στην κορυφή μιας γυάλινης χρωματογραφικής στήλης η οποία είναι πλοιωμένη με ενεργοποιημένο βωξίτη (κοκκομετρία: 60 - 200mesh, ενεργοποίηση 3 ώρες στους 110°C) και silica gel (κοκκομετρία: 70 - 100mesh, ενεργοποίηση: 3 ώρες στους 170°C). Τα μη-αρωματικά συστακά εκλύονται με την προσθήκη πεντανίου, το κλάσμα των αρωματικών εκλύεται με την προσθήκη διαιθυλαιθέρα, χλωροφορμίου και αιθυλικής αλκοόλης [31]. Οι διαλύτες απομακρύνονται με εξάτμιση και τα υπολείματα ζυγίζονται για τον υπολογισμό της συνολικής ποσότητας αρωματικών ως ποσοστό των συνολικών παραχθέντων κλασμάτων.

5. ASTM D1319-08. Πρότυπη μέθοδος προσδιορισμού τύπου υδρογονανθράκων σε υγρά προϊόντα πετρελαίου από ανιχνευτή φθορισμού.

Η μέθοδος αυτή καλύπτει τον προσδιορισμό αρωματικών υδρογονανθράκων με συγκεντρώσεις από 5 - 99%. Ενδείκνυται για κλάσματα πετρελαίου των οποίων η απόσταξη ολοκληρώνεται κάτω από τους 315°C. Η ακρίβεια της μεθόδου έχει υπολογιστεί με καύσιμα που δεν περιέχουν οξυγονούχες ενώσεις. Για κλάσματα με θερμοκρασίες απόσταξης ανώτερες ή κοντινές των προαναφερθέντων ορίων παρατηρούνται λανθασμένα αποτελέσματα διότι στα κλάσματα αυτά δεν γίνεται ορθή έκλυση.

Στη μέθοδο αυτή χρησιμοποιούνται 2 στήλες προσρόφησης. Η μια εξ αυτών χρησιμοποιείται ως πρότυπη και στην άλλη εισάγεται το προς ανάλυση δείγμα. Περίπου 0,75ml δείγματος εισάγονται μέσω σύριγγας εντός της στήλης που είναι κατασκευασμένη από ειδικό γυαλί και γεμάτη με ενεργοποιημένο διοξείδιο του πυριτίου [31]. Ένα μικρό στρώμα του silica gel περιέχει ένα μείγμα από φθορίζουσες χρωστικές. Όταν όλο το δείγμα έχει απορροφηθεί προστίθεται αλκοόλη για την

σταδιακή εκρόφιση του δείγματος από την στήλη. Οι υδρογονάνθρακες διαχωρίζονται σύμφωνα με τις συγγένειες προσρόφησης σε αρωματικά, ολεφίνες και κορεσμένους υδρογονάνθρακες. Ο διαχωρισμός των παραπάνω σε ζώνες γίνεται με την βοήθεια λάμπας υπεριώδους (U.V.) και η ποσοτικοποίηση τους πραγματοποιείται με μέτρηση του μήκους της κάθε ζώνης ως προς το συνολικό.

6. ASTM D5186. Πρότυπη μέθοδος ανάλυσης για προσδιορισμό της περιεκτικότητας σε αρωματικά και πολυκυκλικούς αρωματικούς υδρογονάνθρακες σε καύσιμα ντίζελ και καύσιμα αεροσκαφών με χρωματογραφία υπερκρίσιμου υγρού.

Αυτή η μέθοδος καλύπτει τον προσδιορισμό της συνολικής ποσότητας μονοαρωματικών (MA) και πολυαρωματικών (PNA) υδρογονανθράκων σε καύσιμα κινητήρων ντίζελ, καύσιμα αεροσκαφών και μίγματα κλασμάτων μέσω χρωματογραφίας υπερκρίσιμου ρευστού (SFC). Το εύρος της συγκέντρωσης αρωματικών στα δείγματα στα οποία η μέθοδος αυτή εφαρμόζεται είναι 1 - 75% w/w. Στη χρωματογραφία αυτή, ένα μικρό κλάσμα του δείγματος καυσίμου εκχύεται σε στήλη προσρόφησης που περιέχει SiO_2 και εκλούεται χρησιμοποιώντας υπερκρίσιμο διοξείδιο του άνθρακα ως κινητή φάση. Τα αρωματικά του δείγματος διαχωρίζονται από τα μη αρωματικά και ανιχνεύονται με χρήση ανιχνευτή ιονισμού φλόγας (F.I.D). Η απόκριση σε υδρογονάνθρακες καταγράφεται κατά τη διάρκεια του χρόνου ανάλυσης [31]. Οι χρωματογραφικές περιοχές που αντιστοιχούν στα μονοαρωματικά, πολυαρωματικά και μη αρωματικά συστατικά προσδιορίζονται με ανάλυση προτύπων. Η περιεκτικότητα κατά βάρος καθενός εξ' αυτών υπολογίζεται από το σήμα του ανιχνευτή.

Ο παρακάτω πίνακας (Πίνακας 2.1) παρουσιάζει τις πρότυπες ASTM μεθόδους που χρησιμοποιούνται στον προσδιορισμό της συγκέντρωσης των αρωματικών και τα κλάσματα στα οποία είναι εφαρμόσιμες.

Μέθοδοι ASTM	Βενζίνη	Νάφθα	Καύσιμο Αεροσκαφών	Ντήζελ	Πετρ. Εξ Καύσης & Μαζούτ	Λιπαντικά
D 1319	x	x	X	x		
D 6379		x	X	x		
D 6591			X	x	x	
D 5186			X	x		
D 2549					x	x
D 7419						x

Πίνακας 2.1: Πρότυπες μέθοδοι προσδιορισμού σε πετρελαιοειδή και κλάσματα στα οποία εφαρμόζονται.

2.6.2 Αποτελέσματα Χρωματογραφικής Ανάλυσης των δειγμάτων

Στον παρακάτω πίνακα (Πίνακας) παρουσιάζονται τα αποτελέσματα των μετρήσεων των αρωματικών συστατικών στα 131 δείγματα σε καθεμία από της μεθόδους που εφαρμόστηκαν σε αυτά καθώς και ο χαρακτηρισμός τους σύμφωνα με τον τύπο τους και την προέλευσή τους.

Όνομα Δειγματος	Προέλευση	Τύπος	Αρωμ (%)	D1319 (%)	D5186 (%)	D6591 (%)	D6379 (%)	D7419 (%)	D2549 (%)
MOH_1	MOH	Βενζίνη	27,5						
MOH_4	MOH	Βενζίνη		30,1					
MOH_6	MOH	Μίγμα Ντήζελ		31,2	33,1	34,9			
MOH_10	MOH	Μίγμα Ντήζελ		21,6	22,6	23,0			
MOH_8	MOH	Μίγμα Ντήζελ		32,6					
MOH_5	MOH	Βενζίνη		26,5					
MOH_18	MOH	Βενζίνη		30,5					
MOH_19	MOH	Βενζίνη		23,0					

Όνομα Δειγματοϋ	Προέλευση	Τύπος	Αρωμ (%)	D1319 (%)	D5186 (%)	D6591 (%)	D6379 (%)	D7419 (%)	D2549 (%)
MOH_58	ΜΟΗ	Βενζίνη		35,5					
MOH_22	ΜΟΗ	Βενζίνη		22,8					
MOH_48	ΜΟΗ	Βενζίνη		39,1					
MOH_61	ΜΟΗ	Βενζίνη		20,1					
MOH_29	ΜΟΗ	Βενζίνη		32,8					
MOH_30	ΜΟΗ	Βενζίνη		27,4					
MOH_31	ΜΟΗ	Βενζίνη		29,5					
MOH_32	ΜΟΗ	Βενζίνη		22,6					
MOH_33	ΜΟΗ	Βενζίνη		11,2					
MOH_20	ΜΟΗ	Πετρ.Εσ Καύσης				21,3			
MOH_9	ΜΟΗ	Ντήζελ		22,4	24,4	24,6			
MOH_23	ΜΟΗ	Ντήζελ		19,2	20,0	21,6			
MOH_24	ΜΟΗ	Ντήζελ		21,6	23,2				
MOH_27	ΜΟΗ	Ντήζελ		27,7	28,6	30,0			
MOH_28	ΜΟΗ	Ντήζελ		37,1	36,6				
MOH_35	ΜΟΗ	Ντήζελ		27,7	30,1				
MOH_44	ΜΟΗ	Ντήζελ		38,3	38,9	40,7			
MOH_51	ΜΟΗ	Ντήζελ		21,2	23,9	24,5			
MOH_62	ΜΟΗ	Ντήζελ		29,9	31,1	32,9			
MOH_7	ΜΟΗ	Καύσιμα Αεροσκαφών		16,7		20,3			
MOH_40	ΜΟΗ	Καύσιμα Αεροσκαφών		17,7		21,1	21,0		
MOH_60	ΜΟΗ	Καύσιμα Αεροσκαφών		17,9			20,7		
MOH_25	ΜΟΗ	Καύσιμα Αεροσκαφών		21,3			22,6		
MOH_26	ΜΟΗ	Καύσιμα Αεροσκαφών		17,4			19,2		
MOH_34	ΜΟΗ	Καύσιμα Αεροσκαφών		22,1			24,4		
MOH_36	ΜΟΗ	Καύσιμα Αεροσκαφών		17,3			18,7		
MOH_41	ΜΟΗ	Καύσιμα Αεροσκαφών		16,7					
MOH_42	ΜΟΗ	Καύσιμα Αεροσκαφών		20,0		20,3	22,1		
MOH_43	ΜΟΗ	Καύσιμα Αεροσκαφών		9,9		14,5	11,3		

Όνομα Δειγμάτος	Προέλευση	Τύπος	Αρωμ (%)	D1319 (%)	D5186 (%)	D6591 (%)	D6379 (%)	D7419 (%)	D2549 (%)
ΜΟΗ_2	ΜΟΗ	Νάφθα							
ΜΟΗ_11	ΜΟΗ	Νάφθα							
ΜΟΗ_12	ΜΟΗ	Νάφθα	25,7						
ΜΟΗ_13	ΜΟΗ	Νάφθα	5,8						
ΜΟΗ_14	ΜΟΗ	Νάφθα	24,0						
ΜΟΗ_15	ΜΟΗ	Νάφθα	5,9						
ΜΟΗ_16	ΜΟΗ	Νάφθα	10,6						
ΜΟΗ_17	ΜΟΗ	Νάφθα	5,7						
Μ_9	ΜΟΗ	Απόσταγμα	5,9					22,4	
Μ_10	ΜΟΗ	Απόσταγμα	25,7					30,0	
Μ_11	ΜΟΗ	Υπόλειμμα Απόσταξης						42,6	35,5
Μ_12	ΜΟΗ	Βενζίνη						47,3	
Μ_13	ΜΟΗ	Υπόλειμμα Απόσταξης							
Μ_14	ΜΟΗ	Υπόλειμμα Απόσταξης						61,5	69,3
Μ_15	ΜΟΗ	Μαζούτ						67,2	
Μ_16	ΜΟΗ	Υπόλειμμα Απόσταξης						63,5	63,2
Μ_17	ΜΟΗ	Κεριά						1,5	0,4
Μ_18	ΜΟΗ	Κεριά						1,8	0,5
Μ_19	ΜΟΗ	Κεριά						10,5	12,5
ΜΒ_1	ΜΟΗ	Λιπαντικά						32,0	
ΜΒ_2	ΜΟΗ	Λιπαντικά						25,4	
ΜΒ_3	ΜΟΗ	Λιπαντικά						27,0	
ΜΒ_4	ΜΟΗ	Λιπαντικά						27,5	24,5
ΜΒ_5	ΜΟΗ	Λιπαντικά						23,8	
ΜΒ_6	ΜΟΗ	Λιπαντικά						73,2	
ΜΒ_7	ΜΟΗ	Λιπαντικά						89,8	
ΜΒ_8	ΜΟΗ	Λιπαντικά						45,8	
ΜΒ_9	ΜΟΗ	Λιπαντικά						52,3	
ΜΒ_12	ΜΟΗ	Πετρ.Εσ Καύσης						56,2	54,9
ΜΒ_13	ΜΟΗ	Πετρ.Εσ Καύσης						38,4	

Όνομα Δειγματος	Προέλευση	Τύπος	Αρωμ (%)	D1319 (%)	D5186 (%)	D6591 (%)	D6379 (%)	D7419 (%)	D2549 (%)
MB_15	MOH							44,5	
MB_16	MOH							85,8	
MB_17	MOH	Λιπαντικά						84,3	
DIES_10	EBEXA	Ντήζελ				24,0			
DIES_15	EBEXA	Ντήζελ				24,4			
DIES_21	EBEXA	Ντήζελ				28,5			
DIES_23	EBEXA	Ντήζελ				24,2			
DIES_28	EBEXA	Ντήζελ				23,3			
DIES_31	EBEXA	Ντήζελ				20,1			
DIES_32	EBEXA	Ντήζελ				20,0			
DIES_35	EBEXA	Ντήζελ				22,9			
DIES_38	EBEXA	Ντήζελ				24,5			
DIES_65	EBEXA	Ντήζελ				25,0			
E_11	ELPE	Απόσταγμα							21,5
E_4	ELPE	Μαζούτ							67,0
E_12	ELPE	Μαζούτ							78,9
E_18	ELPE	Υπόλειμμα Απόσταξης							54,4
E_19	ELPE	Μαζούτ							55,1
E_26	ELPE	Απόσταγμα							38,2
E_33	ELPE	Υπόλειμμα Απόσταξης							73,3
E_34	ELPE	Υπόλειμμα Απόσταξης							99,4
E_37	ELPE	Πετρ.Εσ Καύσης							22,9
E_38	ELPE	Υπόλειμμα Απόσταξης							91,0
E_39	ELPE	Πετρ.Εσ Καύσης							32,4
E_40	ELPE	Πετρ.Εσ Καύσης							47,3
E_41	ELPE	Απόσταγμα							41,2
E_44	ELPE	Κεριά							0,2
E_47	ELPE	Μαζούτ							18,0
E_50	ELPE	Απόσταγμα							42,5
E_51	ELPE	Απόσταγμα							39,0

Όνομα Δειγματος	Προέλευση	Τύπος	Αρωμ (%)	D1319 (%)	D5186 (%)	D6591 (%)	D6379 (%)	D7419 (%)	D2549 (%)
E_53	ELPE	Πετρ.Εσ Καύσης							25,0
E_54	ELPE	Απόσταγμα							23,1
E_55	ELPE	Υπόλειμμα Απόσταξης							70,6
E_56	ELPE	Μαζούτ							55,6
E_57	ELPE	Υπόλειμμα Απόσταξης							99,4
E_59	ELPE	Μαζούτ							74,3
E_61	ELPE	Υπόλειμμα Απόσταξης							57,0
E_62	ELPE	Μαζούτ							57,8
E_63	ELPE	Απόσταγμα							30,4
E_67	ELPE	Υπόλειμμα Απόσταξης							51,0
E_74	ELPE	Υπόλειμμα Απόσταξης							35,1
CK_1	EMP	Πετρ.Εσ Καύσης				33,2			
CKH	EMP	Πετρ.Εσ Καύσης				26,3			
HC	EMP	Πετρ.Εσ Καύσης				6,9			
HGO	EMP	Πετρ.Εσ Καύσης				20,3			
HT	EMP	Πετρ.Εσ Καύσης				26,4			
LGO	EMP	Πετρ.Εσ Καύσης				19,8			
MAR_BL	EMP	Πετρ.Εσ Καύσης				25,3			
MGO_1	EMP	Πετρ.Εσ Καύσης				28,0			
MGO_5	EMP	Πετρ.Εσ Καύσης				36,0			
MHC	EMP	Πετρ.Εσ Καύσης				20,9			
MOH_HCN	EMP	Πετρ.Εσ Καύσης				17,9			
MOH_LCN	EMP	Πετρ.Εσ Καύσης				40,2			
MOH_LCO	EMP	Πετρ.Εσ Καύσης				63,7			
RF_03	EMP	Πετρ.Εσ Καύσης				30,9			
RF_73	EMP	Πετρ.Εσ Καύσης				28,3			
SR	EMP	Πετρ.Εσ Καύσης				23,7			
T_26	EMP	Πετρ.Εσ Καύσης				10,6			
U_18	EMP	Πετρ.Εσ Καύσης				24,3			
U_19	EMP	Πετρ.Εσ Καύσης				23,7			

Όνομα Δειγμάτος	Προέλευση	Τύπος	Αρωμ (%)	D1319 (%)	D5186 (%)	D6591 (%)	D6379 (%)	D7419 (%)	D2549 (%)
DIST200	CYCLON1	Απόσταγμα						19,8	
SN150	CYCLON2	Λιπαντικά						15,4	
SN500	CYCLON3	Λιπαντικά						16,8	

2.6.3 Ακρίβεια Πρότυπων Μεθόδων ASTM.

Στον παρακάτω πίνακα παρουσιάζονται οι ακρίβειες των χρησιμοποιούμενων μεθόδων εκφρασμένες με βάση την επαναληψιμότητα(repeatability) και την αναπαραγωγιμότητα (reproducibility).Ως \bar{X} ορίζεται ο μέσος όρος των δειγμάτων που συγκρίνονται.

METHOD	Repeatability	Reproducibility
D1319	1.3	3.7
D2549	1.4	1.5
D5186	0.4	1.8
D6379	$0.129 \bar{X}^{0.667}$	$0.261 \bar{X}^{0.667}$
D6591	0.036 ($\bar{X}+1.5$)	0.116 ($\bar{X}+6.3$)
D7419	0.022($\bar{X}+4$)	0.11 ($\bar{X}+4$)

3.ΦΑΣΜΑΤΟΣΚΟΠΙΑ ΥΠΕΡΥΘΡΟΥ

3.1 Εισαγωγή

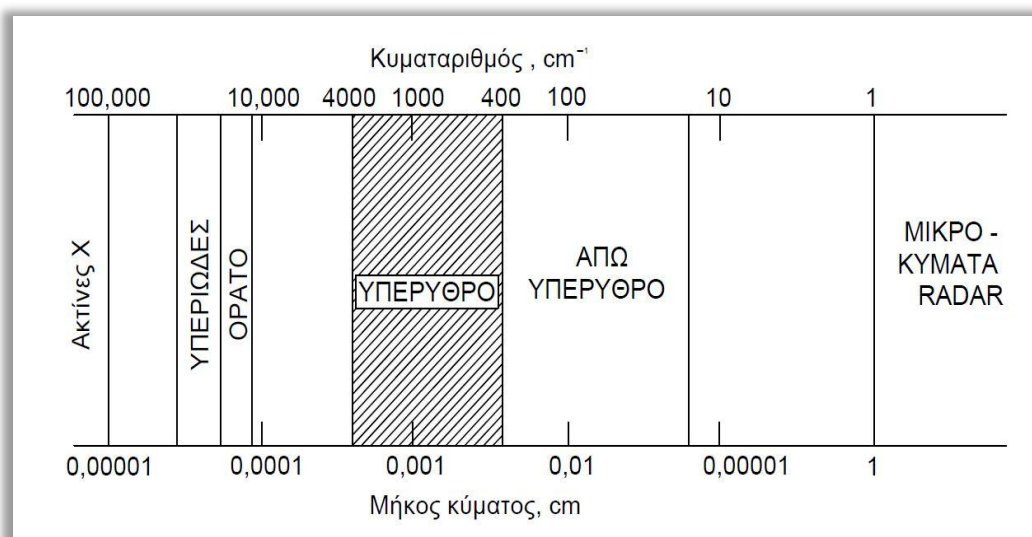
Η φασματοσκοπία υπέρυθρου (infrared spectroscopy, IR) αποτελεί μία αναλυτική τεχνική διερεύνησης της μοριακής σύνταξης (δομής) και ταυτοποίησης των οργανικών, κυρίως, ενώσεων. Κύρια πεδία εφαρμογής της φασματοσκοπίας υπέρυθρου είναι οι τομείς της οργανικής και της φαρμακευτικής χημείας. Σήμερα, το φασματοφωτόμετρο υπέρυθρου αποτελεί ένα αναπόσπαστο όργανο μέτρησης κάθε αναλυτικού εργαστηρίου και κάθε εργαστηρίου βιομηχανίας, που δραστηριοποιείται σε εφαρμογές της οργανικής χημείας. Η υπέρυθρη ακτινοβολία ανακαλύφθηκε το 1800 από τον Sir William Herschel με πειράματα θερμικής αγωγιμότητας. Την εποχή εκείνη η φύση της ακτινοβολίας δεν ήταν ακόμα γνωστή. Ο Herschel διεξήγαγε πειράματα με σκοπό την ανάλυση της ηλιακής ακτινοβολίας με γυάλινο πρίσμα. Κατά την διάρκεια των πειραμάτων, τοποθετώντας θερμόμετρα σε διαδοχικές θέσεις του φάσματος, ανακάλυψε ότι το θερμόμετρο μετά το κόκκινο άκρο σημείωνε την υψηλότερη θερμοκρασία, υποδεικνύοντας την ύπαρξη μεγαλύτερης ροής θερμότητας. Με αυτό το απλό πείραμα ανακαλύφθηκε η υπέρυθρη περιοχή του ηλεκτρομαγνητικού φάσματος και συγχρόνως κατασκευάστηκε το πρώτο υπέρυθρο φασματοφωτόμετρο [1]. Αργότερα, με άλλα πειράματα ο Herschel μέτρησε την απορρόφηση της υπέρυθρης ακτινοβολίας από διάφορες ουσίες. Ωστόσο, δεν ήταν σε θέση να αξιολογήσει την ανακάλυψή του και κυρίως τις πληροφορίες που μπορούσε να πάρει για τη μοριακή δομή των οργανικών ενώσεων από την απορρόφηση αυτού του είδους της ακτινοβολίας. Πριν εκτιμηθεί η σημασία της υπέρυθρης απορρόφησης, έπρεπε να κατανοηθεί καλύτερα η θεωρία και η φύση της ακτινοβολίας. Χρειάστηκε σχεδόν ένας αιώνας για να αναπτυχθούν η απαιτούμενη θεωρία, οι τεχνικές και η οργανολογία της υπέρυθρης ανάλυσης. Στις αρχές του 1903 ο William W. Colbentz, ένας νέος ερευνητής στο Cornell University, βελτίωσε τις πειραματικές τεχνικές και άρχισε τη μέτρηση των φασμάτων απορρόφησης διαφόρων καθαρών ουσιών. Το 1905 δημοσίευσε την πρώτη συλλογή φασμάτων απορρόφησης 131 ενώσεων.

Σήμερα, έναν αιώνα μετά τις εργασίες του Colbentz, η βασική θεωρία της υπέρυθρης φασματοσκοπίας είναι πλήρως κατανοητή[2]. Από το 1905 και μετά οι χημικοί που ασχολούνται με θέματα οργανικής χημείας, θεωρούν την υπέρυθρη φασματοσκοπία σαν μια μέθοδο ταυτοποίησης των ενώσεων και η χημική βιομηχανία

χρησιμοποιεί τα φασματοφωτόμετρα υπερύθρου για την ποσοτική ανάλυση των οργανικών ενώσεων[2].

3.2 Φασματοσκοπία υπερύθρου

Είναι γνωστό ότι, όταν τα μόρια των ενώσεων εκτεθούν σε ακτινοβολία της οποίας το μήκος κύματος βρίσκεται στην περιοχή του υπέρυθρου φάσματος, τότε το μόριο λαμβάνει χώρα απορρόφηση ενός μέρους της ακτινοβολίας και κατ' επέκταση της ενέργειας που αυτή φέρει. Με αυτόν τον τρόπο προκαλούνται διεγέρσεις μεταξύ των διαφόρων ενεργειακών σταθμών του μορίου. Οι διεγέρσεις αυτές περιλαμβάνουν δονήσεις και περιστροφές των δεσμών των ατόμων του μορίου. Ωστόσο, παρά την απορρόφηση IR ακτινοβολίας και την διέγερσή του, το μόριο εξακολουθεί να παραμένει στην θεμελιώδη ηλεκτρονική κατάσταση, δηλαδή τα ηλεκτρόνια δεν αλλάζουν στάθμη [1]. Η περιοχή της υπέρυθρης ακτινοβολίας είναι η περιοχή του φάσματος που εκτείνεται από το ορατό μέχρι τα μικροκύματα, δηλαδή η φασματική περιοχή από 0,7 - 1000μm, ή σε κυματαριθμούς είναι 13300 - 10cm⁻¹ (Σχήμα 3.1.)



Σχήμα 3.1: Η θέση της υπέρυθρης περιοχής στο ηλεκτρομαγνητικό φάσμα.

Η υπέρυθρη περιοχή του φάσματος της ακτινοβολίας υποδιαιρείται σε τρεις επιμέρους περιοχές:

- 1) την περιοχή του **εγγύς υπέρυθρου (NIR, Near Infrared Spectroscopy)**, που εκτείνεται στην περιοχή των μηκών κύματος 0,75 - 2,5μm, ή αλλιώς στην περιοχή των κυματαριθμών 13300 - 4000 cm⁻¹.
- 2) την **θεμελιώδη περιοχή (IR, Mid-Infrared Spectroscopy)**, ή μέσο **υπέρυθρο**, που εκτείνεται στην περιοχή των μηκών κύματος 2,5 - 25μm, ή αλλιώς στην περιοχή των κυματαριθμών 4000 - 400cm⁻¹, που είναι η συνηθέστερα χρησιμοποιούμενη στην Χημεία και
- 3) την περιοχή του **άπω υπέρυθρου (FIR, Far Infrared Spectroscopy)**, που εκτείνεται στην περιοχή των μηκών κύματος από 25 - 1000 μm, ή στην περιοχή των κυματαριθμών 400 -10 cm⁻¹.

Η ευρεία χρήση της φασματοσκοπίας υπέρυθρου στην θεμελιώδη περιοχή οφείλεται στο γεγονός ότι αυτή περιοχή έχει την κατάλληλη ενέργεια για να ενεργοποιήσει τα μόρια και να προκαλέσει δονήσεις των δεσμών των ατόμων τους. Βασική προϋπόθεση για να απορροφήσουν υπέρυθρη ακτινοβολία οι δεσμοί των ατόμων κάποιας ένωσης είναι να μεταβάλλεται η διπολική ροπή τους. Εάν αυτό δε γίνεται, τότε οι δεσμοί δεν θα απορροφήσουν υπέρυθρη ακτινοβολία και η δόνηση θεωρείται ανενεργή στο υπέρυθρο [2]. Όσο μεγαλύτερη είναι η μεταβολή της διπολικής ροπής, τόσο ισχυρότερη είναι η απορρόφηση. Ωστόσο, παρά την απορρόφηση ακτινοβολίας από τους δεσμούς των ατόμων των μορίων, τα άτομα εξακολουθούν να παραμένουν στην θεμελιώδη ηλεκτρονική κατάσταση.

Για γίνει καλύτερα κατανοητό το φαινόμενο της δόνησης των δεσμών μεταξύ των ατόμων ενός μορίου, θα θεωρηθεί ένας απλός διατομικός δεσμός σαν ένα ελατήριο που φέρει δυο μάζες στα άκρα του. Αν το σύστημα του ελατηρίου και των μαζών τεθεί σε παλμική κίνηση, τότε οι δονήσεις του συστήματος αυτού, δύννεται να περιγραφούν από τον νόμο της απλής αρμονικής ταλάντωσης του Hooke:

$$F = k \cdot X_T$$

όπου F: η δύναμη επαναφοράς του ελατηρίου

k: η σταθερά του ελατηρίου και

X_T: η συνολική μετατόπιση του ελατηρίου, ή αλλιώς η επιμήκυνση ή επιβράχυνση του ελατηρίου

Η σταθερά του ελατηρίου k αποτελεί μέτρο της ακαμψίας ενός ελατηρίου. Η συχνότητα στην απλή αρμονική ταλάντωση δίνεται από τον τύπο:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$$

όπου μ η ανηγμένη μάζα που στην περίπτωση του συστήματος του ελατηρίου και των δύο μαζών είναι ο αρμονικός μέσος όρος των μαζών, που δίνεται από τον τύπο:

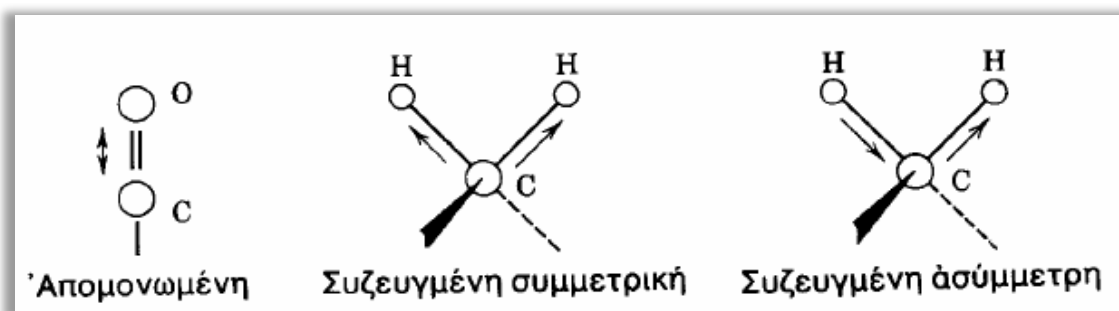
$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$$

Με τον ίδιο ακριβώς τρόπο περιγράφεται και η ταλάντωση δυο ατόμων συνδεδεμένων με δεσμό σε ένα μόριο.

Το είδος και η ένταση των δονήσεων των δεσμών σε ένα μόριο, όταν απορροφούν υπέρυθη ακτινοβολία, είναι χαρακτηριστικά για κάθε είδος δεσμού, όπως και οι περιοχές μηκών κύματος της ακτινοβολίας. Συνεπώς, η απορρόφηση που παρουσιάζεται σε διάφορες περιοχές του υπέρυθρου φάσματος μίας ένωσης, εξαρτάται από το είδος και το αριθμό των δεσμών που περιέχονται στα μόρια της [19]. Με βάση αυτό, γίνεται σαφές πως το φάσμα απορρόφησης μίας ένωσης είναι χαρακτηριστικό της σύστασής της και μπορεί να χρησιμοποιηθεί για την ταυτοποίηση της. Χαρακτηριστικά πεδία εφαρμογής της φασματοσκοπίας υπέρυθρου είναι τα μίγματα οργανικών ενώσεων, όπως το πετρέλαιο, όπου γίνεται ταυτοποίηση των ομάδων των υδρογονανθράκων που περιέχουν.

Τα είδη των δονήσεων που παρουσιάζονται στα μόρια των διαφόρων ενώσεων μπορούν να υποδιαιρεθούν στις ακόλουθες κατηγορίες:

1. Δονήσεις τάσης (stretching vibrations). Οι δονήσεις τάσης είναι αυτές κατά τις οποίες δυο συνδεδεμένα μεταξύ τους άτομα, πάλλονται, μεταβάλλοντας την μεταξύ τους απόσταση, χωρίς ωστόσο να αλλάζουν τον άξονα και τις γωνίες δεσμού. Οι δονήσεις αυτές διακρίνονται στις μεμονωμένες δονήσεις (ο δεσμός O-H) και τις συζευγμένες (η ομάδα μεθυλενίου). Οι συζευγμένες δονήσεις διακρίνονται στις συμμετρικές και στις ασύμμετρες (Σχήμα 3.2).



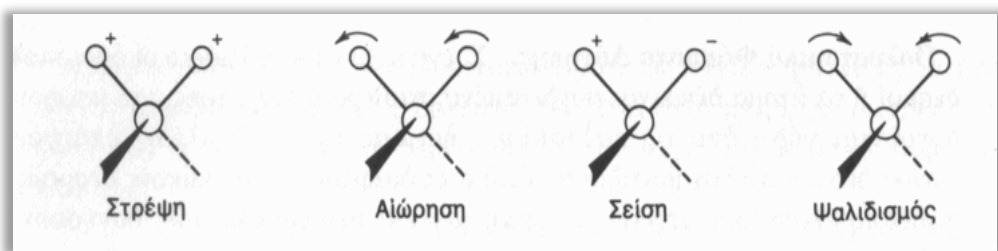
Σχήμα 3.2 : Σχηματικά παραδείγματα δονήσεων τάσης

Οι δονήσεις τάσης απαιτούν γενικά μεγαλύτερες ενέργειες από τις δονήσεις κάμψης, και σημειώνονται με το σύμβολο ν ακολουθούμενο από την χημική ομάδα σε παρένθεση. Για παράδειγμα το $\nu(\text{C}=\text{O})=1600\text{cm}^{-1}$ σημαίνει ότι η θεμελιώδης δόνηση της καρβονυλικής ομάδας παρατηρείται στα 1600cm^{-1} .

2. Δονήσεις κάμψης (bending vibrations). Οι δονήσεις κάμψης χαρακτηρίζονται από μια συνεχή μεταβολή της γωνίας μεταξύ δύο δεσμών. Οι κάμψεις των αρωματικών ομάδων, λ.χ. C-H [26], που λαμβάνουν χώρα στο επίπεδο του πυρήνα του φαινυλίου σημειώνονται με το σύμβολο $\delta(\text{C-H})$, ενώ αυτές που συμβαίνουν εκτός του επιπέδου σημειώνονται με $\gamma(\text{C-H})$. Αυτή η ονοματολογία εφαρμόζεται επίσης στα αλκένια και αλκίνια.

3. Δονήσεις σείσης (wagging vibrations). Οι δονήσεις σείσης παράγονται όταν μια μη γραμμική δομική μονάδα τριών ατόμων πάλλεται εντός του επιπέδου ισορροπίας που σχηματίζεται από τα άτομα και τους δυο δεσμούς. Τέτοιες δονήσεις σημειώνονται με το σύμβολο $\omega(\text{CH}_2)$.

4. Δονήσεις αιώρησης (rocking vibrations). Οι δονήσεις αιώρησης συμβαίνουν όταν η ίδια δομική μονάδα πάλλεται εκτός του επιπέδου ισορροπίας. Το σύμβολο με το οποίο σημειώνουμε αυτόν τον τρόπο δόνησης είναι το $\rho(\text{CH}_2)$.



Σχήμα 3.3 : Σχηματικό διάγραμμα δονήσεων στρέψης, αιώρησης, σείσης και ψαλιδισμού (+ σημαίνει κίνηση προς τα πάνω από το επίπεδο της σελίδας και – προς τα κάτω)

5. Οι δονήσεις συστροφής (twisting vibrations). Συμβαίνουν όταν η ίδια δομική μονάδα περιστρέφεται γύρω από το δεσμό που τη συνδέει με το υπόλοιπο τμήμα του μορίου. Τέτοιες δονήσεις αναφέρονται με το σύμβολο $\tau(\text{CH}_2)$.

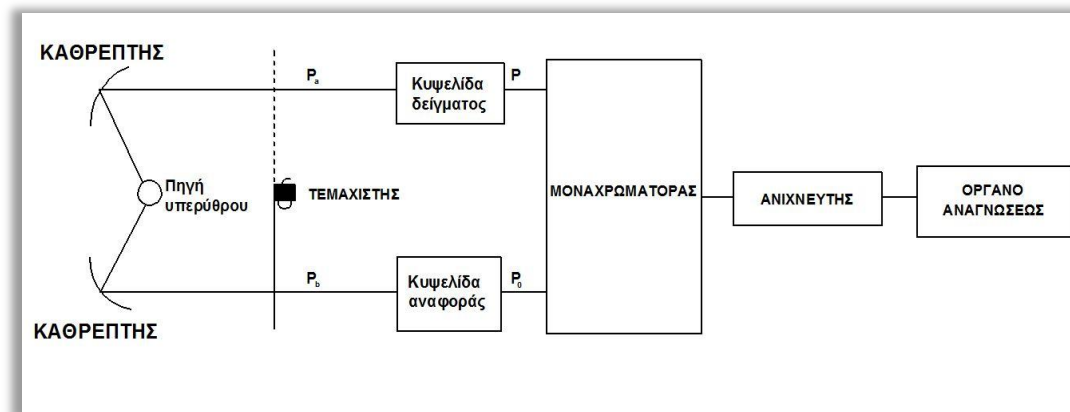
6. Δονήσεις ψαλιδιού ή παραμόρφωσης (scissoring or deformation vibrations)

Παράγονται όταν δυο μη συνδεδεμένα άτομα κινούνται μπρος-πίσω και προς τη μεταξύ τους διεύθυνση. Στις δονήσεις αυτές δίνουμε το σύμβολο $s(\text{CH}_2)$. Έκτός απ' αυτούς τους θεμελιώδεις τρόπους δόνησης, μπορούν να συμβούν αρμονικές και συνδυασμένες δονήσεις[26]. Οι αρμονικές δονήσεις έχουν συχνότητες με ακέραιο πολλαπλάσιο της θεμελιώδους λ.χ. 2ν ή 2δ . Οι συχνότητες των συνδυασμένων δονήσεων είναι το άθροισμα ($\nu+\delta$) ή η διαφορά ($\nu-\delta$) μεταξύ δύο η περισσότερων θεμελιωδών η αρμονικών δονήσεων.

3.3 Οργανολογία υπερύθρου

Περιγραφή λειτουργίας φασματοφωτομέτρων υπερύθρου

Σχηματικά η λειτουργία ενός φασματοφωτομέτρου υπερύθρου παρουσιάζεται στο παρακάτω σχήμα:



Σχήμα 3.4 : Σχηματικό διάγραμμα υπέρυθρου φασματοφωτομέτρου διπλής δέσμης

Η υπέρυθρη ακτινοβολία από την πηγή διαιρείται σε έναν διαιρετή δέσμης ακτινών και με τη βοήθεια κατόπτρων το ένα μέρος της δέσμης διέρχεται από την κυψελίδα αναφοράς, που συνήθως περιέχει αέρα, ενώ το άλλο μέρος της δέσμης από την κυψελίδα του δείγματος. Στην συνέχεια, οι δυο δέσμες κατευθύνονται στον μονοχρωμάτορα, από τον οποίο επιτρέπεται η διέλευση δεσμών ακτινοβολίας συγκεκριμένου μήκους κύματος. Οι δέσμες ακτινών που εξέρχονται από τον μονοχρωμάτορα, κατευθύνονται στον ανιχνευτή, όπου καταγράφεται και μετρείται η ακτινοβολία που προσπίπτει σε αυτόν από την κάθε δέσμη. Τα φασματοφωτόμετρα υπερύθρου χρησιμοποιούν την διπλή δέσμη μηδενισμού. Αυτό γίνεται γιατί με αυτόν τον τρόπο μειώνεται η παρεμποδιστική δράση την υγρασίας και του CO_2 της ατμόσφαιρας.

Ένα φασματοφωτόμετρο υπερύθρου αποτελείται από τις ακόλουθες μονάδες:

Πηγή υπέρυθρης ακτινοβολίας

Ως πηγές υπέρυθρης ακτινοβολίας χρησιμοποιούνται λυχνίες βολφραμίου και οι λυχνίες πυρακτώσεως Nernst ($0,4 - 20\mu m$) και Globar ($1 - 10\mu m$). Η λυχνία

πυρακτώσεως Nernst αποτελείται από οξειδία σπάνιων γαιών ($ZrO_2+Y_2O_3$), σε σχήμα κοίλης ράβδου, που θερμαίνεται ηλεκτρικά στους 1500 - 2000°C, ενώ λυχνία πυρακτώσεως Globar αποτελείται από ράβδο φρυγμένου πυριτικοκαρβιδίου (SiC), που θερμαίνεται στους 1300 - 1500°C[8]. Σήμερα οι λυχνίες αυτές έχουν αντικατασταθεί από πηγές Laser που εκπέμπουν ακτινοβολία μεγάλης και σταθερής έντασης σε όλο το φάσμα εκπομπής.

Μονοχρωμάτορες

Οι μονοχρωμάτορες, ή αλλιώς οι επιλογείς μήκους κύματος, είναι κατασκευασμένοι από υλικά περατά στην υπέρυθρη ακτινοβολία, όπως NaCl ή CsBr. Αντί των δαπανηρών και εύθραυστων πρισμάτων στα σύγχρονα φασματοφωτόμετρα χρησιμοποιούνται ανακλαστικά οπτικά φράγματα και κοίλα κάτοπτρα.

Κελί δείγματος

Το κελί του δείγματος, στην περίπτωση υγρών δειγμάτων, αποτελείται από κρυσταλλικά υλικά, τα οποία είναι περατά από την υπέρυθρη ακτινοβολία. Συνήθως τα κελιά αυτά είναι ζεύγη κρυσταλλικών πλακιδίων από κρυσταλλικό NaCl, KBr και Cd-Se, Zn-Se.

Ανιχνευτές

Οι ανιχνευτές που συνήθως χρησιμοποιούνται στην φασματοσκοπία υπέρυθρου είναι θερμοζεύγη, βολόμετρα ή thermistors και κύτταρα Golay (οπτικοακουστικός ανιχνευτής). Τα θερμοζεύγη ανιχνεύουν τη θερμότητα που παράγεται κατά την απορρόφηση υπέρυθρης ακτινοβολίας από μια μαυρισμένη επιφάνεια (μέλαν σώμα), σύμφωνα με το θερμοηλεκτρικό φαινόμενο. Με τα φωτογωγικά κύτταρα μετράται η αύξηση της αγωγιμότητας ενός λεπτού στρώματος ημιαγωγού, που αποτελείται από ύαλο CdS ή PbS, κατά την πρόσπτωση ακτινοβολίας σε αυτό. Με τα βολόμετρα ή thermistors παρακολουθείται η μεταβολή της ηλεκτρικής αντιστάσεως ενός ελάσματος, με μεγάλο θερμικό συντελεστή, κατά την πρόσπτωση της υπέρυθρης ακτινοβολίας. Το κύτταρο Golay αποτελείται από δοχείο πλήρες αερίου, το οποίο έχει τοιχώματα από ελαστική μεμβράνη. Όταν η ακτινοβολία προσπίπτει επί του δοχείου, διαστέλλεται ο όγκος του αερίου και σαν συνέπεια της διαστολής, μετακινείται ένα διάφραγμα με κατοπτρική επιφάνεια επί της οποίας ανακλάται φωτεινή δέσμη, που παρακολουθείται από φωτοκύτταρο. Η ισχύς της ανακλώμενης δέσμης και συνεπώς

και το σήμα του φωτοκύτταρου μεταβάλλονται σημαντικά με μικρή μετακίνηση του διαφράγματος και έτσι ανιχνεύονται οι ασθενείς ακτινοβολίες

Καταγραφικό σύστημα

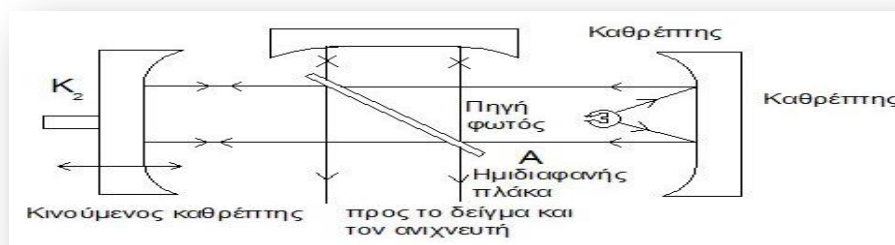
Συνήθως, το καταγραφικό σύστημα αποτελείται από έναν ηλεκτρονικό υπολογιστή, με κατάλληλο λογισμικό, για να καταγράφει το σήμα που λαμβάνεται από τον ανιχνευτή και να το καταχωρεί σε αρχεία για περαιτέρω επεξεργασία.

3.4 Αρχές λειτουργίας φασματοσκοπίας FTIR

Τα φασματοφωτόμετρα υπέρυθρου FTIR ανήκουν στην κατηγορία των φασματοφωτομέτρων, που δεν έχουν επιλογή μήκους κύματος (μονοχρώματα) και χρησιμοποιούν τον μετασχηματισμό Fourier για την δημιουργία του φάσματος. Στο όργανο αυτό αντί του μονοχρώματα, χρησιμοποιείται το σαρωτικό συμβολόμετρο του Michelson με το οποίο αναλύεται η υπέρυθρη ακτινοβολία[2].

Η ανάλυση κατά Fourier ή αλλιώς ο μετασχηματισμός Fourier είναι η ανάλυση μιας μαθηματικής συνάρτησης ή μιας πειραματικής καμπύλης με την μορφή μιας τριγωνομετρικής σειράς. Χρησιμοποιείται ως μέθοδος προσδιορισμού των αρμονικών συστατικών ενός πολύπλοκου περιοδικού κύματος, όπως αυτό που προέρχεται από τα φασματοφωτόμετρα FTIR.

Η φασματοσκοπία FTIR βασίζεται στην καταγραφή του φάσματος με συμβολομετρικές μετρήσεις (interferometric measurements), που υπερτερούν των κοινών μηχανισμών σάρωσης τους φάσματος. Η διάταξη του μηχανισμού του συμβολόμετρου στο φασματοφωτόμετρο FTIR παρουσιάζεται στο σχήμα:



Σχήμα 3.5: Σχηματικό διάγραμμα συμβολόμετρου στο φασματοφωτόμετρο FTIR.

Το συμβολόμετρο Michelson (Michelson Interferometer) είναι μια πλάκα (A) κατασκευασμένη από ημιδιαφανές υλικό, συνήθως KBr, που δεν απορροφά την IR

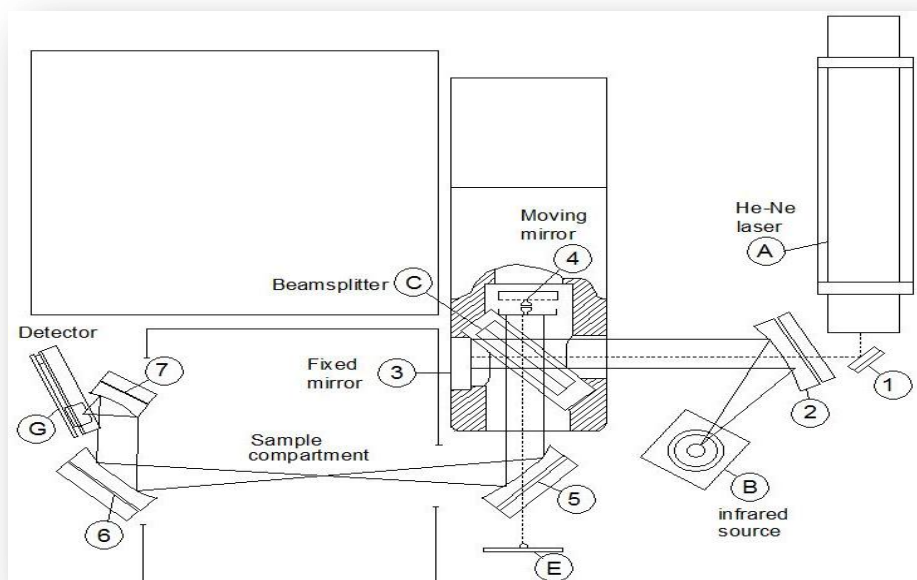
ακτινοβολία, και κατεργασμένη έτσι ώστε να ανακλά το 50% της ακτινοβολίας που προσπίπτει πάνω της. Η πλάκα διαιρεί την ακτινοβολία σε δύο ίσα μέρη αποστέλλοντας το πρώτο μισό στο κάτοπτρο K1 και το άλλο μισό στο K2. Οι δύο δέσμες ανακλώνται στα K1 και K2 και γυρνούν στην πλάκα του διαχωριστή (plate beam splitter) της δέσμης φωτός, όπου και συνενώνονται πάλι σε μια δέσμη φωτός, η οποία διαπερνά το δείγμα της ουσίας και οδηγείται στον ανιχνευτή.

Η μονοχρωματική δέσμη φωτός που εκπέμπεται από την πηγή και συνενώνεται στην πλάκα A, υπόκειται σε θετική ή αρνητική συμβολή ανάλογα με τις αποστάσεις της πλάκας A από τα κάτοπτρα K1 και K2. Εάν οι αποστάσεις αυτές είναι ίσες ή διαφέρουν κατά ένα ακέραιο πολλαπλάσιο του μήκους κύματος της μονοχρωματικής δέσμης, τότε δημιουργείται θετική συμβολή που δίνει μια δέσμη φωτός μεγαλύτερης έντασης. Εάν, όμως οι αποστάσεις της A από τα K1 και K2 δεν είναι ακέραια πολλαπλάσια του μήκους κύματος, τότε οι δέσμες φωτός ανακλώνται στα κάτοπτρα και επανέρχονται στην A, όπου και εξουδετερώνονται.

Στην μονάδα του συμβολομέτρου, το κάτοπτρο K2 κινείται αργά προς την πλάκα A ή απομακρύνεται από αυτή. Με τον τρόπο αυτό ο ανιχνευτής δέχεται ακτινοβολία εναλλασσόμενης έντασης λόγω του φαινομένου της συμβολής. Η πηγή μπορεί να εκπέμπει μια σειρά από μονοχρωματικές ακτινοβολίες (ν_1 , ν_2 κ.τ.λ.), που δημιουργούν στον ανιχνευτή ένα περίπλοκο «μίγμα» ακτινοβολιών με εναλλασσόμενες εντάσεις. Με τη βοήθεια του μετασχηματισμού Fourier και με ένα ηλεκτρονικό υπολογιστή είναι δυνατό να προσδιοριστούν οι συχνότητες και οι εντάσεις των ακτινοβολιών και τελικά να δημιουργηθεί το υπέρυθρο φάσμα.

Όταν η συνενωμένη δέσμη ακτινοβολίας από την πλάκα A του συμβολομέτρου κατευθύνεται στο δείγμα, πριν φθάσει στον ανιχνευτή, οι απορροφήσεις του δείγματος θα εμφανιστούν ως κενά στην κατανομή των συχνοτήτων. Με τον μετασχηματισμό Fourier τα κενά αυτά μετατρέπονται στο κανονικό φάσμα IR. Συνήθως η καταγραφή των σημάτων που δέχεται ο ανιχνευτής γίνεται κάθε 1/1000 του δευτερολέπτου και κάθε πληροφορία συσσωρεύεται σε ένα από τα 1000 σημεία του προγραμματισμένου φάσματος[19]. Ο ηλεκτρονικός υπολογιστής επεξεργάζεται τις πληροφορίες (εντάσεις απορροφήσεων σε διάφορες συχνότητες), εκτελεί τους μετασχηματισμούς Fourier σε λίγα δευτερόλεπτα και παρουσιάζει ένα κανονικό φάσμα υπέρυθρου. Η παραπάνω διαδικασία ολοκληρώνεται σε λίγα δευτερόλεπτα

(10 - 20sec) και μετά από έναν αριθμό σαρώσεων (20 - 100 σαρώσεις) του φάσματος ο «θόρυβος» μειώνεται σημαντικά, ενώ οι εντάσεις των απορροφήσεων βελτιώνονται σε σημαντικό βαθμό.



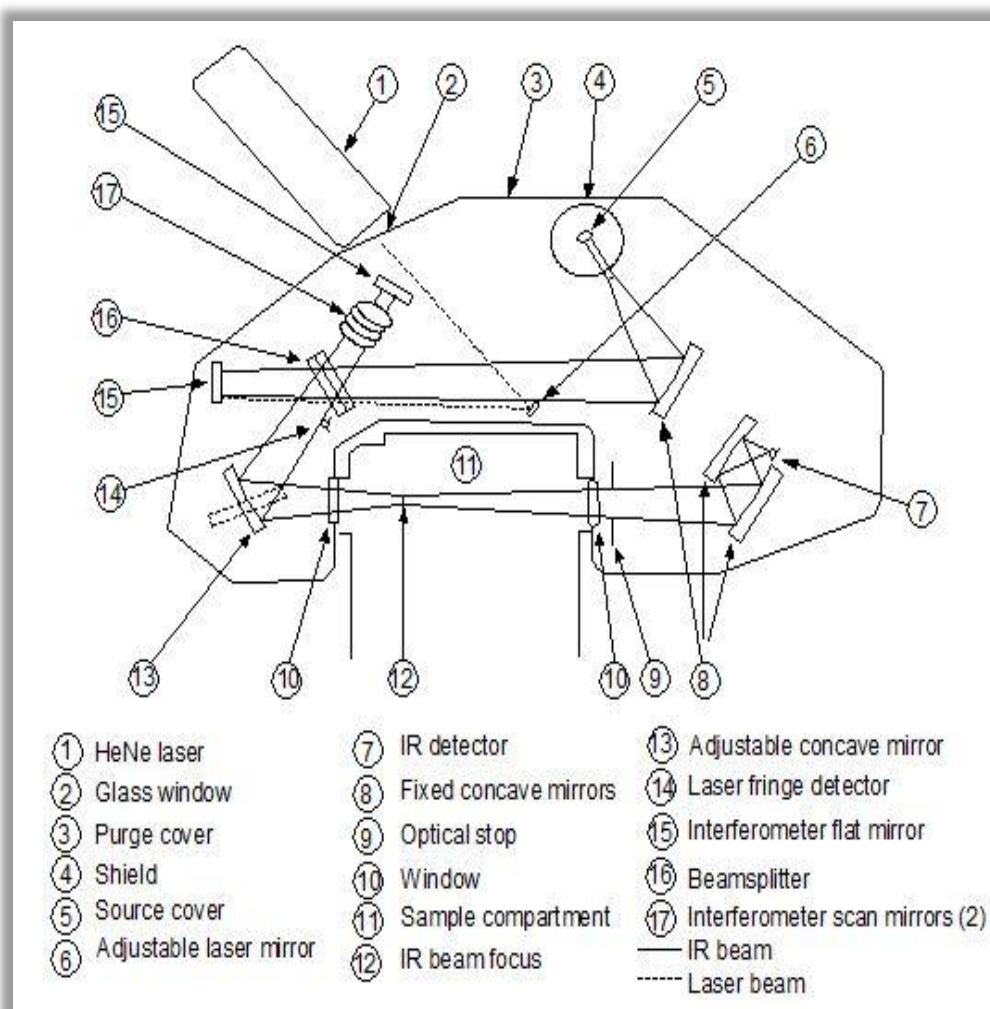
Σχήμα 3.6 : Διάταξη υπέρυθρου φασματοφωτομέτρου FTIR

Το σημαντικότερο πλεονέκτημα του οργάνου είναι η ταχεία σάρωση και καταγραφή που επιτυγχάνει. Ενώ στα συμβατικά φασματοφωτόμετρα υπέρυθρου μετράται η φασματική απορρόφηση σε κάθε μήκος κύματος, στο FTIR σαρώνεται ταυτόχρονα όλη φασματική περιοχή.

3.5 Περιγραφή συλλογής φασματοσκοπικών δεδομένων των δειγμάτων της εργασίας.

Η φασματοσκοπική ανάλυση εκτελέστηκε με το φασματοφωτόμετρο υπέρυθρου FTIR Spectrum 1000 της εταιρείας Perkin-Elmer με ανιχνευτή DGTS (δευτεριωμένης θειϊκής τριγλυκίνης). Τα υγρά δείγματα εισήχθησαν χρησιμοποιώντας ένα οριζόντιο κελί ολικής ανάκλασης HATR (Pike Technologies) με κρύσταλλο σεληνιούχου ψευδαργύρου (ZnSe). Για τα στερεά δείγματα (κεριά, πίτσες) το λεπτό φίλμ πάνω στον κρύσταλλο του κελιού δημιουργούνταν είτε

ζεσταίνοντας το δείγμα είτε διαλύοντας σε κ-εξάνιο ή χλωροφόρμιο, με επακόλουθη εξάτμιση του διαλύτη σε ρεύμα αζώτου. Τα δείγματα αναλύθηκαν στην φασματική περιοχή των $4000 - 650\text{cm}^{-1}$, με βήμα 2cm^{-1} . Κατά την φασματική ανάλυση γίνονταν 20 σαρώσεις (scans).



Σχήμα 3.7 :Σχηματικό διάγραμμα του φασματοφωτομέτρου FTIR Spectrum 1000, της εταιρείας Perkin-Elmer

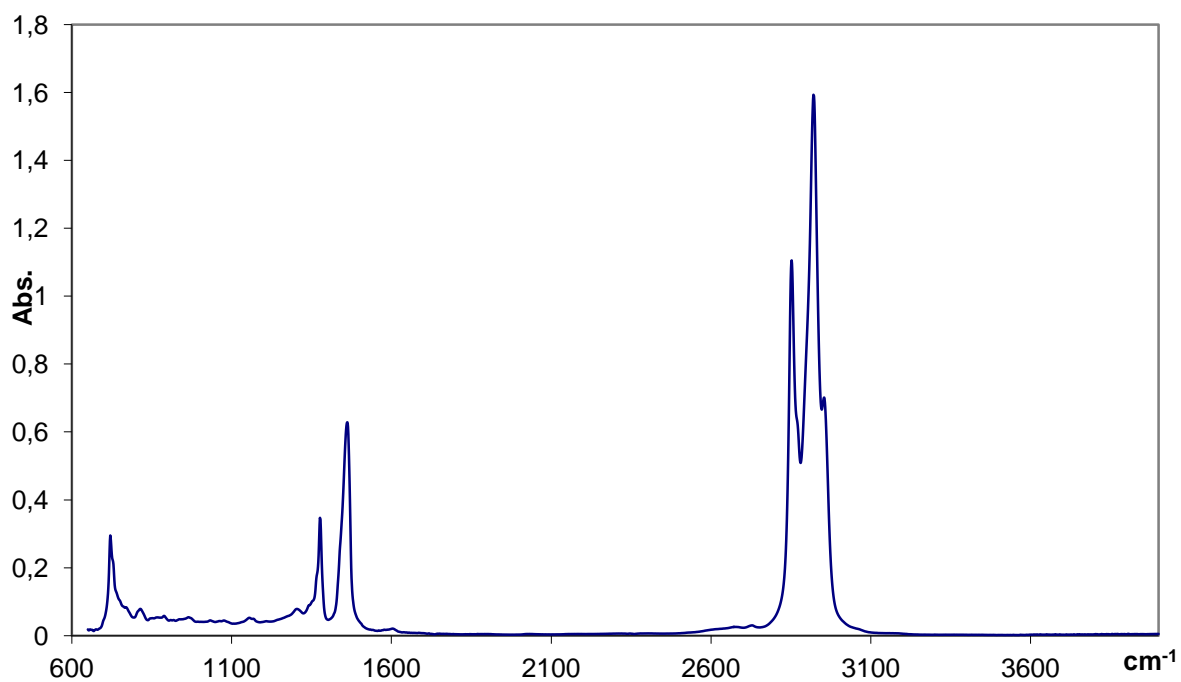


Σχήμα 3.8: Spectrum 1000, της εταιρείας Perkin-Elmer



Σχήμα 3.9: Κελί εισαγωγής δείγματος.

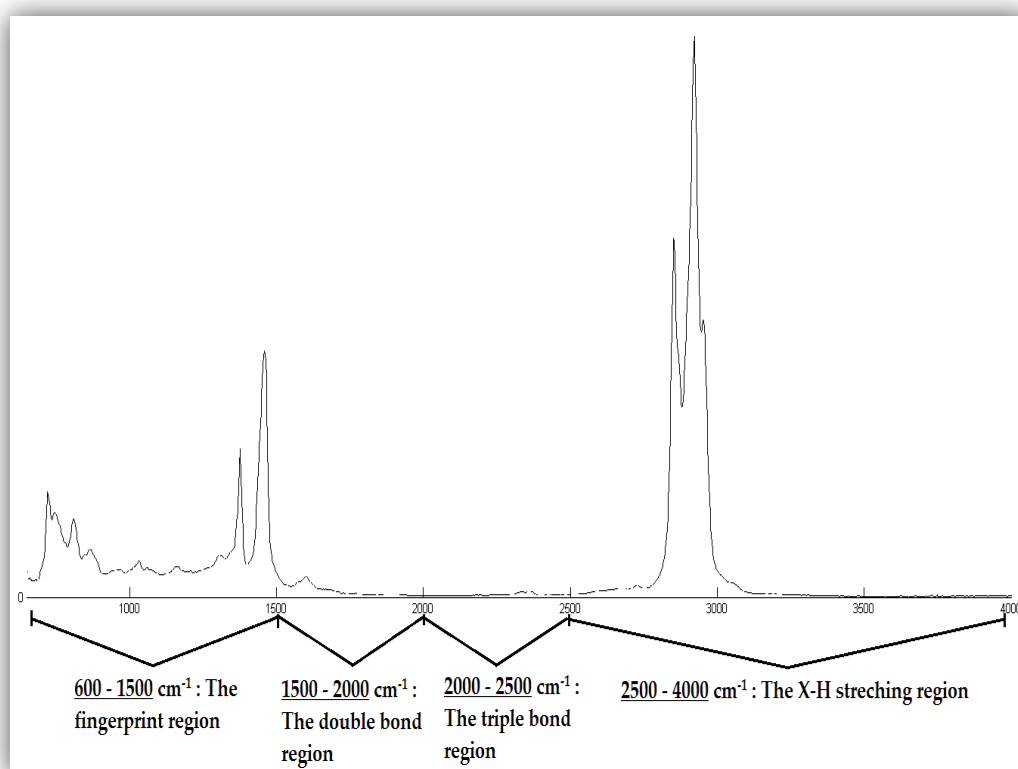
Παρακάτω παρουσιάζεται ένα χαρακτηριστικό φάσμα ενός τυπικού δείγματος (MB4) στην περιοχή $4000 - 600\text{cm}^{-1}$:



Σχήμα 3.10 : Υπέρυθρο φάσμα απορρόφησης ενός τυπικού ντήζελ στην περιοχή $4000 - 600\text{cm}^{-1}$.

3.6 Οι αρωματικοί υδρογονάνθρακες στο φάσμα του υπέρυθρου

Η περιοχή της υπέρυθρης ακτινοβολίας ($4000 - 400\text{cm}^{-1}$) μπορεί να διαιρεθεί κατά προσέγγιση σε τέσσερις περιοχές και η φύση της εκάστοτε ομάδας μπορεί γενικά να καθοριστεί από την περιοχή απορρόφησης του φάσματος στην οποία βρίσκεται. Οι περιοχές αυτές έχουν καθοριστεί ως εξής: Η περιοχή απορρόφησης λόγω της διέγερσης και δόνησης των δεσμών των ατόμων X-H είναι από $4000 - 2500\text{cm}^{-1}$, η περιοχή απορρόφησης λόγω της διέγερσης και δόνησης των τριπλών δεσμών των ατόμων του άνθρακα $\text{C}\equiv\text{C}$ είναι από $2500 - 2000\text{cm}^{-1}$. Για τους διπλούς δεσμούς μεταξύ των ατόμων του άνθρακα $\text{C}=\text{C}$ είναι η περιοχή από $2000 - 1500\text{cm}^{-1}$ ενώ η περιοχή από $1500 - 600\text{cm}^{-1}$ χαρακτηρίζεται ως το αποτύπωμα του μορίου (fingerprinting region).



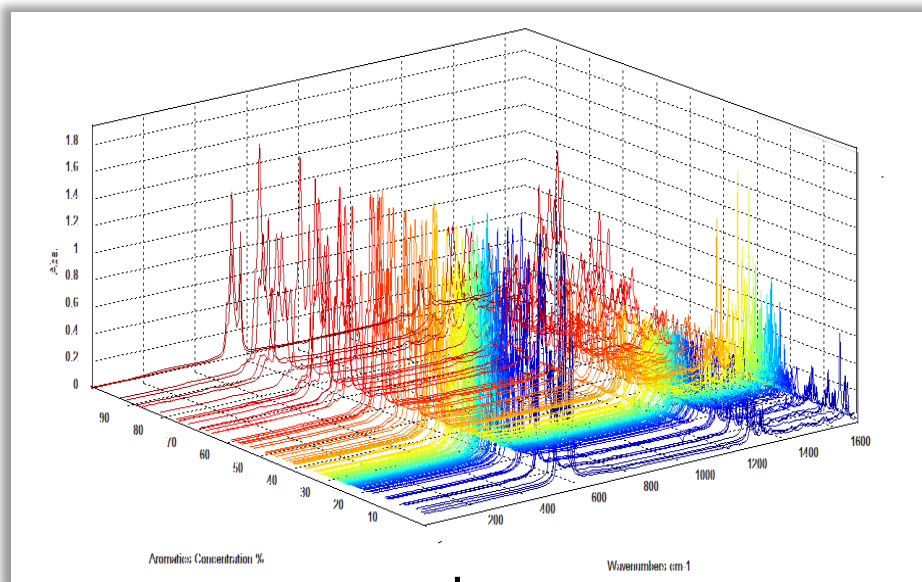
Σχήμα 3.11: Οι 4 περιοχές απορρόφησης του υπέρυθρου φάσματος.

Όπως προαναφέρθηκε οι θεμελιώδεις απορροφήσεις στην περιοχή $2500 - 4000\text{cm}^{-1}$ οφείλονται σε διεγέρσεις των δεσμών O-H, C-H και N-H. Πιο συγκεκρημένα οι δονήσεις των δεσμών μεταξύ των O-H εμφανίζονται στην περιοχή $3700 - 3600\text{cm}^{-1}$ ενώ των N-H δεσμών μεταξύ $3400 - 3300\text{cm}^{-1}$ [7]. Η απορρόφηση που ενδιαφέρει περισσότερο είναι αυτή που προκύπτει από δόνηση των δεσμών C-H (αλειφατικών ενώσεων) που συμβαίνει στην περιοχή $2900 - 3100\text{cm}^{-1}$. Αν ο άνθρακας του δεσμού C-H συνδέεται με διπλό δεσμό ατόμων με άλλο άνθρακα ή με έναν αρωματικό δακτύλιο τότε παρατηρείται άυξηση στην απορρόφηση μεταξύ $3000 - 3100\text{cm}^{-1}$.

Απορροφήσεις που οφείλονται σε δονήσεις των τριπλών δεσμών εμφανίζονται στην περιοχή $2500 - 2000\text{cm}^{-1}$ λόγω των υψηλών σταθερών τάσεων που επικρατούν. Πιο συγκεκριμένα δεσμοί της μορφής $\text{C}\equiv\text{C}$ απορροφούν ανάμεσα στα $2300 - 2050\text{cm}^{-1}$, ενώ δεσμοί της μορφής $\text{C}\equiv\text{N}$ ανάμεσα στα $2300 - 2200\text{cm}^{-1}$. Οι κύριες απορροφήσεις στην περιοχή $2000 - 1500\text{cm}^{-1}$ οφείλονται στις δονήσεις διπλών δεσμών $\text{C}=\text{C}$ και $\text{C}=\text{O}$ (ομάδες καρβονυλίου). Οι απορροφήσεις από δονήσεις δεσμών ομάδων καρβονυλίου είναι οι πιο εύκολες στο να ανιχνευθούν, διότι είναι συνήθως οι πιο έντονες και εξαρτώνται από τον τύπο των δεσμών με τα άλλα άτομα εκτός άνθρακα και οξυγόνου. Εμφανίζονται στην περιοχή από $1830 - 1650\text{cm}^{-1}$. Οι απορροφήσεις δεσμών $\text{C}=\text{C}$ είναι συνήθως ασθενέστερες και εμφανίζονται γύρω από τα 1650cm^{-1} .

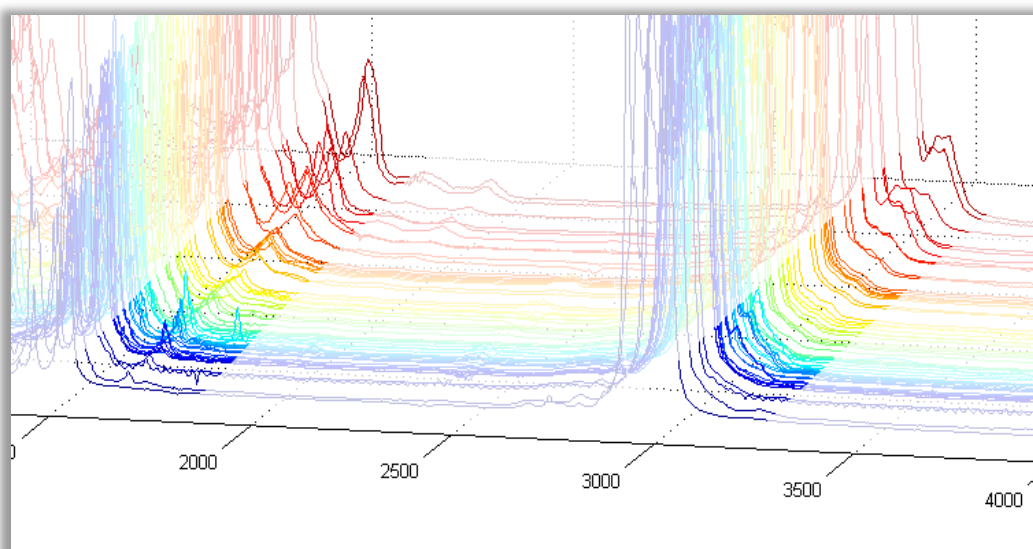
Εως τώρα έχει αναφερθεί ότι κάθε κορυφή ή ομάδα κορυφών σε ένα φάσμα υπερύθρου μπορεί να συσχετιστεί σε μια συγκεκριμένη παραμόρφωση του μορίου ή κίνηση μιας ομάδας ατόμων ή κάμψη ενός συγκεκριμένου δεσμού. Αυτό είναι πιθανό να συμβεί για δεσμούς που έχουν καλή “συμπεριφορά” κατά την δόνηση. Παρόλα αυτά όμως σε πολλές περιπτώσεις η συμπεριφορά των δεσμών δεν είναι “καλή” και παρατηρείται διαφοροποίηση στις εκατοντάδες των κυματαριθμών ακόμα και ανάμεσα σε ίδια μόρια. Αυτό ισχύει κυρίως για την περιοχή $1500 - 650\text{cm}^{-1}$ στην οποία απορροφούν οι περισσότερες σκελετικές δονήσεις και δονήσεις κάμψης και οδηγούν σε μεγάλες αποκλίσεις. Γι’ αυτό το λόγο το φάσμα στην περιοχή αυτή μπορεί να θεωρηθεί ως ένα αποτύπωμα του μορίου.

Παρακάτω παρουσιάζεται το σύνολο των 131 φασμάτων των δειγμάτων πετρελαιοειδών. Τα δείγματα έχουν τοποθετηθεί σύμφωνα με την περιεκτικότητα τους σε αρωματικούς υδρογονάνθρακες.



Σχήμα 3.12: Τα 131 δειγμάτων πετρελαιοειδών.

Όπως φαίνεται στο σχήμα 3.13 στις περιοχές $3000 - 3100\text{cm}^{-1}$ παρουσιάζεται αύξηση της απορρόφησης ανάλογη της περιεκτικότητας των δειγμάτων σε αρωματικά συστατικά.



Σχήμα 3.13: Περιοχές που ανταποκρίνονται με αύξηση απορρόφησης όταν υπάρχει αύξηση αρωματικών.

Σύμφωνα με την διαίρεση των περιοχών του φάσματος που αναφέρθηκε νωρίτερα κάθε μια απο αυτές, αντιστοιχεί σε διαφορετικές ομάδες δεσμών, όμως η συνολική πληροφορία προκύπτει απο την χημειομετρική ανάλυση της απορρόφησης του συνόλου των μηκών κύματος ή απο συνδιασμούς χημειομετρικής ανάλυσης επιμέρους περιοχών του φάσματος.

4. ΠΟΛΥΜΕΤΑΒΛΗΤΕΣ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ

4.1. Εισαγωγή

Για την δημιουργία των μοντέλων πρόβλεψης των ιδιοτήτων, θα γίνει χρήση μεθόδων στατιστικής επεξεργασίας και πολυμεταβλητής ανάλυσης των δεδομένων. Με τις τεχνικές αυτές επιτυγχάνεται η συσχέτιση των δεδομένων απορρόφησης των περιοχών του μέσου υπέρυθρου φάσματος με τις τιμές των ιδιοτήτων των πετρελαϊκών κλασμάτων.

Σε εργασίες που έχουν δημοσιευθεί, για την συσχέτιση των δεδομένων απορρόφησης με τις τιμές ιδιοτήτων των καυσίμων, χρησιμοποιήθηκαν διάφορες πολυμεταβλητές τεχνικές, όπως η πολλαπλή γραμμική παλινδρόμηση MLR (Multiple Linear Regression), η παλινδρόμηση κυρίων συνιστωσών PCR (Principal Component Regression), η μέθοδος των μερικών ελαχίστων τετραγώνων PLS (Partial Least Squares κ.α. Οι τεχνικές αυτές αποτελούν τις σημαντικότερες τεχνικές στον τομέα της πολυμεταβλητής ανάλυσης δεδομένων και είναι οι πλέον διαδεδομένες για την δημιουργία μοντέλων πρόβλεψης ιδιοτήτων. Στο κεφάλαιο αυτό θα γίνει μια συνοπτική αναφορά σε αυτές τις πολυμεταβλητές τεχνικές και στον τρόπο που χρησιμοποιούνται στην κατασκευή των μοντέλων πρόβλεψης ιδιοτήτων.

Για να γίνει καλύτερη η κατανόηση των διαφόρων τεχνικών πολυμεταβλητής ανάλυσης, κρίνεται απαραίτητο να γίνει μια σύντομη αναφορά σε βασικές έννοιες και μεγέθη της στατιστικής.

4.2. Βασικές έννοιες και μεγέθη της στατιστικής

Σε αυτή την παράγραφο θα παρατεθούν εν συντομία οι βασικές στατιστικές έννοιες και μεγέθη, τα οποία είναι απαραίτητα για την επεξεργασία των δεδομένων και για την κατανόηση των πολυμεταβλητών μεθόδων. Τέτοιες έννοιες είναι ο μέσος όρος, το τυπικό σφάλμα, η απόκλιση, η τυπική απόκλιση, το εύρος διακύμανσης, το σφάλμα της πρόβλεψης, το τυπικό σφάλμα της πρόβλεψης κ.α.

Αριθμητικός μέσος ή μέση τιμή (Mean)

Ο αριθμητικός μέσος όρος, ή μέσος όρος, m ενός συνόλου που περιέχει n το πλήθος τιμές x_i , είναι ο λόγος του αθροίσματος του συνόλου των τιμών προς το πλήθος των τιμών του συνόλου. Δίνεται από τον τύπο:

όπου x_i η τιμή της κάθε μεταβλητής του συνόλου και n το πλήθος των τιμών του συνόλου.

Σφάλμα (Error)

Το σφάλμα σ μιας τιμής ενός δείγματος του συνόλου είναι η διαφορά αυτής της τιμής από τον μέσο όρο του συνόλου των δειγμάτων και δίνεται από τον τύπο:

$$\sigma = x_i - m$$

Απόκλιση ή απόλυτο σφάλμα (Deviation ή absolute error)

Η απόκλιση μίας τιμής, ή το απόλυτο σφάλμα, d_i μιας τιμής ενός δείγματος είναι η απόλυτη τιμή της διαφοράς της τιμής αυτής από τον αριθμητικό μέσο όρο ενός συνόλου δειγμάτων και υπολογίζεται από τον τύπο:

$$d_i = |\sigma| = |x_i - m|$$

όπου x_i είναι η τιμή κάθε δείγματος και m ο μέσος όρος του συνόλου των δειγμάτων.

Διακύμανση ή διασπορά (Variance)

Η διακύμανση ή διασπορά V ή s^2 ορίζεται ως ο μέσος όρος των αποκλίσεων των δειγμάτων, από το μέσο όρο των δειγμάτων του συνόλου, υψωμένων στο τετράγωνο:

$$V = \frac{\sum_{i=1}^n (x - m)^2}{n - 1} = s^2$$

Τυπική απόκλιση (Standard Deviation)

Η τυπική απόκλιση αποτελεί μέτρο της διασποράς των μετρήσεων γύρω από το μέσο όρο και ισούται με την τετραγωνική ρίζα της διακύμανσης:

$$s = \sqrt{\frac{\sum_{i=1}^n (x - m)^2}{n - 1}}$$

Για μεγάλο αριθμό μετρήσεων ($n > 30$), όπου ο υπολογιζόμενος μέσος όρος προσεγγίζει τον πραγματικό ισχύει:

$$s = \sqrt{\frac{\sum_{n=1}^n (x - m)^2}{n}}$$

Σχετική τυπική απόκλιση (Relative Standard Deviation)

Η σχετική τυπική απόκλιση S_r είναι η τυπική απόκλιση ανηγμένη ως προς τον μέσο όρο των δειγμάτων του συνόλου:

$$S_r = S/m,$$

ή εκφρασμένη ως ποσοστό

$$S_r(\%) = (S/m) * 100$$

Τυπικό σφάλμα (Standard Error)

Το τυπικό σφάλμα s_m είναι το σφάλμα του μέσου όρου του δείγματος του συνόλου των τιμών των δειγμάτων:

$$s_m = \frac{s}{\sqrt{n}}$$

Εύρος διακύμανσης (Range)

Το εύρος είναι διαφορά της μέγιστης x_{\max} από την ελάχιστη x_{\min} του συνόλου των δειγμάτων και δίνεται από τον τύπο:

$$R = x_{\max} - x_{\min}$$

Τα σημαντικότερα στατιστικά μεγέθη που χρησιμοποιούνται για τον χαρακτηρισμό και την αξιολόγηση των αποτελεσμάτων της πρόβλεψης που προκύπτουν από μοντέλα-αλγορίθμους πρόβλεψης ιδιοτήτων είναι:

Σφάλμα της πρόβλεψης EP (Error of Prediction)

Το σφάλμα της πρόβλεψης σ_i δίνεται από την διαφορά της προβλεπόμενης από το μοντέλο τιμής της ιδιότητας από την πραγματική τιμή :

$$\sigma = y_i^{\text{real}} - y_i^{\text{pred}}$$

όπου y_i^{real} η πραγματική τιμή της ιδιότητας και y_i^{pred} η προβλεπόμενη από το μοντέλο τιμή της ιδιότητας.

Απόλυτο σφάλμα της πρόβλεψης AEP (Absolute Error of Prediction)

Το απόλυτο σφάλμα της πρόβλεψης είναι η απόλυτη τιμή του σφάλματος της πρόβλεψης:

$$|\sigma_i| = |y_i^{\text{real}} - y_i^{\text{pred}}|$$

Όπου y_i^{real} η πραγματική τιμή της ιδιότητας και y_i^{pred} η προβλεπόμενη από το μοντέλο τιμή της ιδιότητας.

Μέσο σφάλμα της πρόβλεψης MEP (Mean Error of Prediction)

Το μέσο σφάλμα της πρόβλεψης είναι η μέση τιμή του σφάλματος της πρόβλεψης και προκύπτει από το λόγο του αθροίσματος των σφαλμάτων της πρόβλεψης $|\sigma_i|$ προς το πλήθος των προβλέψεων n . Δίνεται από τον τύπο:

$$m = \frac{\sum_{i=1}^n \sigma_i}{n} = \frac{\sum_{i=1}^n (y_i^{real} - y_i^{pred})}{n}$$

Μέσο απόλυτο σφάλμα της πρόβλεψης MAEP (Mean absolute Error of Prediction)

Το μέσο απόλυτο σφάλμα της πρόβλεψης $|m|$ είναι το άθροισμα όλων των απόλυτων σφαλμάτων της πρόβλεψης προς τον αριθμό των τιμών n που προβλέφθηκαν:

$$|m| = \frac{\sum_{i=1}^n |\sigma_i|}{n} = \frac{\sum_{i=1}^n |y_i^{real} - y_i^{pred}|}{n}$$

Τυπικό σφάλμα της πρόβλεψης SEP (Standard Error of Prediction)

Το τυπικό σφάλμα της πρόβλεψης είναι η τυπική απόκλιση του σφάλματος της πρόβλεψης και δίνεται από τον τύπο:

$$SEP = \sqrt{\frac{\sum_{i=1}^n (|\sigma_i| - m)^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^n (y_i^{real} - y_i^{pred} - m)^2}{n - 1}}$$

RMSEP (Root-Mean Squared Error of Prediction)

Το μέσο σφάλμα της τετραγωνικής ρίζας της πρόβλεψης RMSEP είναι η τετραγωνική ρίζα του μέσου όρου των τετραγώνων των σφαλμάτων της πρόβλεψης και δίνεται από τον τύπο:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i^{real} - y_i^{pred})^2}{n}}$$

Τυπική απόκλιση της πρόβλεψης SDP (Standard Deviation of Prediction)

Η τυπική απόκλιση της πρόβλεψης είναι μέτρο της διασποράς των απόλυτων σφαλμάτων της πρόβλεψης γύρω από το μέσο απόλυτο σφάλμα της πρόβλεψης:

$$S = \sqrt{\frac{\sum_{i=1}^n (|\sigma_i| - m)^2}{n - 1}}$$

Για μεγάλο αριθμό δειγμάτων $n > 30$ η τυπική απόκλιση της πρόβλεψης είναι :

$$S = \sqrt{\frac{\sum_{i=1}^n (|\sigma_i| - m)^2}{n}}$$

Μέγιστο απόλυτο σφάλμα της πρόβλεψης MaxAEP (Maximum Absolute Error of Prediction)

Το μέγιστο απόλυτο σφάλμα της πρόβλεψης $|\sigma|_{\max}$ δίνεται από τον τύπο:

$$|\sigma|_{\max} = \max\{|\sigma_i|\}$$

Ελάχιστο απόλυτο σφάλμα της πρόβλεψης MinAEP (Minimum Absolute Error of Prediction)

Το ελάχιστο απόλυτο σφάλμα της πρόβλεψης $|\sigma|_{\min}$ δίνεται από τον τύπο:

$$|\sigma|_{\min} = \min\{|\sigma_i|\}$$

Διορθωμένο τετραγωνικό σφάλμα της πρόβλεψης CR-RMSEP η CR-SEP (Corrected-Squared Error of Prediction)

Το διορθωμένο τετραγωνικό σφάλμα της πρόβλεψης είναι η διαφορά μεταξύ του μέσου σφάλματος της τετραγωνικής ρίζας της πρόβλεψης RMSEP και του θορύβου που προκαλούν τα σφάλματα της πρόβλεψης. Δίνεται από τον τύπο:

$$CR - SEP = RMSEP - \left(\frac{s^2 v}{x_{(v,a)}^2} \right)$$

Συντελεστής συσχέτισης R^2 (Correlation Coefficient)

Ο συντελεστής συσχέτισης μεταξύ των n τιμών y_i^{pred} που προκύπτουν από τα μοντέλα πρόβλεψης και των n πραγματικών τιμών y_i^{real} μιας ιδιότητας, δίνεται από την σχέση:

$$R^2 = \frac{\sum_{i=1}^n (y_i^{real} - y_i^{pred})^2}{\sum_{i=1}^n (y_i^{real} m^{real})^2}$$

όπου m^{real} ο μέσος όρος του συνόλου των n πραγματικών τιμών της ιδιότητας.

Τα μεγέθη που παρουσιάστηκαν είναι τα πιο γνωστά και συνήθη στην στατιστική και στην πρόβλεψη των ιδιοτήτων. Με αυτά είναι δυνατόν να αξιολογηθούν τα αποτελέσματα των μοντέλων πρόβλεψης και να συγκριθούν με τα αποτελέσματα άλλων εργασιών.

4.3. Προεπεξεργασία φασματικών δεδομένων

Πριν την χρήση των φασματικών δεδομένων και την αξιοποίησή τους για την εξαγωγή πληροφοριών και συσχετίσεων, πραγματοποιείται η προεπεξεργασία τους με διάφορες διαδικασίες. Τέτοιες διαδικασίες, είναι η διόρθωση γραμμής βάσης (baseline correction), η κανονικοποίηση των φασματικών δεδομένων (normalization-data scaling), η εξομάλυνση του φάσματος (smoothing), η αριθμητική ολοκλήρωση

(integration), η αριθμητική παραγωγή (derivation), η εφαρμογή της ανάλυσης κυρίων συνιστωσών (principal component analysis) στα αρχικά φασματικά δεδομένα κ.α. Πολλές φορές εφαρμόζονται συνδυασμοί αυτών των διαδικασιών.

Με την διόρθωση της γραμμής βάσης (baseline correction) ουσιαστικά μετατοπίζεται ο άξονας της γραμμής βάσης του φάσματος ως προς μία σταθερή τιμή, ορίζοντας μια νέα τιμή ως μηδενική τιμή του φάσματος. Στην περίπτωση της απλής διόρθωσης γραμμής βάσεως μία σταθερή ποσότητα αφαιρείται από κάθε φασματική τιμή.

Στην περίπτωση της διόρθωσης γραμμής βάσεως 2-σημείων, η γραμμή βάσεως αποτελείται από μία ευθεία, που ενώνει δύο σημεία του φάσματος και περιγράφεται από μία γραμμική εξίσωση. Οι νέες τιμές της φασματικής απορρόφησης μετά την διόρθωση προκύπτουν από την αφαίρεση των τιμών της γραμμής βάσης από τις αντίστοιχες αρχικές φασματικές τιμές[8]. Στην περίπτωση της πολλαπλής διόρθωσης της γραμμής βάσης, η διαδικασία είναι ίδια με αυτήν που ισχύει για τα δύο σημεία, μόνο που γίνεται για κάθε περιοχή του φάσματος ξεχωριστά.

Με την κανονικοποίηση (normalization) των τιμών των φασματικών δεδομένων επιτελείται η μετατροπή ενός συνόλου φασματικών τιμών σε ένα νέο σύνολο, οι τιμές του οποίου κυμαίνονται σε ένα συγκεκριμένο πεδίο τιμών, συνήθως στο πεδίο τιμών [0,1].

Μαθηματικά η κανονικοποίηση των φασματικών δεδομένων γίνεται σύμφωνα με τους παρακάτω τύπους:

$$x_{sc} = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

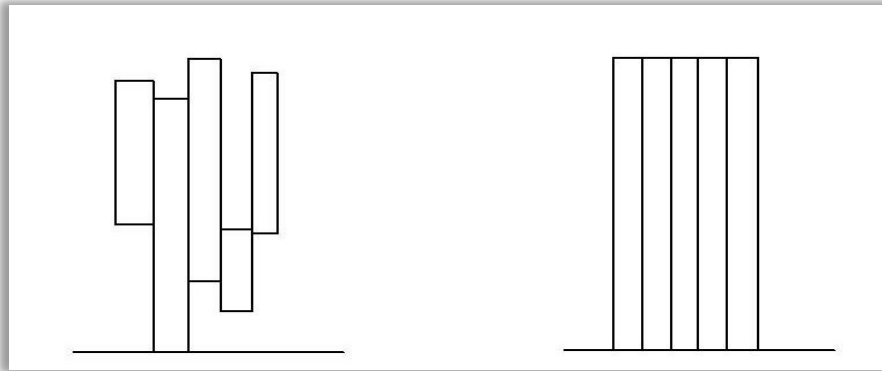
$$x_{sc} = x_i - ST.DEV.$$

$$x_{sc} = \frac{x_i}{\sum_{i=1}^n x_i}$$

$$x_{sc} = \frac{x_i}{x_{max}}$$

όπου x_{sc} η κανονικοποιημένη τιμή της φασματικής απορρόφησης, x_i η αρχική τιμή απορρόφησης ενός δείγματος σε συγκεκριμένο μήκος κύματος, ST.DEV η τυπική απόκλιση, x_{min} και x_{max} η ελάχιστη και η μέγιστη τιμή φασματικής απορρόφησης του

συνόλου των δειγμάτων σε αυτό το μήκος κύματος. Η κανονικοποίηση των φασματικών δεδομένων εφαρμόζεται συνήθως στις περιπτώσεις της δημιουργίας μοντέλων πρόβλεψης και γενικά αλγορίθμων συσχέτισης, διότι προσδίδει ίση «αξία» στα φασματικά δεδομένα κάθε μεταβλητής της απορρόφησης, δηλαδή σε κάθε συγκεκριμένο μήκος κύματος.



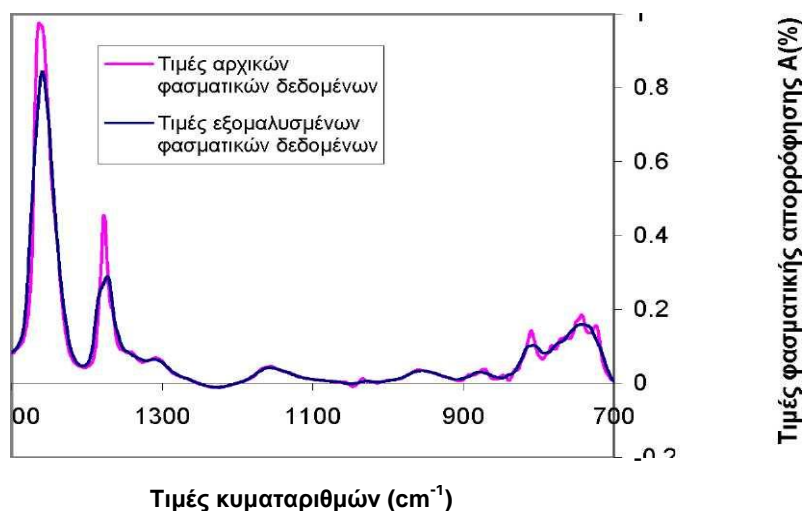
Σχημα 4.1: Σχηματικό διάγραμμα μεταβολής εύρους διακύμανσης δεδομένων που έχουν υποστεί κανονικοποίηση.

Η εξομάλυνση (smoothing) των φασματικών δεδομένων έχει σαν στόχο την εξασθένηση των απότομων μεταβολών στις τιμές των φασματικών δεδομένων. Τα συνήθη φίλτρα εξομάλυνσης είναι του *κινητού μέσου όρου* (moving average), το *Savitsky-Golay*, τα *φίλτρα πολωνύμων* κ.α[9]. Με αυτά επιτυγχάνεται η εξασθένηση του θορύβου των φασματικών δεδομένων, καθώς οι αποκόβονται οι μικρές τυχαίες και απότομες μεταβολές της φασματικής απορρόφησης, που συνήθως συνιστούν θόρυβο.

Χαρακτηριστικά αναφέρεται το φίλτρο του κινητού μέσου, όπου η τιμή της φασματικής απορρόφησης σε μία περιοχή του φάσματος προκύπτει από το μέσο όρο των τιμών της φασματικής απορρόφησης ενός αριθμού τιμών πλησίον και εκατέρωθεν της δεδομένης τιμής. Για ένα φίλτρο κινητού μέσου όπου το οποίο λαμβάνει τις n τιμές εκατέρωθεν της ενδιαφερόμενης φασματικής απορρόφησης x , η τιμή της εξομαλυνθείσας φασματικής απορρόφησης σε μία περιοχή θα δίνεται από τον τύπο:

$$x_s = \frac{x_{1-n} + x_{1-(n-1)} + \dots + x_{i-1} + x_i + x_{i+1} + \dots + x_{i+(n-1)} + x_{i+n}}{2n}$$

Παράδειγμα εξομάλυνσης από την εφαρμογή ενός φίλτρου δίνεται στο παρακάτω σχήμα:



Σχήμα 4.2: Γραφικό αποτέλεσμα εξομάλυνσης φασματικών δεδομένων

Η αριθμητική παραγωγή (derivation) των φασματικών δεδομένων ενισχύει τις μεταβολές των τιμών τους. Συνήθως, χρησιμοποιούνται η 1^η και η 2^η παράγωγος. Όσο μεγαλώνει η τάξη της παραγώγου, τόσο ενισχύονται οι μεγάλες μεταβολές των τιμών των φασματικών δεδομένων σε σχέση με τις μικρότερες. Με αυτό τον τρόπο επιτυγχάνεται μείωση του θορύβου και παράλληλα ενίσχυση του σήματος.

Με την αριθμητική ολοκλήρωση (integration) αθροίζονται οι φασματικές τιμές των δεδομένων σε μικρές περιοχές του φάσματος και έτσι επιτυγχάνεται η ενίσχυση των κορυφών του φάσματος και του λόγου σήματος προς θόρυβο.

Η ανάλυση κυρίων συνιστωσών PCA (principal component analysis) αποτελεί μια πολυμεταβλητή τεχνική ανάλυσης δεδομένων. Με την ανάλυση κυρίων συνιστωσών είναι δυνατόν, η πληροφορία, η οποία είναι συγκεντρωμένη στις φασματικές περιοχές, να συγκεντρωθεί σε λίγες μόνο μεταβλητές. Συχνά

χρησιμοποιείται στην εξαγωγή συσχετίσεων για την κατασκευή μοντέλων πρόβλεψης σε συνδυασμό με άλλες πολυμεταβλητές μεθόδους. Η ανάλυση κυρίων συνιστωσών θα παρουσιαστεί εκτενώς σε επόμενο κεφάλαιο[17].

Στην περίπτωση της παρούσας εργασίας επειδή τα φασματικά δεδομένα είχαν γενικά χαμηλά επίπεδα θορύβου, από τις παραπάνω μεθόδους χρησιμοποιήθηκαν μόνο η κανονικοποίηση του φάσματος καθώς και η ανάλυση των κυρίων συνιστωσών κατά την δημιουργία των μοντέλων πρόβλεψης.

4.4. Πολυπαραμετρικές μέθοδοι επεξεργασίας δεδομένων

4.4.1. Εισαγωγή

Οι πολυμεταβλητές τεχνικές ανάλυσης αποτελούν έναν ταχύτατα εξελισσόμενο τομέα ανάλυσης δεδομένων. Αναφέρονται σε όλες τις στατιστικές μεθόδους που αναλύουν συγχρόνως πολλαπλές μετρήσεις ενός αντικειμένου έρευνας. Πολλές από τις τεχνικές αυτές αποτελούν επέκταση των μονομεταβλητών (ανάλυση της κατανομής μιας μεταβλητής) και των διμεταβλητών τεχνικών ανάλυσης (ανάλυση της διασποράς και της απλής παλινδρόμησης δυο μεταβλητών). Για παράδειγμα η ανάλυση της απλής παλινδρόμησης για μια ανεξάρτητη μεταβλητή μπορεί να επεκταθεί σε επίπεδο πολυμεταβλητής ανάλυσης, ώστε να περιλαμβάνει πολλές ανεξάρτητες μεταβλητές. Ωστόσο υπάρχουν και πολυμεταβλητές τεχνικές που έχουν αναπτυχθεί αποκλειστικά για ανάλυση πολλών μεταβλητών, όπως η ανάλυση κυρίων συνιστωσών (principal component analysis) και η διαχωριστική ανάλυση πολλών μεταβλητών (discriminant analysis).

Οι κυριότερες πολυμεταβλητές τεχνικές ανάλυσης οι οποίες εφαρμόζονται στην επεξεργασία δεδομένων για την κατασκευή μοντέλων πρόβλεψης ιδιοτήτων είναι η πολλαπλή γραμμική παλινδρόμηση MLR (Multiple Linear Regression), η μέθοδος των μερικών ελαχίστων τετραγώνων PLS (Partial Least Squares) και η παλινδρόμηση κυρίων συνιστωσών PCR (Principal Component Regression).

Στο κεφάλαιο αυτό θα γίνει περιγραφή των παραπάνω τεχνικών, καθώς αυτές χρησιμοποιούνται κυρίως στην κατασκευή των μοντέλων πρόβλεψης ιδιοτήτων ενώ δύο από αυτές θα χρησιμοποιηθούν στην εργασία αυτή. Παράλληλα θα

παρουσιασθεί η μέθοδος ανάλυσης κυρίων συνιστωσών PCA (Principal Component Analysis), η οποία αποτελεί την βάση στην οποία στηρίζονται τόσο η παλινδρόμηση των κυρίων συνιστωσών PCR, όσο και μέθοδος των μερικών ελαχίστων τετραγώνων PLS.

Για να κατανοηθούν καλύτερα οι πολυμεταβλητές τεχνικές που θα παρουσιαστούν, θα γίνει αναφορά διαφόρων βασικών χαρακτηριστικών και μεγεθών της πολυμεταβλητής ανάλυσης.

4.4.2 Σημασία της πολυμεταβλητής ανάλυσης και βασικές έννοιες

Στους περισσότερους κλάδους των επιστημών καταγράφονται δεδομένα τα οποία παράγονται κατά τη μελέτη διαφόρων φαινομένων. Τα δεδομένα αυτά στην πλειονοψηφία τους αποτελούν παρατηρήσεις διαφόρων χαρακτηριστικών -μεταβλητών (variables) που έχουν γίνει σε μια ομάδα αντικειμένων-δειγμάτων (samples). Το κάθε αντικείμενο παρακολουθείται για ένα ή περισσότερα από ένα χαρακτηριστικά του. Συνήθως, σαν αποτέλεσμα αυτών των παρατηρήσεων παράγεται ένας μεγάλος όγκος δεδομένων που είναι δύσκολο να εκτιμηθούν και να εξαχθούν συμπεράσματα από αυτά. Ένας απλός τρόπος για την εξαγωγή πληροφοριών από τα δεδομένα αυτά είναι η επεξεργασία κάθε μεταβλητής ξεχωριστά. Αυτός ο τρόπος χαρακτηρίζεται ως μονομεταβλητή ανάλυση και δεν είναι ικανός να αποκαλύψει το σύνολο των πληροφοριών που κρύβονται σε δεδομένα με πολλαπλές μεταβλητές. Σε αυτή την περίπτωση, η πολυμεταβλητή ανάλυση έρχεται να επιλύσει το πρόβλημα καθώς το αντικείμενο της είναι η διερεύνηση της σχέσης που υπάρχει ανάμεσα σε πολλές μεταβλητές. Με αυτή είναι δυνατή η επεξεργασία πολλών, συγχρόνως, μεταβλητών μιας ομάδας αντικειμένων και η μελέτη των μεταξύ τους σχέσεων[20]. Χρησιμοποιείται τόσο για την ανακάλυψη σχέσεων μεταξύ των μεταβλητών, όσο και για την επιβεβαίωση ή την ποσοτικοποίηση των σχέσεων αυτών.

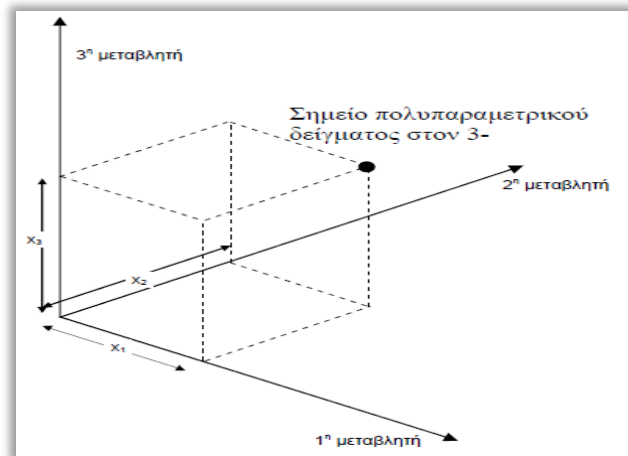
Για να γίνει καλύτερη η κατανόηση της πολυμεταβλητής ανάλυσης θα πρέπει να δοθεί η έννοια του πολυμεταβλητού δείγματος.

Έστω ότι μετρούνται οι τιμές m μεταβλητών ενός δείγματος. Οι τιμές αυτές των μεταβλητών μπορούν να γραφούν με την μορφή ενός πίνακα γραμμής ως:

$$X=[x_1 \ x_2 \ \dots \ x_m]$$

όπου κάθε τιμή x_i αποτελεί την τιμή μιας μεταβλητής του δείγματος.

Αυτός ο πίνακας γραμμής του οποίου κάθε στοιχείο είναι η τιμή μίας μεταβλητής ενός δείγματος αποτελεί ένα πολυμεταβλητό δείγμα. Αν θεωρήσουμε ένα m -διάστατο χώρο του οποίου κάθε διάσταση αντιπροσωπεύει και μια μεταβλητή, τότε το πολυμεταβλητό δείγμα μπορεί να παρασταθεί γραφικά ως ένα σημείο του οποίου οι συντεταγμένες του ως προς κάθε άξονα του συστήματος συντεταγμένων του χώρου, δίνονται από την τιμή της μεταβλητής που εκφράζει αυτός ο άξονας.

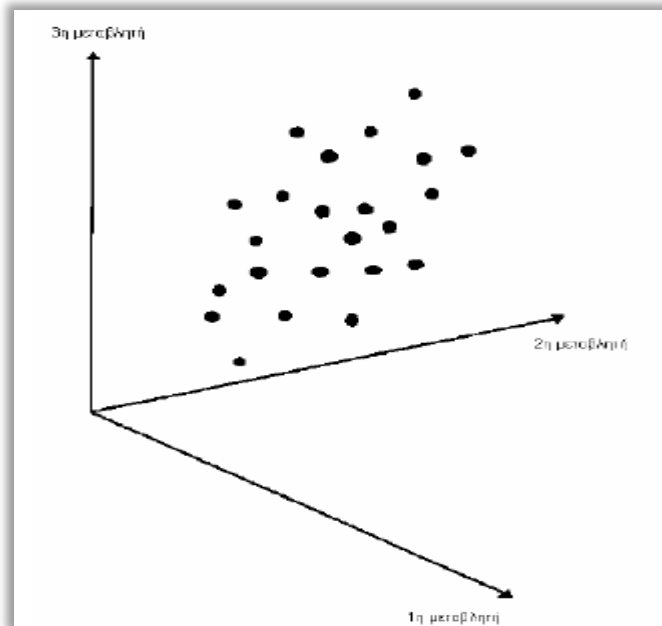


Σχήμα 4.3: Γραφική απεικόνιση ενός πολυμεταβλητού δείγματος στο τρισδιάστατο χώρο.

Αν αντί του ενός δείγματος έχουμε n το πλήθος δείγματα, για τα οποία είναι γνωστές οι τιμές των m μεταβλητών τους, τότε οι τιμές των μεταβλητών μπορούν να οργανωθούν σε ένα πίνακα X διαστάσεων $n \times m$:

$$X_{n \times m} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix}$$

όπου κάθε γραμμή του πίνακα αντιπροσωπεύει ένα δείγμα και κάθε στήλη μία μεταβλητή. Κάθε γραμμή του πίνακα X αποτελεί και ένα πολυμεταβλητό δείγμα, ενώ κάθε στήλη του πίνακα X αποτελεί και μία διάσταση του m -διάστατου χώρου. Συνεπώς ο πίνακας X συνολικά παρίσταται ως ένα σύνολο σημείων στον m -διάστατο χώρο.



Σχήμα 4.4 : Γραφική απεικόνιση ενός συνόλου πολυμεταβλητών δειγμάτων στο τρισδιάστατο χώρο.

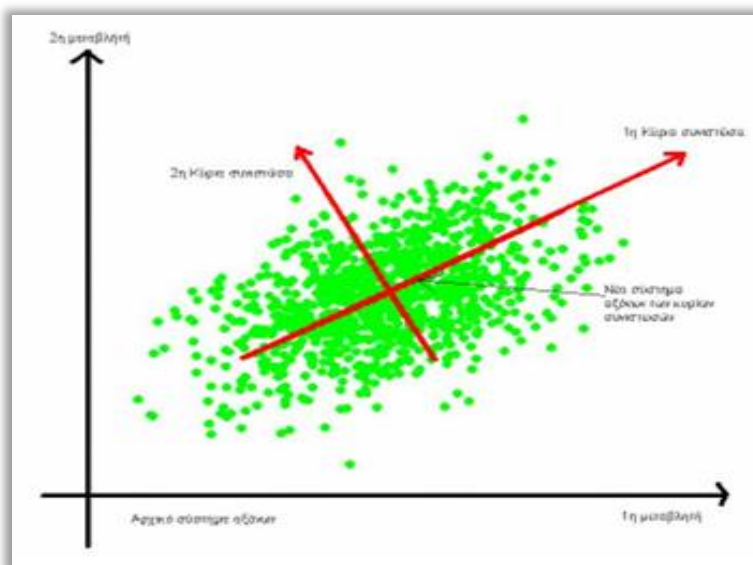
Για κάθε δείγμα που προστίθεται στον πίνακα X αυξάνουν οι γραμμές του, και για κάθε μεταβλητή αυξάνουν οι στήλες του και επομένως οι διαστάσεις του χώρου. Γίνεται σαφές πως για ένα μεγάλο πλήθος δειγμάτων και ιδιοτήτων ο πίνακας X των δεδομένων γίνεται πολύ μεγάλος και η επεξεργασία αυτού του όγκου των δεδομένων πολύ δύσκολη.

Με την πολυμεταβλητή ανάλυση ο όγκος αυτός των δεδομένων είναι δυνατόν να αναλυθεί και αναδειχθεί η πληροφορία που περιέχει. Ανακαλύπτοντας τους βαθμούς συσχέτισης μεταξύ των διαφόρων μεταβλητών, με διάφορες πολυμεταβλητές τεχνικές, είναι δυνατόν να δημιουργηθούν συναρτήσεις που εκφράζουν τις τιμές ορισμένων μεταβλητών σαν συνάρτηση κάποιων άλλων. Ακόμα, μια άλλη δυνατότητα που προσφέρει η ανάλυση αυτή είναι η συρρίκνωση ενός

πίνακα δεδομένων, που μπορεί να γίνει διατηρώντας όσο το δυνατόν μεγαλύτερο μέρος της αρχικής πληροφορίας, που αυτός περιέχει (PCA) .

4.4.3. Ανάλυση Κυρίων Συνιστωσών PCA (Principal Component Analysis)

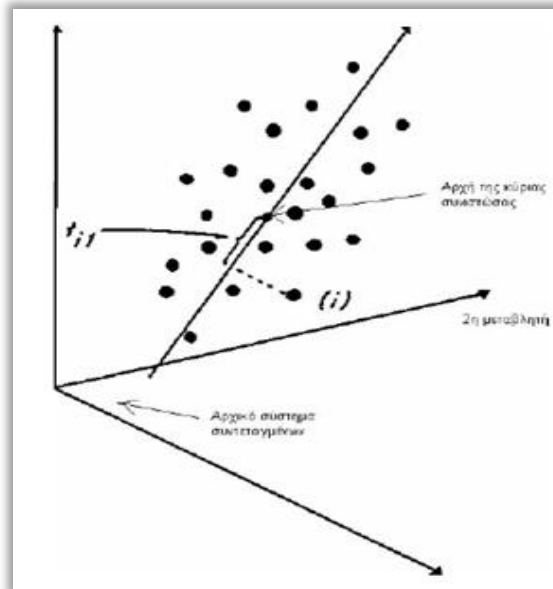
Η Ανάλυση Κυρίων Συνιστωσών είναι μια πολυμεταβλητή μέθοδος στατιστικής επεξεργασίας δεδομένων, με την οποία επιχειρείται ο προσδιορισμός των τυχόν υφιστάμενων γραμμικών συσχετίσεων-τάσεων που υπάρχουν σε ένα μεγάλο αριθμό μεταβλητών και η ερμηνεία τους. Με την τεχνική αυτή επιχειρείται να γίνει συμπίκνωση της πληροφορίας που περιέχεται στις αρχικές μεταβλητές, σε ένα μικρότερο αριθμό νέων μεταβλητών, που είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών, με όσο το δυνατόν μικρότερη απώλεια πληροφορίας[25]. Αυτό επιτυγχάνεται με την προβολή των δεδομένων σε ένα νέο σύστημα αξόνων, οι οποίοι προέρχονται από γραμμικό μετασχηματισμό των αρχικών μεταβλητών. Οι γραμμικοί συνδυασμοί εξάγονται κατά τέτοιο τρόπο, ώστε να είναι ασυσχέτιστοι μεταξύ τους και ο καθένας από αυτούς να εκφράζει όσο το δυνατόν μεγαλύτερο ποσοστό της διασποράς των προηγούμενων μεταβλητών.



Σχήμα 4.5: Γραφική απεικόνιση μετασχηματισμού κυρίων συνιστωσών.

Ο βασικός σκοπός της ανάλυσης των κυρίων συνιστωσών PCA είναι η

μείωση των διαστάσεων ενός συνόλου δεδομένων, που αποτελείται από μεγάλο αριθμό αλληλοεξαρτώμενων μεταβλητών, ενώ συγκρατείται η μέγιστη δυνατή διασπορά τους. Αυτό γίνεται με μετασχηματισμό των δεδομένων σε μια καινούργια ομάδα μεταβλητών, τις κύριες συνιστώσες, οι οποίες δεν συσχετίζονται μεταξύ τους και ταξινομούνται έτσι ώστε, οι πρώτες κύριες συνιστώσες να έχουν όσο το δυνατόν περισσότερη διασπορά από το σύνολο.



Σχήμα 4.6: Γραφική απεικόνιση ενός μετασχηματισμού κυρίων συνιστωσών απο τρισδιάστατο αρχικό σύστημα συντεταγμένων προς το σύστημα μίας κύριας συνιστώσας.

Η ανάλυση κυρίων συνιστωσών αναλύει ένα αρχικό πίνακα δεδομένων X σε ένα γινόμενο δυο πινάκων R και C , δηλαδή:

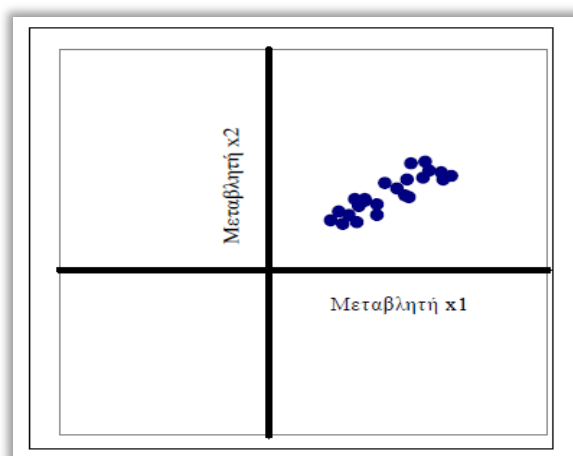
$$X=RC$$

όπου :

C είναι ο πίνακας των loadings factors, του οποίου κάθε γραμμή αποτελεί μία κύρια συνιστώσα, ή αλλιώς, είναι ο πίνακας του οποίου κάθε γραμμή περιέχει τα συνημίτονα κατεύθυνσης μιας κύριας συνιστώσας.

R είναι ο πίνακας των scores factors, του οποίου κάθε στήλη περιέχει τις συντεταγμένες κάθε πολυπαραμετρικού δείγματος δεδομένων πάνω στην γραμμή της κύριας συνιστώσας.

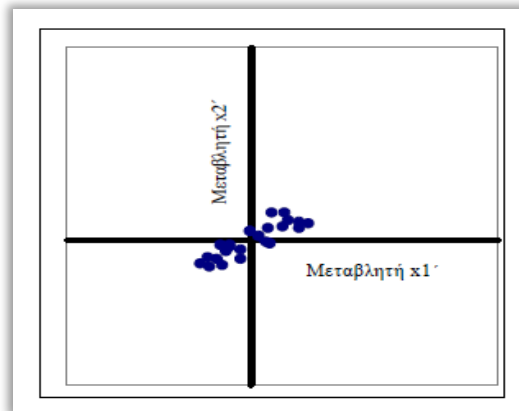
Για να δοθεί μια καλύτερη εικόνα της PCA, θεωρούμε έναν πίνακα δεδομένων, ο οποίος προκύπτει από τις τιμές m μεταβλητών, που έχουν n πολυμεταβλητά δείγματα. Για παράδειγμα οι μεταβλητές θα μπορούσαν να είναι οι m ιδιότητες που μετρούνται σε ένα χημικό δείγμα. Αυτός ο πίνακας δεδομένων έχει διαστάσεις $n \times m$ και κάθε στήλη αυτού αντιπροσωπεύει μία ιδιότητα, ενώ κάθε γραμμή ένα δείγμα. Τα δείγματα του πίνακα μπορούν να απεικονιστούν σε ένα m -διάστατο χώρο, στον οποίο κάθε διάσταση αντιπροσωπεύει μία μεταβλητή, σαν σημεία που οι συντεταγμένες τους από τους άξονες που αντιπροσωπεύουν τις μεταβλητές, είναι οι τιμές των αντιστοιχών μεταβλητών. Αν τα δείγματα που χρησιμοποιούνται είναι σχετικά μεταξύ τους, τότε οι τιμές των μεταβλητών τους θα είναι παραπλήσιες και στον m -διάστατο χώρο θα απεικονίζονται από ένα σμήνος σημείων[29]. Αν θεωρήσουμε ένα επίπεδο του χώρου αυτού, το οποίο ορίζεται από δύο άξονες που αντιπροσωπεύουν δυο μεταβλητές π.χ. x_1 , x_2 , τότε το σύνολο των δειγμάτων θα αντιπροσωπεύεται από ένα σύνολο σημείων, υπό μορφή σμήνους, τα οποία αποτελούν τις προβολές των σημείων των δειγμάτων από τον m -διάστατο χώρο στο επίπεδο αυτό.



Σχήμα 4.7: Γραφική απεικόνιση ενός συνόλου δειγμάτων στο επίπεδο που ορίζουν 2 μεταβλητές

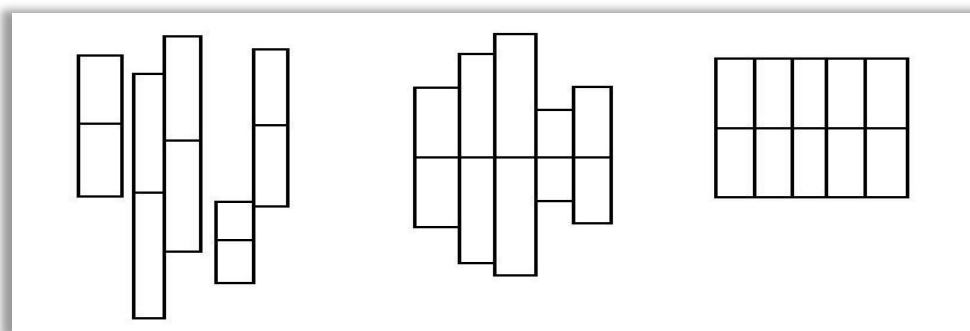
Πριν την εφαρμογή της PCA στον αρχικό πίνακα δεδομένων, θα πρέπει να γίνει μια προεπεξεργασία των αρχικών δεδομένων. Αρχικά από την τιμή της μεταβλητής κάθε στήλης δεδομένων, αφαιρείται ο μέσος όρος των τιμών της στήλης. Με αυτόν τον τρόπο, το σύστημα αξόνων του m -διάστατου χώρου, μεταφέρεται στο κέντρο του σμήνους των σημείων. Αυτή η διαδικασία λέγεται μεσοστάθμιση ή

αλλιώς mean-centering και στο επίπεδο των δυο μεταβλητών η διαδικασία αυτή παριστάνεται από το σχήμα :



Σχήμα 4.8: Γραφική απεικόνιση μεσοσταθμισμένου συνόλου δειγμάτων.

Ο νέος πίνακας των δεδομένων που δημιουργείται, έχει σαν στοιχεία του τις σταθμισμένες ως προς τον μέσο όρο τιμές των μεταβλητών. Στην συνέχεια αυτές οι νέες σταθμισμένες τιμές των μεταβλητών κάθε στήλης, του νέου πίνακα, διαιρούνται με την τυπική απόκλιση του συνόλου των τιμών της στήλης, για να γίνει η διασπορά των τιμών των μεταβλητών κάθε στήλης, ίση με την μονάδα. Με τον τρόπο αυτό εξισώνεται η «συμμετοχή» κάθε μεταβλητής στον προσδιορισμό της θέσης του κάθε δείγματος στον m-χώρο. Αυτή η διαδικασία λέγεται κανονικοποίηση ως προς την διασπορά ή normalization. Με την τελευταία αυτή μετατροπή στον πίνακα των δεδομένων, δημιουργείται ένας νέος πίνακας, κατάλληλος για την εφαρμογή της ανάλυσης κυρίων συνιστωσών. Το σύνολο των διαδικασιών της μεσοστάθμισης και της κανονικοποίησης των τιμών των μεταβλητών ως προς την διασπορά, ονομάζεται autoscaling.



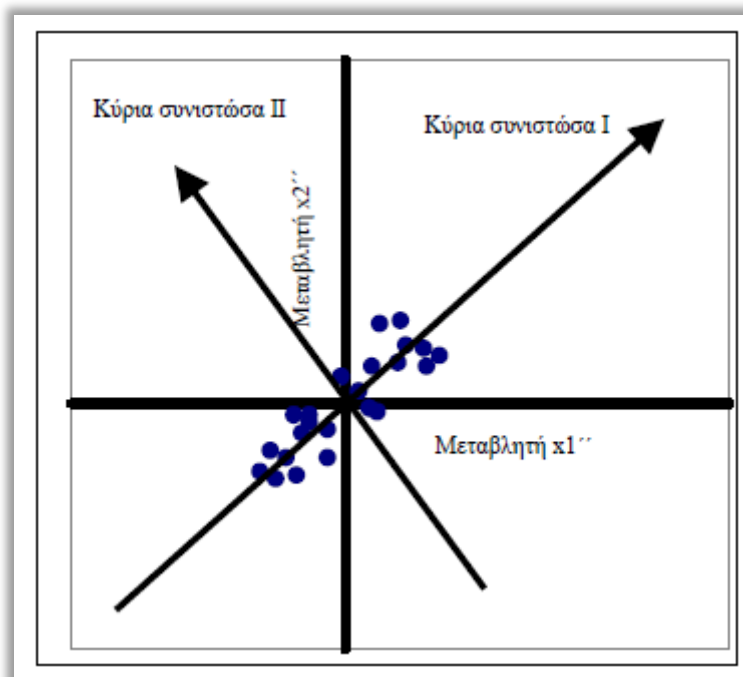
Σχήμα 4.9: Γραφική απεικόνιση της μεταβολής εύρους διακύμανσης δεδομένων με μεσοστάθμιση και autoscaling.

Στον καινούργιο πίνακα, η εφαρμογή της PCA έχει σαν συνέπεια τον μετασχηματισμό του σε ένα γινόμενο πινάκων, δηλαδή :

$$X=RC$$

όπου X ο πίνακας των δεδομένων, R ο πίνακας των scores factors και C ο πίνακας των loadings factors.

Με τον μετασχηματισμό αυτόν το σύστημα αξόνων του m -διάστατου χώρου του οποίου κάθε διάσταση αντιπροσώπευε και μία μεταβλητή, μπορεί να αντικατασταθεί από ένα νέο σύστημα αξόνων το οποίο είναι στραμμένο κατά τις διευθύνσεις των κύριων συνιστωσών. Χαρακτηριστικά βλέπουμε το παράδειγμα των δύο μεταβλητών στο παρακάτω σχήμα:



Σχήμα 4.10: Γραφική απεικόνιση του αποτελέσματος της εφαρμογής της PCA για το παράδειγμα 2 μεταβλητών.

Οι διευθύνσεις των κύριων συνιστωσών δίνονται από τα συνημίτονα κατεύθυνσης, που είναι οι γραμμές του πίνακα των loadings factors, ενώ οι νέες συντεταγμένες των σημείων ως προς τους άξονες των κυρίων συνιστωσών, δίνονται από τον πίνακα των score factors. Ουσιαστικά, με την ανάλυση αυτή του αρχικού πίνακα δεδομένων, γίνεται στροφή του συστήματος συντεταγμένων ως προς τις κατευθύνσεις των κυρίων συνιστωσών. Παρατηρούμε στο δισδιάστατο παράδειγμα

ότι τα περισσότερα δείγματα είναι τοποθετημένα γύρω από τον άξονα που ορίζει η 1η κύρια συνιστώσα και γύρω από τον άξονα που ορίζει η 2η κύρια συνιστώσα. Κάθε γραμμή του πίνακα των score factors περιέχει τις συντεταγμένες του κάθε δείγματος ως προς του άξονες των κυρίων συνιστωσών. Αν αυτό το παράδειγμα επεκταθεί στον m -διάστατο χώρο, τότε αυτό το σμήνος που απαρτίζουν τα δείγματα, δύναται να περιγραφεί από ένα νέο σύστημα αξόνων, το οποίο ορίζουν οι κύριες συνιστώσες. Ο τρόπος με τον οποίο τοποθετούνται οι κύριες συνιστώσες στον χώρο είναι τέτοιος ώστε, να περιγράφεται το σμήνος των δειγμάτων στον χώρο με την μέγιστη δυνατή διασπορά.

Οι κύριες συνιστώσες εξαρτώνται από τις ιδιοτιμές του πίνακα συνδιακύμανσης του αρχικού πίνακα των δεδομένων X , δηλαδή του $COV(X^T X)$. Κάθε ιδιοτιμή αντιστοιχεί σε ένα ιδιοδιάνυσμα, το οποίο περιγράφει μία κύρια συνιστώσα. Επειδή κάθε ιδιοδιάνυσμα αποτελεί γραμμικό συνδυασμό των μεταβλητών του αρχικού πίνακα των δεδομένων, συνεπώς και κάθε κύρια συνιστώσα μπορεί να θεωρηθεί σαν γραμμικός συνδυασμός του αρχικού συστήματος συντεταγμένων, που περιγράφει τον m -διάστατο χώρο. Όσο πιο μεγάλες είναι οι ιδιοτιμές του πίνακα συνδιακύμανσης των αρχικών δεδομένων, τόσο μεγαλύτερη πληροφορία φέρουν οι κύριες συνιστώσες που αντιπροσωπεύουν τα ιδιοδιανύσματα.

Με την ανάλυση του αρχικού πίνακα δεδομένων στο γινόμενο των πινάκων των scores και loadings factors και την περιγραφή των δειγμάτων σαν σημεία του νέου συστήματος συντεταγμένων που ορίζουν οι κύριες συνιστώσες, είναι δυνατόν να γίνει επιλογή των σημαντικότερων κυρίων συνιστωσών, που περιέχουν το μεγαλύτερο ποσοστό πληροφορίας και μπορούν να περιγράψουν κατά το καλύτερο δυνατό το σμήνος των δειγμάτων, χωρίς να υπάρξει σημαντική απώλεια πληροφορίας. Οι κύριες συνιστώσες που θα επιλεγούν εκ του συνόλου αυτών, θα είναι λιγότερες τόσο από τον αριθμό των m αρχικών μεταβλητών, όσο και από το σύνολο των κυρίων συνιστωσών. Παράλληλα, θα πρέπει να περιέχουν αθροιστικά ένα ικανοποιητικό και όσο το δυνατόν μεγαλύτερο ποσοστό της αρχικής πληροφορίας. Το ποσοστό αυτό της αρχικής πληροφορίας προκύπτει από τον βέλτιστο συνδυασμό μεταξύ του αριθμού των κυρίων συνιστωσών και του αθροιστικού ποσοστού πληροφορίας, που φέρουν οι κύριες συνιστώσες που θα επιλεγούν. Από τον συνδυασμό αυτόν, επιδιώκεται να περιγραφεί με όσο το δυνατόν λιγότερες κύριες συνιστώσες, όσο μεγαλύτερο ποσοστό πληροφορίας γίνεται.

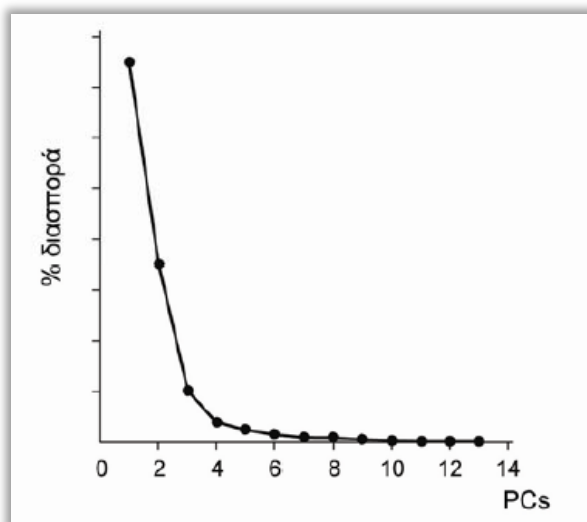
Η επιλογή των κυρίων αυτών συνιστωσών που θα περιγράψουν τον αρχικό πίνακα δεδομένων, είναι δυνατόν να γίνει με την βοήθεια διαφόρων κριτηρίων. Τα σημαντικότερα κριτήρια που χρησιμοποιούνται για το σκοπό αυτό είναι :

- Το κριτήριο των ιδιοτιμών
- Το κριτήριο του ποσοστού διασποράς
- Το κριτήριο του Scree test

Με βάση το **κριτήριο των ιδιοτιμών** ως κύριες συνιστώσες επιλέγονται αυτές, που οι ιδιοτιμές των ιδιοδιανυσμάτων τους έχουν τιμή μεγαλύτερη από την μονάδα

Το **κριτήριο του ποσοστού διασποράς** λαμβάνει ως κύριες συνιστώσες αυτές, που το αθροιστικό ποσοστό της διασποράς των αρχικών δεδομένων, η αλλιώς το ποσοστό της πληροφορίας των αρχικών δεδομένων που αυτές φέρουν, είναι ικανοποιητικό. Δεν υπάρχει κάποια συγκεκριμένη τιμή για το ποσοστό αυτό, αλλά γενικά, στις φυσικές επιστήμες το ποσοστό διασποράς πάνω από 95%, θεωρείται ικανοποιητικό.

Το **κριτήριο του scree test** χρησιμοποιείται για να βρεθεί ο βέλτιστος αριθμός κυρίων συνιστωσών. Σύμφωνα με το κριτήριο αυτό, σχηματίζεται η καμπύλη του διαγράμματος των ιδιοτιμών των ιδιοδιανυσμάτων συναρτήσει των κυρίων συνιστωσών και από την μορφή της επιλέγονται οι κύριες συνιστώσες που θα χρησιμοποιηθούν. Συνήθως, η καμπύλη αυτή αρχίζοντας από τον πρώτο παράγοντα, κατεβαίνει απότομα προς τα κάτω αρχικά, και μετά γίνεται σχεδόν οριζόντια γραμμή. Το σημείο στο οποίο η καμπύλη γίνεται οριζόντια γραμμή θεωρείται ότι δίνει τον μέγιστο αριθμό κυρίων συνιστωσών, που πρέπει να χρησιμοποιηθούν. Από εκεί και πέρα δεν θεωρείται ότι υπάρχει μεγάλη διασπορά.



Σχήμα 4.11: Συμπεριφορά RMSEP με αύξηση των παραγόντων

Στην πράξη συνήθως δεν χρησιμοποιείται μόνο ένα κριτήριο επιλογής κυρίων συνιστωσών, αλλά περισσότερα. Συνήθως, χρησιμοποιείται αρχικά ένα κριτήριο, όπως αυτό των ιδιοτιμών για να γίνει πρώτα η ανάλυση. Έπειτα γίνονται δοκιμές προσθέτοντας ή αφαιρώντας κύριες συνιστώσες και επιλέγεται εκείνος ο αριθμός των κυρίων συνιστωσών, ο οποίος αναπαριστάνει καλύτερα τις σχέσεις μεταξύ των αρχικών μεταβλητών.

Η PCA αποτελεί μία πολύ σημαντική πολυμεταβλητή τεχνική. Με την ικανότητα που έχει να μετατρέπει ένα μεγάλο όγκο δεδομένων, σε έναν σημαντικά μικρότερο και πολύ πιο εύκολα επεξεργάσιμο, παρέχει την δυνατότητα στους ερευνητές να επεξεργάζονται έναν μεγάλο αριθμό δεδομένων και μετρήσεων και να ανακαλύπτουν συσχετίσεις σε αυτά.

4.4.4 Πολλαπλή Γραμμική Παλινδρόμηση

Η μέθοδος της πολλαπλής γραμμικής παλινδρόμησης (Multiple Linear Regression, MLR)) αποτελεί μία από τις απλούστερες και συνηθέστερα εφαρμοζόμενες πολυμεταβλητές μεθόδους στατιστικής επεξεργασίας για ανάπτυξη γραμμικών μοντέλων πρόβλεψης ιδιοτήτων[3]. Αποτελεί στην πραγματικότητα επέκταση, για περισσότερες της μίας ανεξάρτητες μεταβλητές, της μονομεταβλητής πολυωνυμικής παλινδρόμησης των ελαχίστων τετραγώνων. Για πολυμεταβλητό

πληθυσμό n δειγμάτων που εξετάζονται ως προς m ανεξάρτητες μεταβλητές X_j με $j = 1, 2, \dots, m$ και μία εξαρτημένη μεταβλητή Y , το γραμμικό μοντέλο της πολλαπλής γραμμικής παλινδρόμησης δίνεται από την σχέση:

$$y = X \cdot \beta + \varepsilon$$

όπου y το διάνυσμα της μεμονωμένης εξαρτημένης μεταβλητής Y και με στοιχεία τις μετρούμενες τιμές-στόχους.

β το διάνυσμα των συντελεστών των ανεξάρτητων μεταβλητών του πολωνύμου προσέγγισης

και ε το διάνυσμα του υπολειπόμενου σφάλματος του πολωνύμου προσέγγισης

Η εξαρτημένη μεταβλητή αποτελεί γραμμικό συνδυασμό των m ανεξάρτητων μεταβλητών με

συντελεστές πολωνύμου που προσδιορίζονται από την σχέση:

$$\beta = (X^T \cdot X)^{-1} \cdot X^T \cdot y$$

4.4.5 Παλινδρόμηση Κυρίων Συνιστωσών PCR (Principal Component Regression)

Η παλινδρόμηση κυρίων συνιστωσών PCR αποτελεί συνδυασμό των δύο προαναφερθέντων πολυμεταβλητών τεχνικών, δηλαδή της πολλαπλής γραμμικής παλινδρόμησης MLR και της ανάλυσης κυρίων συνιστωσών PCA. Συνδυάζει την δυνατότητα της PCA να μειώνει τον όγκο των δεδομένων με αυτήν της MLR να εξασφαλίζει γραμμικές συσχετίσεις μεταξύ των δεδομένων και των μεταβλητών με απλό και άμεσο τρόπο.

Αναλυτικότερα, ένας πίνακας δεδομένων, με την εφαρμογή της PCA, δύναται να μετασχηματιστεί σε ένα γινόμενο δυο πινάκων, αυτού των scores factors και αυτού των loadings factors. Δεδομένου ότι ο πίνακας των loadings factors αντιπροσωπεύει τις κύριες συνιστώσες, καθώς οι στήλες του είναι τα συνημίτονα κατεύθυνσης των κύριων συνιστωσών, ο πίνακας των scores factors, ουσιαστικά,

αποτελεί τον πίνακα που αντιπροσωπεύει τα αρχικά δεδομένα, ή αλλιώς, τον πίνακα που φέρει την πληροφορία των αρχικών δεδομένων.

Αυτός ο πίνακας των score factors δύναται να συσχετιστεί γραμμικά με άλλα δεδομένα ή μεταβλητές με την βοήθεια της MLR. Συγκεκριμένα, τα στοιχεία ή οι τιμές των μεταβλητών του πίνακα των score factors μπορούν να αποτελέσουν τον πίνακα των ανεξαρτήτων μεταβλητών που μπορούν να συσχετιστούν γραμμικά με τα στοιχεία ή τις τιμές των μεταβλητών ενός άλλου πίνακα, που θεωρούνται ως εξαρτημένες μεταβλητές και ο πίνακας τους, ως πίνακας των εξαρτημένων μεταβλητών. Ο τύπος που δίνει την συσχέτιση μεταξύ του πίνακα των εξαρτημένων και αυτού των ανεξαρτήτων μεταβλητών είναι:

$$Y=RB+E$$

όπου R είναι ο πίνακας των score factors του αρχικού πίνακα δεδομένων, B είναι ο πίνακας των συντελεστών βάρους και E ο πίνακας των υπολοίπων ή των σφαλμάτων.

Από την παραπάνω εξίσωση προκύπτει ότι οι εξαρτημένες μεταβλητές ενός πίνακα μπορούν να εκφραστούν σαν γραμμικός συνδυασμός των μεταβλητών του πίνακα των score factors, που παίζουν τον ρόλο των ανεξαρτήτων μεταβλητών.

Για να εξασφαλιστεί η βέλτιστη γραμμική συσχέτιση μεταξύ των εξαρτημένων μεταβλητών y_i και των στοιχείων του πίνακα των scores factors r_i , αναζητούνται οι κατάλληλοι συντελεστές b_i , που θα συντελέσουν σε αυτό. Σε αυτή την βέλτιστη συσχέτιση μεταξύ ανεξαρτήτων και εξαρτώμενων μεταβλητών, θα πρέπει ο πίνακας των υπολοίπων E στην εξίσωση 4.32 να τείνει ή να είναι μηδενικός, δηλαδή $E=0$. Στην περίπτωση αυτή ο πίνακας B, που εξασφαλίζει την αυτή την βέλτιστη συσχέτιση, δίνεται από την εξίσωση:

$$B=(R^TR)^{-1}R^TY$$

όπου R^T ο ανάστροφος του πίνακα των scores factors

Με την εύρεση του B η συσχέτιση που επιτυγχάνεται μεταξύ των εξαρτημένων μεταβλητών και του πίνακα των scores factors, δίνεται από την εξίσωση:

$$Y=RB+E$$

Η PCR έχει το πλεονέκτημα ότι λύνει τα προβλήματα συγγραμμικότητας που τυχόν υφίσταται στα αρχικά δεδομένα και τα οποία δεν είναι δυνατόν να αντιμετωπιστούν με την MLR. Αυτό γίνεται διότι, μετά την εφαρμογή της PCA στα αρχικά δεδομένα ο πίνακας των scores factors που προκύπτει, περιλαμβάνει «ορθογώνιες» μεταξύ τους μεταβλητές. Παράλληλα, με την απαλοιφή των συνιστωσών με μικρές ιδιοτιμές ιδιοδιανυσμάτων, επιτυγχάνεται και μείωση του θορύβου που αυτές προκαλούν.

Ένα ακόμα πλεονέκτημα της PCR είναι ότι με το να μειώνεται ο όγκος των δεδομένων και των αρχικών μεταβλητών, λόγω της εφαρμογής της PCA, μειώνεται παράλληλα και ο αριθμός των μεταβλητών που απαιτούνται για την κατασκευή των μοντέλων πρόβλεψης των ιδιοτήτων. Έτσι, είναι δυνατόν να κατασκευαστούν πιο σταθερά μοντέλα πρόβλεψης ιδιοτήτων, καθώς απαιτείται μικρότερος αριθμός δειγμάτων για το σκοπό αυτό, και τα μοντέλα που προκύπτουν επιτυγχάνουν πιο γρήγορη και καλύτερη σύγκλιση από αυτά άλλων μεθόδων. Η καλύτερη σύγκλιση επιβεβαιώνεται με τις μικρότερες τιμές σφάλματος, που είναι δυνατόν να παρουσιάζουν τα μοντέλα αυτά σε σχέση με αυτά των άλλων μεθόδων, για τον ίδιο αριθμό δειγμάτων του συνόλου εκπαίδευσης.

Η PCR αποτελεί πολύ σημαντική πολυπαραμετρική τεχνική στην επεξεργασία μεγάλων όγκων δεδομένων, καθώς με την εφαρμογή της PCA, καθίστανται εύκολα διαχειρίσιμοι και δύναται να χρησιμοποιηθούν από την MLR για την εξαγωγή συσχετίσεων.

4.4.6. Μέθοδος των μερικών ελαχίστων τετραγώνων PLS (Partial Least Squares)

Η μέθοδος των μερικών ελαχίστων τετραγώνων PLS αποτελεί μια τεχνική επεξεργασίας δεδομένων, που μοιάζει με την τεχνική της παλινδρόμησης κυρίων συνιστωσών PCR. Κατά την PLS, τα δεδομένα τα οποία περιέχουν περισσότερη πληροφορία τονίζονται περισσότερο από τα δεδομένα που είναι φτωχά σε πληροφορία.

Με την μέθοδο PLS δημιουργούνται μοντέλα που εξασφαλίζουν γραμμικές συσχετίσεις μεταξύ εξαρτημένων και ανεξαρτήτων μεταβλητών. Κατά την συσχέτιση αυτή μεταξύ των εξαρτημένων και των ανεξαρτήτων μεταβλητών, υπολογίζονται με διαδικασίες παλινδρόμησης οι κατάλληλοι συντελεστές των ανεξαρτήτων μεταβλητών, που δίνουν τους βέλτιστους συνδυασμούς συσχέτισης με τις εξαρτημένες[9]. Οι συντελεστές αυτοί υπολογίζονται έτσι ώστε, καθένας από αυτούς να μεγιστοποιεί την συνδιακύμανση μεταξύ των πινάκων των εξαρτημένων και των ανεξαρτήτων μεταβλητών, ή αλλιώς, να μεγιστοποιεί την πληροφορία που μεταφέρεται μεταξύ των εξαρτημένων και των ανεξαρτήτων μεταβλητών. Αν θεωρήσουμε έναν πίνακα εξαρτημένων μεταβλητών Y και έναν πίνακα ανεξαρτήτων μεταβλητών X , τότε κατά την μέθοδο PLS στόχος είναι η κατασκευή ενός μοντέλου που συσχετίζει γραμμικά τις εξαρτημένες με τις ανεξάρτητες μεταβλητές, σύμφωνα με την σχέση:

$$Y=XB$$

Κατά την μέθοδο PLS παράγεται ένας πίνακας συντελεστών βάρους W τέτοιος, ώστε για τον πίνακα των ανεξαρτήτων μεταβλητών X να ισχύει :

$$R=WX.$$

Οι στήλες του W είναι τα διανύσματα στήλης των συντελεστών βάρους, των οποίων το γινόμενο τους με τις στήλες του πίνακα X , παράγουν τον πίνακα των score factors R του X . Αυτοί οι συντελεστές βάρους του πίνακα W υπολογίζονται έτσι ώστε, καθένας από αυτούς να μεγιστοποιεί την συνδιακύμανση μεταξύ του πινάκων των εξαρτημένων και των ανεξαρτήτων μεταβλητών.

Έπειτα, όπως ο πίνακας X των ανεξάρτητων μεταβλητών, έτσι και ο πίνακας των εξαρτημένων μεταβλητών Y μπορεί να γραφτεί ως το γινόμενο των πινάκων R και Q , όπου Q ο πίνακας των loadings factors του Y , δηλαδή :

$$Y=RQ$$

Εάν υπολογιστεί ο Q τότε ο Y γράφεται ως :

$$Y=XB$$

όπου ο πίνακας των συντελεστών δίνεται από την σχέση

$$B=WQ$$

Με τον υπολογισμό του B σχηματίζεται το μοντέλο με το οποίο συσχετίζονται γραμμικά οι πίνακες των εξαρτημένων μεταβλητών με αυτούς των ανεξαρτήτων μεταβλητών. Αν θα θέλαμε να δώσουμε έναν πιο αναλυτικό τύπο για τον πίνακα B θα μπορούσαμε να γράψουμε:

$$B=W(C^TW)^{-1}Q^T$$

όπου C ο πίνακας των loadings factors του X .

Η μέθοδος PLS αποτελεί και αυτή μία πολυμεταβλητή τεχνική που περιλαμβάνει ανάλυση πινάκων και παλινδρόμηση. Παρόλο που είναι πιο δύσκολη η κατανόηση και η επεξήγηση της, είναι πιο γρήγορη από την PCA στην επεξεργασία και την αποσύνθεση των πινάκων. Για τους σκοπούς της πρόβλεψης αποδίδει το ίδιο καλά αποτελέσματα με την PCR .

5. ΑΠΟΤΕΛΕΣΜΑΤΑ

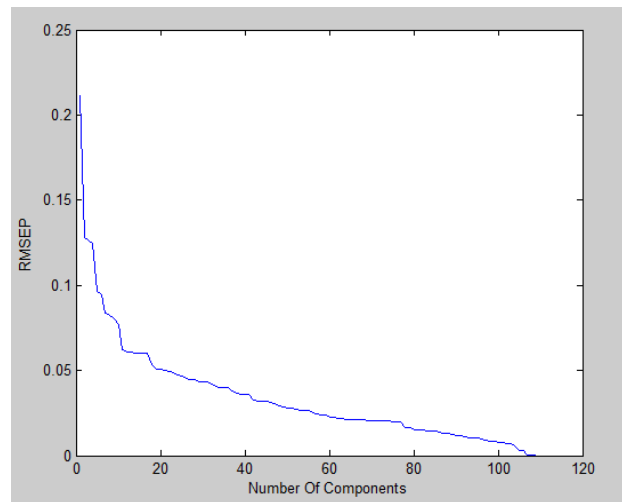
Στο κεφάλαιο αυτό θα παρουσιαστούν αναλυτικά τα χημειομετρικά μοντέλα που αναπτύχθηκαν για την πρόβλεψη της περιεκτικότητας των αρωματικών συστατικών στα δείγματα της παρούσας εργασίας, με βάση τα υπέρυθρα φάσματά τους. Η εκπαίδευση και ο έλεγχος των μοντέλων υλοποιήθηκε σε περιβάλλον Matlab. Στην ανάπτυξη των μοντέλων πρόβλεψης χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS). Συγκεκριμένα αναπτύχθηκαν μοντέλα για επιμέρους ομάδες πετρελαϊκών κλασμάτων, αλλά και για ομάδες δειγμάτων στις οποίες η περιεκτικότητα των αρωματικών είχε προσδιοριστεί με την ίδια αναλυτική μέθοδο. Η ακρίβεια της πρόβλεψης της συγκέντρωσης των αρωματικών, υπολογισμένη με βάση τα δείγματα ελέγχου, στα περισσότερα από τα μοντέλα που αναπτύχθηκαν ήταν συγκρίσιμη εκείνης των πρότυπων εργαστηριακών μεθόδων.

Σε όλα τα μοντέλα έγινε επιλογή ορισμένων τμημάτων του φάσματος, στα οποία υπάρχει έντονη απορρόφηση και προκύπτει πληροφορία για τη χημική σύσταση, ενώ σε ορισμένα από τα μοντέλα πραγματοποιήθηκε προεπεξεργασία των φασματικών δεδομένων. Πιο συγκεκριμένα, κανονικοποίηση ορίζεται η διαδικασία της διαίρεσης κάθε τιμής μεταβλητής ενός δείγματος με το άθροισμα του συνόλου των τιμών – μεταβλητών. Με τον τρόπο αυτό το συνολικό “εμβαδόν” κάθε φάσματος ισούται με τη μονάδα. Με τον όρο κανονικοποίηση ως προς τη μέγιστη τιμή (Pre-max) ορίζεται η διαδικασία διαίρεσης κάθε τιμής μεταβλητής με τη μέγιστη τιμή του κάθε φάσματος [34]. Κάθε μοντέλο δοκιμάζεται με ομάδες, αγνώστων σε αυτό, δειγμάτων τα οποία δηλαδή δεν χρησιμοποιήθηκαν στην ανάπτυξή του για να υπολογιστεί η ικανότητα πρόβλεψής του.

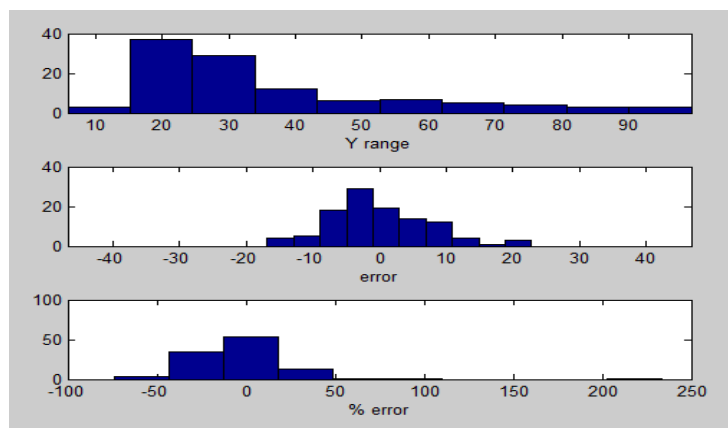
5.1 Μοντέλα με βάση το σύνολο των δειγμάτων

Τα παρακάτω μοντέλα χρησιμοποιούν το σύνολο των δειγμάτων. Σε περιπτώσεις που υπάρχει μέτρηση αρωματικών με παραπάνω από μια μεθόδους επιλέγεται η μέση τιμή τους. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

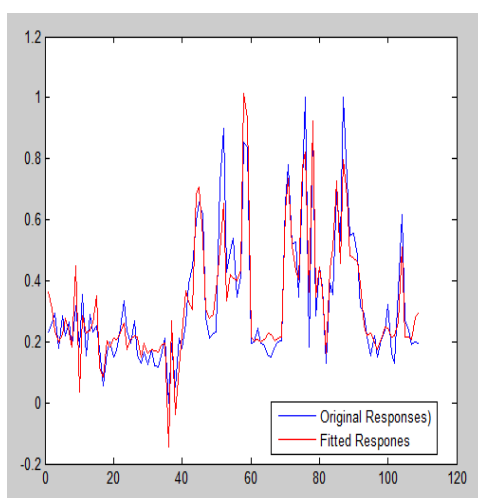
ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ ΜΕ ΜΕΣΕΣ ΤΙΜΕΣ ΑΡΩΜΑΤΙΚΩΝ (PCR1)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	109
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών : 3100 - 2750 cm^{-1} 1800 - 650 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	5.8%
Τυπική απόκλιση	20.51
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	80.64%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.19%
Μέση τιμή απολύτων σφαλμάτων	19.78%
RMSEP	0.08%
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	10
Μέγιστη τιμή αρωματικών(%)	73.2%
Ελάχιστη τιμή αρωματικών(%)	20%
Τυπική απόκλιση	16.44
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	34.37%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	4.1%
Μέση τιμή απολύτων σφαλμάτων	24,25%
RMSEP	0.1



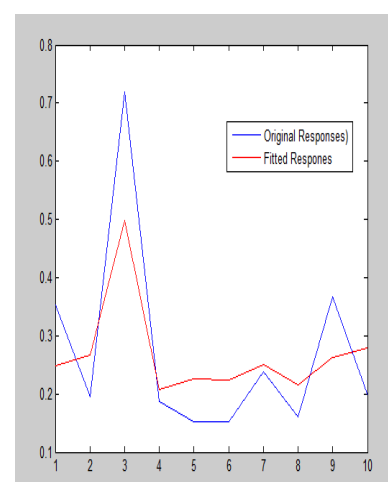
Σχήμα 5.1.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.1.2: Κατανομή ποσοστού αρωματικών και σφάλματος

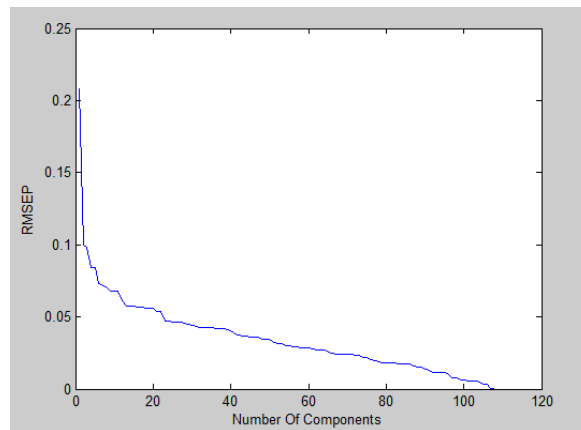


Σχήμα 5.1.3: Δείγματα εκπαίδευσης

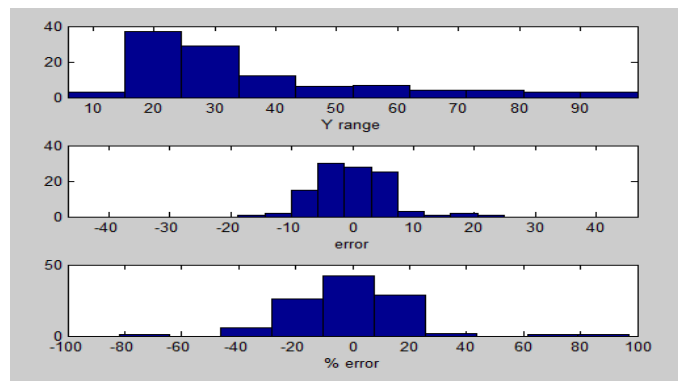


Σχήμα 5.1.4: Δείγματα δοκιμής

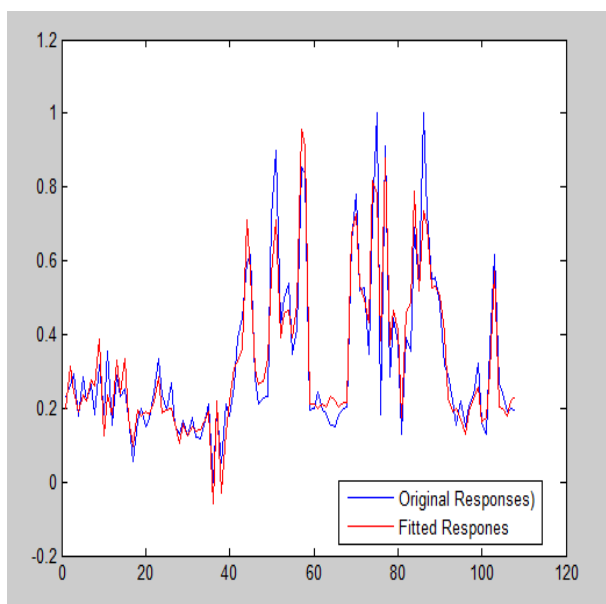
ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ ΜΕ ΜΕΣΕΣ ΤΙΜΕΣ ΑΡΩΜΑΤΙΚΩΝ (PCR2)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	108
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100 - 2750 cm ⁻¹ 1800 - 650 cm ⁻¹
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	5.8%
Τυπική απόκλιση	20.39
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	82%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1%
Μέση τιμή απόλυτων σφαλμάτων	14.75%
RMSEP	0.068%
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	10
Μέγιστη τιμή αρωματικών(%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	18%
Τυπική απόκλιση	16.24
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	21.94%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.95%
Μέση τιμή απόλυτων σφαλμάτων	10.15%
RMSEP	0.04



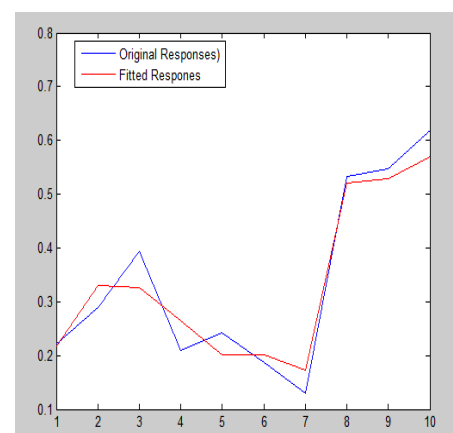
Σχήμα 5.1.5: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.1.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης.

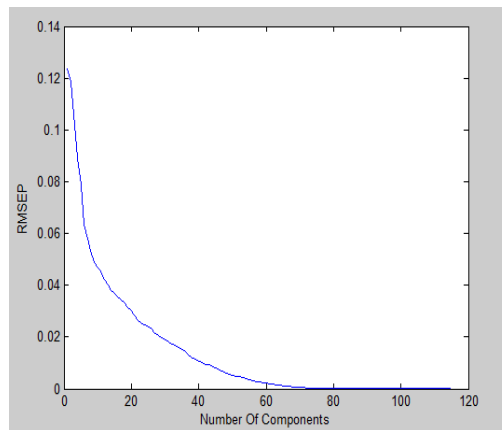


Σχήμα 5.1.7: Δείγματα εκπαίδευσης

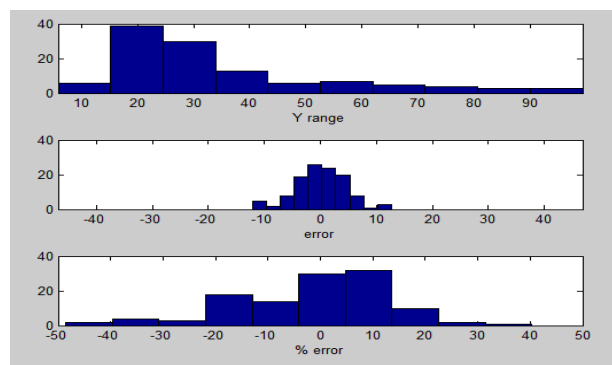


Σχήμα 5.1.8: Δείγματα δοκιμής

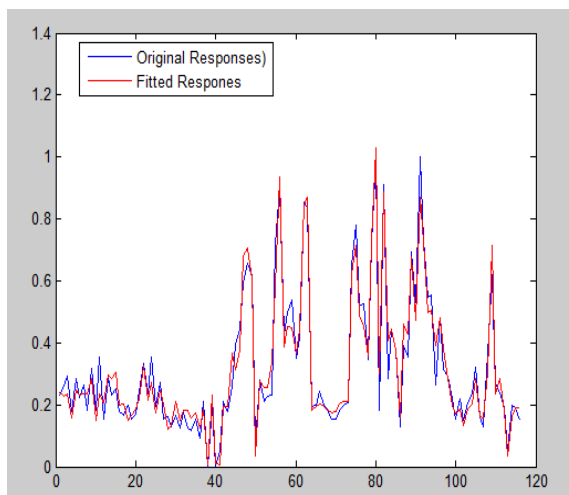
ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ ΜΕ ΜΕΣΕΣ ΤΙΜΕΣ ΑΡΩΜΑΤΙΚΩΝ (PLS1)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	116
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm^{-1} 1800-650 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	5.8%
Τυπική απόκλιση	20.4
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	48.5%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.03%
Μέση τιμή απολύτων σφαλμάτων	11.4%
RMSEP	0.049
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	12
Μέγιστη τιμή αρωματικών(%)	42.5%
Ελάχιστη τιμή αρωματικών(%)	17.3%
Τυπική απόκλιση	7.9
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	31%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1%
Μέση τιμή απολύτων σφαλμάτων	11.1%
RMSEP	0.05



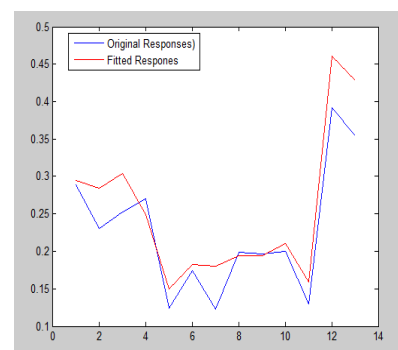
Σχήμα 5.1.9: Συμπεριφορά *RMSEP* κατά την αύξηση των παραγόντων



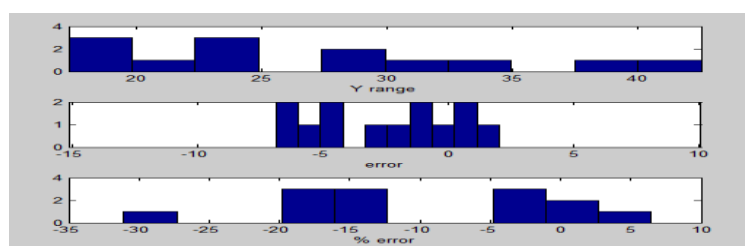
Σχήμα 5.1.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.1.11: Δείγματα εκπαίδευσης

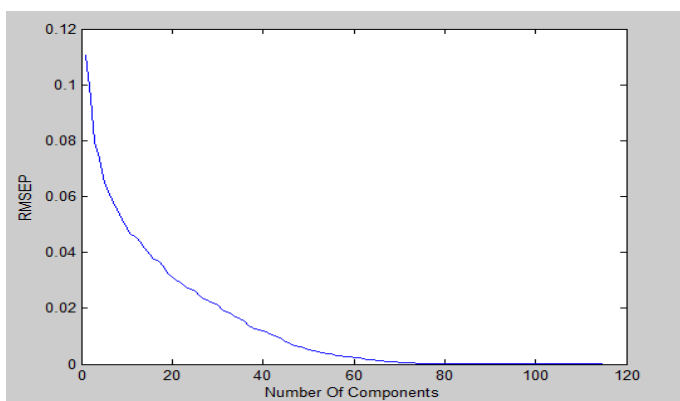


Σχήμα 5.1.12: Δείγματα δοκιμής

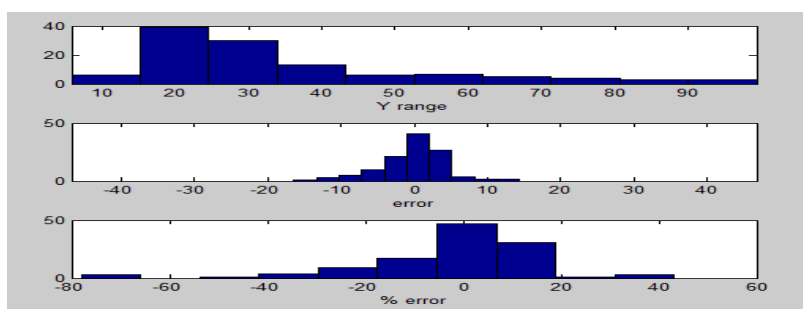


Σχήμα 5.1.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

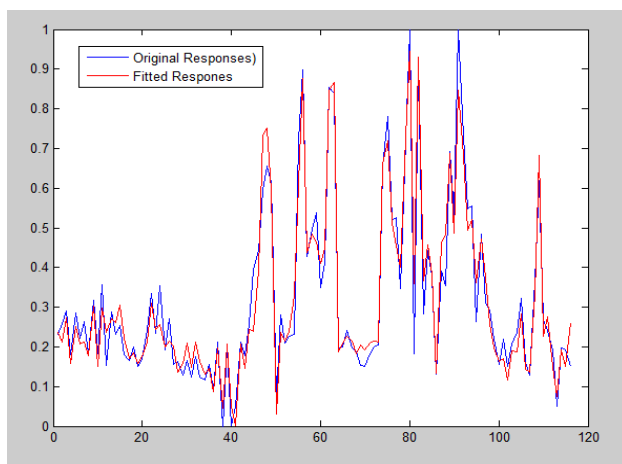
ΣΥΝΟΛΟ ΔΕΙΓΜΑΤΩΝ ΜΕ ΜΕΣΕΣ ΤΙΜΕΣ ΑΡΩΜΑΤΙΚΩΝ (PLS2)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	116
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm^{-1} 1800-650 cm^{-1}
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	5.8%
Τυπική απόκλιση	20.4
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	71.8%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.59%
Μέση τιμή απολύτων σφαλμάτων	12.5%
RMSEP	0.055
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	11
Μέγιστη τιμή αρωματικών(%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	17.3%
Τυπική απόκλιση	13
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	71.8%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	4%
Μέση τιμή απολύτων σφαλμάτων	15.8%
RMSEP	0.064



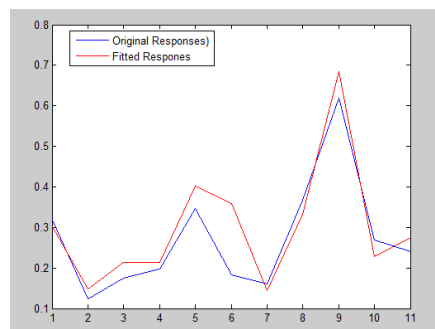
Σχήμα 5.1.14: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



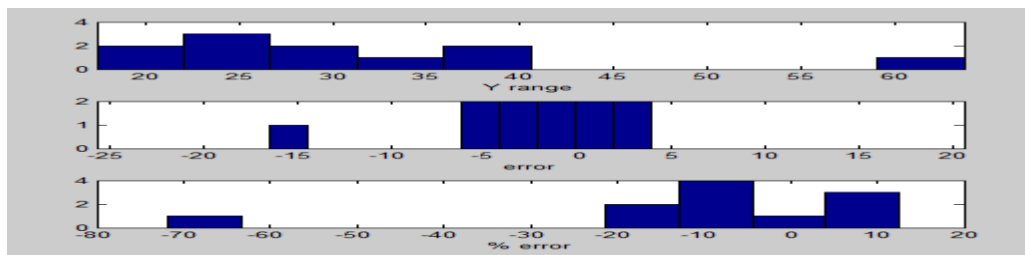
Σχήμα 5.1.15: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.1.16: Δείγματα εκπαίδευσης



Σχήμα 5.1.17: Δείγματα δοκιμής

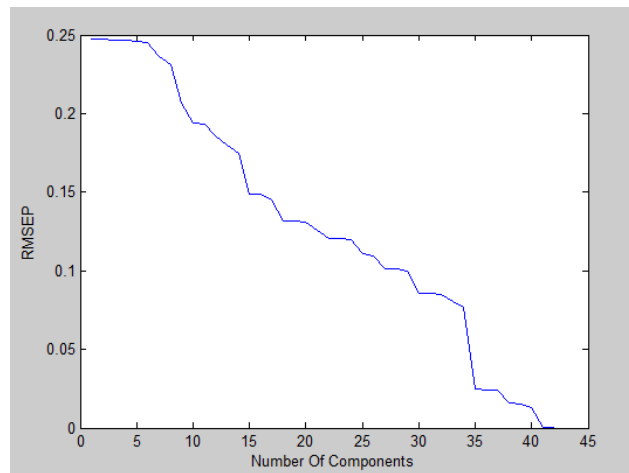


Σχήμα 5.1.18: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

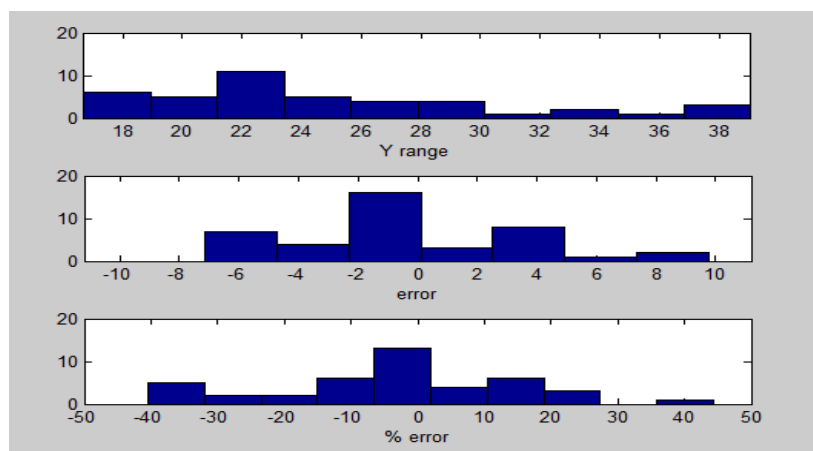
5.2 Μοντέλα Πρότυπης Αναλυτικής Μεθόδου ASTM D1319

Τα παρακάτω μοντέλα βασίζονται στο σύνολο των δειγμάτων που έχουν αναλυθεί με την πρότυπη χρωματογραφική μέθοδο ASTM D1319. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

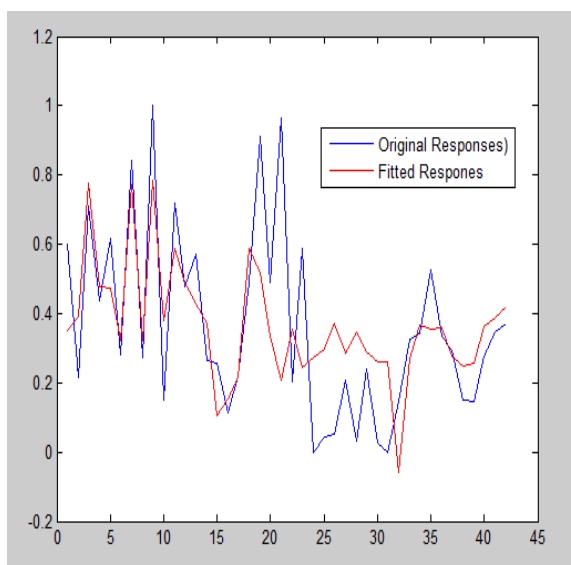
ASTM D1319 (PCR3)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	42
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2800 cm^{-1} 1800-700 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	39%
Ελάχιστη τιμή αρωματικών(%)	16.7%
Τυπική απόκλιση	5.97
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	44.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.5%
Μέση τιμή απολύτων σφαλμάτων	14.4%
RMSEP	0.206
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	6
Μέγιστη τιμή αρωματικών(%)	38.3%
Ελάχιστη τιμή αρωματικών(%)	16.7%
Τυπική απόκλιση	7.6
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	44.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.8%
Μέση τιμή απολύτων σφαλμάτων	18.6 %
RMSEP	0.34



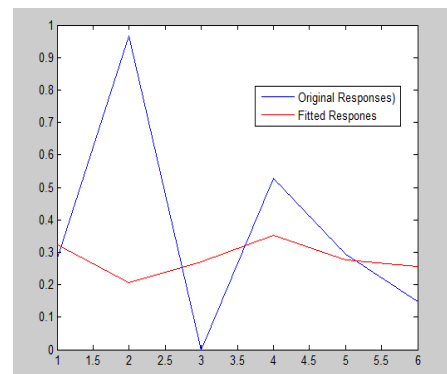
Σχήμα 5.2.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.2.2: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

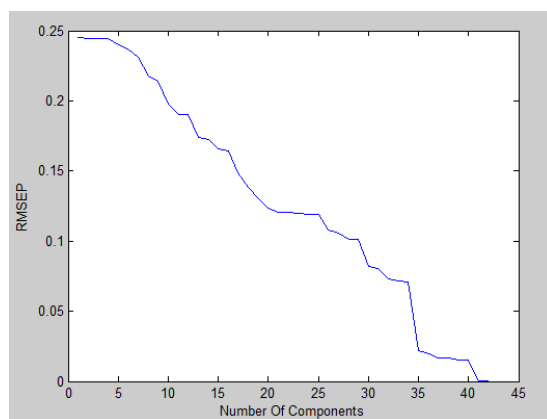


Σχήμα 5.2.3: Δείγματα εκπαίδευσης

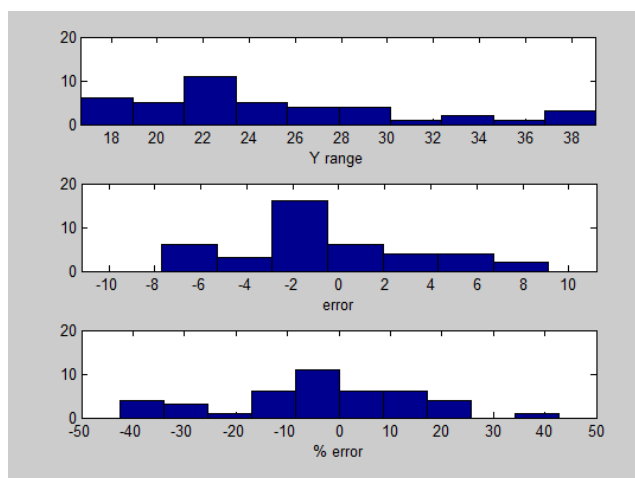


Σχήμα 5.2.4: Δείγματα δοκιμής

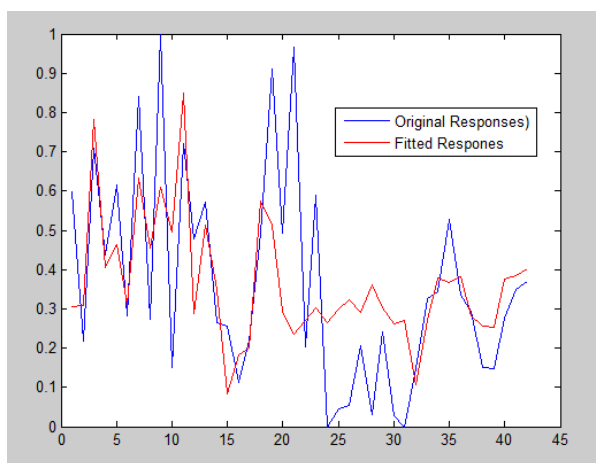
ASTM D1319 (PCR4)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	42
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2800 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: PRE-MAX
Μέγιστη τιμή αρωματικών (%)	39%
Ελάχιστη τιμή αρωματικών(%)	16.7%
Τυπική απόκλιση	5.97
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	42.6%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.2%
Μέση τιμή απολύτων σφαλμάτων	14.8%
RMSEP	0.214
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	28.5%
Ελάχιστη τιμή αρωματικών(%)	16.7%
Τυπική απόκλιση	4.9
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	42.5%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.1%
Μέση τιμή απολύτων σφαλμάτων	20.5 %
RMSEP	0.206



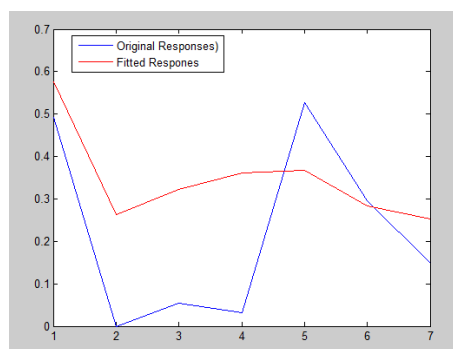
Σχήμα 5.2.5: Συμπεριφορά $RMSEP$ κατά την αύξηση των παραγόντων



Σχήμα 5.2.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

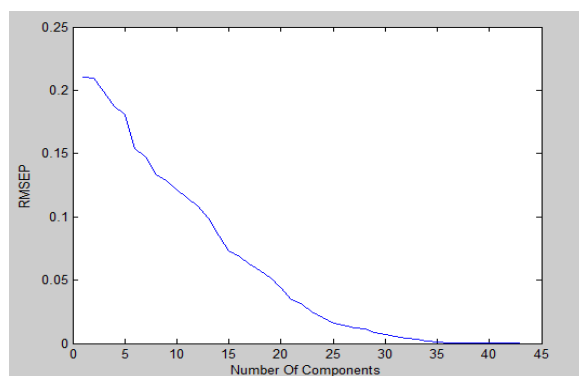


Σχήμα 5.2.7: Δείγματα εκπαίδευσης

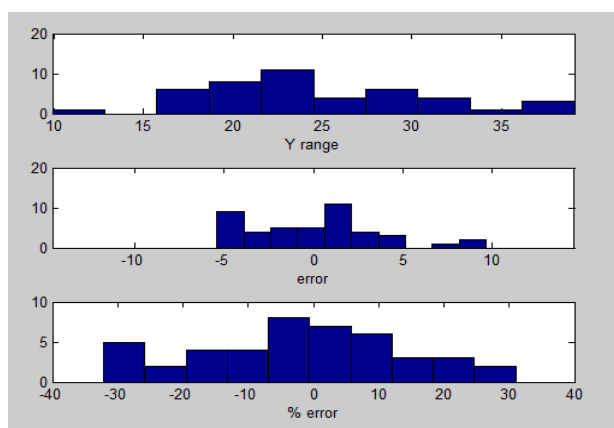


Σχήμα 5.2.8: Δείγματα δοκιμής

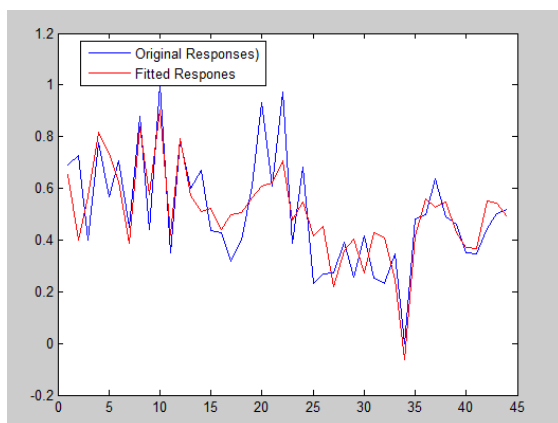
ASTM D1319 (PLS3)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	44
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2800 cm^{-1} 1800-700 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	39%
Ελάχιστη τιμή αρωματικών(%)	9.9%
Τυπική απόκλιση	6.3
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	30.5%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.1%
Μέση τιμή απολύτων σφαλμάτων	12.8%
RMSEP	0.128
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	30.1%
Ελάχιστη τιμή αρωματικών(%)	17.9%
Τυπική απόκλιση	4.47
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	30.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	2.8%
Μέση τιμή απολύτων σφαλμάτων	15.3%
RMSEP	0.127



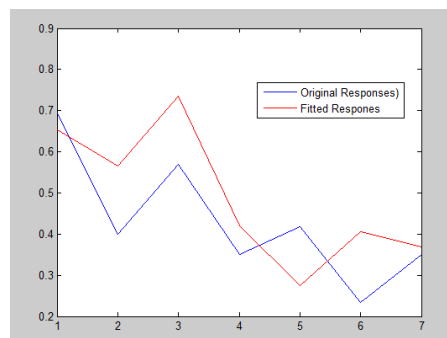
Σχήμα 5.2.9: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



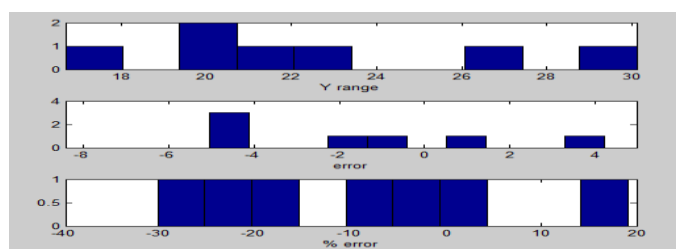
Σχήμα 5.2.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.2.1: Δείγματα εκπαίδευσης

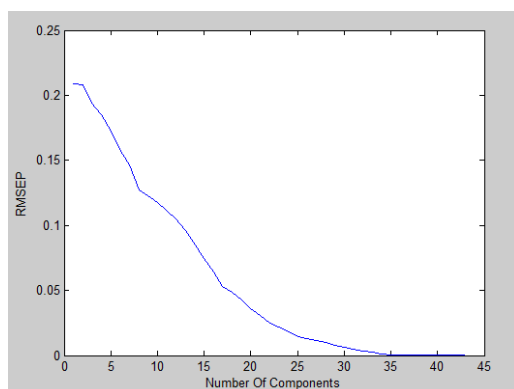


Σχήμα 5.2.12: Δείγματα δοκιμής

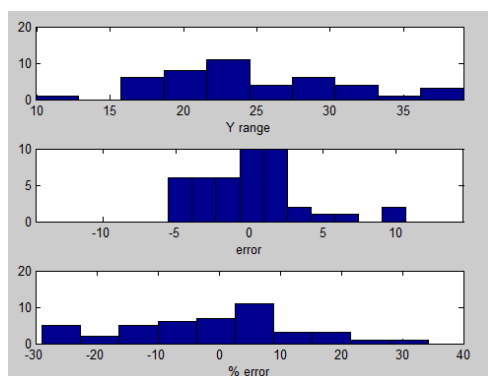


Σχήμα 5.2.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

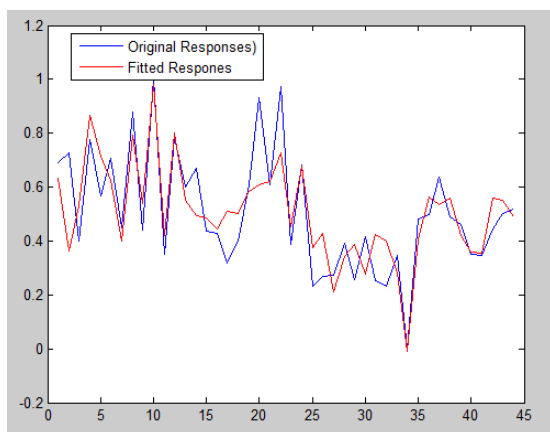
ASTM D1319 (PLS4)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	44
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2800 cm^{-1} 1800-700 cm^{-1}
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	39%
Ελάχιστη τιμή αρωματικών(%)	9.9%
Τυπική απόκλιση	6.3
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	34.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.7%
Μέση τιμή απολύτων σφαλμάτων	11.5%
RMSEP	0.122
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	6
Μέγιστη τιμή αρωματικών(%)	27.7%
Ελάχιστη τιμή αρωματικών(%)	16.7%
Τυπική απόκλιση	4.5
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	24.6%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.25%
Μέση τιμή απολύτων σφαλμάτων	13.5%
RMSEP	0.108



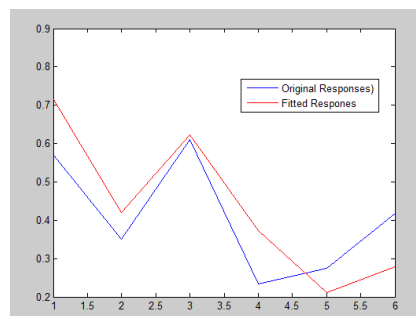
Σχήμα 5.2.14: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



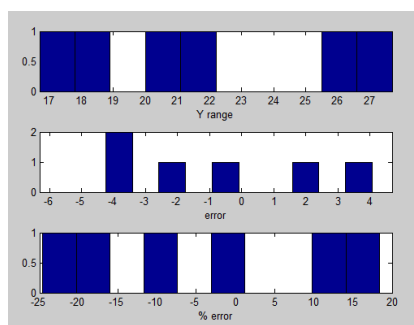
Σχήμα 5.2.15: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.2.16: Δείγματα εκπαίδευσης



Σχήμα 5.2.17: Δείγματα δοκιμής

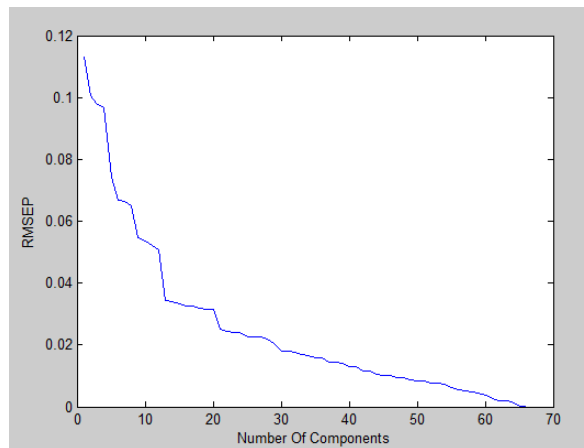


Σχήμα 5.2.18: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

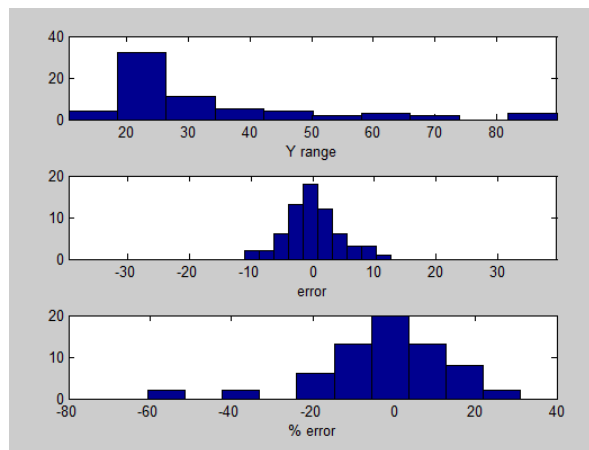
5.3 Μοντέλα με βάση τα δείγματα που έχουν αναλυθεί με πρότυπες αναλυτικές μεθόδους HPLC.

Τα παρακάτω μοντέλα βασίζονται στο σύνολο των δειγμάτων που έχουν αναλυθεί με τις πρότυπες μεθόδους υγρής χρωματογραφίας υψηλής πίεσης (HPLC). Αυτές είναι οι ASTM D6379, D6591, D7419. Η ιδιαιτερότητα του μοντέλου αυτού είναι πως τα δείγματα εκπαίδευσης περιέχουν πολλές διαφορετικές ομάδες πετρελαϊκών κλασμάτων. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

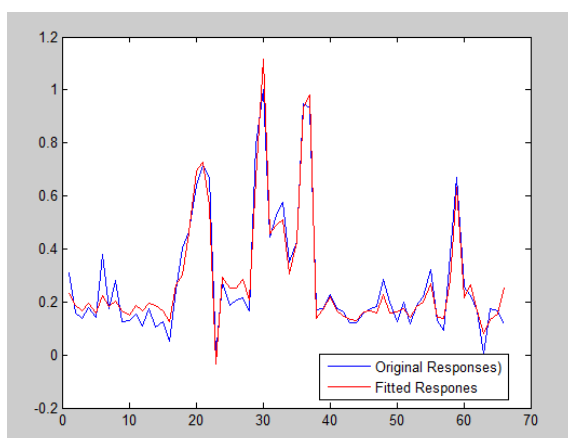
HPLC (PCR5)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	66
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700 cm^{-1} 1800-700 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	89.8%
Ελάχιστη τιμή αρωματικών (%)	10.5%
Τυπική απόκλιση	18
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	60.4%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.06%
Μέση τιμή απολύτων σφαλμάτων	11.8%
RMSEP	0.054
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	6
Μέγιστη τιμή αρωματικών (%)	63.7%
Ελάχιστη τιμή αρωματικών (%)	17.9%
Τυπική απόκλιση	16.26
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	18.9%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	4.7%
Μέση τιμή απολύτων σφαλμάτων	10.2%
RMSEP	0.035



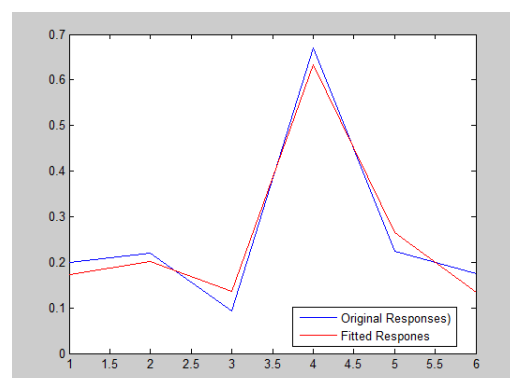
Σχήμα 5.3.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.3.2: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

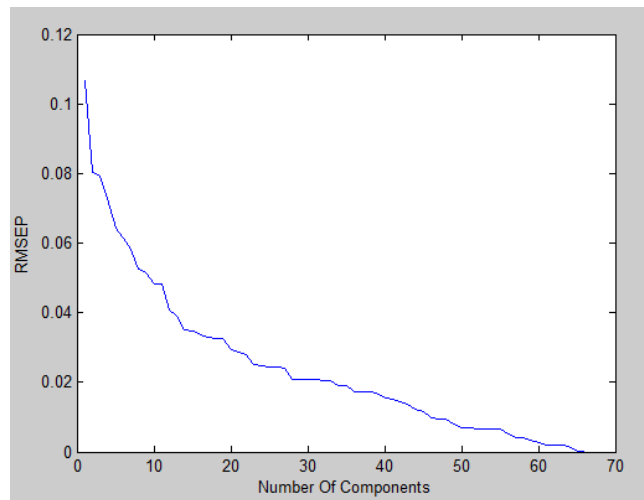


Σχήμα 5.3.3: Δείγματα εκπαίδευσης

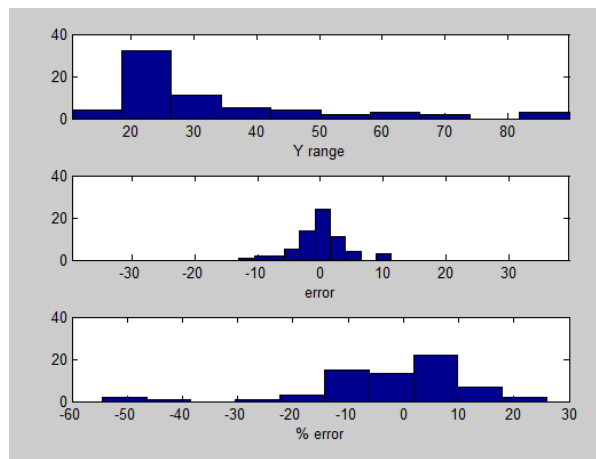


Σχήμα 5.3.4: Δείγματα δοκιμής

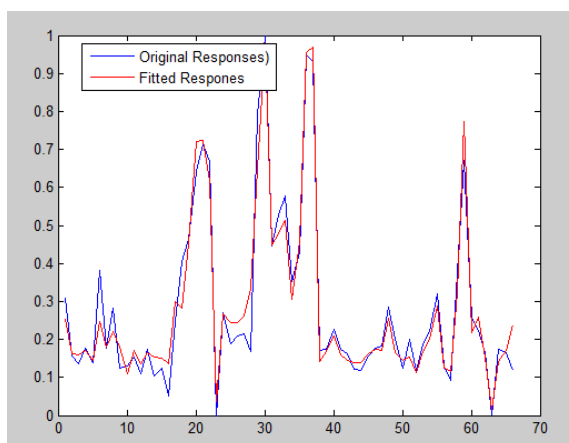
HPLC (PCR6)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	66
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	89.8%
Ελάχιστη τιμή αρωματικών(%)	10.5%
Τυπική απόκλιση	18
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	54.5%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.4%
Μέση τιμή απολύτων σφαλμάτων	10.2%
RMSEP	0.052
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	73.2%
Ελάχιστη τιμή αρωματικών(%)	19.8%
Τυπική απόκλιση	19.42
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	15.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	2.1%
Μέση τιμή απολύτων σφαλμάτων	5.8%
RMSEP	0.054



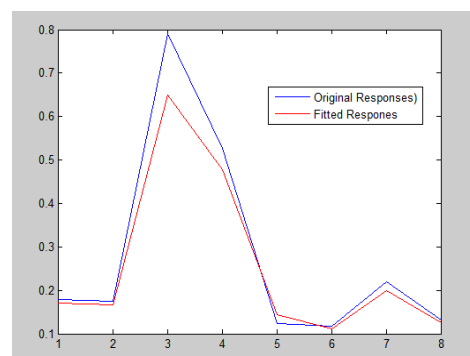
Σχήμα 5.3.5: Συμπεριφορά $RMSEP$ κατά την αύξηση των παραγόντων



Σχήμα 5.3.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

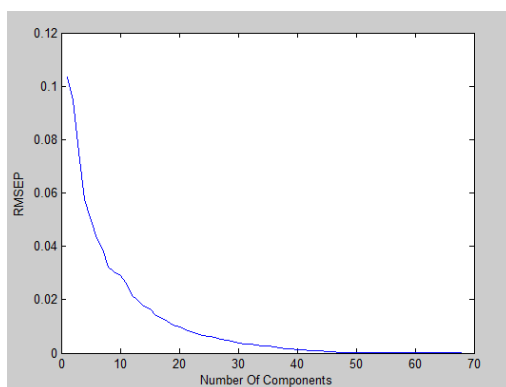


Σχήμα 5.3.7: Δείγματα εκπαίδευσης

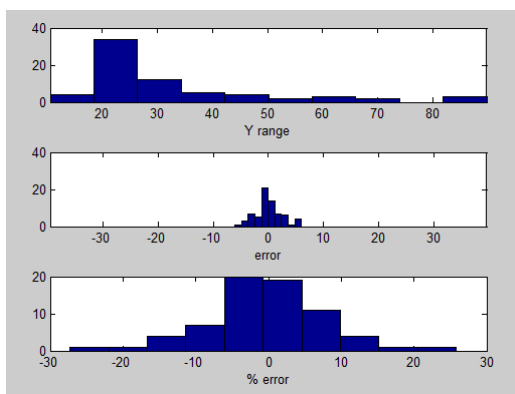


Σχήμα 5.3.8: Δείγματα δοκιμής

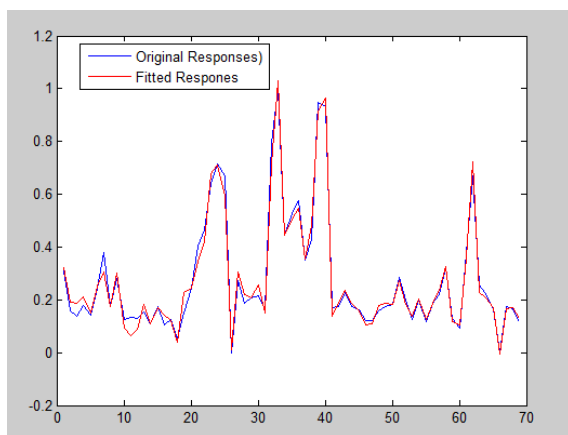
HPLC (PLS5)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	69
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	89.8%
Ελάχιστη τιμή αρωματικών(%)	10.5%
Τυπική απόκλιση	17.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	27.2%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.2%
Μέση τιμή απολύτων σφαλμάτων	6%
RMSEP	0.03
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	30%
Ελάχιστη τιμή αρωματικών(%)	20.3%
Τυπική απόκλιση	3.16
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	9.5%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.5%
Μέση τιμή απολύτων σφαλμάτων	3.05%
RMSEP	0.013



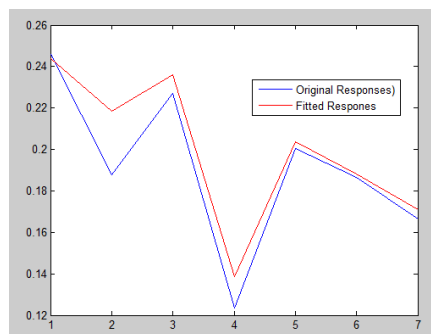
Σχήμα 5.3.9: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



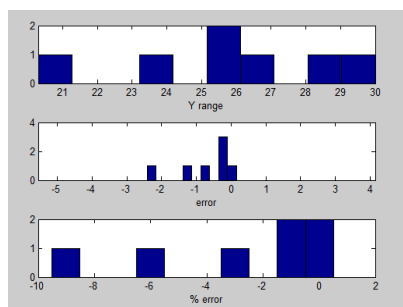
Σχήμα 5.3.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.3.11: Δείγματα εκπαίδευσης

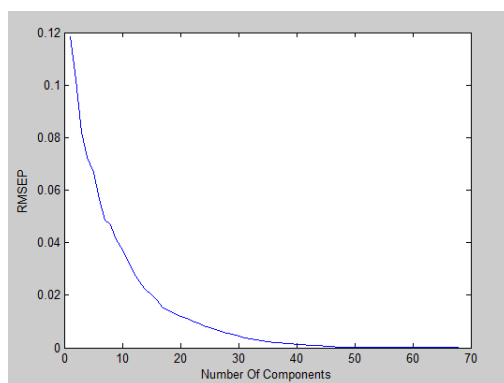


Σχήμα 5.3.12: Δείγματα δοκιμής

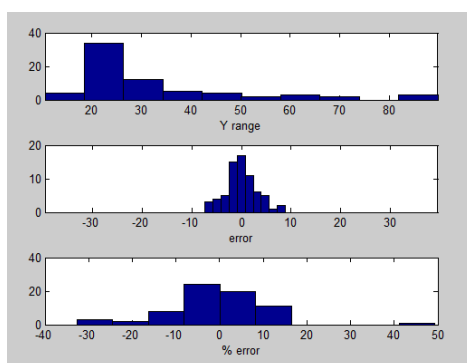


Σχήμα 5.3.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

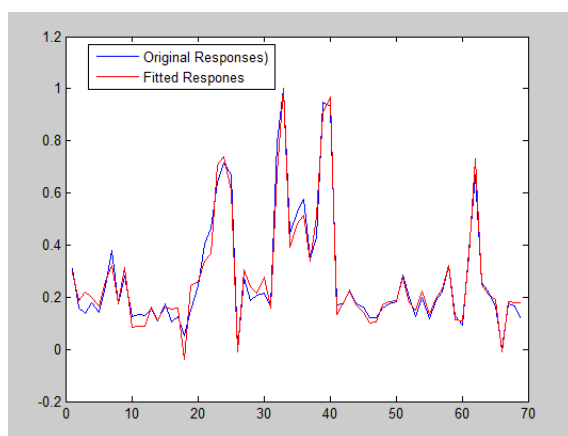
HPLC (PLS6)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	69
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: PRE-MAX
Μέγιστη τιμή αρωματικών (%)	89.8%
Ελάχιστη τιμή αρωματικών(%)	10.5%
Τυπική απόκλιση	17.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	49.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.08%
Μέση τιμή απολύτων σφαλμάτων	8.4%
RMSEP	0.04
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	63.5%
Ελάχιστη τιμή αρωματικών(%)	21.6%
Τυπική απόκλιση	15
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	15.8%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	2.6%
Μέση τιμή απολύτων σφαλμάτων	7.7%
RMSEP	0.035



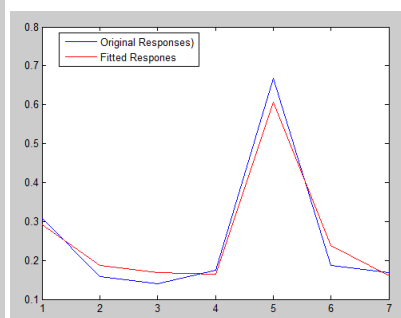
Σχήμα 5.3.14: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



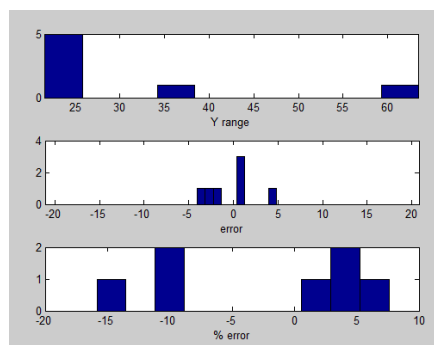
Σχήμα 5.3.15: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.3.16: Δείγματα εκπαίδευσης



Σχήμα 5.3.17: Δείγματα δοκιμής



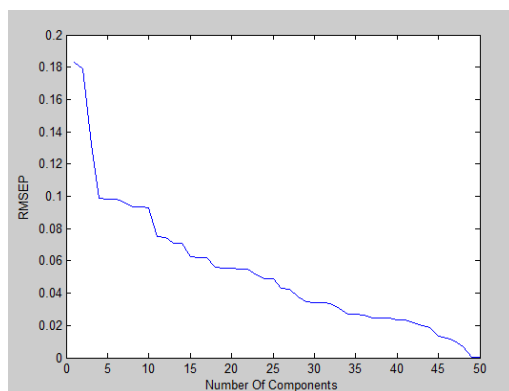
Σχήμα 5.3.18: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

5.4 Μοντέλα με βάση τα δείγματα που έχουν αναλυθεί με τις πρότυπες αναλυτικές μεθόδους ASTM D7419, D2549.

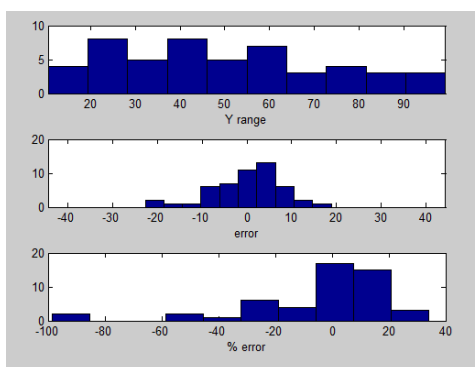
Τα παρακάτω μοντέλα βασίζονται στο σύνολο των δειγμάτων που έχουν αναλυθεί με τις πρότυπες χρωματογραφικές μεθόδους ASTM D7419 και D2549. Οι μέθοδοι αυτοί πιστοποιούνται για την ανάλυση βαρύτερων κλασμάτων πετρελαιοειδών και γιαυτό το λόγο έχουν επιλεγεί, ως δείγματα εκπαίδευσης του μοντέλου, έλαια, λιπαντικά και κατάλοιπα. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

ASTM D7419 ΚΑΙ D2549 (PCR7)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	50
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700 cm^{-1} 1800-750 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	10.5%
Τυπική απόκλιση	23.6
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	94%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.2%
Μέση τιμή απολύτων σφαλμάτων	17%
RMSEP	0.093
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	30.4%
Τυπική απόκλιση	23.5%
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	27.4%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.2%

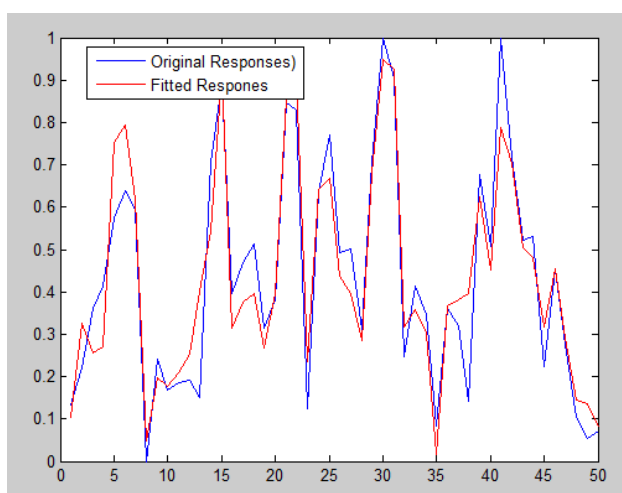
Μέση τιμή απόλυτων σφαλμάτων	8.9%
RMSEP	0.09



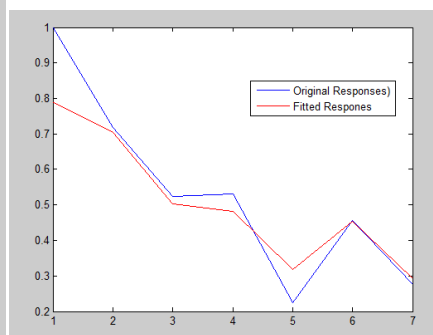
Σχήμα 5.4.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.4.2: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

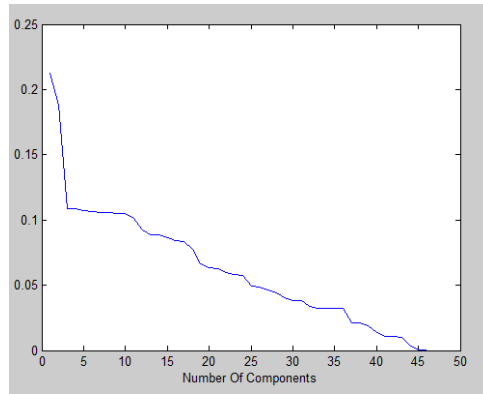


Σχήμα 5.4.3: Δείγματα εκπαίδευσης

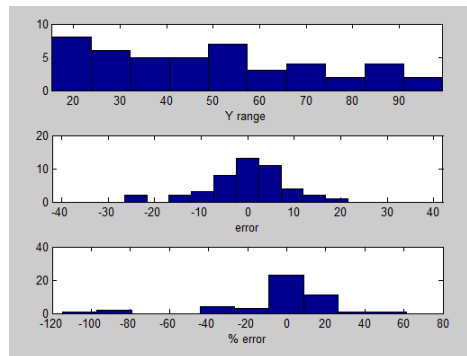


Σχήμα 5.4.4: Δείγματα δοκιμής

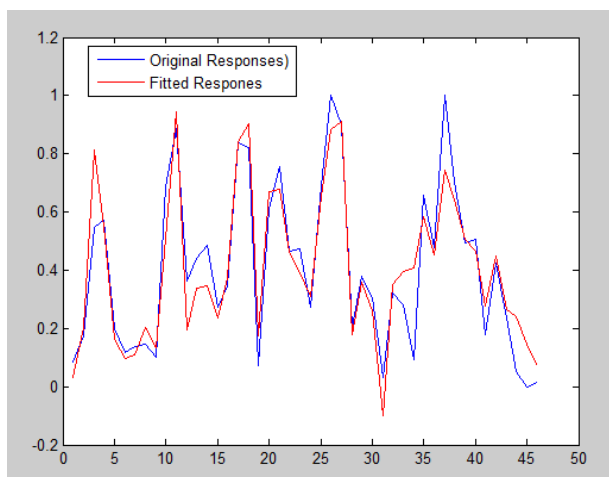
ASTM D7419 & D2549 (PCR8)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	46
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700 cm ⁻¹ 1800-750 cm ⁻¹
	Προεπεξεργασία: PRE-MAX
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	15.4%
Τυπική απόκλιση	23.8
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	80%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.3%
Μέση τιμή απολύτων σφαλμάτων	18%
RMSEP	0.105
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	73.2%
Ελάχιστη τιμή αρωματικών(%)	23.8%
Τυπική απόκλιση	16.43%
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	20.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	5.1%
Μέση τιμή απολύτων σφαλμάτων	11.1%
RMSEP	0.07



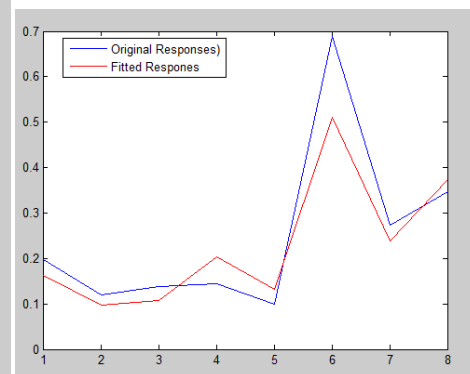
Σχήμα 5.4.5: Συμπεριφορά $RMSEP$ κατά την αύξηση των παραγόντων



Σχήμα 5.4.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

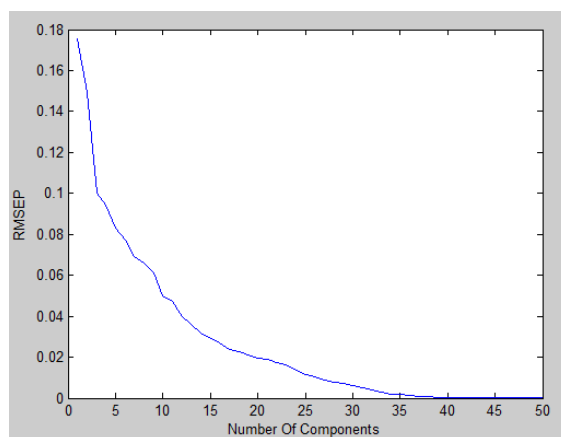


Σχήμα 5.4.7: Δείγματα εκπαίδευσης

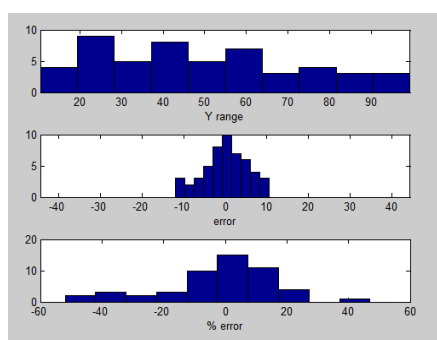


Σχήμα 5.4.8: Δείγματα δοκιμής

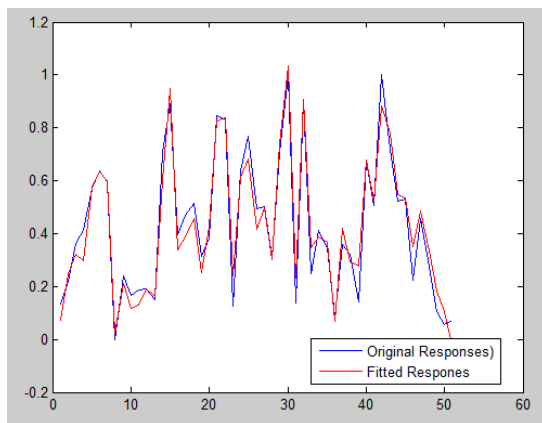
ASTM D7419 & D2549 (PLS7)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	51
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700cm ⁻¹ 1800-750 cm ⁻¹
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4 %
Ελάχιστη τιμή αρωματικών(%)	10.5 %
Τυπική απόκλιση	23.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	51.9%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.17%
Μέση τιμή απόλυτων σφαλμάτων	12.5%
RMSEP	0.061
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	9
Μέγιστη τιμή αρωματικών(%)	73.3%
Ελάχιστη τιμή αρωματικών(%)	15.4%
Τυπική απόκλιση	19.2
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	48.7%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	2.1%
Μέση τιμή απόλυτων σφαλμάτων	19.6%
RMSEP	0.066



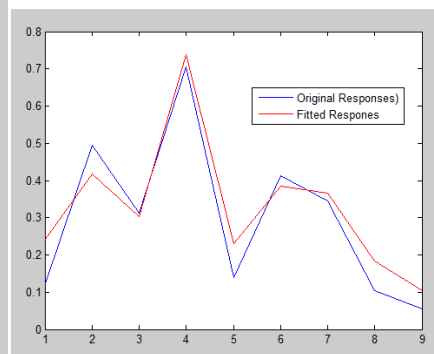
Σχήμα 5.4.9: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



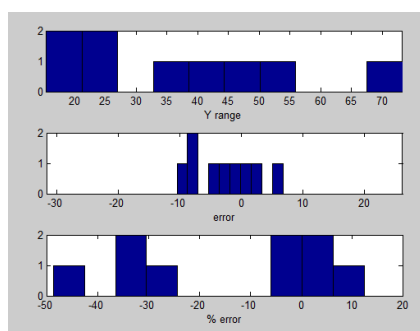
Σχήμα 5.4.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.4.11: Δείγματα εκπαίδευσης

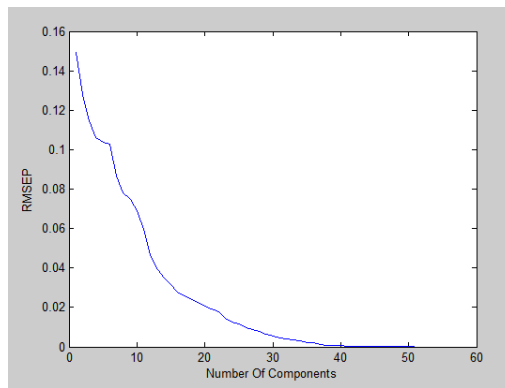


Σχήμα 5.4.12: Δείγματα δοκιμής

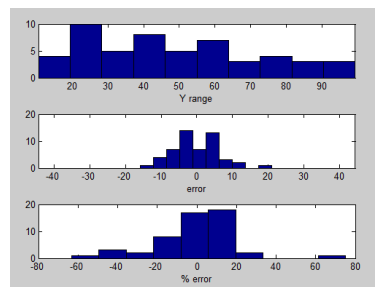


Σχήμα 5.4.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

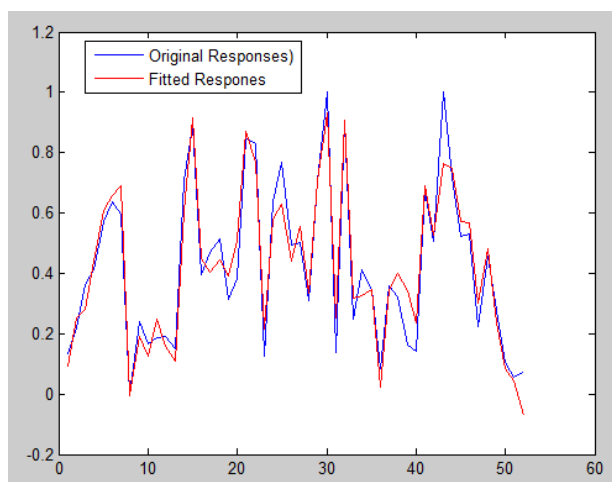
ASTM D7419 & D2549 (PLS8)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	51
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2700cm ⁻¹ 1800-750 cm ⁻¹
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	99.4 %
Ελάχιστη τιμή αρωματικών(%)	10.5 %
Τυπική απόκλιση	23.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	75%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.05%
Μέση τιμή απολύτων σφαλμάτων	14.2%
RMSEP	0.075
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	67%
Ελάχιστη τιμή αρωματικών(%)	15.4%
Τυπική απόκλιση	18.1
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	75%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.6%
Μέση τιμή απολύτων σφαλμάτων	20.9%
RMSEP	0.075



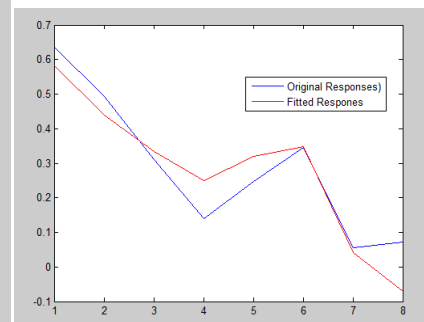
Σχήμα 5.4.14: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



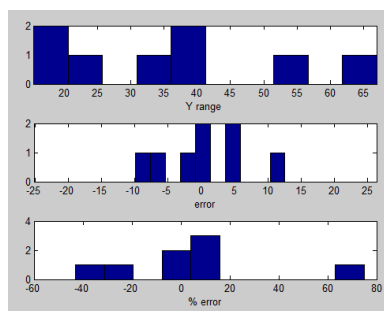
Σχήμα 5.4.15: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.4.16: Δείγματα εκπαίδευσης



Σχήμα 5.4.17: Δείγματα δοκιμής



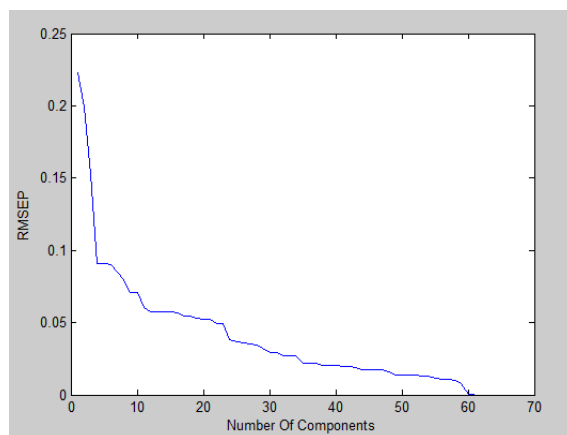
Σχήμα 5.4.18: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

5.5 Μοντέλα με βάση τα δείγματα που έχουν αναλυθεί με πρότυπες αναλυτικές μεθόδους ASTM D1319, D2549

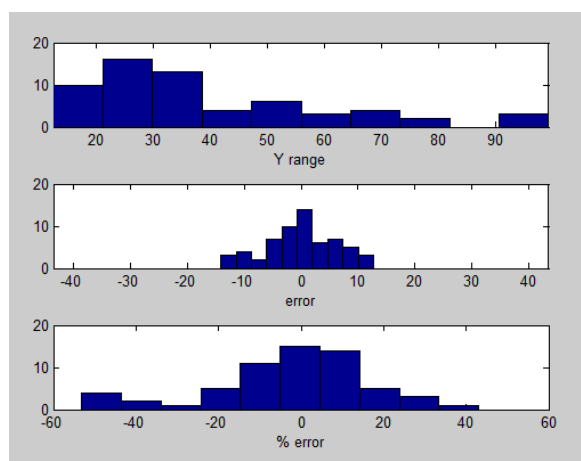
Τα παρακάτω μοντέλα βασίζονται στο σύνολο των δειγμάτων που έχουν αναλυθεί με τις πρότυπες χρωματογραφικές μεθόδους ASTM D1319 και D2549, οι οποίες αποτελούν μεθοδολογίες ανοικτής στήλης. Οι μέθοδοι αυτοί χρησιμοποιούνται για την ανάλυση μεσαίων κλασμάτων πετρελαιοειδών και γιαυτό το λόγο έχουν επιλεχθεί, ως δείγματα εκπαίδευσης του μοντέλου, κάυσιμα αεροσκαφών, ντίζελ και ορισμένα έλαια. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

ΑΝΟΙΧΤΗΣ ΣΤΗΛΗΣ (PCR9)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	61
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2650 cm^{-1} 1800-750 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	21.5
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	53.1%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.2%
Μέση τιμή απολύτων σφαλμάτων	14.8%
RMSEP	0.071
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	78.9%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	25.4
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	28%

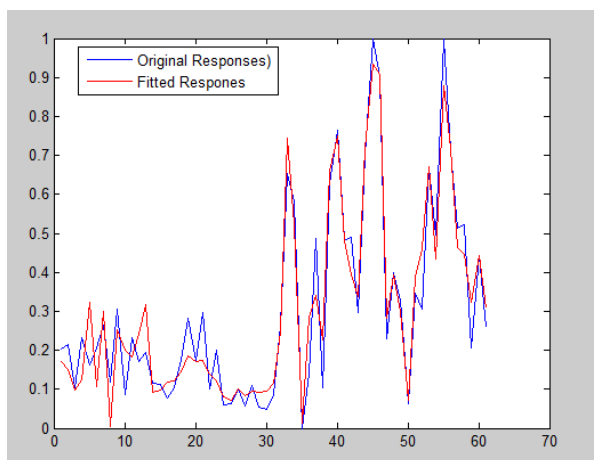
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.7%
Μέση τιμή απόλυτων σφαλμάτων	10.7 %
RMSEP	0.059



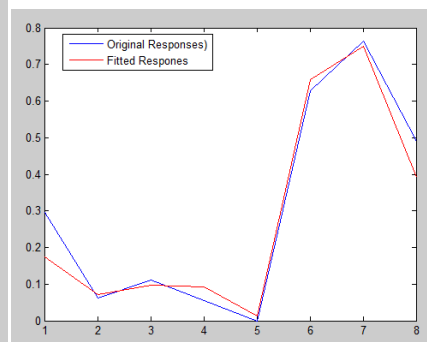
Σχήμα 5.5.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.5.2: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

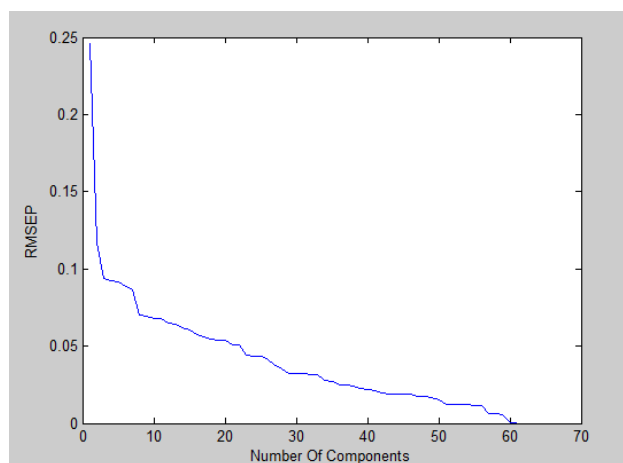


Σχήμα 5.5.3: Δείγματα εκπαίδευσης

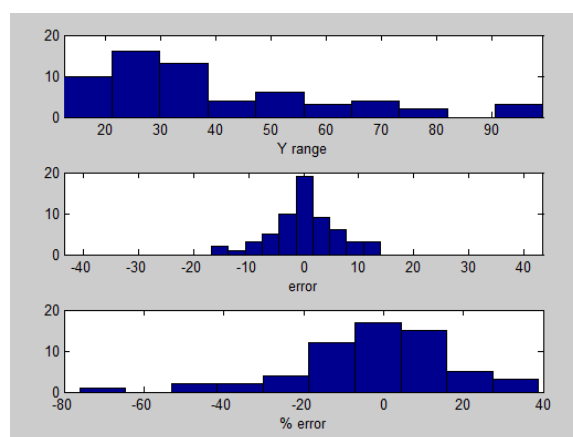


Σχήμα 5.5.4: Δείγματα δοκιμής

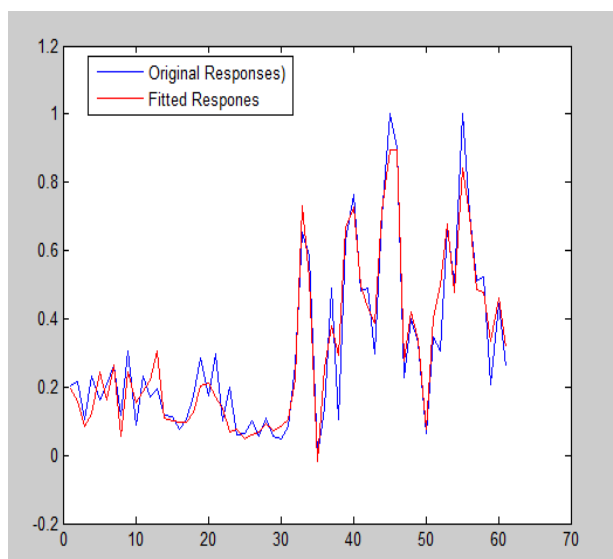
ΑΝΟΙΧΤΗΣ ΣΤΗΛΗΣ (PCR10)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	61
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2650 cm^{-1} 1800-750 cm^{-1}
	Προεπεξεργασία: κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	21.5
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	76%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.2%
Μέση τιμή απολύτων σφαλμάτων	13.9%
RMSEP	0.07
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	6
Μέγιστη τιμή αρωματικών(%)	55.6%
Ελάχιστη τιμή αρωματικών(%)	32.4%
Τυπική απόκλιση	9.2
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	42.7%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	2%
Μέση τιμή απολύτων σφαλμάτων	13.6 %
RMSEP	0.093



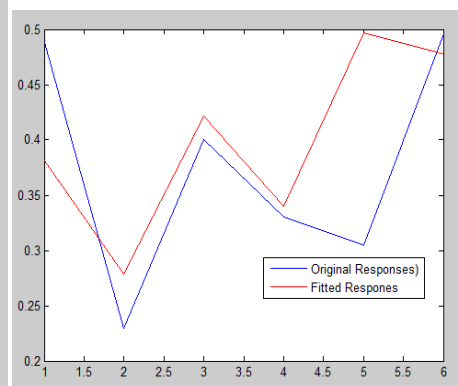
Σχήμα 5.5.5: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.5.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

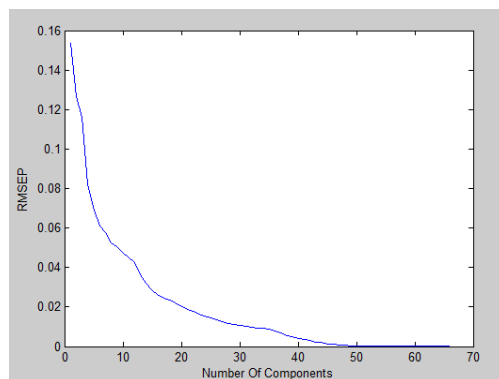


Σχήμα 5.5.7: Δείγματα εκπαίδευσης

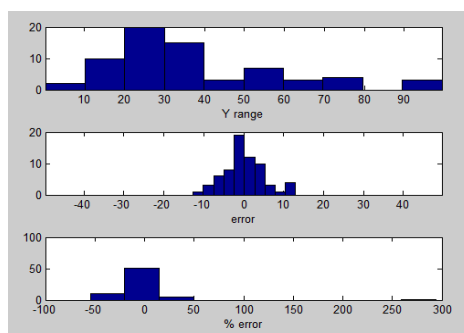


Σχήμα 5.5.8: Δείγματα δοκιμής

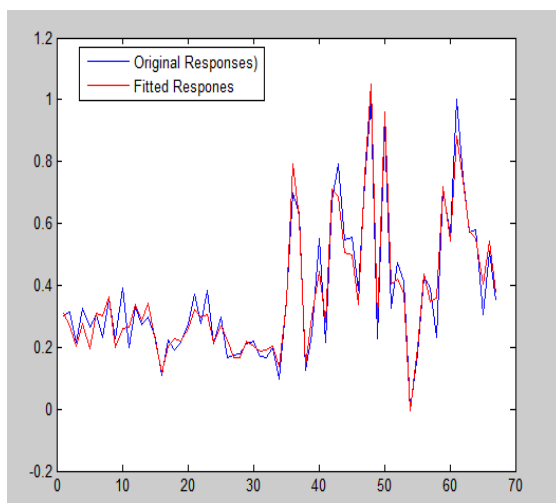
ΑΝΟΙΧΤΗΣ ΣΤΗΛΗΣ (PLS9)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	67
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2650 cm^{-1} 1800-750 cm^{-1}
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	0.2%
Τυπική απόκλιση	21.8%
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	300%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.6%
Μέση τιμή απόλυτων σφαλμάτων	16.2%
RMSEP	0.050
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	38.2%
Τυπική απόκλιση	19.8
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	13.4%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	2.4%
Μέση τιμή απόλυτων σφαλμάτων	7.8%
RMSEP	0.056



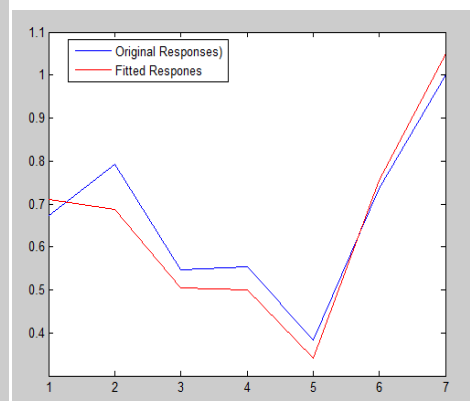
Σχήμα 5.5.9: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



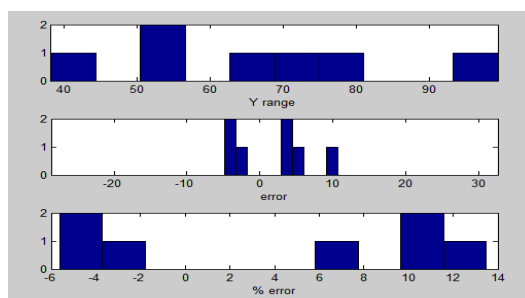
Σχήμα 5.5.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.5.11: Δείγματα εκπαίδευσης

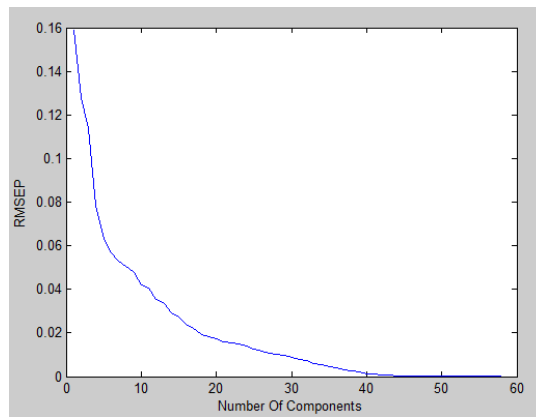


Σχήμα 5.5.12: Δείγματα δοκιμής

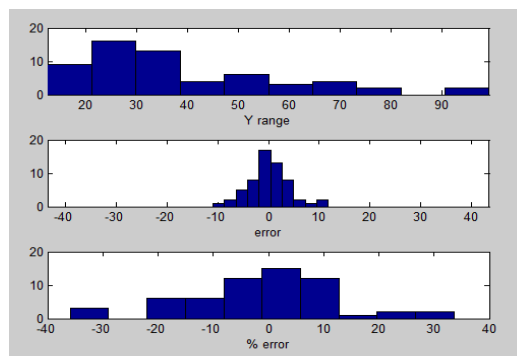


Σχήμα 5.5.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

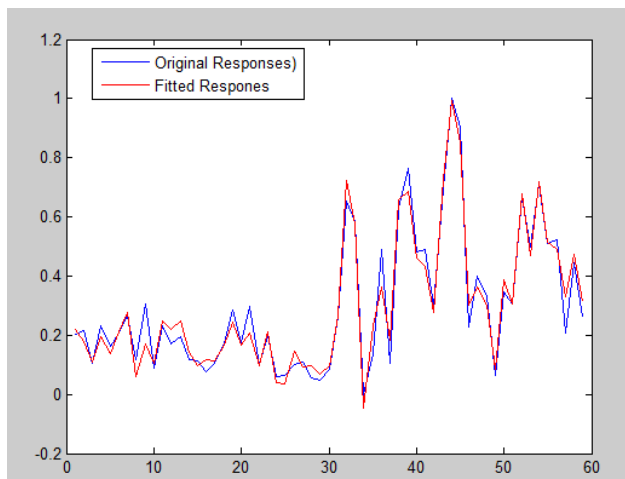
ΑΝΟΙΧΤΗΣ ΣΤΗΛΗΣ (PLS10)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	67
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2650 cm^{-1} 1800-750 cm^{-1}
	Προεπεξεργασία: κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	20.2%
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	35.9%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.08%
Μέση τιμή απολύτων σφαλμάτων	9.85%
RMSEP	0.048
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	57.8 %
Ελάχιστη τιμή αρωματικών(%)	24.5%
Τυπική απόκλιση	13.3%
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	35.9%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	4.1%
Μέση τιμή απολύτων σφαλμάτων	13.7%
RMSEP	0.066



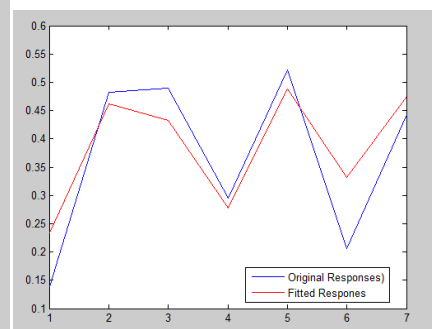
Σχήμα 5.5.14: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



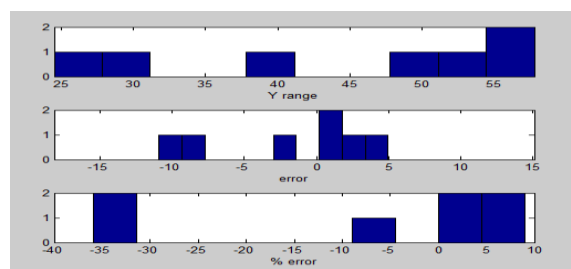
Σχήμα 5.5.15: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.5.16: Δείγματα εκπαίδευσης



Σχήμα 5.5.17: Δείγματα δοκιμής

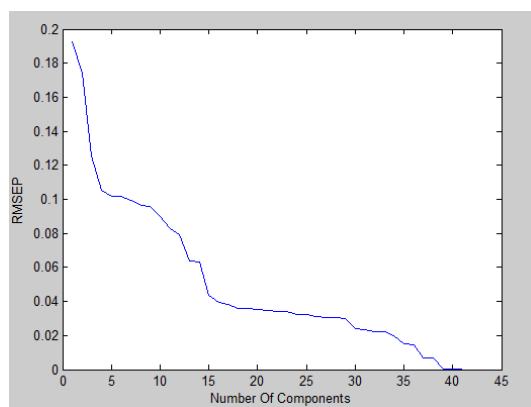


Σχήμα 5.5.18: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

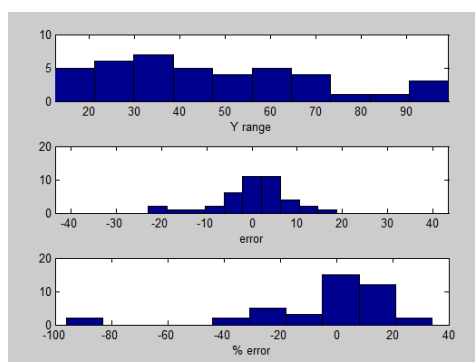
5.6 Μοντέλα με βάση τα δείγματα με σημείο βρασμού μεγαλύτερο των 400°C.

Τα παρακάτω μοντέλα βασίζονται στο σύνολο των δειγμάτων που έχουν σημείο βρασμού μεγαλύτερο των 400°C. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

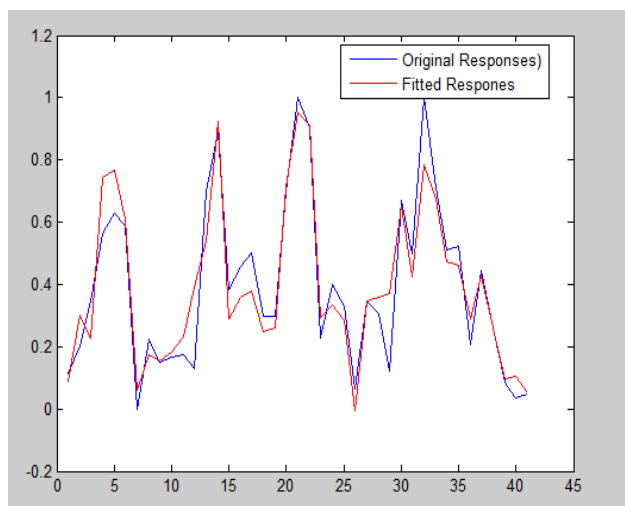
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΠΑΝΩ ΑΠΟ 400° C (PCR11)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	41
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	23.6
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	95.6%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.5%
Μέση τιμή απολύτων σφαλμάτων	16.9%
RMSEP	0.096
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	6
Μέγιστη τιμή αρωματικών(%)	74.3%
Ελάχιστη τιμή αρωματικών(%)	30.4%
Τυπική απόκλιση	16.1
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	23.2%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.6%
Μέση τιμή απολύτων σφαλμάτων	7.2%
RMSEP	0.047



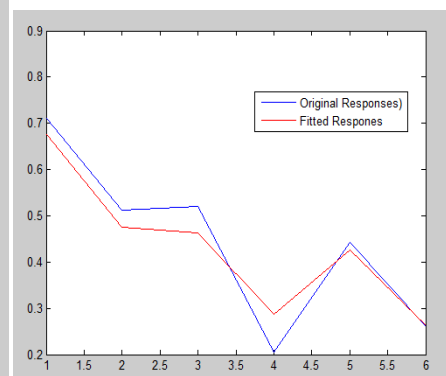
Σχήμα 5.6.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.6.2: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

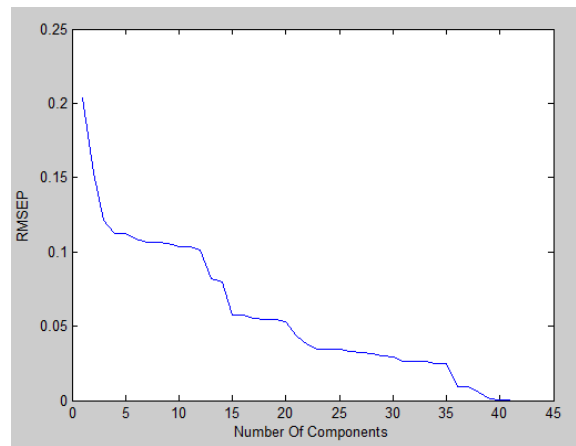


Σχήμα 5.6.3: Δείγματα εκπαίδευσης

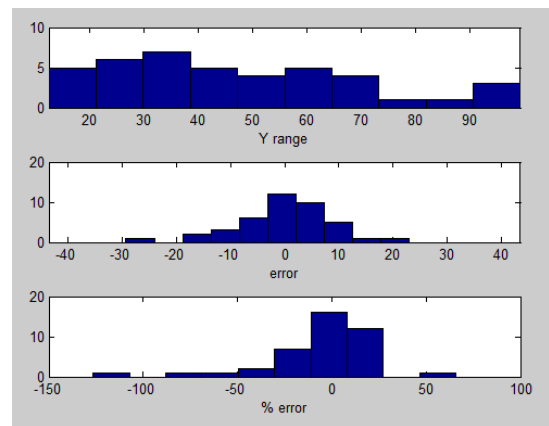


Σχήμα 5.6.4: Δείγματα δοκιμής

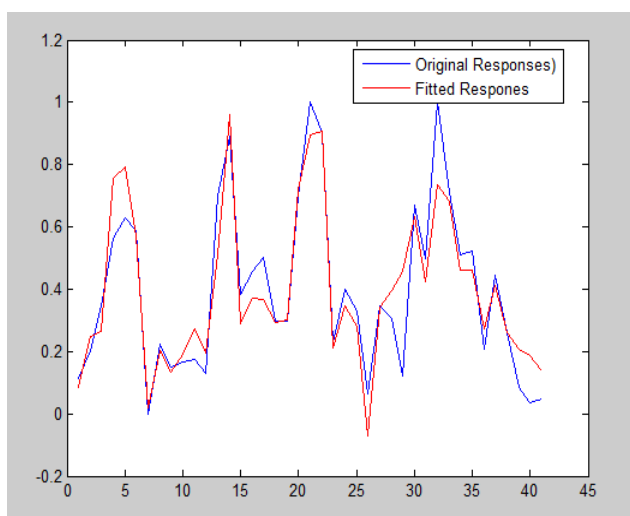
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΠΑΝΩ ΑΠΟ 400° C (PCR12)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	41
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: PRE-MAX
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	23.6
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	126%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.2%
Μέση τιμή απολύτων σφαλμάτων	19.2%
RMSEP	0.106
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	74.3%
Ελάχιστη τιμή αρωματικών(%)	30.4%
Τυπική απόκλιση	14.8
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	19.3%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.4%
Μέση τιμή απολύτων σφαλμάτων	8.05%
RMSEP	0.05



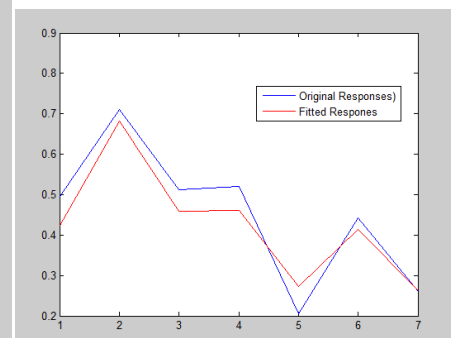
Σχήμα 5.6.5: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.6.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

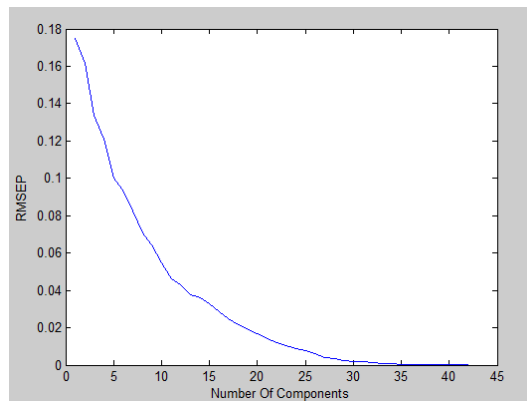


Σχήμα 5.6.7: Δείγματα εκπαίδευσης

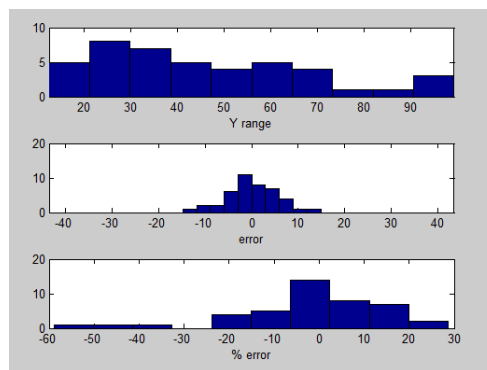


Σχήμα 5.6.8: Δείγματα δοκιμής

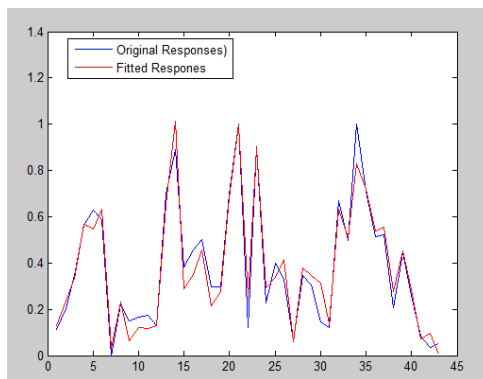
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΠΑΝΩ ΑΠΟ 400° C (PLS11)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	43
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	23.6
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	58.9%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.1%
Μέση τιμή απολύτων σφαλμάτων	11.8%
RMSEP	0.065
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	74.3%
Ελάχιστη τιμή αρωματικών(%)	30.4%
Τυπική απόκλιση	23.6
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	20%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.1%
Μέση τιμή απολύτων σφαλμάτων	6.1%
RMSEP	0.038



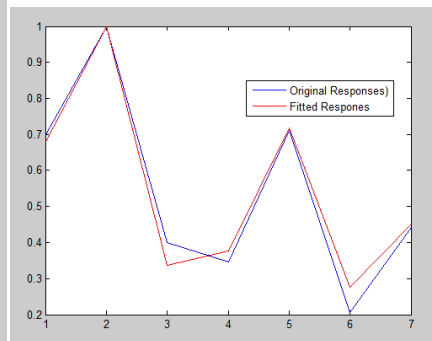
Σχήμα 5.6.9: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



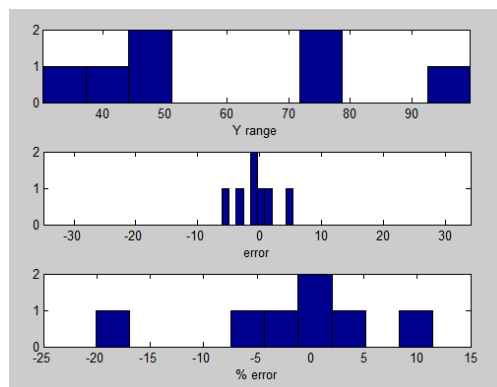
Σχήμα 5.6.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.6.11: Δείγματα εκπαίδευσης

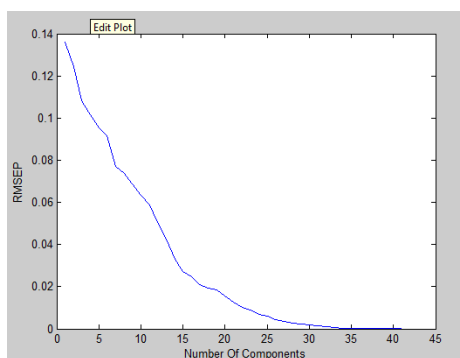


Σχήμα 5.6.12: Δείγματα δοκιμής

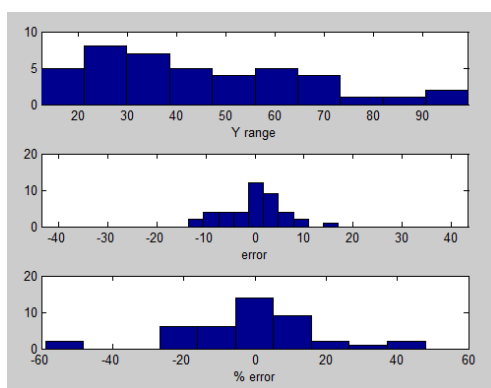


Σχήμα 5.6.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

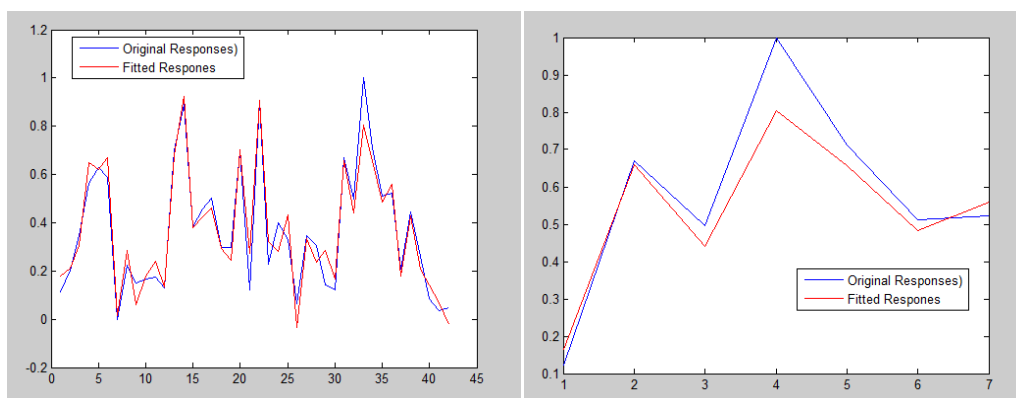
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΠΑΝΩ ΑΠΟ 400° C (PLS12)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	42
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	12.5%
Τυπική απόκλιση	22.3
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	58.8%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.3%
Μέση τιμή απολύτων σφαλμάτων	13.8%
RMSEP	0.069
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	7
Μέγιστη τιμή αρωματικών(%)	99.4%
Ελάχιστη τιμή αρωματικών(%)	23.1%
Τυπική απόκλιση	23.1
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	17%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.3%
Μέση τιμή απολύτων σφαλμάτων	8.4%
RMSEP	0.083



Σχήμα 5.6.14: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων

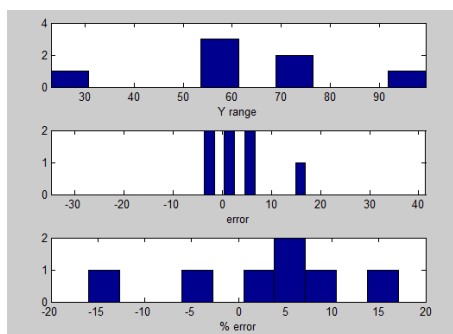


Σχήμα 5.6.15: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.6.16: Δείγματα εκπαίδευσης

Σχήμα 5.6.17: Δείγματα δοκιμής

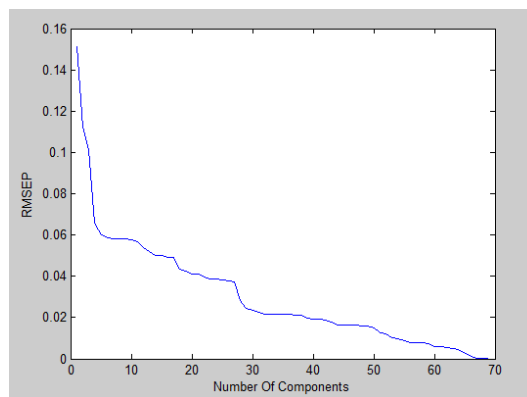


Σχήμα 5.6.18: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

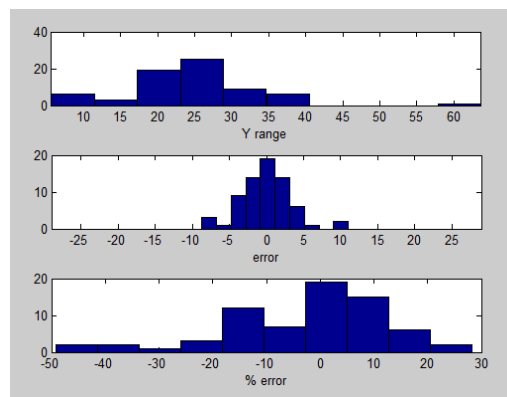
5.7 Μοντέλα με βάση τα δείγματα με σημείο βρασμού μικρότερο των 400°C.

Τα παρακάτω μοντέλα βασίζονται στο σύνολο των δειγμάτων που έχουν σημείο βρασμού μικρότερο των 400°C. Χρησιμοποιήθηκαν οι χημειομετρικές τεχνικές της παλινδρόμησης κυρίων συνιστωσών (Principal Components Regression, PCR) και των μερικών ελαχίστων τετραγώνων (Partial Least Squares, PLS).

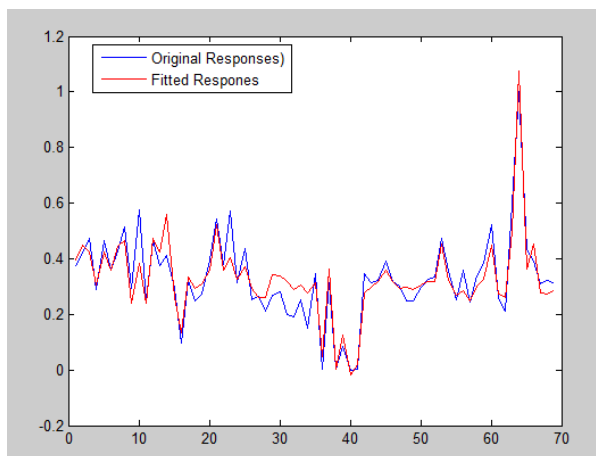
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΩΣ 400° C (PCR13)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	69
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	5.7%
Τυπική απόκλιση	8.9
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	49.2%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.09%
Μέση τιμή απολύτων σφαλμάτων	11.2%
RMSEP	0.058
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	17.9%
Τυπική απόκλιση	15.2
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	15.9%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	6.6%
Μέση τιμή απολύτων σφαλμάτων	12.2%
RMSEP	0.063



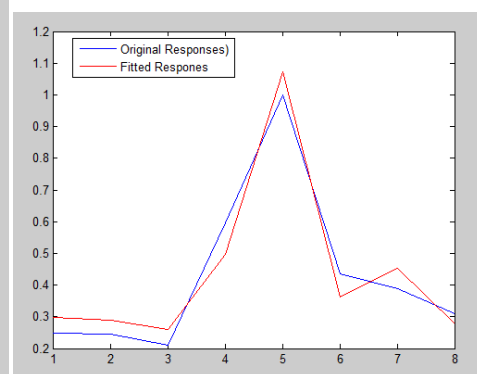
Σχήμα 5.7.1: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.7.2: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

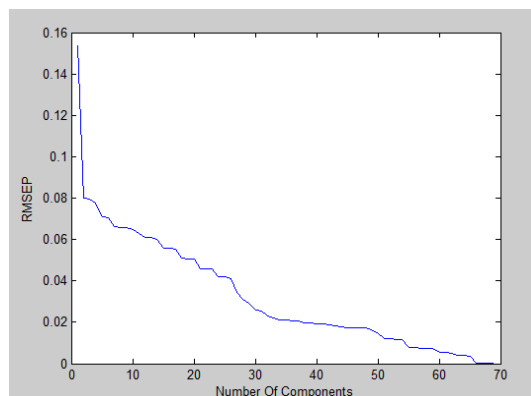


Σχήμα 5.7.3: Δείγματα εκπαίδευσης

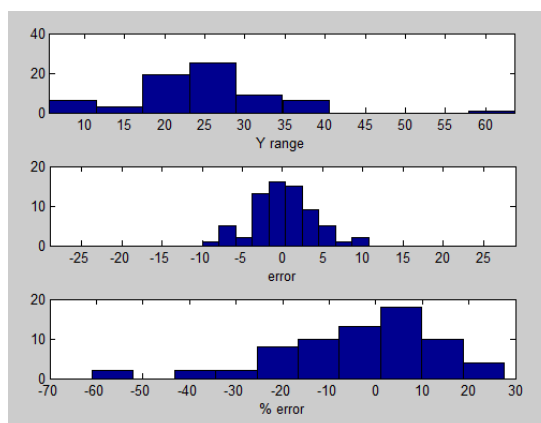


Σχήμα 5.7.4: Δείγματα δοκιμής

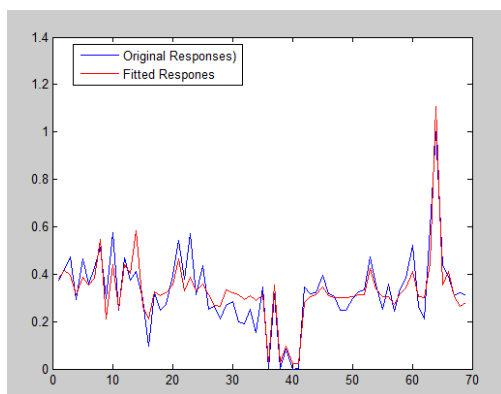
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΩΣ 400° C (PCR14)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	69
Αλγόριθμος	PCR
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: PRE-MAX
Μέγιστη τιμή αρωματικών (%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	5.7%
Τυπική απόκλιση	8.9
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	60.8%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.52%
Μέση τιμή απολύτων σφαλμάτων	13%
RMSEP	0.066
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	40.2%
Ελάχιστη τιμή αρωματικών(%)	19.8%
Τυπική απόκλιση	7.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	21.5 %
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.6%
Μέση τιμή απολύτων σφαλμάτων	11.7%
RMSEP	0.074



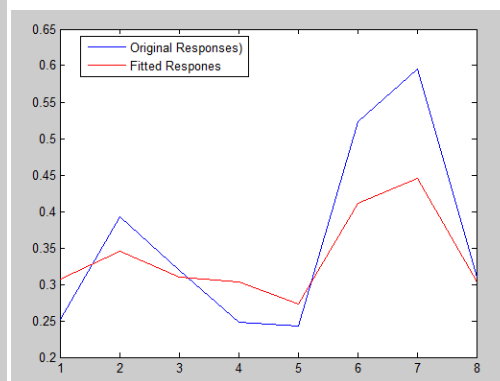
Σχήμα 5.7.5: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



Σχήμα 5.7.6: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης

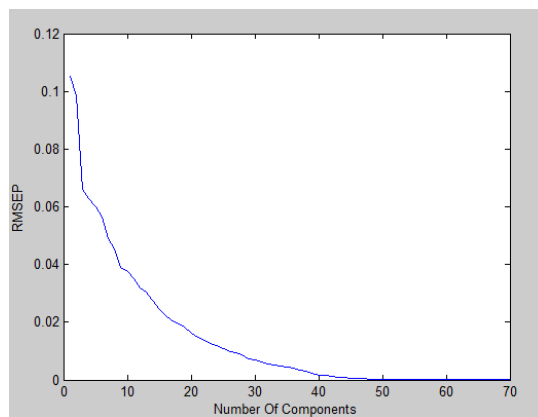


Σχήμα 5.7.7: Δείγματα εκπαίδευσης

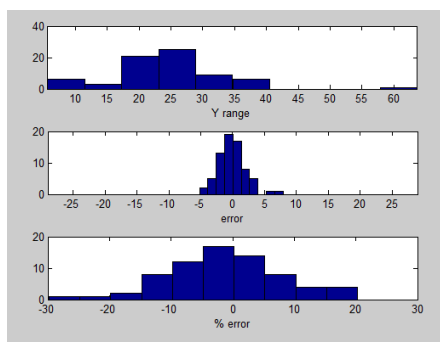


Σχήμα 5.7.8: Δείγματα δοκιμής

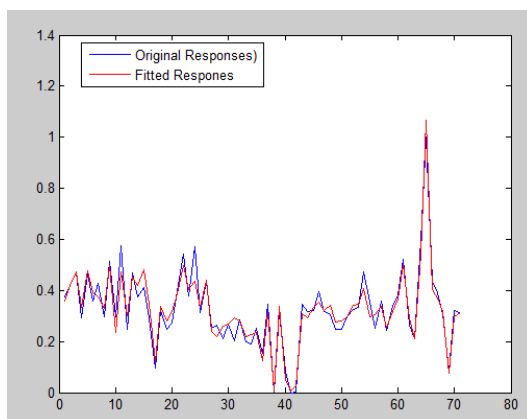
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΩΣ 400° C (PLS13)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	71
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: -
Μέγιστη τιμή αρωματικών (%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	5.7%
Τυπική απόκλιση	8.8
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	29.8%
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.13%
Μέση τιμή απολύτων σφαλμάτων	7.5%
RMSEP	0.039
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	8
Μέγιστη τιμή αρωματικών(%)	40.2%
Ελάχιστη τιμή αρωματικών(%)	17.9%
Τυπική απόκλιση	7.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	8.55 %
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.7%
Μέση τιμή απολύτων σφαλμάτων	4.5%
RMSEP	0.027



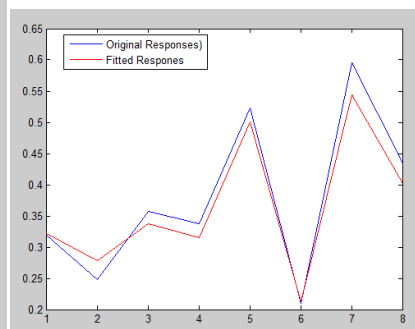
Σχήμα 5.7.9: Συμπεριφορά RMSEP κατά την αύξηση των παραγόντων



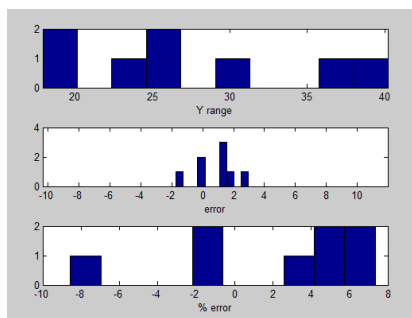
Σχήμα 5.7.10: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.7.11: Δείγματα εκπαίδευσης

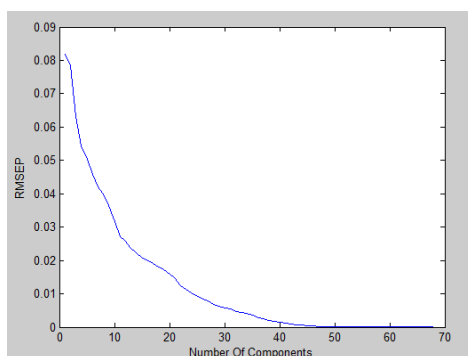


Σχήμα 5.7.12: Δείγματα δοκιμής

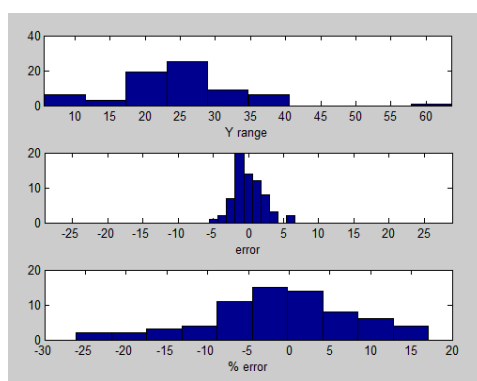


Σχήμα 5.7.13: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής

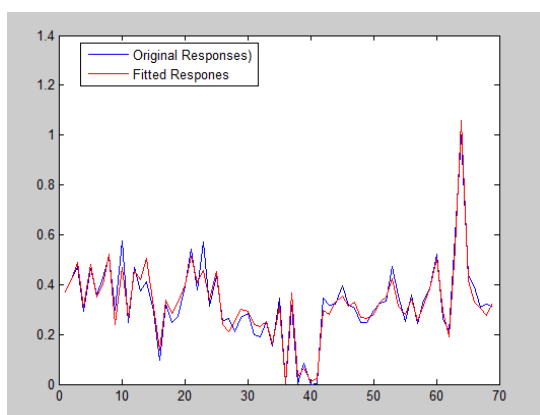
ΔΕΙΓΜΑΤΑ ΜΕ ΣΗΜΕΙΟ ΒΡΑΣΜΟΥ ΩΣ 400° C (PLS14)	
Εκπαίδευση Μοντέλου	
Αριθμός Δειγμάτων	69
Αλγόριθμος	PLS
Προεπεξεργασία	Επιλογή μεταβλητών: 3100-2750 cm ⁻¹ 2400-2300 cm ⁻¹ 1800-700 cm ⁻¹
	Προεπεξεργασία: Κανονικοποίηση
Μέγιστη τιμή αρωματικών (%)	63.7%
Ελάχιστη τιμή αρωματικών(%)	5.7%
Τυπική απόκλιση	8.9
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	26 %
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	0.11%
Μέση τιμή απολύτων σφαλμάτων	7.1%
RMSEP	0.036
Ελεγχος Μοντέλου	
Αριθμός δειγμάτων	9
Μέγιστη τιμή αρωματικών(%)	28.5%
Ελάχιστη τιμή αρωματικών(%)	20%
Τυπική απόκλιση	2.7
Αριθμός Παραγόντων	9
Μέγιστο απόλυτο σφάλμα πρόβλεψης	7.7 %
Ελάχιστο απόλυτο σφάλμα πρόβλεψης	1.2%
Μέση τιμή απολύτων σφαλμάτων	4.3%
RMSEP	0.020



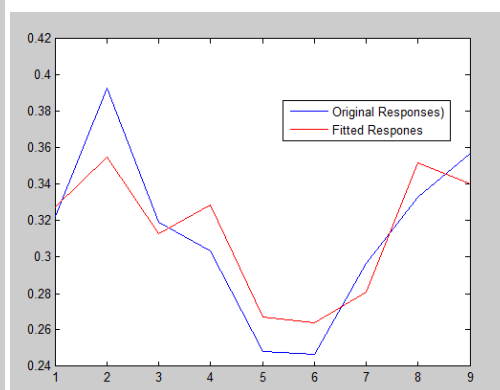
Σχήμα 5.7.15: Συμπεριφορά $RMSEP$ κατά την αύξηση των παραγόντων



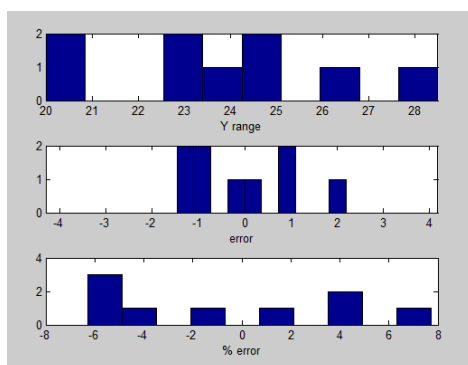
Σχήμα 5.7.16: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων εκπαίδευσης



Σχήμα 5.7.17: Δείγματα εκπαίδευσης



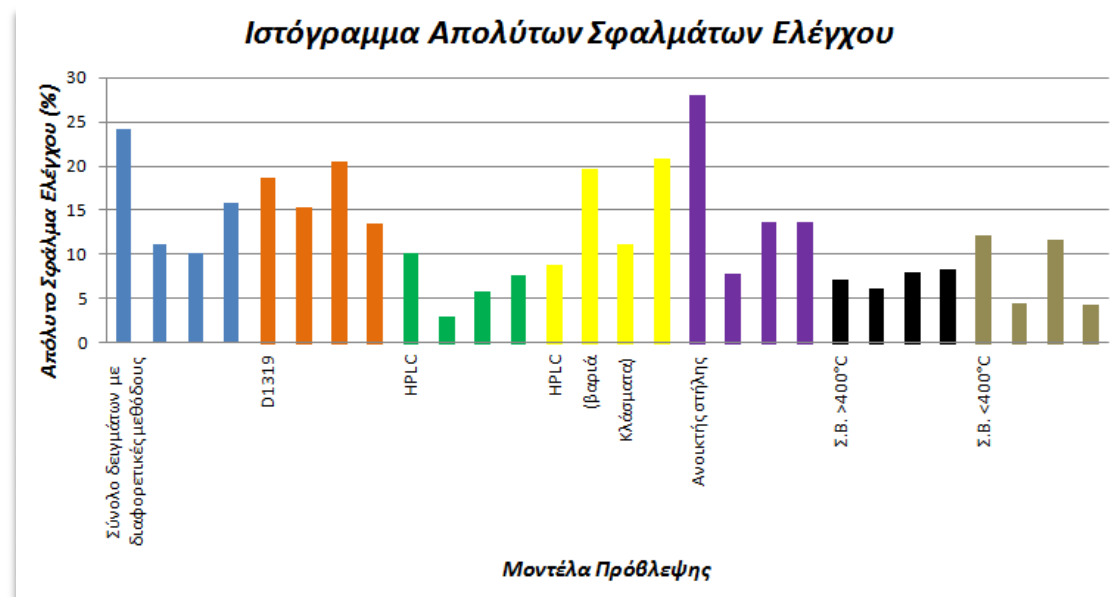
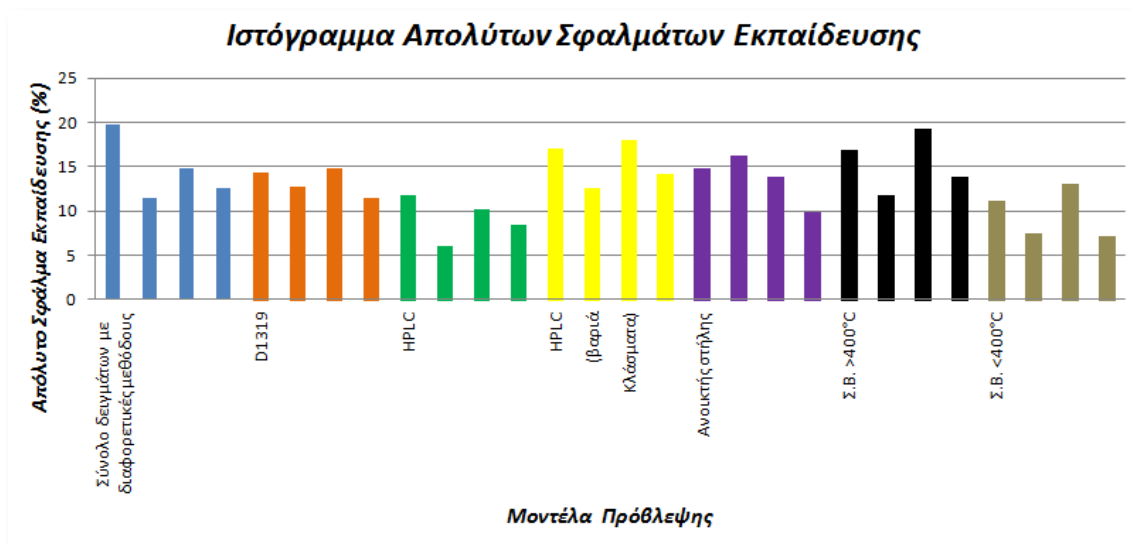
Σχήμα 5.7.18: Δείγματα δοκιμής



Σχήμα 5.7.19: Κατανομή ποσοστού αρωματικών και σφάλματος των δειγμάτων δοκιμής.

Ακολουθεί συγκεντρωτικός πίνακας των αποτελεσμάτων καθώς και ιστογράμματα των απολύτων σφαλμάτων πρόβλεψης των δειγμάτων εκπαίδευσης αλλά και ελέγχου για κάθε μοντέλο.

Μέθοδος/Μοντέλο	Μοντέλο	Αρ.Δειγμ.	Προεπεξερ.	Τοπ.Απόκ.	Απολ. Σφ.Εκπαίδ. (%)	Απολ. Σφ.Πρόβλ. (%)	RMSEP	Αρ.Παραγ.
Σύνολο δειγμάτων με διαφορετικές μεθόδους	<i>PCR1</i>	109	-	20.51	19.78	24.25	0.102	9
	<i>PLS1</i>	116	-	20.40	11.40	11.10	0.050	9
	<i>PCR2</i>	108	Ομαλοπ.	20.39	14.75	10.15	0.040	9
	<i>PLS2</i>	116	Ομαλοπ.	20.40	12.50	15.80	0.064	9
D1319	<i>PCR3</i>	42	-	5.97	14.40	18.60	0.340	9
	<i>PLS3</i>	44	-	6.30	12.80	15.30	0.127	9
	<i>PCR4</i>	42	Pre-max	5.97	14.80	20.50	0.206	9
	<i>PLS4</i>	44	Ομαλοπ.	6.30	11.50	13.50	0.108	9
HPLC	<i>PCR5</i>	66	-	18.00	11.80	10.20	0.035	9
	<i>PLS5</i>	69	-	17.70	6.00	3.05	0.013	9
	<i>PCR6</i>	66	Ομαλοπ.	18.00	10.20	5.80	0.054	9
	<i>PLS6</i>	69	Pre-max	17.70	8.40	7.70	0.035	9
HPLC (βαριά Κλάσματα)	<i>PCR7</i>	50	-	23.60	17.00	8.90	0.090	9
	<i>PLS7</i>	51	-	23.70	12.50	19.60	0.066	9
	<i>PCR8</i>	46	Pre-max	23.80	18.00	11.10	0.070	9
	<i>PLS8</i>	51	Ομαλοπ.	23.70	14.20	20.90	0.075	9
Ανοικτής στήλης	<i>PCR9</i>	61	-	21.50	14.80	28.00	0.059	9
	<i>PLS9</i>	67	-	21.80	16.20	7.80	0.056	9
	<i>PCR10</i>	61	Ομαλοπ.	21.50	13.90	13.60	0.093	9
	<i>PLS10</i>	67	Ομαλοπ.	20.20	9.85	13.70	0.066	9
Σ.Β. >400°C	<i>PCR11</i>	41	-	23.60	16.90	7.20	0.047	9
	<i>PLS11</i>	43	-	23.60	11.80	6.10	0.038	9
	<i>PCR12</i>	41	Pre-max	23.60	19.20	8.05	0.050	9
	<i>PLS12</i>	42	Ομαλοπ.	22.30	13.80	8.40	0.083	9
Σ.Β. <400°C	<i>PCR13</i>	69	-	8.90	11.20	12.20	0.063	9
	<i>PLS13</i>	71	-	8.80	7.50	4.50	0.027	9
	<i>PCR14</i>	69	Pre-max	8.90	13.00	11.70	0.074	9
	<i>PLS14</i>	69	Ομαλοπ.	8.90	7.10	4.30	0.020	9



6. ΑΞΙΟΛΟΓΙΣΗ ΜΟΝΤΕΛΩΝ-ΣΥΜΠΕΠΑΣΜΑΤΑ

Τα μοντέλα που κατασκευάστηκαν δείχνουν να εκπαιδεύονται. Αυτό σημαίνει πως υπάρχει πληροφορία στο φάσμα παρά τις ευρείς κατανομές σφαλμάτων, που προκύπτουν από τις λιγότερο αξιόπιστες χρωματογραφικές αναλύσεις που μετρούν το ποσοστό των αρωματικών συστατικών.

Πιο συγκεκριμένα σε μοντέλα που χρησιμοποιήθηκε το σύνολο των δειγμάτων κατά την εκπαίδευση, ή ένα μεγάλο μέρος τους, υπήρχαν μεγάλα ποσοστά σφάλματος και αυτό συνέβη γιατί υπήρχαν διαφορετικοί τύποι δειγμάτων, τα οποία είχαν αναλυθεί με διαφορετικές μεθόδους. Ο τρόπος που ορίζει η κάθε αναλυτική μέθοδος την ομάδα αρωματικών συστατικών είναι διαφορετικός διότι κάθε μια εξ αυτών στηρίζεται σε συγκεκριμένο μηχανισμό προσδιορισμού.

Αναμενόμενο ήταν στα βαριά κλάσματα να είναι δυσκολότερος ο διαχωρισμός των αρωματικών συστατικών από τα μη αρωματικά καθώς τα πρώτα εμφανίζονται σε μεγάλα μόρια που είναι συνδεδεμένα με μακρυές παραφινικές αλυσίδες και ναφθενικούς δακτυλίους που καθιστούν δύσκολο τον σαφή καθορισμό αυτών. Επιπλέον σε περιπτώσεις που υπάρχουν πολλές ακραίες τιμές η ποιότητα των αναλυτικών δεδομένων μειώνεται.

Αποτέλεσμα των παραπάνω ήταν να μην γίνει εφικτή η δημιουργία ενός καθολικού μοντέλου για όλους τους τύπους δειγμάτων και για όλες τις αναλυτικές μεθόδους.

Επισημαίνεται πως τα καλύτερα μοντέλα (εκπαίδευσης – ελέγχου) ήταν τα μοντέλα που αναπτύχθηκαν με βάση τα ελαφριά κλάσματα με σημείο βρασμού μικρότερο των 400°C καθώς και τα μοντέλα που αναπτύχθηκαν με βάση τα δείγματα που είχαν αναλυθεί με τις πρότυπες μεθόδους ASTM HPLC. Εν πολλύς αναμενόμενο να είναι τα παραπάνω τα καλύτερα από τα μοντέλα διότι στα ελαφρα κλάσματα ο διαχωρισμός αρωματικών και μη αρωματικών είναι σαφής και οι πρότυπες HPLC ως μέθοδοι ενόργανης ανάλυσης είναι καλά προτυποποιημένες και δίνουν ακριβή και επαληθεύσιμα αποτελέσματα. Αντίθετα οι μέθοδοι ανοιχτής στήλης ειδικά εφαρμοζόμενες σε βαριά κλάσματα αναμένεται να είναι λιγότερο ακριβής.

Τέλος αξίζει να σημειωθεί πως με μικρότερο αριθμό δειγμάτων ιδίου τύπου είναι εφικτό να αναπτυχθούν μοντέλα πιο εξειδικευμένα, δηλαδή ικανά να προβλέπουν ικανοποιητικά, ποσοστά αρωματικών συστατικών για δείγματα ιδίου τύπου με αυτά της εκπαίδευσης. Ενώ για να δημιουργηθεί ένα καθολικό μοντέλο πρόβλεψης θα πρέπει να υπάρχουν πολύ περισσότερα δείγματα διαφορετικού τύπου για την εκπαίδευσή του.

Βιβλιογραφία

- [1] Βαλαβανίδης Αθ.Π. : Φασματοσκοπία Οργανικών Ενώσεων (2006).
- [2] Βαλαβανίδης Αθ.Π. : Βασικές Αρχές Μοριακής Φασματοσκοπίας Και Εφαρμογές Στην Οργανική Χημεία (1990).
- [3] Μίτζα Ανδριαννα : Μελέτη της Επίδρασης της Μίξης Κλασμάτων στη Διαμόρφωση του Αριθμού Οκτανίων της Βενζίνης (2006).
- [4] Πασαδάκης Νικολαος : Ορυκτά Καύσιμα (2010).
- [5] Πασαδάκης Νικολαος: Φυσικοχημικός Χαρακτηρισμός Ενεργειακών Πρώτων Υλών (2004).
- [6] Πασαδάκης Νικολαος: Οργανική Γεωχημεία Ιζηματογενών Πετρωμάτων (2004).
- [7] Παπαγεωργίου Κωνσταντίνος : Spectral Analysis & classification of oil products using near-IR.
- [8] Σουρλίγκας Σταύρος : Πρόβλεψη Ιδιοτήτων του Ντήζελ με Χημειομετρική Επεξεργασία Υπερύθρων Φασματοσκοπικών Δεδομένων (2004).
- [9] Σταθερόπουλος Μ., Παππά Α.: Χημειομετρία, Σημειώσεις Μαθήματος, Θέματα Σύγχρονης Αναλυτικής Χημείας, ΕΜΠ (1998).
- [10] Χατζάκης Γ. , Ραυτόπουλος Δ. : Πρόβλεψη Ιδιοτήτων Μεταχειρισμένων Λιπαντικών Με FTIR.
- [11] Χατζηιωάννου Θ.Μ., Κούππαρη Μ.Α.: «Ενόργανη Ανάλυση», Πανεπιστήμιο Αθηνών (1997).
- [12] Clifford Jones : Hydrocarbons, Physical Properties and their Relevance to Utilisation (2010).
- [13] Dyroff V. George : Manual on Significance of Petroleum Products (1989).
- [14] Hanselman Duane & Littlefield Bruce : The Student Edition of Matlab (1997).
- [15] Fodor G., Kohl K.: Analysis of middle distillates fuels by midband infrared Spectroscopy (1993).
- [16] Geladi P., Kowalski B : Partial least-squares regression: A tutorial», Laboratory for chemometrics and center for process analytical chemistry, University of Washington, Analytica Chimica Acta, Volume 185(1986).
- [17] Geladi P. Martens H.: A calibration tutorial for spectral data. Part 1. Data pretreatment and principal component regression using Matlab, Journal of near

infrared spectroscopy, Volume 4, Pages 225-242 (1996), NIR Publications 1996

[18] Geladi P. Martens H., Hadjiiski L., Hopke P.: «A calibration tutorial for spectral data. Part 1. Partial least squares regression using Matlab and some neural network results», Journal of near infrared (1996)

[19] Kemp W : «Organic spectroscopy», The Macmillan Press, London (1991).

[20] Lindsay I Smith : A tutorial on Principal Components Analysis (2002).

[21] Massart D.L, B.G.M. Vandeginste, S.M. Deming, Y. Michotte, L. Kaufman : Chemometrics: a textbook.

[22] Pecsock R., Shields D., Cairns T., McWilliam I: «Σύγχρονες μέθοδοι στη χημική ανάλυση», Απόδοση στα Ελληνικά: Σ. Βολιώτης (1980).

[23] Rasmus Bro : Multivariate calibration. What is in chemometrics for the analytical chemist. Rasmus Bro. . Analytica Chimica Acta 500 (2003).

[24] Riazi M.R. : Characterization and Properties of Petroleum Fractions (2005).

[25] Shlens j. : A Tutorial on Principal Component Analysis (1993).

[26] Speight James G. :The chemistry and technology of petroleum Second edition, revised and expanded (2002).

[27] Speight James G. : Handbook of Petroleum Analysis (2001).

[28] Stuart Barbara H. : Infrared spectroscopy: fundamentals and applications (2004).

[29] Wold S., Esbensen K., Geladi P.: Principal component Analysis, Chemometrics and Intelligent Laboratory Systems, Volume 2 (1987).

[30] Xin Bao, Liankui Dai :Partial least squares with outlier detection in spectral analysis: A tool to predict gasoline properties (2009).

[31] American Society for Testing and Materials (ASTM): Annual book of ASTM standards (1995).

[32] Agilent technologies (2003):

<http://www.plusources.com/upfile/proimage/20131271757066808.pdf>

<http://www.chem.agilent.com/Library/datasheets/Public/820623-001d.pdf>

[33] Aromatic Hydrocarbon. Retrieved from Wikipedia.org :

http://en.wikipedia.org/wiki/Aromatic_hydrocarbon

[34] MATLAB : «Neural Network Tutorial », Matlab Help Documentation, Version 6.5 R13