



### Τεχνική Αναφορά ΤΔ004.01

## Κατάστρωση χημειομετρικών μεθοδολογιών ποσοτικής ανάλυσης Βασικές αρχές, ορθή πρακτική και συνήθη σφάλματα

Η ανάπτυξη προσιτών και αξιόπιστων τεχνικών δονητικής φασματοσκοπίας (ATR-IR, NIR, Raman, κλπ) επιτρέπει την αξιοποίηση των φασμάτων για την ποσοτική ανάλυση ή τον προσδιορισμό ιδιοτήτων των υλικών που μελετώνται, παρακάμπτοντας την εφαρμογή χρονοβόρων και δαπανηρών αναλυτικών μεθόδων. Στην απλούστερη περίπτωση, το προς ανάλυση συστατικό του δείγματος χαρακτηρίζεται από τουλάχιστον μία φασματική ταινία, που είναι αρκούτσως ισχυρή και δεν επικαλύπτεται από τις άλλες ταινίες του φάσματος, ούτε αλληλεπιδρά με αυτές. Η ένταση της ταινίας (που εκφράζεται με την απορρόφηση στο χαρακτηριστικό μήκος κύματος) εξαρτάται γραμμικά από την συγκέντρωση σύμφωνα με τον γνωστό νόμο Beer-Lambert

$$A = \epsilon bc$$

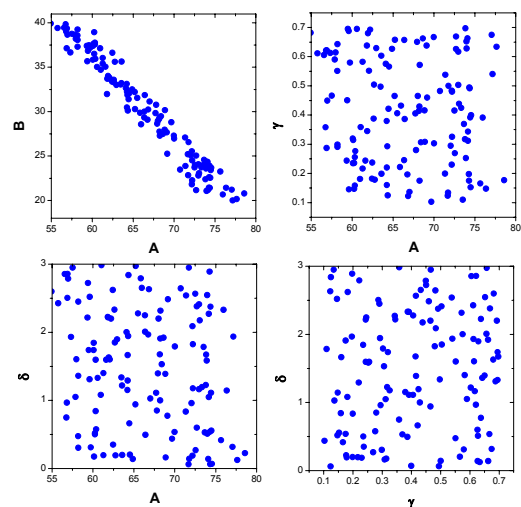
όπου **A** η μετρούμενη απορρόφηση,  $\epsilon$  ο συντελεστή απορρόφησης της συγκεκριμένης ταινίας, **b** το πάχος του δείγματος και **c** η συγκέντρωση του συστατικού που απορροφά. Η μεθοδολογία ολοκληρώνεται με την βαθμονόμηση του συντελεστή  $\epsilon$  με την βοήθεια ενός ή περισσοτέρων δοκίμιων αναφοράς.

Όστοςο συνηθέστατα τα δεδομένα και οι απαιτήσεις είναι συνθετότερα. Το υλικό μπορεί να αποτελείται από περισσότερες πρώτες ύλες, οι χαρακτηριστικές ταινίες των οποίων επικαλύπτονται σε βαθμό που δεν είναι δυνατή η διάκρισή τους. Ακραία, κάθε σημείο του φάσματος ενσωματώνει πληροφορία από το σύνολο των συστατικών του υλικού, τα οποία επιπλέον μπορεί να αλληλεπιδρούν μεταξύ τους. Η βαθμονόμηση απαιτεί πλέον **n** δοκίμια αναφοράς (όπου **n** πολύ μεγαλύτερο του αριθμού των συστατικών), παύει να είναι μονοδιάστατη και απαιτεί χημειομετρική επίλυση σε χώρο πολλών διαστάσεων.

Εστω ότι αναζητείται φασματοσκοπική χημειομετρική μεθοδολογία ποσοτικής ανάλυσης μιας ομάδας προϊόντων τα οποία βασίζονται στην απλή μίξη 4 πρώτων υλών και διαφοροποιούνται μόνο ως προς την σχετική αναλογία τους. Όλες οι πρώτες ύλες διαθέτουν σύνθετη φασματοσκοπική υπογραφή. Στο

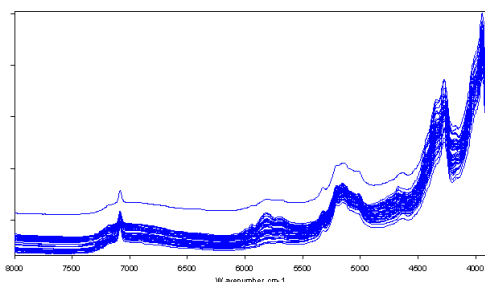
παράδειγμα, δύο από τις πρώτες ύλες (στο εξής **A** και **B**) βρίσκονται σε υψηλό ποσοστό, και το άθροισμά τους φτάνει το 97-99% κβ. Οι υπόλοιπες δύο ( $\gamma$  και  $\delta$ ) αποτελούν πρόσθετα που χρησιμοποιούνται σε πολυ μικρότερες συγκεντρώσεις, αλλά η παρουσία τους είναι καθοριστική για τις ιδιότητες των προϊόντων.

Είναι προφανές ότι η βαθμονόμηση δεν μπορεί να στηριχθεί αποκλειστικά σε δοκίμια πραγματικών προϊόντων, μια και αυτά αντιπροσωπεύουν εντοπισμένες «νησίδες» μέσα σε ένα ευρύ πεδίο μεταβολής των συγκεντρώσεων. Απαιτείται η παρασκευή επαρκούς αριθμού συνθετικών εργαστηριακών δειγμάτων, με αναλογίες συστατικών που δεν θα είναι εξαρτημένες μεταξύ τους. Στο συγκεκριμένο παράδειγμα, η προϋπόθεση αυτή μπορεί (και πρέπει) να τηρηθεί αυστηρά για τα πρόσθετα  $\gamma$  και  $\delta$ , αλλά οι συγκεντρώσεις των κύριων συστατικών **A** και **B** αναγκαστικά θα παρουσιάζουν υψηλή συσχέτιση (αφού το άθροισμά τους προσεγγίζει το 100%). Οι συνέπειες της συσχέτισης των συγκεντρώσεων **A** και **B** θα γίνουν εμφανείς αργότερα.



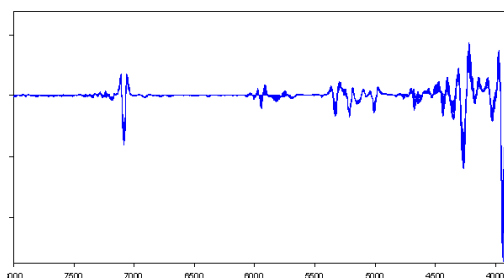
**Σχήμα 1.** Η συγκεντρώσεις των επι μέρους συστατικών στα δείγματα βαθμονόμησης πρέπει να είναι ανεξάρτητες μεταξύ τους. Στο παράδειγμα, αυτό επιτυγχάνεται για τα πρόσθετα  $\gamma$  και  $\delta$ , αλλά όχι για το ζεύγος των κύριων συστατικών **A** και **B**.

Το επόμενο στάδιο περιλαμβάνει την επιλογή της φασματοσκοπικής τεχνικής (με κριτήριο την ευαισθησία, την ταχύτητα κλπ) και την συλλογή των σχετικών φασμάτων (Σχήμα 2).



**Σχήμα 2.** Φάσματα απορρόφησης στο εγγύς υπέρυθρο (NIR) των δειγμάτων αναφοράς.

Συνηθέστατα, τα φάσματα που θα χρησιμοποιηθούν για την κατάστρωση της μεθοδολογίας PLS υφίστανται μαθηματική προεπεξεργασία (π.χ. υπολογισμός της 2<sup>ης</sup> παραγώγου όπως στο Σχ. 3) που αποσκοπεί στην αφαίρεση τυχαίων χαρακτηριστικών και τον εμπλουτισμό τους με πληροφορία σχετική με τον ζητούμενο προσδιορισμό. Σε κάθε περίπτωση η προεπεξεργασία που θα επιλεγεί πρέπει να διατηρεί επαρκή πληροφορία για το σύνολο των συστατικών που ενδιαφέρουν.



**Σχήμα 3.** Η 2<sup>η</sup> παράγωγος των φασμάτων του Σχ. 2 χρησιμοποιείται για να εξαλείψει μεταβολές της απορρόφησης υποβάθρου (background) που οφείλονται σε μικροδιαφορές στην τοποθέτηση του δείγματος στο φωτόμετρο.

Η κατάστρωση ποσοτικών μεθοδολογιών που θα επιτρέψει την πρόβλεψη της σύστασης με βάση τα φάσματα διευκολύνεται πλέον από την ύπαρξη προσιτών πακέτων λογισμικού, ορισμένα εκ των οποίων ενσωματώνονται στο λειτουργικό λογισμικό των φωτομέτρων. Με αναφορά στην παρακάτω σχηματική εξίσωση,

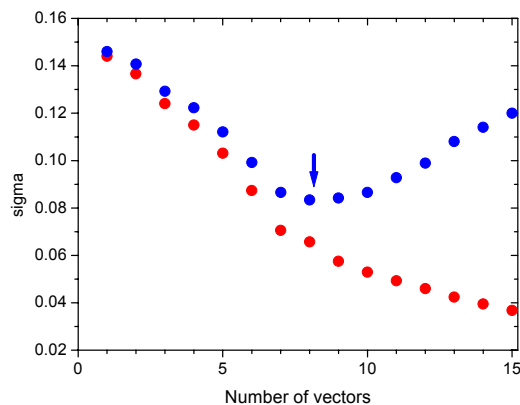
$$\begin{matrix} 1 \\ \boxed{y} \\ n \end{matrix} = \begin{matrix} m \\ \boxed{X} \\ n \end{matrix} \begin{matrix} 1 \\ \boxed{b} \\ m \end{matrix} + \begin{matrix} 1 \\ \boxed{e} \\ n \end{matrix}$$

αναζητείται ο αλγόριθμος **b**, ο οποίος εφαρμοζόμενος στον πίνακα **X** των φασματοσκοπικών δεδομένων (**n** φάσματα x **m** σημεία ανά φάσμα), αποδίδει τον πίνακα των γνωστών (ή ανεξάρτητα μετρημένων) συγκεντρώσεων **y**, ελαχιστοποιώντας το σφάλμα **e**.

Σε μία από τις κοινότερες χημειομετρικές μεθοδολογίες, γνωστή με τα αρχικά PLS (Partial Least Squares ή Projection to Latent Spectra), η επίλυση περιλαμβάνει την ανάπτυξη του πίνακα **X** σε γραμμικό συνδυασμό ανυσμάτων βάσης (**p<sub>i</sub>**), πολλαπλασιασμένους με κατάλληλους συντελεστές (**t<sub>i</sub>**):

$$\begin{matrix} m \\ \boxed{X} \\ n \end{matrix} = \begin{matrix} 1 \\ \boxed{t_1} \\ n \end{matrix} \begin{matrix} m \\ \boxed{p_1} \\ 1 \end{matrix} + \begin{matrix} 1 \\ \boxed{t_2} \\ n \end{matrix} \begin{matrix} m \\ \boxed{p_2} \\ 1 \end{matrix} + \dots + \begin{matrix} 1 \\ \boxed{t_r} \\ n \end{matrix} \begin{matrix} m \\ \boxed{p_r} \\ 1 \end{matrix}$$

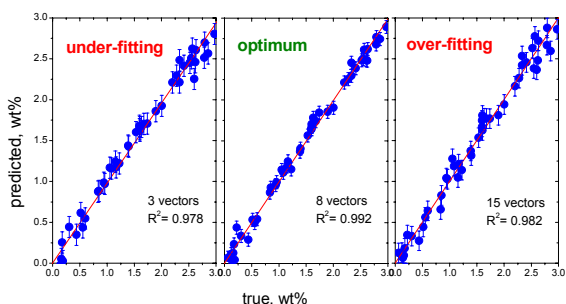
Μαθηματικά, το σφάλμα **e** μειώνεται αυξανόμενου του αριθμού των ανυσμάτων βάσης (Σχ. 4, κόκκινα σημεία) και μηδενίζεται ασυμπτωτικά όταν αυτός φτάσει τον αριθμό των δειγμάτων αναφοράς (**n**). Ωστόσο, μια τέτοια βαθμονόμηση είναι τετριμμένη, ασταθής και αναξιόπιστη διότι στηρίζεται στο σύνολο των φασματικών χαρακτηριστικών ανεξάρτητα από το αν αυτά σχετίζονται με την συγκέντρωση του συστατικού ή αποτελούν τυχαίο θόρυβο.



**Σχήμα 4.** Το σφάλμα της πρόβλεψης της συγκέντρωσης του συστατικού **δ** συναρτήσει του αριθμού των ανυσμάτων βάσης. Ο βέλτιστος αριθμός ανυσμάτων στην συγκεκριμένη περίπτωση είναι 8 (βέλος). Η καμπύλη με κόκκινο χρώμα αποτελεί τετριμμένη λύση.

Αναζητείται συνεπώς ο βέλτιστος αριθμός ανυσμάτων βάσης, αυτός δηλαδή που αξιοποιεί σε μέγιστο βαθμό την φασματική πληροφορία που είναι ουσιαστική για την πρόβλεψη (no underfitting), χωρίς να επηρεάζεται από εκείνη που δεν σχετίζεται με το ζητούμενο (no overfitting). Η εύρεση του βέλτιστου αριθμού ανυσμάτων προϋποθέτει

τον διαχωρισμό των δειγμάτων / φασμάτων αναφοράς σε δύο διακριτά υποσύνολα βαθμονόμησης και ελέγχου. Αυτό εξασφαλίζει ότι τα τυχαία χαρακτηριστικά των φασμάτων ελέγχου δεν θα συμπεριληφθούν στα κριτήρια (ανύσματα) της πρόβλεψης και έχει σαν συνέπεια την μη γραμμική εξάρτηση του σφάλματος της πρόβλεψης από τον αριθμό των ανυσμάτων βάσης (Σχ. 4, μπλέ σημεία). Ο βέλτιστος αριθμός ανυσμάτων προσδιορίζεται με σαφήνεια από τη θέση του ελαχίστου σφάλματος. Στην ίδια θέση, ο συντελεστής συσχέτισης μεταξύ πραγματικών και προβλεπόμενων συγκεντρώσεων του συστατικού  $\delta$  παρουσιάζει μέγιστο (Σχ. 5).



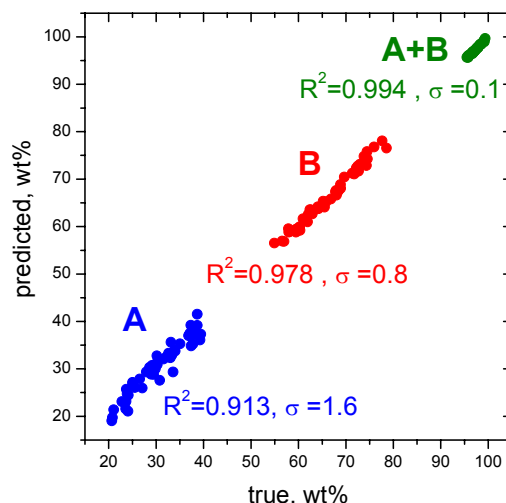
**Σχήμα 5.** Συσχετισμός προβλεπόμενων και πραγματικών τιμών της συγκέντρωσης του συστατικού  $\delta$  με 3, 8 και 15 ανύσματα βάσης. Είναι φανερή η βελτιστοποίηση του συσχέτισης με ενδιάμεσο αριθμό ανυσμάτων βάσης (πρβλ. Σχ. 4).

Εντελώς παρόμοια εικόνα προκύπτει από την μεθοδολογία για τον προσδιορισμό της συγκέντρωσης του συστατικού  $\gamma$ : Αυτή βελτιστοποιείται με 6-διάστατη ανυσματική βάση επιτρέποντας πρόβλεψη με σφάλμα  $\sigma=0.03$  % κβ και συντελεστή συσχέτισης  $R^2=0.974$ .

Ενδιαφέρον παρουσιάζει η απόπειρα κατάστροφης μεθοδολογίας για τα κύρια συστατικά **A** και **B**, η συγκέντρωση των οποίων παρουσιάζει, όπως είδαμε, υψηλό συσχέτισμο. Ο αλγόριθμος PLS δεν δυσκολεύεται να προβλέψει τις συγκεντρώσεις των δύο συστατικών με φαινομενικά καλή ακρίβεια (Σχ. 6, μπλέ, κόκκινο). Αυτό που δεν

είναι προφανές στον μη εξοικειωμένο χρήστη είναι ότι οι ανυσματικές βάσεις που αναπτύσσονται για τα συστατικά αυτά είναι, επίσης, ισχυρά συσχετισμένες μεταξύ τους. Το αποτέλεσμα είναι ότι οποιοδήποτε σφάλμα στον προσδιορισμό του ενός συστατικού (π.χ. υποεκτίμηση της συγκέντρωσης του **A**), «διαρρέει» στον προσδιορισμό του άλλου (οδηγώντας στην υπερεκτίμηση της συγκέντρωσης του **B**).

Αντίθετα, το άθροισμα των συγκεντρώσεων **A+B** δεν παρουσιάζει συσχέτισμο με τις συγκεντρώσεις των συστατικών μειοψηφίας  $\gamma$  και  $\delta$ , και μπορεί να αντιμετωπισθεί σαν ανεξάρτητη μεταβλητή. Στο Σχ. 5 συγκρίνονται οι προβλέψεις που αφορούν τις συγκεντρώσεις **A**, **B** και **A+B**. Παρά το μικρό εύρος των τιμών **A+B**, η πρόβλεψη παρουσιάζει εξαιρετικά υψηλό  $R^2$  και σφάλμα (0.1 %κβ) πολύ χαμηλότερο από τα αντίστοιχα σφάλματα των επι μέρους συστατικών (1.6 και 0.8 %κβ.), επιτρέποντας έμμεσα τον ακριβέστατο προσδιορισμό της συνολικής συγκέντρωσης των προσθέτων  $\gamma+\delta$ .



**Σχήμα 6.** Συντελεστές συσχέτισης για την χημειομετρική πρόβλεψη της συγκέντρωσης των κύριων συστατικών **A** και **B**, και του αθροίσματός τους **A+B**.

Το Εργαστήριο Φασματοσκοπικών Εφαρμογών του ΙΟΦΧ/ΕΙΕ παρέχει ερευνητικές υπηρεσίες σε βιομηχανικούς φορείς που περιλαμβάνουν μεταξύ άλλων: τον χαρακτηρισμό (ταυτοποίηση, conformity) πρώτων υλών, ενδιάμεσων και προϊόντων με γρήγορο και μη καταστρεπτικό τρόπο, ανά παρτίδα, ακόμα και ανά περιέκτη, την ανάπτυξη ειδικών μεθοδολογιών ποιοτικού & ποσοτικού ελέγχου που στηρίζονται σε φασματοσκοπικές τεχνικές (συνήθως φασματοσκοπία NIR) και χημειομετρία, καθώς και τον σχεδιασμό εμπειρών αυτοματοποιημένων συστημάτων για την on line παρακολούθηση της παραγωγής. Επίσης παρέχει συμβουλευτικές υπηρεσίες για την διείσδυση σε νέες αγορές μέσω της ανάπτυξης προϊόντων και διεργασιών υψηλής προστιθέμενης αξίας, συμμετέχει σε δράσεις τεχνομεταιας, εκπαίδευσης και κατάρτισης προσωπικού, και αναλαμβάνει συμβουλευτικό ρόλο στην αξιολόγηση νέων τεχνολογιών ή αποτελεσμάτων τρίτων. Για τις υπηρεσίες αυτές το Εργαστήριο αναζητεί, και σε πολλές περιπτώσεις δημιουργεί, την ζήτηση της αγοράς μέσω των συστηματικών επαφών με στελέχη του βιομηχανικού R & D, της υλοποίησης διερευνητικών έργων, της ανάληψης έργων μεσοπρόθεσμης διάρκειας και, τέλος, της ανάληψης υπεργολαβικών μακροχρόνιας συνεργασίας.

Για περισσότερες πληροφορίες σχετικά με τις βιομηχανικές εφαρμογές της δονητικής φασματοσκοπίας επισκεφθείτε την ιστοσελίδα <http://www.eie.gr/nhrf/institutes/tpci/researchteams/mspc/mspc-asl-gr.html> ή επικοινωνήστε με τους ερευνητές του ΕΙΕ Β. Γκιώνη (210-7273820, [vagonis@eie.gr](mailto:vagonis@eie.gr)) ή Γ. Χρυσικό (210-7273819, [gdchryss@eie.gr](mailto:gdchryss@eie.gr))