

Strategie di ottimizzazione di metodi cromatografici

Uso dell'analisi multivariata

di Maria Carla Gennaro, Emilio Marengo

La cromatografia liquida, con la sua notevole flessibilità, pone spesso problemi di scelta delle migliori condizioni operative, per ottenere un'efficace risoluzione di miscele di analiti. I parametri che devono essere fissati vanno dalla composizione della fase mobile, al suo pH, al flusso, a molti altri. I migliori metodi per affrontare questo problema sono costituiti dalle tecniche di ottimizzazione fornite dall'approccio chemiometrico. Nel presente articolo vengono proposti alcuni esempi di applicazioni di tecniche di ottimizzazione sia continua (metodi di pianificazione sperimentale e reti neurali), sia discontinua (metodo del Simplex), alla risoluzione di problemi cromatografici.



Hewlett-Packard

S spesso scorrendo metodi di cromatografia liquida riportati in letteratura si rimane colpiti dalla complessità della composizione di alcune fasi mobili, che sembrano quasi ricette magiche: non solo è elevato il numero dei componenti, ma anche molto variabili le concentrazioni relative.

Questa complessa composizione, in contrasto con il suggerimento generale che vale per tutti i "cromatografi" che propongono nei loro metodi fasi mobili troppo complesse, "KISS" ("Keep It Simple, Stupid!"), deriva verosimilmente dal fatto che spesso ci si ispira, nella messa a punto di una separazione, ad un metodo proposto in letteratura. Il metodo viene poi adattato, per migliorarne la risoluzione o per adeguarlo al caso in studio, aggiungendo qualche nuovo componente: si ottimizza la concentrazione di quest'ultimo, lasciando però inalterata la composizione precedente, sia dal punto di vista qualitativo sia quantitativo. Tutti i componenti erano davvero necessari e le concentrazioni forse migliorabili? Si incontrano talora, tra chi lavora in cromatografia liquida, degli scienziati spesso anche dotati di invidiabili "mani d'oro" in grado di garantire sempre dati di elevatissima qualità, che hanno anche doti divinatorie e formulano su due piedi con estrema disinvoltura composizioni di fasi mobili, che poi risultano subito vincenti. Ma... tutti gli altri? Devono ricorrere a procedure di ottimizzazione.

Un progetto di ottimizzazione di tipo OVAT (One-Variable-at-A-Time) che modifica in modo sistematico il valore di una variabile tenendo fissi i valori di tutte le altre e va a vederne l'effetto sulla risposta è un metodo che prevede anzitutto un numero molto elevato di esperimenti. Ciò non solo impiega tempo e denaro, ma inoltre corre il rischio che durante l'esecuzione degli esperimenti di ottimizzazione, che sistematicamente prendono in considerazione tutte le combinazioni possibili, il sistema in esame vari, ad esempio varino le prestazioni della colonna cromatografica, rendendo inutile il lavoro. Inoltre comunque non viene in

questo trattamento considerata la possibilità, spesso frequente, di interazioni tra variabili. Allo scopo di poter considerare, per una loro ottimizzazione, l'effetto delle variabili implicate in modo simultaneo, occorre far uso di metodi di analisi multivariata e in particolare di metodi di pianificazione degli esperimenti (o *experimental design*), che permettono di trarre informazioni sul comportamento del sistema da un numero abbastanza limitato di esperimenti rappresentativi, adeguatamente distribuiti nello spazio delle variabili. Il metodo trova buona applicazione in cromatografia liquida quando la fase mobile contiene più componenti o quando richiede ad esempio la correzione del pH; in gascromatografia può essere utile nella ottimizzazione delle reazioni di derivatizzazione pre-colonna. Anche il processo di estrazione dell'analita dal campione può essere avvantaggiato da una procedura di ottimizzazione della resa di estrazione. Utilizzando l'analisi multivariata vi sono due principali approcci che si possono seguire: l'ottimizzazione simultanea o quella discontinua. Nell'ottimizzazione simultanea si progetta un piano fattoriale in cui si valuta l'effetto dei fattori rilevanti che influiscono sul comportamento del sistema. Dalle risposte ottenute si costruisce un modello di regressione, che correla la risposta sperimentale (ad esempio il tempo di ritenzione) alle diverse variabili o a loro interazioni e che permette la scelta delle condizioni ottimali del metodo, la predizione di risultati in determinate condizioni ma anche suggerimenti sui meccanismi di ritenzione e sull'effetto di diversi fattori sulla risposta. Il metodo discontinuo invece tende ad ottenere un'ottimizzazione in tempi più rapidi ma non permette di costruire un modello che correla risposta sperimentale alle variabili: pertanto non fornisce informazioni sul sistema e sull'importanza relativa delle variabili che regolano il metodo.

Ottimizzazione mediante metodo discontinuo

L'applicazione di questo metodo di ottimizzazione presuppone che si abbiano condizioni di partenza già individuate e che tutte le variabili che si considerano siano importanti. Si eseguono al-

M.C. Gennaro, E. Marengo, Università del Piemonte Orientale "Amedeo Avogadro". gennaro@mfn.unipmn.it

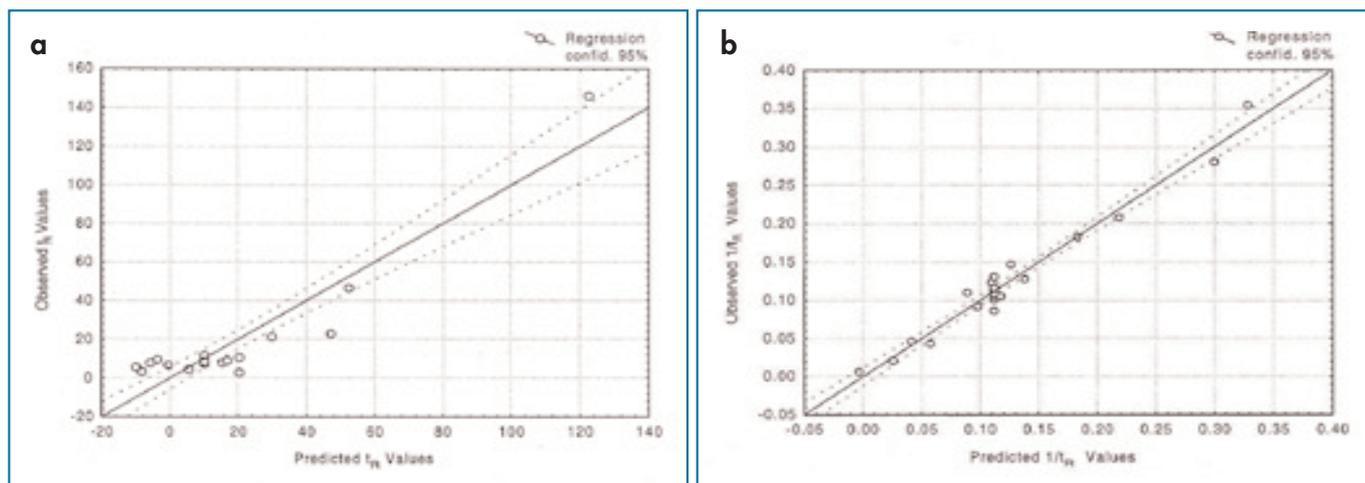


Figura 1 - Tempi di ritenzione osservati e predetti, a) senza applicare la trasformazione di Box-Cox, b) applicando la trasformazione di Box-Cox

cuni esperimenti in condizioni opportune e in funzione del loro risultato si progettano gli esperimenti successivi, in modo analogo a quando si procede in laboratorio a un'ottimizzazione del tipo "try and see". L'ottimizzazione secondo un metodo discontinuo più nota è quella basata sul *Simplex*. Se p è il numero di fattori sperimentali da ottimizzare si inizia da un set di $(p+1)$ esperimenti, detto appunto *Simplex*. L'ottimizzazione si basa sulle successive proiezioni dell'esperimento che dà la peggiore risposta rispetto al centroide degli altri. L'esperimento che dà la peggiore risposta è via via eliminato e sostituito dal nuovo esperimento programmato dalla proiezione. Si procede fino a convergenza, ossia quando, per eliminazione della peggiore risposta, il nuovo esperimento suggerito dalla proiezione non fornisce un risultato migliore. Questo metodo è stato utilizzato per l'ottimizzazione di un metodo in cromatografia in fase inversa con reagente di interazione ionica che si proponeva la separazione di 19 conservanti comunemente utilizzati in prodotti per l'igiene personale e la cosmesi e caratterizzati da struttura chimica molto diversa. Si è scelto questo metodo per ottenere una risposta più rapida rispetto al metodo di *experimental design*, dato l'elevato numero di analiti che si intendeva separare. Lo scopo era quindi di ottenere la massima risoluzione possibile tra tutte le coppie di picchi adiacenti: quindi la risposta da ottimizzare è composta di più singole risposte.

Una risposta multicriterio si può affrontare calcolando una funzione globale in grado di descrivere la qualità del risultato complessivo che tiene conto degli aspetti rilevanti del sistema: l'ottimizzazione di questa funzione fornisce le condizioni ottimali che contemporaneamente ottimizzano tutte le risposte. Si definisce una funzione di desiderabilità D :

$$D = \left(\prod_{i=1, n} d_i \right)^{\frac{1}{n}}$$

in cui è d_i la desiderabilità dell' i -esimo criterio ossia lo score normalizzato nell'intervallo tra 0 e 1, attribuito al risultato ottenuto per l' i -esimo criterio. La singola desiderabilità d_i è una equazione matematica, che correla i valori assunti dalla risposta i (risoluzione tra due picchi adiacenti) ad un punteggio normalizzato tra 0 e 1. Essendo D la media geometrica delle desiderabilità singole, assume il valore zero anche quando solo una di queste ha punteggio zero. Il massimo valore di D garantisce le migliori

condizioni del sistema rispetto a tutte le d_i . L'applicazione del metodo *Simplex* e della funzione di desiderabilità ci ha permesso la separazione di 14 conservanti dei 19 considerati [1].

Metodo continuo di ottimizzazione secondo pianificazione degli esperimenti

Il metodo di pianificazione degli esperimenti si basa sulla esecuzione di esperimenti pianificati a due livelli per ogni variabile e permette di valutare l'effetto dei fattori e di loro interazioni. Gli esperimenti richiesti sono 2^p dove p è il numero delle variabili che si intende considerare. Quando p sia molto elevato, si effettuano disegni frazionari che comportano un numero proporzionalmente minore di esperimenti. Attraverso il calcolo di una t -test è possibile individuare le variabili e le interazioni tra variabili che sono significative e individuare invece le variabili che non influenzano sulla risposta, almeno nel dominio sperimentale considerato, come ad esempio la presenza di un componente della fase mobile, la correzione della temperatura o del pH. Un aspetto non secondario è spesso quello di una semplificazione del metodo, quando questo venga trasferito alla fase applicativa. Prima di porre interpretazioni sui fattori rilevanti è necessario valutare l'abilità del modello a descrivere i dati sperimentali: questa operazione può essere eseguita attraverso un sistema *leave-one-out* in cui alternativamente ogni esperimento è lasciato da parte e la sua risposta calcolata mediante un modello elaborato sulle risposte degli altri esperimenti, oppure attraverso l'analisi dei residui. Occorre ancora verificare la linearità del si-

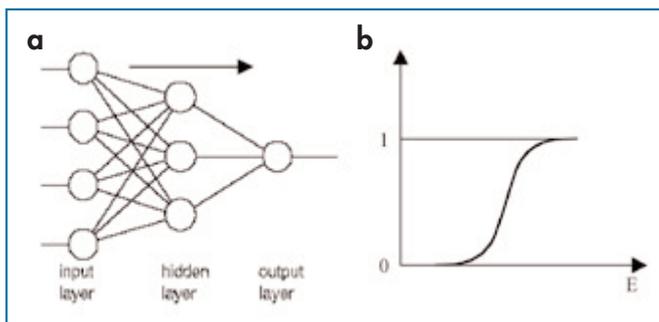


Figura 2 - a) Schema di rete neurale back-propagation e b) tipica funzione di trasferimento di tipo sigmoide

stema per escludere possibili effetti quadratici, effettuando un *F-test* che confronta la differenza tra risultati sperimentali e calcolati al centro del dominio sperimentale. Se si osservano deviazioni significative si deve procedere all'esecuzione di altri esperimenti addizionali (generalmente secondo uno star-design) che prevedano almeno tre livelli per ogni variabile.

Questi passi successivi dell'ottimizzazione secondo pianificazione sperimentale sono ormai praticamente codificati e descritti in vari testi [2-5]. Tuttavia possono verificarsi casi particolari che essenzialmente dipendono dal singolo sistema chimico considerato in cui questi metodi presentano dei limiti. Nei paragrafi seguenti verranno riportati due esempi tipici.

Trasformazione di Box-Cox

Nella separazione di una miscela di 21 solfonati aromatici (benzen- e naftalen-solfonati diversamente sostituiti) in cromatografia IIR-RP-Hplc [6], le risposte ottenute in termini di tempi di ritenzione nelle diverse condizioni sperimentali previste dal disegno, non mostravano una distribuzione omogenea a causa dell'ampio intervallo di variazione, come mostrato in Figura 1a, che riporta a titolo di esempio per il 4-ammino-1-naftalensolfonato, i tempi di ritenzione sperimentali in funzione di quelli calcolati. Mentre le coordinate spaziali nel dominio delle variabili in cui effettuare gli esperimenti vengono scelte, come suggerito dai piani fattoriali, in modo omogeneo e rappresentativo, le relative risposte dipendono dal sistema chimico. Nel caso presente pertanto il modello più che su una serie di dati che coprono un certo intervallo, sembra essere stato costruito su due dati. Per ovviare a questa difficoltà, si è ricorsi ad una trasformazione non lineare della risposta secondo l'equazione di Box-Cox, nella quale facendo variare un unico parametro λ , si può valutare l'effetto di più trasformazioni e trovare quella più adatta al caso specifico, che può essere una trasformazione $y^{-\lambda}$ (per $\lambda < 0$), $\log y$ (for $\lambda = 0$), to y^{λ} (per $\lambda > 0$). Il valore ottimale di λ è quello che rende minima la somma dei residui. La equazione di trasformazione è:

$$y^{(\lambda)} = \frac{y^{\lambda} - 1}{\lambda \cdot y^{\lambda-1}}$$

dove \bar{y} è la media geometrica delle risposte, ossia

$$\bar{y} = (y_1 \cdot y_2 \cdot \dots \cdot y_n)^{1/n}$$

Verificata la bontà del modello (Figura 1b) si applica quindi la trasformazione inversa e i dati vengono espressi nella variabile originale, cioè i tempi di ritenzione. Usando i modelli di regressione è poi possibile identificare le condizioni ottimali di risoluzione della miscela.

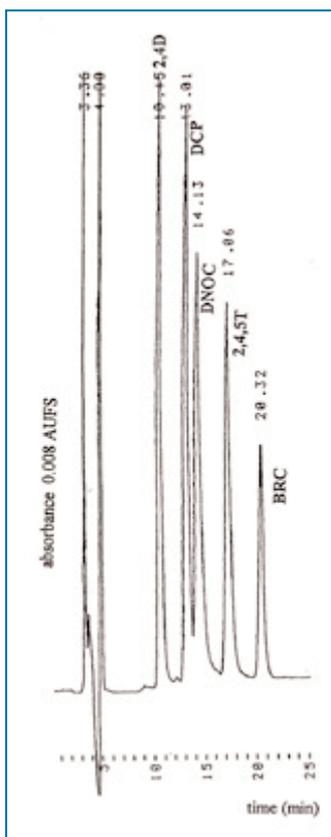


Figura 3 - Separazione di 5 pesticidi in cromatografia RP-Hplc con reagente di interazione ionica, ottenuta mediante l'ausilio delle reti neurali artificiali

Reti neurali artificiali

Un altro caso di ottimizzazione in cui la pianificazione degli esperimenti secondo piani fattoriali si è dimostrata non adeguata si è verificato nella separazione IIR-RP-Hplc di una miscela di pesticidi con proprietà acide, basiche e neutre [7]. Mentre le specie più polari ovviamente risentono di variazioni di pH e di concentrazione del reagente di interazione, la specie meno polare veniva influenzata solo dalla percentuale di solvente organico nella fase mobile e, se questa non variava, il suo tempo di ritenzione rimaneva invariato. Questo è in accordo con il fatto che il meccanismo di ritenzione delle specie è diverso, prevalentemente di interazione per le specie polari, di fase inversa per la specie meno polare. Evidentemente il metodo di pianificazione degli esperimenti non è sufficientemente flessibile. Per ottenere l'ottimizzazione della separazione è stato infatti necessario utilizzare un algoritmo basato sull'uso delle reti neurali artificiali. In particolare si è usato un algoritmo in *back propagation*. In uno strato interno ogni neurone è associato a una variabile, quindi si considerano strati di neuroni intermedi (i cosiddetti *hidden layers*) e uno strato esterno in cui ogni neurone è associato a una risposta (Figura 2). Il segnale si muove dallo strato più interno verso l'esterno, trasferendo via via ai neuroni degli strati successivi una porzione di segnale che ha accumulato. La porzione che viene trasferita è regolata da una funzione di trasferimento simile a quella riportata in Figura 2.

Il segnale viene accumulato applicando un peso moltiplicatore w : i pesi sono ottimizzati durante il processo detto di training; l'algoritmo di *back-propagation* permette la ottimizzazione successiva dei pesi associati a ogni coppia di neuroni connessi finché la rete è in grado di fornire il corretto out-put per quel determinato input. In questo processo ogni esperimento del training-set è via via presentato alla rete e i pesi vengono corretti nella direzione di una diminuzione dell'errore. Le iterazioni vengono condotte fino ad avere risultati soddisfacenti in base all'errore ammesso come accettabile. Mediante l'uso di neuroni, che hanno imparato a comprendere come, cambiando alcune condizioni della fase mobile, un analita della miscela fosse più trattenuto, un altro meno trattenuto e uno non venisse influenzato, è stato possibile ottenere un buon metodo per la separazione simultanea di 5 pesticidi, di cui è riportato il cromatogramma in Figura 3.

Bibliografia

- [1] E. Marengo *et al.*, *Journal of Chromatography A*, 2001, **39** (8), 339.
- [2] G.E.P. Box *et al.*, *Statistics for Experimenters*, Wiley, New York, 1978.
- [3] S.N. Deming, S.L. Morgan, *Experimental Design: A Chemometric Approach*, 2nd Ed., Elsevier, Amsterdam, 1993.
- [4] D.L. Massart *et al.*, *Chemometrics a Textbook*, Elsevier, Amsterdam, 1988.
- [5] R. Carlson, *Optimization in Organic Synthesis*, Elsevier, Amsterdam, 1993.
- [6] E. Marengo *et al.*, *Chemometrics and Intelligent Laboratory System*, 2000, **53**(1-2), 56.
- [7] E. Marengo *et al.*, *Journal of Chromatography A*, 1998, **799**, 47.