

Metodologie chemiometriche e loro applicazioni allo studio della corrosione

G. Luciano, P. Traverso, P. Letardi

Scopo di questo articolo è introdurre il lettore all'applicazione di metodologie chemiometriche - nello specifico Design of Experiment (DoE), classificazione e tecniche di modelling - che correntemente non sono ampiamente diffuse nell'ambito dello studio della corrosione.

Attraverso la descrizione di alcuni esempi sarà dimostrato come queste metodologie possono essere applicate con l'intento di migliorare la qualità del workflow di un laboratorio di ricerca.

PAROLE CHIAVE: corrosione, modellazione, Principal Component Analysis (PCA), Response Surface Methodology (RMS), Partial Least Squares (PLS)

INTRODUZIONE

Il seguente articolo introduce allo studio di una selezione di metodologie chemiometriche: 1) Design of Experiment, 2) classificazione e 3) modelling. Attraverso alcuni esempi applicativi, l'obiettivo è quello di mostrare come l'utilizzo di queste metodologie possa essere applicato con profitto allo studio della corrosione. La Chemiometria è definita dalla IUPAC come: "La scienza che correla le misure svolte su un sistema o processo chimico allo stato del sistema grazie all'applicazione di metodi statistici o matematici". Esistono molte applicazioni che gravitano intorno a questo ambito, visto che l'applicazione di una corretta metodologia statistica al trattamento dei dati ottenuti con diverse sperimentazioni, dovrebbe essere una priorità per ogni laboratorio. Se ipotizziamo di stilare un elenco delle principali classi di metodologie che fanno parte della chemiometria, dobbiamo considerare almeno le seguenti quattro:

- 1) Signal processing
- 2) Design of Experiments (DoE) e Response Surface Modelling (RMS)
- 3) Pattern recognition
- 4) Calibration

Il signal processing, che riguarda l'analisi dei segnali ottenuti da strumentazioni analitiche, non è affrontato in questo articolo poiché la letteratura disponibile sull'argomento è già molto vasta [3]. La rassegna si concentrerà invece sulla descrizione delle metodologie di DoE, nonché sull'elaborazione dei dati ottenuti attraverso l'utilizzo di diverse tecniche di modelling (partial least squares) e classificazione, grazie all'analisi delle componenti principali (PCA).

DESIGN OF EXPERIMENTS

(DoE - Response Surface Modelling)

Nelle metodologie di ricerca classiche, l'obiettivo principale è quello di definire una legge che goda delle proprietà di una categoria assoluta ad un determinato livello di conoscenza: in altre parole, che sia corretta o non corretta.

In presenza di molteplici fattori il sistema diventa complesso a tal punto che non è facilmente ottenibile una descrizione matematica "esatta" che tenga conto di tutti i fattori in gioco. In questo caso si ricorre ad un modello polinomiale la cui qualità può

essere migliorata a seconda del grado utilizzato. Applicare questo genere di modelli permette di studiare problemi di ottimizzazione in cui interviene un elevato numero di fattori [4].

I principali campi di utilizzo del DoE sono:

- a) *Screening*. In questo tipo di design si cerca di individuare i fattori che maggiormente contribuiscono al successo di un dato processo. A questo scopo è principalmente utilizzato il factorial design [2].
- b) *Ottimizzazione*. È la metodologia più comunemente applicata all'analisi dei sistemi chimici. Il DoE è impiegato con l'obiettivo di trovare un optimum, ossia una risposta ottimale rispetto ai fattori considerati, nel tempo più breve possibile [2].
- c) *Risparmio di tempo*. Un'accurata preparazione di un esperimento permette di ridurre il numero delle prove da svolgere rispetto a quelle richieste applicando un approccio convenzionale "One Variable At Time (OVAT)". Questa è la caratteristica principale che giustifica il diffuso impiego del DoE in applicazioni industriali.
- d) *Modelling quantitativo*. Ogni qual volta è necessario ricavare un modello matematico dagli esperimenti svolti, l'applicazione del DoE permette di ottenere un modello qualitativamente superiore a quello che si può ricavare con un approccio convenzionale.

Se guardiamo alla corrosione come ad un "fenomeno di deterioramento delle caratteristiche di un materiale, usualmente un metallo, a seguito dell'interazione tra il materiale e l'ambiente" [5,6], processo che coinvolge contemporaneamente un'ampia varietà di parametri quali acidità, concentrazione di sostanze chimiche, temperatura, natura di rivestimenti protettivi, sostanze inibenti, ecc., e se si vuole individuare una possibile correlazione con i differenti tipi di risposta quali ad esempio, impedenza del sistema, perdita di massa o, ancora, variazione di alcune proprietà meccaniche, l'applicazione del DoE risulta essere una scelta ottimale, anche in previsione di uno studio più approfondito che segua un prescreening del sistema in esame.

Descriveremo di seguito due dei design più utilizzati, ovvero il factorial design e il fractional factorial design.

Factorial Design

L'approccio del design of experiment fattoriale non ha una diretta connessione con il fattoriale propriamente inteso (ad esempio 4!). Idealmente, in un disegno fattoriale sono prese in considerazione tutte le possibili combinazioni dei livelli dei fattori.

A fine puramente illustrativo, considereremo un disegno fatto-

Giorgio Luciano, Pierluigi Traverso, Paola Letardi
CNR- ISMAR, via De Marini 6, I-16149 Genova

riale con due fattori e due livelli per fattore, indicato come 2^2 . In generale, il numero totale delle misure è calcolato come s^k , dove s indica il numero dei livelli e k il numero dei fattori. Nel caso di due soli livelli indicheremo i valori come “low” e “high”, correlandoli ad un parametro quantitativo.

In un design 2^2 avremo quindi quattro combinazioni dei livelli “high” e “low” dei due fattori. Queste combinazioni sono convenzionalmente riportate usando la presenza (assenza) di lettere minuscole per rappresentare i livelli basso ed alto dei corrispondenti fattori designati dall'appropriata lettera maiuscola; ad esempio a indica che il fattore A è a livello massimo, mentre B al minimo. Quando tutti i livelli sono a livello minimo si usa la denominazione (1) che viene utilizzata anche in presenza di più di due fattori.

L'analisi della varianza per il modello in considerazione risulta essere:

$$Y_{ij} = \mu + A_i + B_j + (AB)_{ij} \quad i=1,2 \quad j=1,2 \quad (\text{eq.1})$$

dove Y_{ij} è la risposta quando l' i -esimo livello di A è usato in combinazione con il j -esimo livello di B , e “1” e “2” rappresentano per ciascun fattore il livello minimo e massimo, rispettivamente. Nell'equazione (1) non compare il termine di errore che andrebbe aggiunto in caso di ripetizione del design. In questo caso l'equazione (1) si trasforma come segue:

$$Y_{ij} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk} \quad i=1,2 \quad j=1,2 \quad k=1,2,\dots,c \quad (\text{eq.2})$$

In Tab. I, il livello massimo è indicato da (+1) e il livello più basso come (-1). Da notare che il prodotto vettoriale di ogni combinazione è 0, il che indica che è ortogonale. Questo implica che l'effetto di ciascun fattore ed il prodotto dei due fattori, possono essere stimati in modo indipendente. Gli effetti A e B sono gli effetti principali ed il prodotto tra i due (AB) è definito come l'effetto di interazione.

Fractional Factorial Design

Il valore numerico di 2^k aumenta rapidamente all'aumentare di k . Un disegno fattoriale 2^k ci permette di stimare 2^{k-1} effetti; ma quando k è maggiore di 3 o 4, gran parte di questi effetti sono interazioni di ordine superiore che potrebbero non risultare significative (usualmente le interazioni di ordine superiore risultano essere meno importanti); per questo motivo un full factorial design sarebbe dispendioso in termini di costo sperimentale. Un fractional factorial design, come ci suggerisce il nome, permette di eseguire un numero ridotto di esperimenti rispetto ad un full factorial design, con l'obiettivo di non “sprecare” esperimenti collocati in punti che permetterebbero di stimare effetti di interazioni molto probabilmente non significative.

Un fractional factorial design a due livelli è scritto nella forma generica 2^{k-p} , dove k denota il numero dei fattori e la frazione del design completo 2^k da svolgere è pari a $1/2^p$. La risoluzione del fractional factorial design indica gli effetti che possono essere stimati dal design. Risoluzione III indica che possono essere stimati solo gli effetti principali in assenza di interazioni. Risoluzione IV indica che sono stimabili solo gli effetti princi-

Combinazione	A	B	AB
(1)	-1	-1	+1
a	+1	-1	-1
b	-1	+1	-1
ab	+1	+1	+1

TAB. I **Factorial Design 2^2 .**
Factorial Design 2^2 .

pali in presenza di interazioni a due fattori (considerando che interazioni di maggior ordine non esistano) e di ordine V significa che possono essere stimati sia gli effetti principali che le interazioni a due variabili in assenza di interazioni di ordine maggiore di due. Ad esempio utilizzeremo la notazione 2^{6-2} per rappresentare un fractional factorial design che rappresenta $1/4$ del fully design 2^6 che ha risoluzione IV.

Ciò significa che considerare una minor rilevanza dell'interazione fra i diversi parametri potrebbe, in questo esempio, ridurre il numero di misure da eseguire da 64 a 16 [7]

Approssimazione attraverso Response Functions

Generalmente parlando, siamo interessati a trovare una relazione funzionale tra i parametri che abbiamo studiato ed il nostro sistema, collegando il valore di una risposta y (o la sua media) a delle variabili quantitative ξ attraverso una relazione del tipo:

$$E = \eta = f(\xi_1, \xi_2, \dots, \xi_k) = f(\xi) \quad (\text{eq.3})$$

Assumendo che la natura di queste funzioni non sia nota, possiamo utilizzare una approssimazione polinomiale. Per fare ciò è conveniente utilizzare non l'attuale misura numerica, ma delle variabili codificate x_i dove

$$x_i = (\xi_i - \xi_{i0}) / S_i$$

dove $(\xi_{i0} - S_i)$ è la zona di interesse e ξ_{i0} il centro della variabile di interesse.

Ad esempio se ξ_i è la temperatura e la nostra regione di interesse è $115 \pm 10^\circ\text{C}$, il valore codificato sarà $x_i = (\xi_i - 115) / 10$.

In generale otterremo quindi un polinomio (utilizzando le variabili codificate) che sarà la combinazione dei prodotti e delle potenze delle x . Possiamo pensare al nostro polinomio come ad una serie di Taylor, sviluppo della reale funzione teorica $f(\xi)$, che è troncata dopo n -esimo ordine. Da notare quindi che:

- maggiore è il grado dell'approssimazione e più la nostra serie si avvicinerà alla funzione teorica;
- minore è la regione R in cui utilizziamo l'approssimazione e migliore sarà l'approssimazione stessa.

Riportando in grafico le funzioni otteniamo un riscontro grafico come quello riportato in Fig. 1 e Fig. 2 [8].

Per meglio comprendere l'applicazione di uno dei più comuni tipi di DoE allo studio della corrosione ci riferiamo al lavoro di Hajeeh [9] nel quale gli autori hanno studiato la corrosione di tubi in lega di alluminio, ottone e acciaio al carbonio. Scopo della ricerca è stato quello di identificare, per ciascuna lega, le variabili che maggiormente influenzano il processo corrosivo, determinandone nel contempo l'ordine di grandezza e l'interazione tra le variabili maggiormente significative.

A questo scopo sono state considerate le seguenti variabili: temperatura, condizioni di aerazione oltre alla concentrazione di urea, inibitore, solfuri e cloruri.

Un'approccio di tipo “convenzionale” consisterebbe nel variare un singolo parametro alla volta, comportando quindi un consistente lavoro sperimentale, senza ottenere alcuna indicazione sull'interazione tra i diversi fattori coinvolti. Gli autori hanno quindi applicato un fractional factorial design (esaminando due livelli per ciascun fattore), considerando un valore di 2^{6-1} per un totale di 32 esperimenti eseguiti per ciascuna lega in esame.

Il piano sperimentale (experimental plan) è mostrato in Tab. II. L'importanza di ciascun fattore è stata discriminata applicando metodologie di tipo statistico.

Lo studio ha messo in evidenza che per le leghe di acciaio al carbonio, i parametri più significativi sono:

- a) concentrazione di ossigeno
- b) concentrazione di solfuri

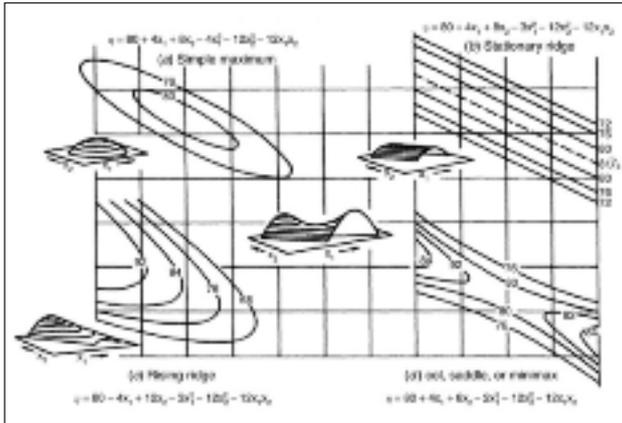


FIG. 1 Esempio di tipi di superficie definite da un polinomio di secondo ordine per due variabili codificate x_1 e x_2 [8].

Some examples of types of surfaces defined by second-order polynomials in two predictor variables x_1 and x_2 .

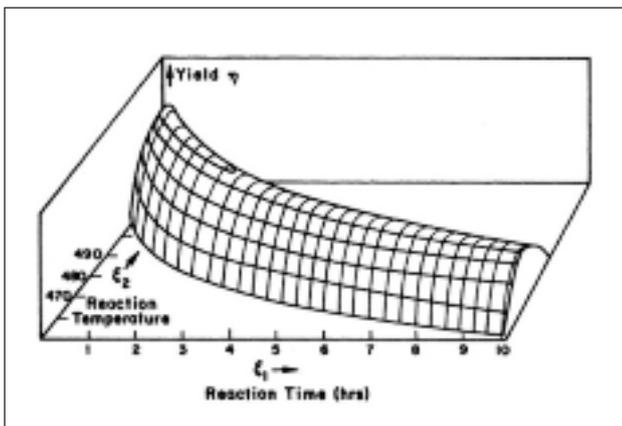


FIG. 2 Esempio di superficie di risposta tra risposta η e variabili ξ_1 e ξ_2 [8].

A theoretical response surface showing a relationship between response η and variables ξ_1 and ξ_2 [8].

Variabile	Low level	High level
Temperatura	30°C	45°C
Ossigeno	Deaerated	Aerated
Urea	0.0 ppm	3.5 ppm
Inibitore	0.0 ppm	50.0 ppm
Solfuri	0.0 ppm	2.0 ppm
Cloruri	19,500 ppm	23,000 ppm

TAB. II *Experimental plan riportato in Hajeesh [9].*

Experimental plan as shown in Hajeesh [9].

- c) temperatura
 - d) interazioni a due termini (ad es. temperatura × ossigeno, urea × ossigeno, ecc.)
- mentre per le leghe di bronzo:
- a) inibitore
 - b) interazioni a tre termini (ad es. ossigeno × urea × inibitore + temperatura × solfuro × cloruro, temperatura × inibitore × solfuro + urea × inibitore × cloruro, ecc)

Grazie quindi all'utilizzo del DoE gli autori del lavoro hanno po-

tuto valutare il peso di ogni singola variabile coinvolta nello studio della corrosione in acqua di mare delle tre diverse leghe.

In Tabella III sono riassunti i risultati, pubblicati nella letteratura scientifica, di alcune applicazioni del DoE. È da sottolineare il fatto che anche qualora l'utilizzo del DoE non implicasse necessariamente una riduzione del numero di esperimenti da eseguire, tale approccio permette comunque di individuare un modello matematico che studia le interazioni tra variabili. Per un'analisi critica dell'applicazione dell'approccio One Variable At a Time (OVAT) e di quello Multivariate si rimanda altrove [8].

PATTERN RECOGNITION

Il principale obiettivo della pattern recognition è la classificazione, supervised e unsupervised, degli oggetti in un numero di categorie o classi. Tra i vari framework in cui la pattern recognition è stata tradizionalmente formulata, l'approccio statistico è il più studiato ed applicato nella pratica. Il design per i sistemi di classificazione richiede particolare attenzione alla definizione delle classi e alla loro rappresentazione, all'estrazione e selezione delle proprietà degli oggetti che classifichiamo, ai metodi di selezione e alle loro prestazioni, nonché agli opportuni test da studiare. Gli oggetti da analizzare possono essere costituiti da immagini, segnali o da ogni tipo di misura da classificare a seconda dell'applicazione in esame. Generalmente ci riferiamo a questi oggetti con il termine di pattern [18-20].

Esistono diverse metodologie di classificazione. In questo articolo ci concentreremo su uno dei metodi più comunemente utilizzati, ovvero l'analisi delle componenti principali.

Principal Component Analysis (PCA)

L'analisi delle componenti principali è stata formalizzata da Pearson nel 1901 [21], sebbene i suoi fondamenti fossero già noti ai matematici e fisici con il nome di analisi agli autovalori.

L'idea centrale della PCA è quella di ridurre la dimensionalità di un set di dati in cui abbiamo una serie di termini matematici mediante una trasformazione lineare ortogonale in un nuovo sistema di coordinate. Poiché la PCA permette di conoscere il peso dei singoli fattori coinvolti in un sistema sotto studio, essa è particolarmente utile quando ci occupiamo di problemi di classificazione.

Grazie alla PCA è possibile costruire una relazione matematica empirica dei dati in esame come descritto dalla seguente equazione:

$$A = T_k V_k^T + \epsilon \quad (\text{eq.4})$$

dove T_k è la matrice $n \times k$ delle componenti principali e V_k è la matrice $m \times k$ degli autovettori (anche indicati come scores P). Gli autovettori possono essere usati per formare un set di basi ortonormali per A . Gli autovettori sono anche chiamati "loadings", "abstract factors" o "eigenspectra". Una interpretazione fisica dei vettori non è sempre possibile.

Le colonne T sono chiamate scores e sono mutualmente ortogonali ma non normalizzate. Possono essere usate per formare un set di vettori base colonna di A . Per ogni sorgente di varianza indipendente ci aspettiamo di trovare una singola componente. La prima componente è la direzione in cui è rappresentato il massimo della varianza dei dati. La seconda componente è ortogonale alla prima e descrive la varianza rimanente. Il numero delle PCA estratte è uguale al numero delle variabili osservate che contengono una quantità di informazione proporzionale al loro ordine di estrazione. Possiamo partizionare la matrice dei dati in una somma di k matrici di singolo rango $n \times m$ [22]:

$$A = t_1 v_1 T + t_2 v_2 T + \dots + t_k v_k T + \epsilon \quad (\text{eq.5})$$

Un esempio di PCA è descritta in [23], dove si mettono in rela-

Riferimento bibliografico	Metodo	Fattori studiati	Valori considerati per ciascun fattore	Numero minimo di esperimenti (*)	Esperimenti svolti
Hajeesh [9]	Fractional factorial	7	6 a 3 livelli e 2 ad un livello	Come minimo $3^6=729$	64
Mondel [10]	Full factorial	3	2	$2^3=9$	8
Rodriguez [11]	Full factorial	2	4	$2^4+2^4=32$	32
Dowey [12]	Full factorial	4	2	$2^4=16$	16 ed 8 (x2 set)
Dreel [13]	D-Optimal	4 (x2 set)	3	$4^3+4^3=128$	27 espansi a 46
Fernandez [14]	Full factorial	4	3	$3^4=81$	8
Li [15]	Uniform	5	(10 o 5)	243	10
Öktem [16]	Full factorial	5	3	$3^5=243$	243
Tryland[17]	Vari	5	3	$3^5=243$	$27+27+32=86$

TAB. III *Paragone tra numero di esperimenti richiesti utilizzando e non utilizzando l'experimental design. * il numero degli esperimenti da eseguire, non applicando un design of experiment, è calcolato considerando 3 livelli per ciascun fattore ed ipotizzando che l'utente decida di eseguire tutte le possibili combinazioni di esperimenti per il caso considerato. In genere questo è il caso più ottimistico dal momento che lo sperimentatore inizia studiando un parametro soltanto, impiegando più di tre livelli, e dopo averlo "ottimizzato" decide di fissarlo e di procedere con gli altri parametri in esame, tralasciando le interazioni del sistema considerato.*

Comparison of the number of required experiments using or not using experimental design.

**The number of experiments, when experimental design is not applied, can be calculated considering three levels for each factor and under the hypothesis that the user would perform all the possible combinations of experiments for the given problem. This should be considered an optimistic view, since generally the experimenter starts to study ONE parameter at more than 3 levels, then "optimises" it, fixes it, and finally proceeds along the study of the others, leaving out all the interactions for the system considered.*

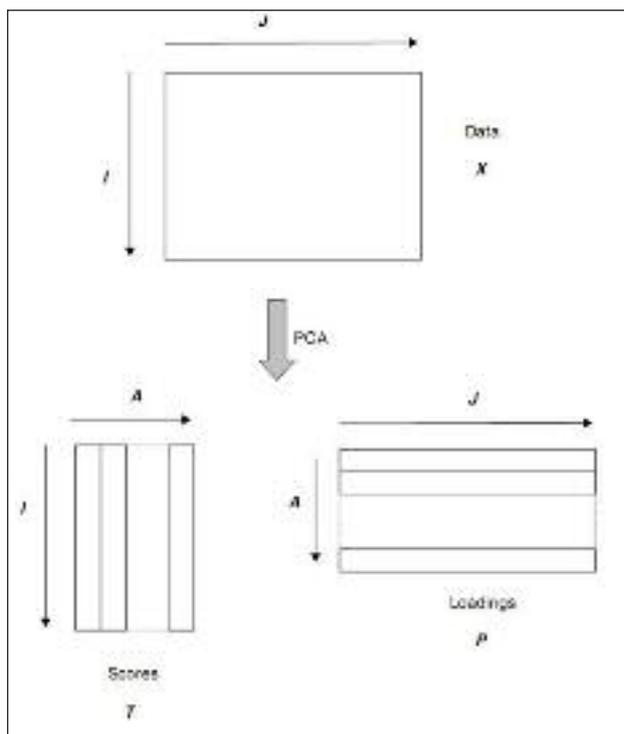


FIG. 3 *Rappresentazione schematica della PCA [2]. Schematic representation of PCA workflow [2].*

zione le condizioni ambientali con la formazione di prodotti di corrosione; per questo scopo sono stati raccolti trenta campioni da un monumento in bronzo ed analizzate mediante diffrattometria dei raggi X (XRD). Attraverso la PCA sono state analizzate la quantità di atacamite, brocantite, quarzo, calcite, gesso e antlerite, oltre all'orientamento delle superfici ed alle condizioni ambientali (esposizione, quantità di pioggia, vento). Delle 10 variabili considerate 5 includono più dell'80% della variabilità.

Il grafico delle componenti 1 e 2 è riportato in Fig. 4

In base all'analisi PCA, i campioni analizzati sono stati suddivisi in 3 gruppi collegati alle variabili in esame secondo lo schema seguente:

Gruppo 1: Gesso, Orientamento

Gruppo 2: Atacamite, Antlerite, Quarzo, Protezione dalla Pioggia

Gruppo 3: Brocantite, Calcite, Pioggia, Vento.

In questo caso l'applicazione della PCA ha permesso di quantificare l'influenza dei parametri in esame fornendo risultati di rilevante importanza per i professionisti nel settore della conservazione di manufatti metallici di interesse storico-artistico.

CALIBRAZIONE

La calibrazione multivariata è il processo di apprendimento sul come combinare diverse fonti di informazione al fine di risolvere problemi di selezione ed ottenere, in modo automatico, i campioni non significativi all'interno di uno specifico dataset (outlier).

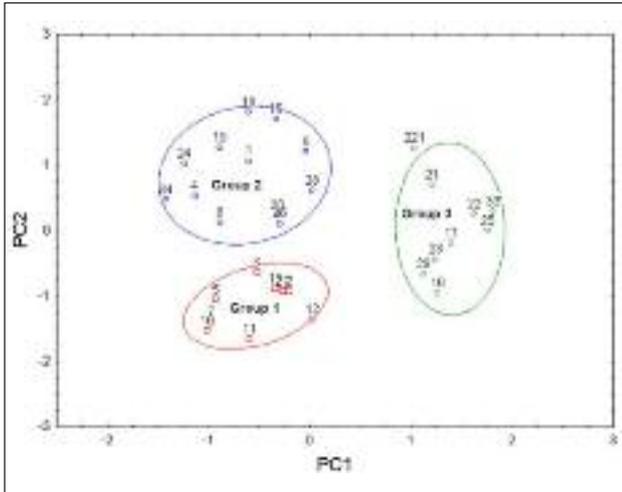


FIG. 4 Esempio di score plot nell'analisi PCA che mostra come la PCA può essere utilizzata per l'individuazione di cluster in un dataset [23].

Example of score plot that show how PCA can be used to spot cluster in a dataset [23].

La tecnica della calibrazione multivariata è applicata all'analisi di un'ampia tipologia di misure analitiche quali UV, Vis, IR, (trasmissione, riflettanza, fluorescenza), diffrazione dei raggi X, NMR, MS, analisi termica, cromatografia (GC, HPLC), elettroforesi, analisi di immagine (tomografia, microscopia).

In questa sezione si introdurrà una delle tecniche più comunemente utilizzate in ambito chemiometrico: la PLS (partial least squares), una variazione della least squares methodology.

La regressione PLS è una generalizzazione della regressione multipla (MLR) ed è di particolare interesse in quanto permette di analizzare dati caratterizzati da forte rumore e numerose variabili X strettamente correlate tra loro, considerando nel contempo diverse variabili di risposta Y. La regressione PLS è un metodo utilizzato per la costruzione di un modello bi-lineare dove dall'informazione originale (matrice X dei predittori), sono estratte alcune variabili latenti, dette componenti principali (PC). Per il calcolo delle variabili latenti sono utilizzati i dati contenuti in Y, sottomatrice delle risposte, garantendo che le prime componenti calcolate siano le più significative per la predizione delle risposte stesse. La regressione ai minimi quadrati parziali consente di ottenere un'elevata stabilità del modello che risulta quindi utilizzabile sia quando il rapporto tra oggetti e variabili è inferiore a uno, sia in presenza di variabili molto correlate tra loro, sia in presenza di più risposte. Tutte queste caratteristiche sono, in effetti, presenti nel problema analitico qui studiato: uno spettro corrisponde alla lettura di assorbanza per qualche centinaio di valori di lunghezza d'onda (numerosi variabili) laddove gli standard sottoposti a misura sono qualche decina; una banda di assorbimento corrisponde a letture di assorbanza fortemente correlate; infine le miscele solide analizzate contengono sempre diversi componenti. Il metodo di regressione PLS permette la costruzione di un unico modello per tutte le lunghezze d'onda e non tanti modelli quante sono le lunghezze d'onda stesse. I risultati dei metodi di proiezione come la regressione PLS dipendono dalla scalatura dei dati: se questa è eseguita nella maniera appropriata, si può focalizzare il modello sulle variabili Y più importanti e usare l'esperienza per dare maggiore peso alle variabili X, che sono portatrici del più alto numero di informazioni. In letteratura sono riportate le basi scientifiche dell'utilizzo della normalizzazione come pretrattamento di dati spettroscopici e la sua larga applicazione in questo campo [24].

Nord et al [25] hanno utilizzato la PLS nella valutazione della relazione tra lo stato di conservazione di manufatti bronzei e le condizioni ambientali. Principale scopo è stato quello di identificare il ruolo dei singoli fattori chimico-fisici nel fenomeno di deterioramento. I risultati ottenuti, studiando 82 parametri, hanno mostrato che l'acidità del suolo, l'aerazione, la presenza di sali, ceneri ed umidità, accelerano il degrado e che la propagazione del processo corrosivo si è realizzata in modo preponderante negli ultimi secoli.

CONCLUSIONI

Sebbene non siano di uso comune, le tecniche multivariate hanno diverse applicazioni nell'ambito dello studio della corrosione, con conseguente possibile risparmio di risorse materiali ed economiche.

E' da sottolineare comunque che le condizioni richieste per l'applicazione di queste tecniche (ed in particolare DoE e regressione multivariata) comportano alcune limitazioni:

- trattandosi di tecniche esplorative molto semplici da applicare, uno dei rischi è quello di utilizzarle senza eseguire un'attenta analisi dei risultati ottenuti;
- l'applicazione di una regressione ad un dataset non statisticamente significativo può condurre ad una modellazione fuorviante.

Si riportano di seguito alcuni dei principali vantaggi nell'applicazione di queste tecniche:

- identificazione nel sistema in esame delle variabili più significative a livello statistico;
- possibile valutazione dell'interazione tra i parametri considerati;
- risparmio di tempo in fase di progettazione ed analisi dei dati;
- creazione di un modello matematico tra risposte in esame e parametri selezionati.

Va tenuto conto del tempo richiesto per una opportuna formazione del personale all'uso di queste tecniche [26].

BIBLIOGRAFIA

- S. Wold, Chemometrics; what do we mean with it, and what do we want from it? Chemometrics and Intelligent Laboratory Systems 30 (1995) 109-115.
- R.G. Brereton, Chemometrics: data analysis for the laboratory and chemical plant, Wiley, 2003.
- V. Madisetti, The Digital Signal Processing Handbook, Second Edition - 3 Volume Set, CRC Press, 2009.
- Z.R. Lalic, Design of Experiments in Chemical Engineering: A Practical Guide, 1st ed., Wiley-VCH, 2005.
- S.D. Cramer, ASM Handbook: Volume 13B: Corrosion: Materials, 10th ed., ASM International, 2005.
- B. S. J. Covino, ASM Handbook: Corrosion: Fundamentals, Testing, and Protection, 10th ed., ASM International, 2003.
- T.P. Ryan, Modern experimental design, Wiley-Interscience, 2007.
- G.E.P. Box, N.R. Draper, Response Surfaces, Mixtures, and Ridge Analyses, Hoboken, NJ, USA, John Wiley & Sons, Inc., 2007.
- M. Hajeer, Estimating corrosion: a statistical approach, Materials & Design. 24 (2003) 509-518.
- D.P. Mondel, S. Das, B.K. Prasad, Study of erosive-corrosive wear characteristics of an aluminium alloy composite through factorial design of experiments, Wear 217 (1998) 1-6.
- J. Rodriguez, A. Martin, R. Fernandez, J.E. Fernandez, An experimental study of the wear performance of NiCrBSi thermal spray coatings, Wear 255 (2003) 950-955.
- S. Dowey, B. Rähle, A. Matthews, Performance analysis of coated tools in real-life industrial experiments using statistical techniques, Surface and Coatings Technology 99 (1998) 213-221.
- S. Dreer, R. Krismer, P. Wilhartitz, Multidimensional optimisation of process parameters by experimental design for the deposition of aluminium and silicon oxynitride films with predictable com-

- position, *Surface and Coatings Technology* 114 (1999) 29-38.
- [14] J. E. Fernández, M. del Rocío Fernández, R. Vijande Diaz, R. Tucho Navarro, Abrasive wear analysis using factorial experiment design, *Wear* 255 (2003) 38-43.
- [15] J.F. Li, H. Liao, B. Norm, C. Cordier, G. Maurin, J. Foct, et al., Uniform design method for optimization of process parameters of plasma sprayed TiN coatings, *Surface and Coatings Technology* 176 (2003) 1-13.
- [16] H. Öktem, T. Erzurumlu, H. Kurtaran, Application of response surface methodology in the optimization of cutting conditions for surface roughness, *Journal of Materials Processing Technology* 170 (2005) 11-16.
- [17] T. Tryland, O.S. Hopperstad, M. Langseth, Design of experiments to identify material properties, *Materials & Design* 21 (2000) 477-492.
- [18] A. K. Jain, R. P.W. Duin, J. Mao, *IEEE Transactions on pattern analysis and machine intelligence* 22 (2000) 4-37.
- [19] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, 4th ed., Academic Press, 2008.
- [20] C.C. Lee, F. Mansfeld, Automatic classification of polymer coating quality using artificial neural networks, *Corrosion Science* 30 (1999) 439-461
- [21] K. Pearson, On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine* 2 (1901) 559-572.
- [22] P. Gemperline, *Practical Guide To Chemometrics*, Second Edition, CRC, 2006.
- [23] K. Polikreti, V. Argyropoulos, D. Charalambous, A. Vossou, V. Perdikatsis and C. Apostolaki, Tracing correlations of corrosion products and microclimate data on outdoor bronze monuments by principal component analysis, *Corrosion Science* 51 (2009), 2416-2422
- [24] H. Martens, T. Næs, *Multivariate Calibration*, 1st ed., Wiley, 1992.
- [25] A.G. Nord, E. Mattsson and K. Tronner, Factors influencing the long-term corrosion of bronze artefacts in soil, *Protection of Metals* 41 (2005), pp. 309-316.
- [26] G. Luciano, P. Traverso, P. Letardi, Applications of chemometric tools in corrosion studies, *Corrosion Science*. 52 (2010) 2750-2757

Abstract

Chemometric techniques and their application to the study of corrosion

Keywords:

corrosion, modeling, Principal Component Analysis (PCA), Response Surface Methodology (RMS), Partial Least Squares (PLS)

The aim of this paper is to introduce the reader to the applications of chemometric techniques (i.e. Design of Experiment (DoE), classification and modelling techniques) in the field of corrosion science, which currently are not commonly spread. By means of key examples, it will be shown how they can be fruitfully applied and enhance the quality of a laboratory workflow.

Chemometrics is defined by IUPAC as "the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods". It is currently applied in diverse field of chemistry by using the following methodologies: 1) signal processing, 2) Design of Experiment (DoE), 3) pattern recognition and 4) calibration. The present paper will mainly focus on the last three, since the available literature in signal processing is wide and well established [1,2,3].

After dealing with the mathematical aspects of DoE and the response surface modelling [7], we will show a key application with references to other valuable examples [8]. As highlighted in another paper [7], chemometrics plays a major role in the screening phase of a test as it leads to reduce the number of the needed experiments, when compared with the univariate approach. Further, it allows deriving a mathematical relation among the selected parameters that can be also plotted in the form of a contour surface for a better representation of the results.

We will analogously proceed to describe the pattern recognition throughout an example of its application to the study of a bronze statue. Once pattern recognition is applied, corrosion products are grouped as a function of microclimates, thus giving useful information about the conservation state and the relationships among microclimatic factors and degradation process.

Finally, we will discuss theory and applications of calibration methodologies, principally the partial least square technique [24] that leads to define a mathematical model based on 5 parameters out of 82 as mainly responsible for the state of conservation of ancient bronze findings.