

*La riproduzione anche parziale del presente documento  
deve essere autorizzata per iscritto da  
Direzione  
CAP HOLDING S.p.A.*

## **CAP HOLDING S.p.A**



**Indagine sensoriale strumentale per la rilevazione di  
dispersioni odorigene di provenienza da Impianto di  
depurazione sito in Canegrate (MI)**

a cura di

**PCA Technologies S.r.l.**  
*Monitoraggi Ambientali*  
**Via Gajo n. 12 , 20015 Parabiago (MI)**  
[\*\*info@pcatechnologies.com\*\*](mailto:info@pcatechnologies.com)

**Totale pagine relazione + allegato n. 29**

**Vs. Riferimento: Ordine n.1502026 del 28 Luglio 2015**  
**CIG: Z0F158EF19**

## Premessa

### Obiettivo

Lo scopo della indagine è la verifica della presenza sul territorio di emissioni provenienti da Impianto di depurazione sito in Canegrate, Via Cascinette (MI).

La procedura di indagine ha richiesto il campionamento e la determinazione delle caratteristiche sensoriali di ogni possibile e/o probabile emissione in atmosfera presente all'interno dell'impianto di depurazione (la loro impronta sensoriale), la successiva elaborazione chemiometrica per la costruzione della libreria specifica delle sorgenti campionate e il monitoraggio in esterno con installazione di n. 2 analizzatori sensoriali modello PEN3-Meteo completi di centralina meteorologica, posizionati in due punti diversi del territorio circostante.

### Le fasi operative applicate nella seguente indagine

**1. Memorizzazione e caratterizzazione delle fonti di possibile emissione all'interno dell'impianto di depurazione (mappatura sorgenti):**

Questa operazione prevede la determinazione delle caratteristiche sensoriali (la loro "impronta digitale") di tutte le possibili e/o probabili emissioni odorigene all'interno dell'impianto di depurazione allo scopo di costruire l'archivio di riferimento delle possibili e/o probabili sorgenti di dispersione odorigena, confrontarle con le caratteristiche dell'aria "inodore" e permettere il loro successivo riconoscimento durante il monitoraggio. L'archivio di riferimento o libreria (denominato "pattern", permetterà di eseguire i calcoli di elaborazione necessari al riconoscimento delle stesse sul territorio circostante durante il periodo di monitoraggio.

**2. Monitoraggio dell'aria e valutazione della presenza sul territorio di dispersioni odorigene provenienti dall'impianto di depurazione.**

Memorizzate ed archiviate le impronte sensoriali delle emissioni (definite sorgenti), la fase successiva è l'installazione di n. 2 installazioni di PEN-Meteo e l'inizio del monitoraggio in continuo. Gli strumenti sono stati installati presso due abitazioni private e in posizione da permettere non solo il campionamento dell'aria ambiente ma anche la corretta rilevazione della velocità e direzione del vento. La caratterizzazione e il riconoscimento delle dispersioni rilevate è stata eseguita mediante elaborazioni di data mining (chemiometria) come vedremo di seguito.

### Il Metodo di misura sensoriale strumentale

Le misure sensoriali strumentali oggetto di questo lavoro sono state eseguite mediante l'analizzatore sensoriale Airsense modello PEN3 (naso elettronico), in grado di determinare le impronte olfattive di un campione di tipo aeriforme, liquido e solido.

L'analizzatore sensoriale simula il processo mentale di memorizzazione e riconoscimento del sistema olfattivo umano. Come i nostri recettori mandano un messaggio elettrico al nostro cervello, nello strumento i sensori inviano un segnale

elettrico al computer. A questo livello si ha l'elaborazione, la memorizzazione, la caratterizzazione ed il riconoscimento delle caratteristiche del campione incognito.

I sensori utilizzati dallo strumento sono di tipo MOS (Metal Oxide Sensors) trattati opportunamente con diversi layer di metallo e opportune temperature per essere produrre segnali alla presenza di diverse classi di composti volatili. Caratteristica molto importante e peculiare dello strumento utilizzato è la programmazione di ogni sensore ad una specifica e diversa temperatura di lavoro (nel range 150-500°C), caratteristica che permette di ampliare il range di rilevazione delle sostanze volatili responsabili della percezione di odore.

L'utilizzo di questa tecnica di misura ha l'obiettivo di classificare l'odore, in sostanza memorizzare l'impronta digitale del campione che è la risultante di una miscela di innumerevoli composti, spesso presenti a concentrazioni bassissime ma con un'alta soglia di percezione olfattiva; in altre parole i sensori, essendo sensibili verso classi di diversi analiti, consentono di "leggere", attraverso i segnali ottenuti ai sensori, ***l'impronta digitale caratteristica e peculiare*** di un campione.

### **Il processo di misura dell'analizzatore sensoriale PEN3**

PEN3 è composto da dieci sensori MOS termoregolati singolarmente nel range 150-500°C, da una pompa di aspirazione del campione, da una pompa di aspirazione in controcorrente di "aria zero" e da un software di elaborazione chemiometrica installato su un personal computer.

Il "flusso della misura" avviene come di seguito:

- Vengono definiti i parametri di analisi.
- Allo Start dell'operatore lo strumento esegue una serie di operazioni che assicurano l'assenza, all'interno della cella di misura e sui sensori del campione precedente.
- Concluso il ciclo di "Aria zero" e quello di "Calibrazione" viene aspirato il campione all'interno della cella di misura.
- Conclusa la misura, l'analizzatore provvede ad eseguire un'analogica operazione di pulizia e calibrazione dei sensori.

"Aria zero" (per pulizia sensori): tale operazione è eseguita attraverso l'aspirazione in controcorrente di aria purificata mediante opportuna filtrazione su filtro a carboni attivi.

### **Il processo di elaborazione dei dati**

Eseguite successive repliche su ogni campione di riferimento (sorgenti), esse vengono trattate mediante algoritmi finalizzati alla ricerca delle componenti (caratteristiche) che differenziano i diversi campioni fra loro (e le loro diverse impronte sensoriali).

Per questo primo passaggio è utilizzata l'elaborazione definita PCA (Principal Component Analysis, metodo esplorativo) che permette di avere una visione d'insieme delle impronte ottenute, semplificando il lavoro di scelta delle funzioni discriminanti. Seguirà poi l'analisi multivariata LDA che è il passaggio intermedio

prima della memorizzazione delle impronte ottenute. Questi passaggi hanno l'obiettivo di memorizzare le caratteristiche dei campioni in funzione della migliore discriminazione ottenuta e validare i dati ottenuti, passaggio questo fondamentale per qualsiasi misura di tipo strumentale e per costruire un archivio di riferimento verificabile, controllabile a posteriori e soprattutto statisticamente valido. Con il modello PEN3 in uso è possibile utilizzare elaborazioni di tipo Euclidico, Mahalanobis o di Correlazione Pearson, l'analisi discriminante DFA e/o la regressione PLS che determinano, con algoritmi diversi la classe di appartenenza del campione incognito o quantificare, secondo gli obiettivi dell'indagine (per esempio in Unità Olfattometriche) le caratteristiche dei campioni incogniti e permettere un efficace ed attendibile riconoscimento di una dispersione odorigena.

**LA SEQUENZA DEI PASSAGGI OPERATIVI PER LA MAPPATURA DELLE SORGENTI  
(costruzione archivio di riferimento)  
E MONITORAGGIO SUL TERRITORIO (riconoscimento sorgente dispersiva)**

1. Campionamento delle arie in emissione dalle sorgenti di possibile emissione odorigena all'interno dell'impianto
2. Misura delle caratteristiche sensoriali dei campioni sorgente prelevati .
3. Validazione dei dati risultanti (mediante elaborazione LDA-Linear Discriminant Analysis)
4. Costruzione della libreria di riferimento che contiene tutte le caratteristiche sensoriali delle sorgenti ed verificare la capacità discriminante del sistema di misura (elaborazione PCA-Principal Component Analysis).
5. Installazione di n. 2 Environmental Odor Unit PEN3-Meteo in area sensibile (stazione di rilevamento completa di centralina meteorologica per acquisizione di velocità e direzione del vento) per monitoraggio 24 ore su 24.
6. Classificazione e riconoscimento delle eventuali dispersioni rilevate ai bersagli e facenti parte della libreria sorgenti, campionate all'interno dell'impianto di depurazione (elaborazione di Correlazione Pearson) .
7. Risultati finali , considerazioni e conclusioni

**1) Campionamento delle arie in emissione**

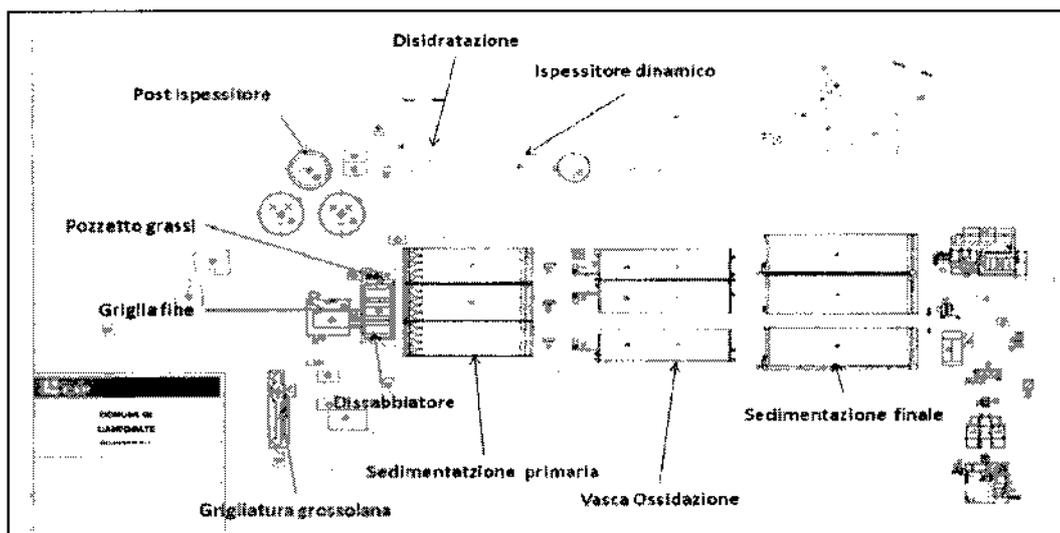
I campionamenti di aria sorgente sono stati eseguiti mediante pompa a polmone in sacchetti di Nalophan da 12 lt:

sulle sorgenti areali passive (vasche) il campionamento è stato eseguito con l'ausilio di cappa tipo windtunnel in grado di simulare la condizione atmosferica di flusso parallelo senza rimescolamento verticale ; tutti gli altri prelievi sono stati eseguiti con campionamento diretto in ambiente e/o area interna capannone.

Di seguito i punti all'interno dell'impianto sottoposti a campionamento e successiva misura mediante analizzatore PEN3

Le sorgenti campionate all'interno dell'impianto
Aria fondo ambiente inodore (*)
Ispezzatore dinamico
Post ispezzatore
Vasca sedimentazione primaria
Vasca sedimentazione secondaria
Vasca Ossidazione
Dissabbiatura
Disidratazione
Griglia fine
Griglia grossolana
Pozzetto grassi

(\*) Il campione di aria "inodore" utilizzata come riferimento è stata campionata in zona non influenzata dall'impianto e in assenza di percezione odore da parte dei presenti (in località Parabiago, Via Gajo).



Tutti i prelievi relativi alle "sorgenti" depuratore sono stati eseguiti all'interno delle aree durante le normali e quotidiani trattamenti e operazioni di lavoro.

N.B. : è importante notare che, ad esclusione delle vasche, tutti gli altri punti sottoposti a prelievo sono aree confinate e sottoposte ad aspirazione e trattamento mediante filtri a carbone attivi prima della loro emissione in atmosfera; per considerare anche eventuali emissioni diffuse e/o saturazione dei filtri a carbone, i nostri prelievi sono stati eseguiti all'interno del capannone e prima del trattamento di abbattimento e filtrazione.

## 2) Misura mediante analizzatore sensoriale PEN3 Airstense dei campioni "sorgente"

I campioni "sorgente" sono stati sottoposti entro 4 ore dal campionamento (per evitare qualsiasi modificazione chimico-fisica del campione) all'analizzatore PEN3 per la determinazione della loro impronta digitale, ogni campione è stato sottoposto a tre repliche di misura in sequenza random.

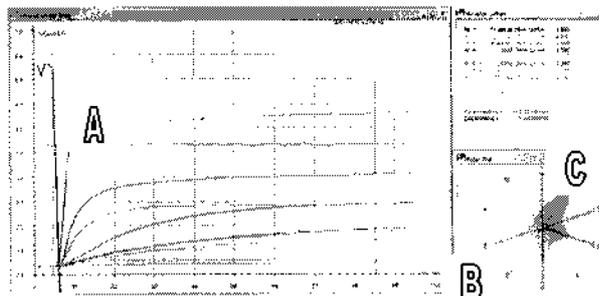
I parametri strumentali utilizzati sono i seguenti:

- Flusso aspirazione campione: 400 ml/min
- Tempo di misura: 100 sec.
- Tempo di campionamento data di variazione: ogni secondo
- Flusso pulizia: 600 ml/min
- Tempo di pulizia intermedio: 500 sec.

Per ogni misura effettuata viene determinata "l'impronta sensoriale" del campione (profilo dei segnali ottenuti dai 10 diversi sensori), seguirà la Validazione dei dati in ingresso (le impronte memorizzate) utilizzando il metodo LDA Linear discriminant Analysis e secondo la procedura di "cross validation- leaving one-leaving out".

A lato il tipico andamento dei sensori durante una misura:

- ↓ nel quadrante A la variazione dei sensori,
- ↓ nel quadrante B l'espressione grafica dell'impronta sensoriale del campione,
- ↓ nel quadrante C i segnali grezzi ottenuti ai sensori



## 3) Validazione dei dati mediante LDA (Linear Discriminant Analysis)

Per la costruzione di un archivio di riferimento delle sorgenti attendibile è fondamentale, a garanzia del risultato, la validazione dei dati in ingresso per avere conferma che ogni ripetizione di misura di ciascun campione appartenga realmente alla classe assegnata. Questo processo permette di verificare che per esempio, la 1° misura eseguita sul campione "Emissione biofiltro" abbia la stessa "impronta sensoriale" (identiche caratteristiche) anche nelle successive repliche eseguite sulla stessa emissione (verificare l'appartenenza di Classe<sup>(\*)</sup> attribuita).

(\*) Nelle analisi chemiometriche si fa uso del termine "Classi" per rappresentare il gruppo di misure e campioni aventi analoghe caratteristiche (o provenienza). Per esempio, tutte le misure eseguite sul campione prelevato all'emissione del biofiltro saranno definite Classe "Emissione biofiltro", le misure eseguite sui campioni di aria prelevati sulle vasche saranno definite classe "Vasche" e così di seguito.

La validazione dei dati è stata effettuata nel modo *Cross Validation* (o validazione incrociata), in cui vengono ripercorse tutte le tappe eseguite per la costruzione del pattern primario. Dopo aver elaborato i dati con PCA, si ricalcola il modello con l'esclusione di un oggetto alla volta (una misura) e di un oggetto ogni k oggetti

(*leave-more-out*), predicendo i valori della risposta per tutti gli oggetti esclusi dal modello: si procede come di seguito per ogni classe:

- ▶ eliminare la misura n. 1
- ▶ ricalcolare il modello utilizzando solo le misure rimaste;
- ▶ riassegnare la misura n.1 ad una classe in base al modello ricalcolato sulle misure rimaste;
- ▶ eliminare il n. 2;
- ▶ ricalcolare nuovamente il modello sulle rimaste e procedere alla nuova assegnazione della misura n. 2;
- ▶ proseguire eliminando uno alla volta tutte le misure fino al completamento e riassegnandole di volta in volta in base al modello ricalcolato senza la misura eliminata.

Al termine del procedimento avremo ricalcolato  $n$  volte il modello e verificato così l'attendibilità del modello costruito .

In questo caso abbiamo eseguito l'elaborazione di validazione LDA sugli n. 10 campioni prelevati alle sorgenti e n. 1 campioni di aria ambiente inodore (n. 11 Classi = 11 sorgenti) assegnando alla stessa classe le 3 repliche di misura dello stesso campione. Tutte le misure sono state assegnate alla classe corretta

#### **4) Costruzione del pattern di riferimento sorgenti (libreria) (archivio di riferimento emissioni - elaborazione PCA).**

Eseguiti i passaggi operativi visti sopra per la verifica della attendibilità dei dati di misura, si procede alla costruzione della libreria di riferimento (pattern) che include tutte le sorgenti campionate, utilizzando la tecnica di elaborazione di analisi multivariata PCA "*Principal Component Analysis*".

*L'analisi PCA è finalizzata ad estrarre la massima informazione possibile contenuta in una struttura di dati multivariati (i dati ottenuti dai 10 sensori dell'analizzatore PEN), sintetizzandola in poche combinazioni lineari delle variabili stesse in modo tale da ridurre la dimensionalità dei dati ottenuti da ogni misura e determinare il valore numerico che meglio esprime le caratteristiche di un campione rispetto gli altri considerati.*

Questo metodo viene frequentemente impiegato nella prima fase di elaborazione dei dati e serve a dare una visione generale del problema, a capire le relazioni tra gli oggetti e/o le classi considerate e fornire una indicazione preliminare sul ruolo delle variabili, mettendo eventualmente in luce la possibilità di eliminarne alcune che, essendo strettamente correlate tra loro, portano informazioni simili e possono quindi essere considerate ridondanti.

Dal punto di vista geometrico, la PCA è un processo di rotazione dei dati originali, effettuato in modo che il primo nuovo asse (che costituirà la prima componente principale PC#1) sia orientato nella direzione di massima varianza dei dati, il secondo

sia perpendicolare al primo e sia nella direzione della successiva massima varianza dei dati, e così di seguito per tutti i  $p$  dei nuovi assi ( $p$  = numero dei sensori in uso). Il numero di questi nuovi assi (le componenti principali, PC1, PC2,.....PCx) sarà quindi pari al numero di variabili originali (nel nostro caso avremo dieci componenti principali = numero dei sensori in uso). Le loro direzioni rispetto alle direzioni degli assi originali (le variabili originali) vengono determinate dagli "autovettori", che sono i vettori del nuovo spazio, espressi da coefficienti (loadings) compresi tra  $\pm 1$ .

A ciascun autovettore è associato un autovalore  $m$  che ne rappresenta la varianza spiegata: in generale gli autovalori più grandi sono associati ad informazione rilevante (utile, modellante, interpretabile, ecc.), mentre gli autovalori più piccoli sono associati a una variabilità dovuta a rumore o a informazione secondaria o non rilevante. Poiché, come si è già detto, gli autovettori sono gli assi relativi alle direzioni di massima varianza, in ordine via via decrescente, la prima componente principale PC#1 sarà in grado di spiegare la maggior percentuale di varianza, la seconda ne spiegherà un po' meno, la terza ancora di meno e così via, fino a che le ultime componenti contribuiranno a spiegare poco o nulla della variabilità presente nei dati in esame. In questi casi è quindi possibile eliminare parte della variabilità residua (e quindi anche parte del "rumore" che accompagna l'informazione rilevante) prendendo in considerazione solo un numero  $m$  di componenti, minore del numero delle variabili originali  $p$ .

Il metodo PCA fornisce come risultato due matrici: *matrice dei "loadings"* e *matrice degli "score"*, matrici numeriche fondamentali e utili per la costruzione della libreria di riferimento sorgenti. La matrice di "Loading" rappresenta l'importanza di ciascuna variabile originale in quel autovettore e ci permette di capire quali sono i sensori dell'analizzatore sensoriale maggiormente discriminanti e utili nelle successive elaborazioni. La matrice degli "Score", diversamente dai loadings, i cui valori sono limitati tra  $\pm 1$ , hanno valore medio uguale a zero, ma possono assumere valori numerici qualsiasi: gli scores rappresentano quindi le nuove coordinate degli oggetti nello spazio delle componenti principali.

### ***I grafici risultanti dalla elaborazione PCA***

Al fine di interpretare al meglio i risultati di questa indagine sensoriale è opportuno ribadire che ogni grafico deve essere letto in considerazione delle classi in esso contenute; un esempio banale ma efficace:

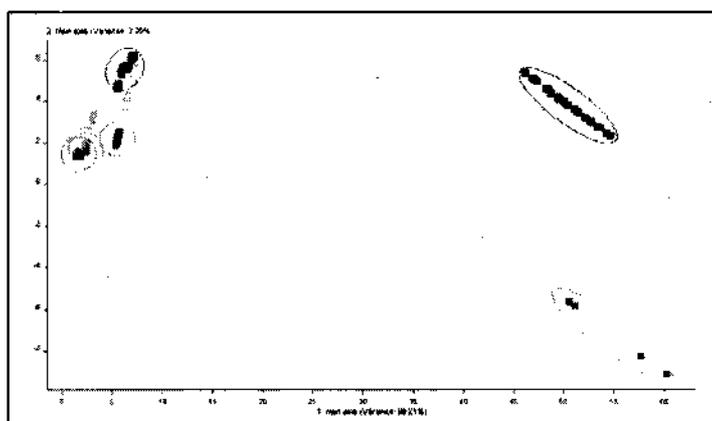
*"" in una sala riunioni devono entrare X persone che siederanno attorno a un tavolo; queste si disporranno secondo il loro peso (es. gerarchia, peso, dimensione, etc.) occupando tutto lo spazio a loro disposizione (grafico PCA). Se nella sala sopraggiunge un'altra persona, tutte le altre dovranno cambiare la loro disposizione nello spazio per rispettare le gerarchie, il peso, etc. che la nuova situazione impone. Se l'ultimo arrivato è l'Amm. Delegato (gerarchia) che è molto grasso (volume + peso) costringerà tutti gli altri a comprimersi fra loro per trovare posto ed evidenziare così l'importanza (gerarchia + peso + volume) che spetta al nuovo arrivato "" : questo è il motivo dell'affermazione che "ogni grafico deve essere letto in considerazione delle classi in esso contenute".*

Un grafico prodotto da elaborazione PCA mostra il “peso” delle diverse impronte sensoriali determinate sui campioni prelevati e ne caratterizza le mutue differenze. Ogni classe contiene tre differenti misure eseguite sul prelievo. Il grafico è costruito su due componenti principali PC#1 e PC#2 (\*) che mostrano graficamente la capacità discriminante del sistema.

**PC#1** è definita Prima Componente Principale (asse X del grafico) ed esprime, in termini percentuali, quale è la relazione migliore o la differenza meglio spiegata tra i diversi campioni.

**PC#2** è definita Seconda Componente Principale (asse Y del grafico) ed è la direzione, normale rispetto alla PC#1, lungo la quale la rimanente variazione (non spiegata da PC#1) è massima.

Prendendo in esame il **grafico (Principal Component Analysis)** che riporta i campioni prelevati alle sorgenti all’interno del depuratore, si notino le dimensioni degli assi:



↓ Asse X, (prima componente principale PC#1) valore 96,21%

↓ Asse Y, (seconda componente principale PC#2) valore 3,20 %.

I valori degli assi, indicano che il calcolo effettuato dal sistema chemiometrico è riuscito a spiegare il 99,41% delle caratteristiche dei campioni in esso contenuti, di cui 96,21% attraverso la componente PC#1 e 3,20 % attraverso la componente PC#2. Il restante a 100% è ripartito tra la terza, la quarta, via via fino alla decima componente principale. L’obiettivo è costruire un modello chemiometrico con la migliore separazione (discriminazione) delle diverse impronte sensoriali delle fonti di emissione.

Come sopra spiegato, sul grafico 1 sono raffigurati i campioni posizionati su un piano bidimensionale (PC1/PC2) in funzione delle reciproche differenze (o similitudini) relative. Appare chiara la risoluzione e la conseguente separazione (le impronte sensoriali delle sorgenti sono diverse) .

Ogni sorgente di emissione è rappresentata da un cluster compatto (nuvola compatta di dati rappresentati in uno spazio bi-tridimensionale), costituito da tutti i dati ad esso relativi. La compattezza del cluster testimonia la ripetibilità del sistema e la coerenza dei dati costituenti il cluster stesso (valori incoerenti verrebbero visualizzati con lo stesso colore del cluster d’appartenenza, ma sarebbero posti lontani da esso).

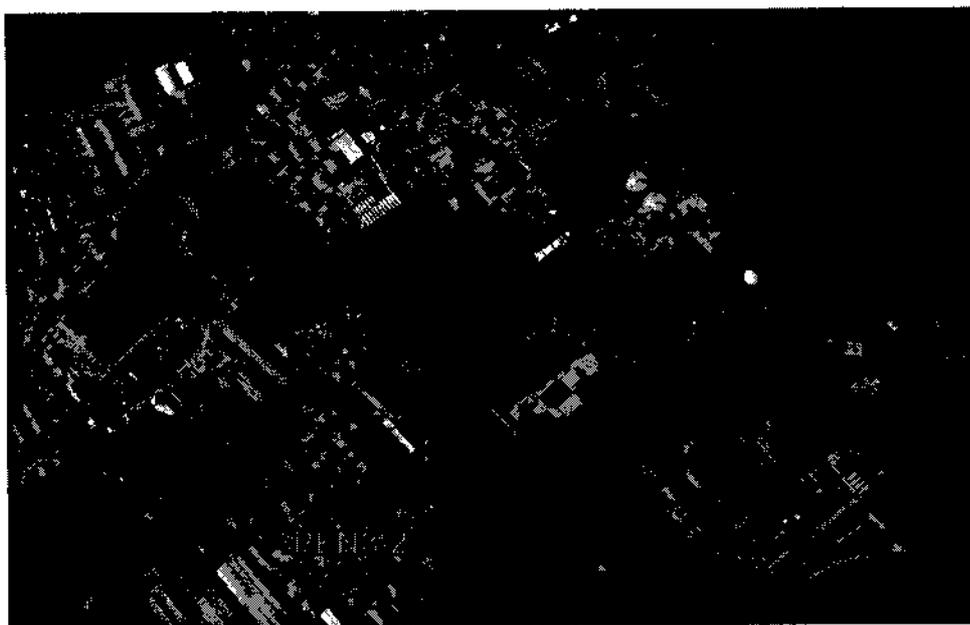
Poiché la distanza di ogni cluster dall'aria di fondo ambiente (classe di riferimento) rappresenta la differenza che tale classe ha nei confronti dell'aria "inodore", è possibile affermare che, più è grande tale distanza, più elevata è l'intensità odorigena e la quantità dei composti volatili presenti nell'aria campionata alla sorgente.

*Tutti questi passaggi di misura, validazione ed elaborazioni chemiometriche hanno come unico scopo la costruzione di un archivio di riferimento delle emissioni statisticamente attendibile, valido e verificabile a posteriori, utilizzando software chemiometrici di facile reperibilità e di utilizzo comune. Fatto ciò si può passare successivamente alla fase di monitoraggio sul territorio con il riconoscimento (predizione) di una dispersione odorigena.*

#### **5) Installazione di PEN3-Meteo e monitoraggio 24 ore su 24**

Per la fase di monitoraggio lo strumento esegue il campionamento diretto dell'aria memorizzando i dati dei segnali dei sensori ogni secondo.

- ✦ Il primo strumento di rilevamento odori PEN3-Meteo (vedi in mappa PEN3-1) è stato installato presso Via Cascinette Canegrate; altezza installazione a circa 6 metri di altezza ; distanza da confine depuratore 10-12 mt. , coordinate al suolo 45°34'31.55 N – 8°55'35.26 E). Il **monitoraggio è iniziato alle ore 16.43 del 1 Luglio 2015 e si è concluso alle ore 16.12 del 20 Luglio 2015** per un totale di 20 giorni consecutivi.
- ✦ il secondo strumento (vedi in mappa PEN3-2) è stato installato presso abitazione Via Bellini , Canegrate a circa 6 metri di altezza; distanza da confine area depuratore 410 metri ; coordinate al suolo 45°34'14.40 N – 8°56'20.06 E). il **monitoraggio è iniziato alle ore 17.08 del 30 Giugno 2015 e si è concluso alle ore 11.26 del 20 Luglio 2015** per un totale di 21 giorni consecutivi.



Ogni stazione di misura è stata programmata per eseguire un campionamento di aria ambiente ogni secondo per cicli di misura di 100 secondi seguite da 500 secondi di pulizia della cella di misura e dei sensori con aria zero. Durante il monitoraggio gli strumenti utilizzano tutti i sensori disponibili alle temperature programmate dall'operatore. I dati di misura della velocità e direzione del vento sono misurati nello stesso momento del campionamento e misura.

## 6) Classificazione, caratterizzazione (riconoscimento) delle dispersioni rilevate nel punto di misura.

Per il riconoscimento degli eventi registrati è stato utilizzato il metodo di analisi multivariata "Correlazione Pearson" (\*), metodo di uso comune e matematicamente fondato. A ulteriore verifica della coerenza dei risultati ottenuti, nel periodo di monitoraggio, sono stati rilevati anche velocità e direzione del vento ( $\Delta$ ): questi parametri sono molto utili per la verifica del fenomeno di dispersione sul territorio degli odori e per la ulteriore conferma della provenienza cardinale delle dispersioni e coerenza del risultato prodotto.

( $\Delta$ ) I dati meteo (direzione e velocità del vento) sono stati rilevati dalla centralina incorporata nella Stazione di rilevamento odori, a circa 6 metri dal suolo.

(\*) *Analisi di Correlazione, comunemente conosciuta come "coefficiente di correlazione Pearson" che ha lo scopo di "definire il grado di similarità tra due oggetti che risulteranno tanto più simili tra loro quanto le variabili che li descrivono avranno valori correlati". Questo metodo permette di svincolarsi dall'intensità dei segnali e di elaborare i dati (le impronte sensoriali) secondo il loro profilo (variabili correlate). La scelta di questo metodo è necessaria perché è importante prendere in considerazione la inevitabile diluizione a cui viene sottoposto il campione originale nella sua diffusione sul territorio e la conseguente diminuzione della sua intensità di segnale.*

## 7) Risultati finali e conclusioni

Si riportano di seguito tabelle complete degli eventi rilevati nelle due stazioni di misura, completi di direzione e velocità del vento durante evento

### **Stazione di misura PEN3-2, Via Bellini : eventi rilevati**

Data	Inizio evento	Direzione vento - velocità	Classe di attribuzione (riconoscimento)
30 Giugno	22.05	Calma senza vento	Sconosciuta
1 Luglio	21.14	NNE – 1 m/s	Sconosciuta
4 Luglio	20.36	SSO – bava di vento	Sconosciuta
<b>6 Luglio</b>	<b>21.36</b>	<b>SSO- 2,10 m/s</b>	Sconosciuta
9 Luglio	7.22	N – 6,70 m/s	Sconosciuta
10 Luglio	10.25	N – 3.60 m/s	Sconosciuta
<b>13 Luglio</b>	<b>20.40</b>	<b>SO – 1.50 m/s</b>	Sconosciuta
14 Luglio	20.05	SO – bava di vento	Sconosciuta

**Stazione di misura PEN3-1, Via Cascinette . : eventi rilevati**

Data	Inizio evento	Direzione vento - velocità	Classe di attribuzione (riconoscimento)
6 Luglio	21.16	SSO - 2.10 m/s	Sconosciuta
13 Luglio	22.50	SO - 1.50 m/s	Sconosciuta

N.B. sono stati riportati esclusivamente gli eventi di intensità superiore al massimo di segnale ottenuto sul campione di aria ambiente inodore + 50 % di G/GO

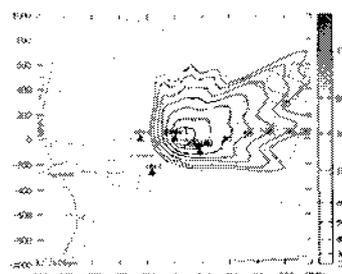
n.d.r.

**Il depuratore**

Ad esclusione delle vasche (OX, sedimentazione, ispessitori) tutte le altre zone sottoposte a prelievo sono zone confinate, in aspirazione e sottoposte a trattamento di abbattimento emissioni mediante filtri a carboni attivi.

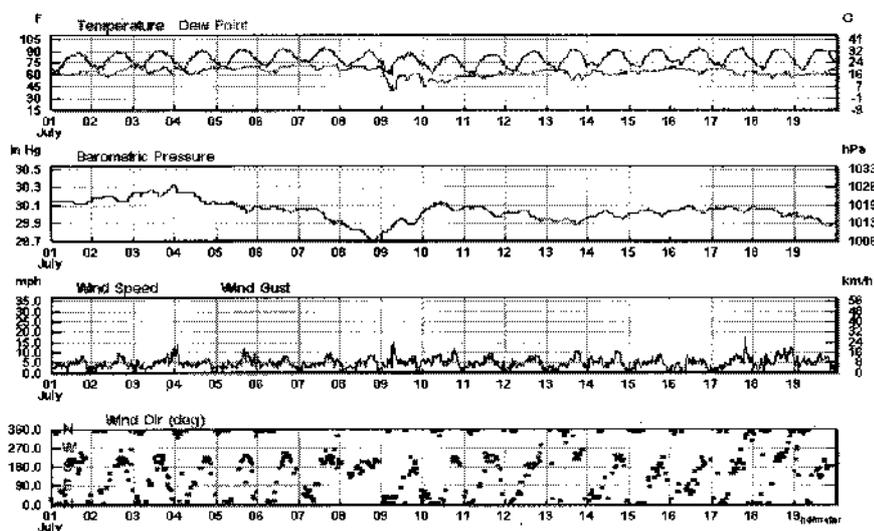
Le vasche di sedimentazione primaria, secondaria e di ossidazione posseggono una elevata superficie di scambio ma bassa emissione superficiale di composti odorigeni; sui campioni prelevati su queste vasche è stata calcolata anche la concentrazione di odore in emissione espressa in  $OU_E/m^3$  (massimo valore determinato  $52 OU_E/m^3$ ).

Utilizzando il dato di concentrazione, la superficie complessiva delle vasche ( $13176 m^2$ ) e i dati meteorologici del periodo è stato sviluppato un modello di dispersione di tipo stazionario (AERMOD) che descrive l'andamento del profilo verticale ed orizzonale delle concentrazioni all'interno della stable Boudary Layer (strato a contatto con il terreno) mediante una funzione ibrido-gaussiana Aermom) e un sistema di modellazione atmosferica integrato, costituito da 3 moduli specifici per simulazioni a raggio limitato (short range dispersion - scelto per la vicinanza dei recettori) di emissioni da fonti stazionarie e due moduli relativi al pre-processamento meteo (Aermet) e pre-processamento del territorio (Aermap).



**Il risultato al modello ci conferma la pressoché irrilevante influenza odorigena delle vasche sul territorio circostante:** alla distanza di 25 metri il valore di unità olfattometriche  $OU_E/m^3$  risultano inferiori alla soglia di percettibilità media dell'odore ( $5-20 OU_E/m^3$ ).

Di seguito andamento dei parametri meteo nel periodo di indagine 1 - 20 Luglio 2015



### Gli eventi rilevati

La totalità degli eventi registrati risultano essere di provenienza da emissione non presente nella nostra libreria (sconosciuti). L'elaborazione di correlazione Pearson applicata per il riconoscimento della sorgente ci fornisce sempre sorgente sconosciuta e con un livello di capacità predittiva da DFA (least probability) superiore al 90% (minimo range di accettabilità da noi richiesto per il riconoscimento 75%).

Il periodo di monitoraggio è stato caratterizzato da situazioni a basso trasporto aereo (velocità del vento media del periodo 1,60 m/s) e in queste condizioni le dispersioni odorigene ricadono in prossimità della sorgente stessa; in tale situazione stocastica il bersaglio di Via Cascinette dovrebbe essere quello maggiormente colpito e influenzato dalle eventuali dispersioni odorigene del depuratore **ma questo non avviene in considerazione degli eventi rilevati in tale sito e dalle loro caratteristiche.**

Il recettore di Via Bellini risulta essere quello maggiormente colpito da presenza di odore e, per il motivo sopra citato, la sorgente dispersiva si può ritenere locata nelle immediate vicinanze. Inoltre si nota nel caso specifico, una permanenza temporale mediamente lunga dell'odore, probabilmente dovuta dalla particolare conformazione della abitazione e ovviamente dalle condizioni stocastiche sopra menzionate.

Nelle giornate del 6 e 13 Luglio abbiamo condizioni di trasporto aereo leggermente superiore alla media del periodo (1.50 – 2.10 m/s) e in queste giornate si rilevano, ad orari coerenti e **con trasporto aereo Sud-Sud-Ovest – SO**, dispersioni di elevata similarità in entrambe gli strumenti installati.

### **La dispersione prevalente rilevata ai recettori**

In allegato si riportano le “impronte digitali” (espressione grafica della matrice numerica ottenuta ai sensori durante le misure) delle sorgenti campionate all’interno dell’impianto e quella rilevata ai recettori e classificata “Sconosciuta”; questo per mostrare, in forma grafica e per semplicità, la differenza.

### **Conclusioni**

Nel periodo di indagine 30 Giugno – 20 Luglio 2015 non si sono mai rilevate dispersioni oggettivamente riconducibili ad aree di pertinenza del depuratore



PCA Technologies S.r.l.  
Il Direttore Tecnico  
Dr. F. Crivelli

*Segue teoria delle elaborazione applicate nel presente studio*

**Le elaborazioni chemiometriche utilizzate nella indagine depuratore Canegrate  
periodo 30 Giugno- 20 Luglio 2015**

Nella determinazione dell'odore e delle dispersioni odorigene, l'utilizzo dei cosiddetti "nasi elettronici" a sensori MOS termoregolati singolarmente, è oggi unico mezzo per eseguire misure in continuo 24 ore su 24 ed ottenere misure oggettive e non influenzabili da fattori esterni. Questa tecnica di misura è oggi da considerarsi come "Migliore Tecnologia disponibile" nelle valutazioni dell'impatto odorigeno sul territorio. È importante però rispettare alcune procedure rigorosamente scientifiche e comportamentali che confermino la bontà delle misure effettuate e i calcoli utilizzati; per esempio:

- ↓ Controllare e Validare i dati in ingresso (le misure effettuate) con metodi solidi di calcolo come per esempio il metodo Leave one - leave out (vedi LDA-Linear Discriminant Analysis)
- ↓ Utilizzare metodi di elaborazione matematicamente solidi e riconosciuti validi come nel nostro caso le tecniche PCA, LDA, Correlazione Pearson, etc.)

Di seguito riportiamo, perché utile sia sotto l'aspetto formativo ed esplicativo di quanto fatto, i principi e la teoria dei metodi di elaborazione utilizzati nella indagine eseguita per conto di Provincia Autonoma di Trento, Agenzia per la depurazione, Servizio Gestione degli Impianti.

***Cosa è l'analisi multivariata o chemiometria:*** è una scienza che si basa sull'uso di metodi matematici e statistici per la elaborazione di una grande massa di dati.

Dove le variabili sono numerose e anche correlate tra loro, l'utilizzo dei metodi chemiometrici aiuta a fornire una visione globale del problema, evidenziando le relazioni tra le variabili considerate e l'importanza relativa di ciascuna di esse nell'ambito di un determinato problema, e può inoltre mettere in evidenza le relazioni tra i campioni in base alla loro distribuzione nello spazio multi-dimensionale (similitudine) descritto dall'insieme delle variabili (segnali ottenuti ai sensori di PEN3). Questi metodi sono alla base delle misure sensoriali e permettono i seguenti passaggi:

- ↓ Esplorazione iniziale dei dati
- ↓ Evidenziazione dell'esistenza di gruppi omogenei di campioni non classificati a priori (cluster analysis)
- ↓ Creazione di modelli matematici per la predizione di risposte quantitative (regressione)
- ↓ Creazione di modelli matematici per la predizione di risposte qualitative (classificazione).

Le variabili che si presentano nel trattamento di un problema multivariato possono essere di natura differente e possono, quindi, venire espresse da diverse unità di misura. Nel nostro caso, per la caratterizzazione di un campione (emissione aeriforme), le variabili sono i segnali dei sensori espressi in G/GO, il suo nome, la

classificazione di provenienza , etc. , tutte variabili queste con unità di misura o diverso peso.

La maggior parte dei metodi chemiometrici richiede quindi che venga effettuato un pretrattamento dei dati per eliminare l'effetto delle diverse unità di misura e dare a ciascuna variabile lo stesso peso a priori. In assenza di tale trattamento, le variabili espresse da numeri più grandi o che assumono valori in grandi intervalli avrebbero un peso maggiore (ad esempio, maggiore varianza) di variabili espresse da numeri piccoli o definiti in un intervallo di valori limitato. Per evitare che ciò si verifichi, è necessario trasformare tutte le variabili in modo da uniformarne le unità di misura. Il più comune metodo di scalatura è "autoscaling", che trasforma tutte le variabili in modo che ciascuna di esse abbia una media uguale a zero e una varianza unitaria. Altri metodi comunemente usati sono la centratura, in cui ciascuna variabile viene centrata intorno al proprio valor medio, e la scalatura di intervallo, in cui ogni variabile viene riscalata tra zero e uno.

***La prima fase di elaborazione dei dati derivanti dalle misure sulle emissioni campionate :***

***"Analisi delle componenti principali PCA (analisi esplorativa)"***

PCA è finalizzata ad estrarre la massima informazione possibile contenuta in una struttura di dati multivariati, sintetizzandola in poche combinazioni lineari delle variabili stesse. Questo metodo viene frequentemente impiegato nella prima fase di elaborazione dei dati e serve a dare una visione generale del problema, a capire le relazioni tra i campioni e/o le classi considerate ed a fornire un'indicazione preliminare sul ruolo delle variabili, mettendo eventualmente in luce la possibilità di eliminarne alcune che, essendo strettamente correlate tra loro, portano informazioni simili e possono quindi essere considerate ridondanti.

Dal punto di vista geometrico la PCA consiste in un processo di rotazione dei dati originali effettuato in modo che il primo nuovo asse (la prima componente principale PC#1) sia orientato nella direzione di massima varianza dei dati, il secondo sia perpendicolare al primo e sia nella direzione della successiva massima varianza dei dati e così di seguito per tutti i p nuovi assi. Il numero di questi nuovi assi (le componenti principali, PC#X) sarà quindi pari al numero di variabili originali (il numero dei sensori utilizzati, nel nostro caso n. 10 sensori).

***Metodo di classificazione***

Nel metodo di classificazione i gruppi dei campioni sono chiamati "classi" e preesistono agli oggetti che consideriamo (ciascun oggetto è a priori assegnato alla sua classe di appartenenza). In questo caso si opera quindi su insiemi di dati ove per ciascun campione è nota la classe di appartenenza (per esempio buono o cattivo, accettato o rifiutato, uscita biofiltro o emissione a camino, etc.).

L'obiettivo dell'analisi di classificazione è la verifica dell'esistenza di differenze tra le classi in funzione delle variabili considerate e la formulazione di un modello che sia in grado di assegnare ciascun campione alla classe cui esso appartiene. Se il modello è stato ottenuto da un insieme di dati di cui sono note le classi con certezza (training set), il potere predittivo del modello può essere verificato utilizzando un insieme di

dati anch'essi con classe nota (evaluation set), i cui campioni vengono classificati dal metodo secondo il modello precedentemente calcolato.

Infine, se il modello ottenuto è accettato, esso può essere utilizzato per l'assegnazione di oggetti di classe ignota alla classe che meglio li rappresenta. In questo modo possiamo trovare modelli per classificare per esempio individui sani e individui malati, emissione da azienda X da azienda Y sulla base di una serie di parametri. Esistono vari metodi di classificazione. Uno dei principali è l'Analisi Discriminante, nelle sue diverse varianti (Linear Discriminant Analysis, Quadratic Discriminant Analysis, Regularized Discriminant Analysis), che è considerata una tecnica di tipo parametrico e che pertanto presuppone una distribuzione multinormale delle variabili. Nel nostro caso è stata utilizzata l'analisi discriminante LDA. Esistono anche tecniche di tipo non parametrico (es.: KNN, K-Nearest Neighbour, anche questa parte integrante di PEN-Meteo), che prendono in considerazione solo le distanze tra gli oggetti nello spazio n-dimensionale, indipendentemente dalla loro distribuzione.

### ***Linear Discriminant Analysis, LDA (Validazione dei dati in ingresso)***

Il metodo LDA presuppone che tutte le variabili siano distribuite normalmente e che le matrici di varianza/covarianza siano simili per le diverse classi. Il metodo stima un'unica matrice di covarianza pesata e calcola una funzione discriminante lineare che separa gli oggetti per ottenere la massima discriminazione possibile tra i centri delle classi e la minima tra gli oggetti appartenenti alla stessa classe. I risultati dell'analisi possono essere visualizzati graficamente in modo analogo a quanto presentato relativamente all'Analisi delle Componenti Principali, cioè rappresentando gli "scores" in uno spazio a due dimensioni individuato da due funzioni discriminanti (generalmente le prime due che contengono la maggior parte dell'informazione). Questo metodo presenta il vantaggio di essere matematicamente ben fondato e può funzionare bene anche quando la distribuzione delle variabili non è perfettamente normale. Gli oggetti (campioni) vengono classificati in base alla seguente regola: un oggetto  $X$  viene classificato nella classe  $A$  se

$$p(A|X) > p(k|X) \quad \text{per tutti i } k \neq A \text{ e } k = 1, \dots, G \text{ dove}$$

$p(A|X)$  è la *probabilità a posteriori* che l'oggetto  $X$  appartenga alla classe  $A$  e  $G$  è il numero totale di classi. Questa probabilità viene calcolata dalla regola di Bayes:

$$p(A|X) = \frac{P_g f(X|A)}{\sum_k P_k f(X|k)}$$

dove  $P_g$  è la *probabilità a priori* della classe  $A$  e  $f(X|A)$  è la densità di probabilità che una classe  $A$  contenga l'oggetto  $X$ . La densità di probabilità è generalmente ignota e deve essere stimata dagli oggetti del training set. Di fatto, un oggetto  $X$  viene classificato nella classe  $A$  se è minima l'espressione

$$D_g(X) = (X - \bar{X}_g)^T S^{-1} (X - \bar{X}_g) + \ln |S_g| - 2 \ln P_g$$

Dove  $Dg(X)$  è detto **discriminant score**,  $S^{-1}g$  è l'inversa della matrice di covarianza della classe  $g$  e  $Xg$  è il corrispondente **centroide di classe**.

### **Metodi di cluster analysis**

La cluster analysis ha come fine quello di individuare, all'interno di un insieme, dei gruppi di campioni che presentano caratteristiche simili in funzione di determinate variabili. La similarità tra campioni appartenenti allo stesso gruppo (cluster) è sempre maggiore della similarità tra qualsiasi campione del gruppo con tutti gli altri campioni non appartenenti a quel gruppo. I metodi di cluster analysis presuppongono quindi che non esistano classi stabilite a priori ma, al contrario, hanno come obiettivo quello di ricercare l'eventuale esistenza di raggruppamenti "naturali". L'esistenza dei gruppi viene valutata in base alle caratteristiche degli oggetti di ciascun cluster. Il livello di similarità tra  $n$  oggetti distribuiti in uno spazio  $p$ -dimensionale (dove  $p$  è il numero delle variabili) viene valutato in base alla loro distanza: si assume cioè che due campioni siano tra loro simili se la loro distanza è piccola, e che siano dissimili se la loro distanza è grande. È quindi possibile calcolare le distanze tra i diversi campioni utilizzando una tra le numerose misure di distanza disponibili (distanza Euclidea, di Mahalanobis, di **Correlazione Pearson**, di Minkowski, ecc.) e trasformare una misura di distanza in una misura di similarità (sempre compresa tra zero e uno) con delle semplici trasformazioni. È questo uno dei modi di classificazione utilizzati da PEN-Meteo attraverso il software WinMuster.

### **Base teorica dei metodi di calcolo predittivi utilizzati in questo lavoro**

Durante il monitoraggio eseguito nel territorio di Canegrate per il riconoscimento e la classificazione delle dispersioni rilevate al bersaglio è stata applicata l'analisi di Correlazione, metodo di calcolo comunemente conosciuto come "Coefficiente di correlazione Pearson". Il coefficiente di correlazione Pearson è utilizzato per descrivere il grado di correlazione tra due variabili ed assume sempre valori compresi tra -1 e +1. Esso ha lo scopo di definire il grado di similarità tra due oggetti che risulteranno tanto più simili tra loro quanto le variabili che li descrivono avranno valori correlati. Il coefficiente di correlazione (lineare) di Pearson tra due variabili aleatorie (\*) o due variabili statistiche  $X$  e  $Y$ , è definito come la loro covarianza (°) divisa per il prodotto delle deviazioni standard (Λ) delle due variabili:

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

dove  $\sigma_{xy}$  è la covarianza tra  $X$  e  $Y$   
 e  $\sigma_x, \sigma_y$  sono le due deviazioni standard

Il coefficiente assume valori compresi tra -1 e +1

(\*) **variabile aleatoria o variabile casuale**: risultato numerico di un esperimento quando questo non è prevedibile con certezza (detto anche "non deterministico")

(°) **covarianza**: indice che misura la contemporaneità della variazione lineare di due variabili

(Λ) **deviazioni standard**: misura di variabilità di una popolazione o di una variabile casuale, derivato direttamente dalla varianza che ha la stessa unità di misura dei valori osservati)

Nel caso di indipendenza lineare il coefficiente assume valore zero, mentre non vale la conclusione opposta, ovvero dal coefficiente nullo non si può desumere l'indipendenza lineare. L'ipotesi di assenza di autocorrelazione (*autocorrelazione = grado di dipendenza spaziale tra i valori assunti da una variabile campionata; se è dimostrata autocorrelazione tra due valori al cambiare delle peculiarità di uno di essi varierà anche l'altro*) è più restrittiva ed implica quella di indipendenza fra due variabili. Valori prossimi a +1 (o -1) possono essere misurati anche in presenza di relazioni non lineari. Valori positivi vengono misurati in presenza di correlazione lineare positiva (per esempio  $y=a+bx$ , dove  $b>0$ ), mentre valori negativi vengono misurati in presenza di correlazione lineare negativa (es.:  $y=a+bx$ , con  $b<0$ ). Gli indici di Pearson di  $n$  variabili, possono essere presentati in una matrice (matrice = tabella rettangolare contenente dei numeri) di correlazione. Si tratta di una tabella a doppia entrata, che è una matrice quadrata (matrice quadrata = matrice con numero uguale di righe e colonne) di dimensione  $[n,n]$  avente nelle righe e colonne le variabili oggetto di studio. La matrice è diagonale (matrice diagonale = matrice quadrata in cui solamente i valori della diagonale principale possono essere diversi da zero), poiché l'indice di correlazione di una variabile con sé stessa è unitario, ed è simmetrica. Infatti, vale che:  $\rho_{xx} = 1$  e  $\rho_{yx} = \rho_{xy}$

### Analisi Discriminante DFA

L'analisi discriminante DFA, utilizzata da PEN3, viene condotta per definire una modalità di assegnazione dei casi a differenti gruppi, in funzione di una serie di variabili fra di loro correlate. I gruppi sono già definiti al momento dell'analisi, pertanto l'interesse è rivolto a **definire un modello che consenta di assegnare un nuovo caso ad un gruppo predefinito**, in funzione di un certo numero di variabili. Tramite l'analisi discriminante è possibile definire un modello matematico che ci consenta di collocare una eventuale nuova ed incognita impronta olfattiva sulla base di quelle conosciute.

Nell'analisi discriminante si trova una combinazione lineare di variabili che consente di calcolare il coefficiente di discriminazione (D):

$$D = b_0 + b_1x_1 + \dots + b_nx_n$$

dove:

$D$  = coefficiente discriminante

$b_0$  = costante

$x_i$  =  $n$ -ma variabile indipendente

$b_i$  =  $n$ -mo coefficiente delle funzioni disc.

Il numero di funzioni discriminanti ottenibili è uguale a  $(k-1)$  dove  $k$  è il numero dei gruppi. Per semplicità considereremo solo il caso della discriminazione fra due gruppi, ma le considerazioni fatte sono estendibili alla discriminazione fra più gruppi. Il metodo di calcolo della funzione discriminante è quello dei **minimi quadrati**, analogo a quello usato per la regressione lineare multipla, che consente di ottenere, per i valori di  $D$ , una variabilità minima all'interno dei gruppi e massima fra i gruppi.

Per due gruppi (A, B) essendo  $k=2$  esisterà una sola funzione discriminante i cui coefficienti sono dati dalla risoluzione della equazione seguente:

$$b = \Delta X \cdot W$$

dove:

$$b = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{pmatrix}$$

$$\Delta X = \begin{pmatrix} x_{1A} - x_{1B} \\ x_{2A} - x_{2B} \\ \vdots \\ x_{nA} - x_{nB} \end{pmatrix}$$

$W$  = matrice di dispersione comune

$x_{iA}$  = media della  $i$ -ma variabile del gruppo A

$x_{iB}$  = media della  $i$ -ma variabile del gruppo B

I valori medi dei coefficienti di discriminazione per i due gruppi sono calcolabili nel seguente modo:

$$\bar{D}_A = b_0 + b_1 x_{1A} + \dots + b_n x_{nA}$$

$$\bar{D}_B = b_0 + b_1 x_{1B} + \dots + b_n x_{nB}$$

con:

$x_{iA}$  = media della  $i$ -ma variabile indep. del gruppo A

$x_{iB}$  = media della  $i$ -ma variabile indep. del gruppo B

Le varianze del coefficiente di discriminazione per i due gruppi sono così calcolabili:

$$s^2_{D_A} = b \cdot X'_A X_A \cdot b$$

$$s^2_{D_B} = b \cdot X'_B X_B \cdot b$$

dove:

$X'_A X_A$  = matrice di dispersione del gruppo A

$X'_B X_B$  = matrice di dispersione del gruppo B

La soglia discriminante deve essere tanto più vicina ad una delle due medie tanto minore e' la deviazione standard del gruppo corrispondente, pertanto viene calcolata nel seguente modo:

$$D_0 = \frac{\bar{D}_A s_{D_B} + \bar{D}_B s_{D_A}}{s_{D_A} + s_{D_B}}$$

Di conseguenza, un nuovo elemento (impronta olfattiva di campione incognito) viene assegnato al gruppo in funzione della soglia discriminante: se il valore di  $D$  e' superiore alla soglia viene assegnato al gruppo con la media di  $D$  più alta, viceversa se il valore e' più piccolo. All'assegnazione di un elemento ad una classe può essere assegnata una probabilità, in funzione del valore dello scarto standardizzato fra il valore della media di gruppo ed il valore della soglia discriminante (% DFA) .