ELSEVIER

Chemometrics and Intelligent Laboratory Systems 50 (2000) 1–18

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Tutorial

The Mahalanobis distance

R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart *

ChemoAC, Pharmaceutical Institute, Department of Pharmacology and Biomedical Analysis, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium

Abstract

The theory of many multivariate chemometrical methods is based on the measurement of distances. The Mahalanobis distance (MD), in the original and principal component (PC) space, will be examined and interpreted in relation with the Euclidean distance (ED). Techniques based on the MD and applied in different fields of chemometrics such as in multivariate calibration, pattern recognition and process control are explained and discussed. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Mahalanobis distance; Euclidean distance; Principal components

Contents

* Corresponding author. Tel.: +32-2-477-4737; fax: +32-2-477-4735; E-mail: fabi@vub.vub.ac.be

0169-7439/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0169-7439(99)00047-7

1. Introduction

Multivariate chemometrical techniques are often based on the measurement of distances between objects. The most commonly used distance measures are the Euclidean distance (ED) and the Mahalanobis distance (MD) [1]. Both distances can be calculated in the original variable space and in the principal component (PC) space. The ED is easy to compute and interpret, but this is less the case for the MD.

In the original variable space, the MD takes into account the correlation in the data, since it is calculated using the inverse of the variance–covariance matrix of the data set of interest. However, the computation of the variance–covariance matrix can cause problems. When the investigated data are measured over a large number of variables (e.g., NIR spectra), they can contain much redundant or correlated information. This so-called multicollinearity in the data leads to a singular or nearly singular variance–covariance matrix that cannot be inverted. A second limitation for the calculation of the variance–covariance matrix is that the number of objects in the data set has to be larger than the number of variables. For these reasons, it is clear that in many cases, feature reduction is needed. This can be done by, e.g., selecting a small number of meaningful variables. The MD (and the ED) can also be calculated using a smaller number of latent variables (PCs) obtained after PC analysis (PCA) instead of the original variables. In this case, the MD, however, does not need to correct for the covariance between the variables, since PCs are by definition orthogonal (uncorrelated). However, the way each of the residual PCs is weighted in the computation of the distance must be taken into account. This will be explained in more detail in the next sections.

In the field of multivariate calibration, the MD is used for different purposes, namely: for the detection of outliers [2,3], the selection of calibration samples from a large set of measurements [4] and for investigating the representativity between two data sets [5,6]. In process control, the MD is used for, e.g., the Hotelling's T^2 test [7,8]. In pattern recognition, the MD is applied in clustering techniques such as the *k*-Nearest Neighbour method (kNN) [9], in discrimination techniques such as linear, quadratic and regularised discriminating analysis (LDA, QDA and RDA) [10] and in class modelling techniques such as UNEQ (multivariate normal class model assuming an individual dispersion of each class) [11], EQ (multivariate normal class model assuming equal dispersion of each class) [12] and modifications of Soft Independent Modelling of Class Analogy (SIMCA) [13].

A good comprehension of this distance is therefore useful. We will try to clarify the relationship between the ED and the MD calculated in the original variable space and the PC space.

2. Graphical interpretation of the MD

The MD and ED will first be illustrated with a simple example in two dimensions, x_1 and x_2. The two first columns of the simulated data in Table 1 were used. The ED towards the center of the data can be calculated for each of the n objects as

$$\mathrm{ED}_i = \sqrt{(x_{i1} - \bar{x}_1)^2 + (x_{i2} - \bar{x}_2)^2} \quad \text{for } i = 1 \text{ to } n, \tag{1}$$

where x_{i1} and x_{i2} are the values of the object i for, respectively, variables x_1 and x_2 and $\bar{x}_1$ and $\bar{x}_2$ the means of the n values measured at, respectively, $x_{.1}$ and $x_{.2}$. For the first object, the ED towards the center of the variable space is computed as

$$\mathrm{ED}_1 = \sqrt{(4-6)^2 + (3-5.350)^2} = 3.086.$$

Table 1
The simulated data with four variables ($x_1, \ldots, x_4$) and the same data after column-centering

Object number (i)	x_1	x_2	x_3	x_4	x_1 centered	x_2 centered	x_3 centered	x_4 centered
1	4.00	3.00	1.00	2.00	−2.000	−2.350	−2.125	−1.245
2	5.00	4.00	2.00	3.50	−1.000	−1.350	−1.125	0.255
3	8.00	7.00	3.00	4.00	2.000	1.650	−0.125	0.755
4	8.00	6.00	5.00	4.00	2.000	0.650	1.875	0.755
5	9.00	7.00	2.00	3.00	3.000	1.650	−1.125	−0.245
6	6.00	3.00	5.00	3.00	0.000	−2.350	1.875	−0.245
7	6.00	5.00	3.00	2.50	0.000	−0.350	−0.125	−0.745
8	10.00	8.00	2.00	3.00	4.000	2.650	−1.125	−0.245
9	2.00	3.00	1.50	3.40	−4.000	−2.350	−1.625	0.155
10	4.00	4.00	3.00	3.00	−2.000	−1.350	−0.125	−0.245
11	6.00	6.00	6.00	4.00	0.000	0.650	2.875	0.755
12	6.50	4.50	0.00	2.00	0.500	−0.850	−3.125	−1.245
13	9.00	8.00	5.00	5.00	3.000	2.650	1.875	1.755
14	4.00	5.00	1.00	1.00	−2.000	−0.350	−2.125	−2.245
15	4.00	6.00	3.00	5.00	−2.000	0.650	−0.125	1.755
16	6.00	7.00	2.00	4.00	0.000	1.650	−1.125	0.755
17	2.50	4.50	6.00	4.00	−3.500	−0.850	2.875	0.755
18	5.00	5.50	8.00	3.00	−1.000	0.150	4.875	−0.245
19	7.00	5.50	1.00	2.50	1.000	0.150	−2.125	−0.745
20	8.00	5.00	3.00	3.00	2.000	−0.350	−0.125	−0.245
$\bar{x}$	6.00	5.350	3.125	3.245				

This can be repeated for the remaining objects (see Table 2). Fig. 1a shows the centered data plotted in the original variable space. The variables x_1 and x_2 are clearly correlated, since all points appear on a line. The circles represent equal EDs to the center point of the data.

Table 2
The ED and MD of each object towards the centroid of the set of simulated data computed using only the first two original variables x_1 and x_2

Object number (i)	ED	MD
1	3.0859	1.5464
2	1.6800	0.9122
3	2.5928	1.0814
4	2.1030	0.9652
5	3.4238	1.3576
6	2.3500	2.2133
7	0.3500	0.3296
8	4.7982	1.8947
9	4.6392	1.8278
10	2.4130	0.9549
11	0.6500	0.6122
12	0.9862	1.0640
13	4.0028	1.7186
14	2.0304	1.0983
15	2.1030	1.8097
16	1.6500	1.5540
17	3.6017	1.8037
18	1.0112	0.7664
19	1.0112	0.5629
20	2.0304	1.5710

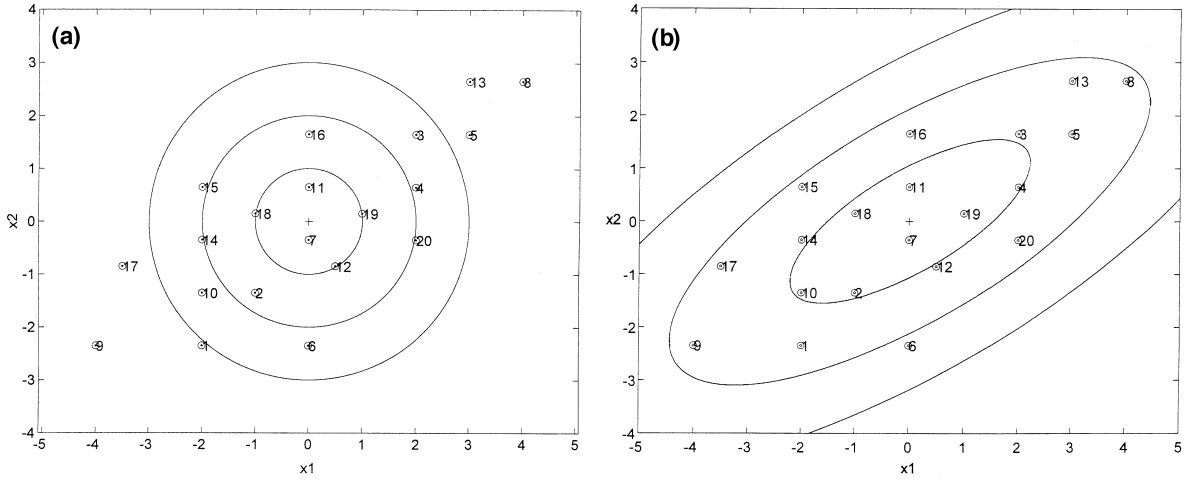


Fig. 1. (a) Plot of the simulated data for two variables x_1 and x_2 together with the circles representing equal EDs towards the center point. (b) Plot of the simulated data for two variables x_1 and x_2 together with the ellipses representing equal MDs towards the center point.

To be able to compute the MD, first the variance–covariance matrix $\mathbf{C}_x$ is constructed:

$$\mathbf{C}_x = \frac{1}{(n-1)}(\mathbf{X}_c)^T(\mathbf{X}_c), \tag{2}$$

where $\mathbf{X}$ is the data matrix containing n objects in the rows measured for p variables. $\mathbf{X}_c$ is the column-centered data matrix $(\mathbf{X} - \overline{\mathbf{X}})$. In the case of two variables, x_1 and x_2, the variance–covariance matrix is

$$\mathbf{C}_x = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \tag{3}$$

where σ_1^2 and σ_2^2 are the variances of the values of, respectively, the first and second variable; $\rho_{12}\sigma_1\sigma_2$ is the covariance between the two variables.

For our example, the variance–covariance matrix is equal to

$$\mathbf{C}_x = \begin{bmatrix} 4.921 & 2.500 \\ 2.500 & 2.397 \end{bmatrix},$$

with $\rho_{12} = \frac{2.5}{\sqrt{4.921}\sqrt{2.397}} = 0.728$. The MD for each object $\boldsymbol{x}_i$ is then

$$\mathrm{MD}_i = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})\mathbf{C}_x^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T}, \tag{4}$$

with

$$\mathbf{C}_x^{-1} = \begin{bmatrix} \sigma_2^2/\det(\mathbf{C}_x) & -\rho_{12}\sigma_1\sigma_2/\det(\mathbf{C}_x) \\ -\rho_{12}\sigma_1\sigma_2/\det(\mathbf{C}_x) & \sigma_1^2/\det(\mathbf{C}_x) \end{bmatrix},$$

where $\det(\mathbf{C}_x) = \sigma_1^2\sigma_2^2(1-\rho_{12}^2)$ is the determinant of the variance–covariance matrix.

For an object $\boldsymbol{x}_i$ measured in two variables, x_1 and x_2, Eq. (4) can be rewritten, since

$$\left[(x_1-\bar{x})(x_2-\bar{x})\right]\mathbf{C}_x^{-1}=\left[\frac{\sigma_2^2(x_1-\bar{x})-(x_2-\bar{x})\rho_{12}\sigma_1\sigma_2}{\det(\mathbf{C}_x)}\;\frac{\sigma_1^2(x_2-\bar{x})-(x_1-\bar{x})\rho_{12}\sigma_1\sigma_2}{\det(\mathbf{C}_x)}\right]$$

and

$$\begin{aligned}
&\left[(x_1-\bar{x})(x_2-\bar{x})\right]\mathbf{C}_x^{-1}\begin{bmatrix}(x_1-\bar{x})\\(x_2-\bar{x})\end{bmatrix}\\
&=\frac{\sigma_2^2(x_1-\bar{x})^2-(x_2-\bar{x})(x_1-\bar{x})\rho_{12}\sigma_1\sigma_2}{\det(\mathbf{C}_x)}+\frac{\sigma_1^2(x_2-\bar{x})^2-(x_1-\bar{x})(x_2-\bar{x})\rho_{12}\sigma_1\sigma_2}{\det(\mathbf{C}_x)}\\
&=\frac{\sigma_2^2(x_1-\bar{x})^2(1-\rho_{12}^2)+\sigma_1^2(x_2-\bar{x})^2-2(x_1-\bar{x})(x_2-\bar{x})\rho_{12}\sigma_1\sigma_2+\sigma_2^2(x_1-\bar{x})\rho_{12}^2}{\sigma_1^2\sigma_2^2(1-\rho_{12}^2)}\\
&=\frac{(x_1-\bar{x})^2}{\sigma_1^2}+\frac{(x_2-\bar{x})^2}{\sigma_2^2(1-\rho_{12}^2)}-2\frac{(x_1-\bar{x})(x_2-\bar{x})\rho_{12}}{\sigma_1\sigma_2(1-\rho_{12}^2)}+\frac{\rho_{12}^2(x_1-\bar{x})^2}{\sigma_1^2(1-\rho_{12}^2)}\\
&=\frac{(x_1-\bar{x})^2}{\sigma_1^2}+\left(\frac{x_2-\bar{x}}{\sigma_2\sqrt{1-\rho_{12}^2}}-\frac{\rho_{12}(x_1-\bar{x})}{\sigma_1\sqrt{1-\rho_{12}^2}}\right)^2
\end{aligned}$$

so that

$$\mathrm{MD}_i=\sqrt{\left(\frac{x_{i1}-\bar{x}_1}{\sigma_1}\right)^2+\left[\left\{\left(\frac{x_{i2}-\bar{x}_2}{\sigma_2}\right)-\rho_{12}\left(\frac{x_{i1}-\bar{x}_1}{\sigma_1}\right)\right\}\frac{1}{\sqrt{1-\rho_{12}^2}}\right]^2}. \tag{5}$$

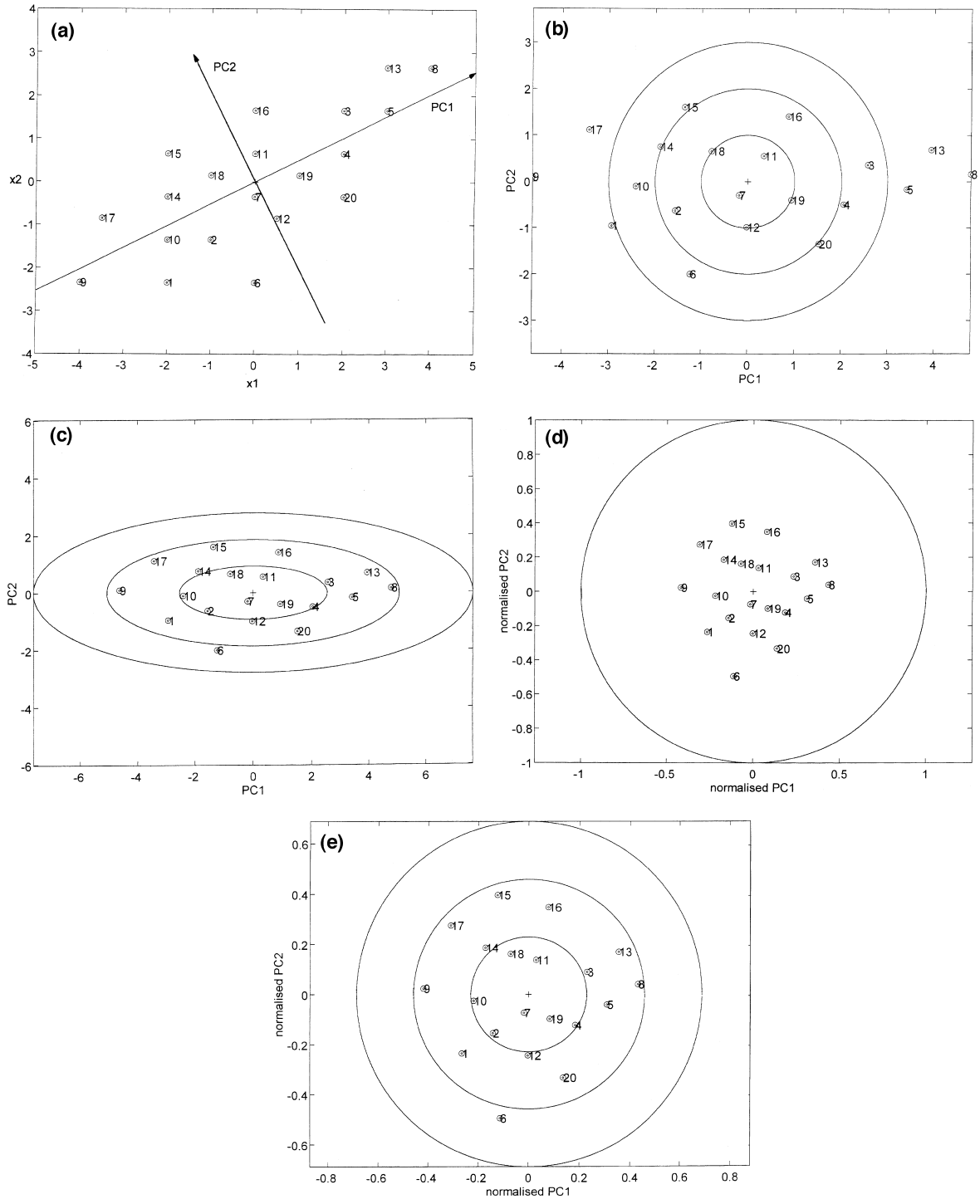
This expression shows that the part of the second variable which is already explained by the first variable is subtracted. In other words, the MD corrects for the correlation within the data. When the two variables in Eq. (5) are uncorrelated ($\rho_{12}=0$), the equation is reduced to the formula for the ED in Eq. (1). The MD computed using Eq. (5), for the first object is:

$$\mathrm{MD}_1=\sqrt{\left(\frac{-2}{\sqrt{4.921}}\right)^2+\left[\left\{\left(\frac{-2.350}{\sqrt{2.397}}\right)-0.728\left(\frac{-2}{\sqrt{4.921}}\right)\right\}\frac{1}{\sqrt{1-(0.728)^2}}\right]^2}=1.5464.$$

The ellipses in Fig. 1b represent equal MDs from the center point of the cloud in the original $x_{.1}$, $x_{.2}$ space. Points 6 and 10 are, respectively, at 2.21 and less than 1 MD unit away from the center point, while their EDs from the center, are about equal (2.35 and 2.41). This example illustrates the effect of taking into account the variance–covariance matrix of the data points. Since point 6 is lying in a direction where there is less correlation, the chance that a new measurement would be on that side of the cloud is smaller than the possibility that it would be in the position of point 10. The MD takes into account this probability due to the correlation within the two variables and attributes to point 6 a larger distance than to point 10, while the ED does not.

Let us now look what happens when we calculate the ED and the MD after performing PCA. As mentioned before, there can be clear advantages for working in the PC space. Let us again start by considering the simplest case, namely: when using only the first two variables of the data in Table 1. Accordingly, two PCs can be constructed of which the first one is in the direction of the largest variance within the data. The second PC is or-

thogonal to the first PC and is constructed through the center point of the cloud (see Fig. 2a), if we work on column-centered data. Fig. 2b shows the same data in what we will call the (unnormalised) PC space: the struc-



ture in the data remains the same, only the axes are rotated in the direction of the largest variance within the data. The points on the new axes now represent the scores of each object on PC_1 and PC_2.

Due to the way PCA is carried out, PC_1 always explains a larger amount of the total variance in the data than PC_2. Therefore, in the PC space as shown in Fig. 2b, the scores on each PC_1 and PC_2 are weighted according to the amount of the variance explained by PC_1 and PC_2. We can, if we prefer, calculate the ED and MD in this PC space, instead of in the original space. The circles in Fig. 2b show equal EDs from the center point of the cloud (which is also the center point of the PC space). Fig. 2c shows that equal MDs are represented by ellipses in the PC space. Point 6 is again more than two distance units away from the center point, while point 10 is only within one distance unit. The results are the same as for the MD in the original space (Fig. 1b). As will be shown in the numerical example and in the next paragraph, this is only true when one uses all PCs for the calculation of the MD.

When one divides the scores on each PC by their weight, i.e., the amount of the total variance in the data explained, the ''unweighted'', ''equally weighted'' or ''normalised'' scores are obtained. The normalised scores define the normalised PC space. Along each PC, the points have now the same variance.

The MD computed in the normalised PC space leads to circles around the center point, since the scores along each PC have now equal variance (Fig. 2d). On the plot, it can be seen that point 10 is less than one distance unit and point 6 at more than two distance units from the center. We observe that the MD computed in the normalised PC space is equal to the MD computed in the unnormalised PC space which will be confirmed in the next paragraph.

The ED computed in the normalised PC space also leads to circles around the center point. Fig. 2e shows only the circle for the ED = 1. The ED computed in the normalised PC space is clearly not equal to the ED computed in the (unnormalised) PC space (see Fig. 2b). As will be shown in the next paragraph, the ED computed in the normalised PC space is, except for a constant factor, equal to the MD computed in that space.

If one would represent the lines of equal EDs as computed in the unnormalised PC space in the normalised PC space, they would appear as ellipses. This is because the ED computed in the (unnormalised) PC space does not give equal importance to each of the PCs, but weighs each PC according to the amount of the total variance in the data explained by it. Since PC_1 is always constructed in the direction of the largest variance within the data, it is attributed a larger contribution (weight) than PC_2 for the calculation of the ED in the unnormalised PC space (see also the equations in Section 3).

3. Statistical interpretation of the MD

Chemical measurements contain random errors due to, e.g., the sampling, the sample pretreatment, the detector, etc. These random errors will tend to a normal distribution as the number of measurements becomes larger [14]. For a single variable x, the general form of the normal distribution is written as [15]:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)}, \quad (6)$$

with μ and σ the true mean and standard deviation of the measurements of $\boldsymbol{x}$ $(1 \times p)$. In other words, the function $f(x)$ is a density function describing the measurements x when they are univariately normal distributed.

Fig. 2. (a) Plot of the simulated data for two variables x_1 and x_2 together with the constructed PC_1 and PC_2. (b) Plot of the simulated data projected in the latent variable space, PC_1 vs. PC_2, or score plot together with the circles representing equal EDs towards the center point. (c) Score plot of the simulated data together with the ellipses representing equal MDs towards the center point. (d) Plot of the simulated data projected in the normalised PC space, PC_1 vs. PC_2, with the circles representing equal EDs towards the center point computed in this space. (e) Plot of the simulated data projected in the normalised PC space, PC_1 vs. PC_2, with the circles representing equal MDs towards the center point computed in this space.

The density function has a Gaussian shape. The normalisation factor $1/(\sigma\sqrt{2\pi})$ is used to standardise so that the area under the curve is always equal to 1. The multivariate form of the normal distribution is a direct generalisation of the univariate normal distribution [16]:

$$f(\boldsymbol{x}_i) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2}(2\pi)^{p/2}} \mathrm{e}^{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})^T}, \tag{7}$$

with $\boldsymbol{\Sigma}$ the true variance–covariance matrix of $\mathbf{X}$. Eq. (7) compared to the density function of a univariate normal distribution (Eq. (6)) shows that in the exponential, the MD (Eq. (8)) is used, which is generalised from the univariate $(x-\mu)^2/\sigma^2$. In the denominator, the variance–covariance matrix $\boldsymbol{\Sigma}$ is used instead of σ^2. The density function $f(\boldsymbol{x})$ in Eq. (7), visualised in two dimensions, has the shape of an ellipse. To find points in a p-dimensional space with an equal density, one can take the natural logarithm of both sides of Eq. (7), so that after rearranging, Eq. (8) is obtained:

$$\mathrm{MD}_i^2 = (\boldsymbol{x}_i - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})^T = a \ (\text{constant}). \tag{8}$$

The ED is computed as:

$$\mathrm{ED}_i^2 = (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T \tag{9}$$

and does not take correlation into account.

4. Distance measures and their relationship

4.1. ED and MD in the original space

A data matrix $\mathbf{X}$ $(n \times p)$, containing n objects $\boldsymbol{x}_i$ measured by p variables, is considered.

The ED between the ith row vector $\boldsymbol{x}_i$ $(1 \times p)$ of $\mathbf{X}$ and the mean row vector $\bar{\boldsymbol{x}}$ $(1 \times p)$ of $\mathbf{X}$ is calculated in the original space as

$$\mathrm{ED}_i^{\mathrm{o}} = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T} \quad \text{for } i = 1 \text{ to } n, \tag{10}$$

while the MD is calculated as

$$\mathrm{MD}_i^{\mathrm{o}} = \sqrt{(\boldsymbol{x}_i - \bar{\boldsymbol{x}})\mathbf{C}_x^{-1}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T} \quad \text{for } i = 1 \text{ to } n, \tag{11}$$

where $\mathbf{C}_x$ is the variance–covariance matrix (see Eq. (2)).

4.2. PC analysis

We will consider PCA after pre-treatment by column-centering. The data matrix $\mathbf{X}$ $(n \times p)$ is column-centered by subtracting the column mean $\bar{x}_j$ from each column vector $\boldsymbol{x}_j$ $(n \times 1)$.

A basic equation in this context is the singular value decomposition (SVD).

$$\mathbf{X}_{\mathrm{c}} = \mathbf{U}(n \times a)\,\boldsymbol{\Lambda}(a \times a)\mathbf{V}^T(a \times p) \quad \text{where } a = n-1 \text{ if } n \le p \text{ or } a = p \text{ if } n > p. \tag{12}$$

$\boldsymbol{\Lambda}$ is the diagonal matrix of the singular values λ_i, for $i = 1$ to a, which describe the amount of the total variance explained by each PC. They are sorted such that $\lambda_1 > \lambda_2 > \ldots \lambda_a$.

U is the matrix which contains for each object a row vector of what we have called the normalised scores, the product of $\mathbf{U}\mathbf{\Lambda}$ is the score matrix $\mathbf{T}$ $(n \times a)$ with the row vectors containing the (unnormalised) scores. $\mathbf{V}$ $(p \times a)$ is the loadings matrix. If the first r PCs are retained, Eq. (6) can be rewritten as:

$$\mathbf{X}_c = \mathbf{U}(n \times r)\,\mathbf{\Lambda}(r \times r)\mathbf{V}^T(r \times p) + \mathbf{E}(n \times p) \quad \text{for } r < a, \tag{13}$$

where **E** is the matrix containing the residuals. Many methods were proposed to calculate the number of significant PCs, but we will not go into this subject here. Jackson [17] and Malinowski and Howery [18] provide an overview of some methods.

The SVD was applied on the simulated data shown in Table 1 using all four variables (see Table 3).

4.3. The ED and MD in the PC space

The definitions for the ED and MD in Eqs. (10) and (11) can be applied using either the unnormalised or the normalised scores.

Table 3
The result of applying the SVD algorithm on the column-centered simulated data using all four variables

U =

−0.3020	−0.2050	0.1631	−0.0416
−0.1497	−0.0857	−0.0102	−0.2989
0.2332	−0.0362	−0.1372	−0.0693
0.2120	0.1632	0.1288	−0.1750
0.2819	−0.1874	0.0327	0.0647
−0.0899	0.1775	0.4837	−0.3137
−0.0271	−0.0333	0.1386	0.1283
0.4032	−0.2090	−0.0085	0.1712
−0.4263	−0.0699	−0.1913	−0.2156
−0.2162	0.0253	0.0131	−0.0222
0.0777	0.3112	−0.0146	0.0577
−0.0595	−0.3618	0.1423	−0.0216
0.3927	0.1705	−0.1848	−0.1429
−0.2187	−0.2277	0.0047	0.6311
−0.0974	0.0831	−0.5317	−0.1709
0.0726	−0.0897	−0.4030	0.0696
−0.2517	0.3909	−0.1573	0.0837
−0.0053	0.5086	0.2061	0.4185
0.0440	−0.2574	0.0467	0.0414
0.1263	−0.0672	0.2788	−0.1944

S =

11.1717	0	0	0
0	9.3529	0	0
0	0	4.4101	0
0	0	0	2.8220

V =

0.8239	−0.2211	0.4512	−0.2622
0.5311	0.0194	−0.6331	0.5628
0.1472	0.9423	0.2699	0.1323
0.1324	0.2505	−0.5681	−0.7727

The ED between the vector $\boldsymbol{x}_i$ $(1 \times p)$ and the mean vector $\bar{\boldsymbol{x}}$ $(1 \times p)$ in the unnormalised PC space is calculated as

$$\mathrm{ED}_i^{\boldsymbol{t}} = \sqrt{\sum_{j=1}^{r} t_{ij}^2} = \sqrt{\boldsymbol{t}_i \boldsymbol{t}_i^T} \quad \text{for } i = 1 \text{ to } n. \tag{14}$$

Using the normalised score vectors $\boldsymbol{u}_i$ $(1 \times r)$ of the normalised score matrix $\mathbf{U}$ $(n \times r)$ for r important PCs, the ED in the normalised PC space is calculated as

$$\mathrm{ED}_i^{\boldsymbol{u}} = \sqrt{\sum_{j=1}^{r} u_{ij}^2} = \sqrt{\boldsymbol{u}_i \boldsymbol{u}_i^T} \quad \text{for } i = 1 \text{ to } n. \tag{15}$$

In analogy with the ED, the MD in the unnormalised and normalised PC space are calculated using, respectively, the unnormalised ($\mathbf{T}$) or normalised ($\mathbf{U}$) score matrix.

In the unnormalised PC space, the variance–covariance matrix and the MD are calculated as

$$\boldsymbol{C}_{\boldsymbol{t}} = \boldsymbol{T}^T \boldsymbol{T} / (n-1) \tag{16}$$

$$\mathrm{MD}_i^{\boldsymbol{t}} = \sqrt{\boldsymbol{t}_i \mathbf{C}_{\boldsymbol{t}}^{-1} \boldsymbol{t}_i^T} \quad \text{for } i = 1 \text{ to } n. \tag{17}$$

In the normalised PC space,

$$\mathbf{C}_{\boldsymbol{u}} = \boldsymbol{U}^T \boldsymbol{U} / (n-1) \tag{18}$$

$$\mathrm{MD}_i^{\boldsymbol{u}} = \sqrt{\boldsymbol{u}_i \mathbf{C}_{\boldsymbol{u}}^{-1} \boldsymbol{u}_i^T} \quad \text{for } i = 1 \text{ to } n. \tag{19}$$

4.4. Relations between the ED and the MD in the original and the normalised and unnormalised PC space

□ Since $(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$ is equal to $\boldsymbol{u}_i \boldsymbol{\Lambda} \mathbf{V}^T$ after PCA on the mean centered data matrix $\mathbf{X}_c$, Eq. (10) can be written as

$$\mathrm{ED}_i^{o} = \sqrt{(\boldsymbol{u}_i \boldsymbol{\Lambda}) \mathbf{V}^T \mathbf{V} (\boldsymbol{u}_i \boldsymbol{\Lambda})^T} = \sqrt{(\boldsymbol{u}_i \boldsymbol{\Lambda}) \mathbf{I} (\boldsymbol{u}_i \boldsymbol{\Lambda})^T} = \sqrt{(\boldsymbol{u}_i \boldsymbol{\Lambda})(\boldsymbol{u}_i \boldsymbol{\Lambda})^T} = \sqrt{\boldsymbol{t}_i \boldsymbol{t}_i^T}, \tag{20}$$

where $\mathbf{I}$ is the identity matrix. This formula shows that the ED in the original space is equal to the ED in the unnormalised PC space when all (a) PCs are selected.

□ The variance–covariance matrix for the calculation of the MD in the unnormalised PC space (Eq. (16)) can be simplified to

$$\mathbf{C}_{\boldsymbol{t}} = (\mathbf{U}\boldsymbol{\Lambda})^T (\mathbf{U}\boldsymbol{\Lambda}) / (n-1) = \boldsymbol{\Lambda}^2 / (n-1) \tag{21}$$

so that Eq. (17) can be rewritten as

$$\mathrm{MD}_i^{\boldsymbol{t}} = \sqrt{(\boldsymbol{u}_i \boldsymbol{\Lambda}) \mathbf{C}_{\boldsymbol{t}}^{-1} (\boldsymbol{u}_i \boldsymbol{\Lambda})^T} = \sqrt{n-1} \sqrt{(\boldsymbol{u}_i \boldsymbol{\Lambda}) \boldsymbol{\Lambda}^{-2} \left(\boldsymbol{\Lambda}^T \boldsymbol{u}_i^T\right)} = \sqrt{n-1} \sqrt{\boldsymbol{u}_i \boldsymbol{u}_i^T} = \sqrt{n-1} * \mathrm{ED}_i^{\boldsymbol{u}}. \tag{22}$$

The variance–covariance matrix for the normalised scores (Eq. (18)) can be simplified to

$$\mathbf{C}_{\boldsymbol{u}} = \mathbf{I} / (n-1) \tag{23}$$

so that Eq. (19) can be rewritten as

$$\mathrm{MD}_i^{\boldsymbol{u}} = \sqrt{n-1} \sqrt{\boldsymbol{u}_i \mathbf{I}^{-1} \boldsymbol{u}_i^T} = \sqrt{n-1} \sqrt{\boldsymbol{u}_i \boldsymbol{u}_i^T} = \sqrt{n-1} * \mathrm{ED}_i^{\boldsymbol{u}}. \tag{24}$$

Table 4
The computed ED and MD in the original variable space (ED^o and MD^o), the (unnormalised) PC space (ED^t and MD^t) and the normalised PC space (ED^u and MD^u) for the simulated data of Table 1 using all four variables

Object number (i)	ED_i^o	ED_i^t for all PCs	ED_i^u for all PCs	MD_i^o	MD_i^t for all PCs	MD_i^u for all PCs
1	3.9482	3.9482	0.4019	1.7519	1.7519	1.7519
2	2.0379	2.0379	0.3452	1.5049	1.5049	1.5049
3	2.7034	2.7034	0.2817	1.2277	1.2277	1.2277
4	2.9169	2.9169	0.3447	1.5025	1.5025	1.5025
5	3.6122	3.6122	0.3462	1.5092	1.5092	1.5092
6	3.0163	3.0163	0.6099	2.6584	2.6584	2.6584
7	0.8326	0.8326	0.1937	0.8442	0.8442	0.8442
8	4.9344	4.9344	0.4854	2.1159	2.1159	2.1159
9	4.9180	4.9180	0.5193	2.2636	2.2636	2.2636
10	2.4286	2.4286	0.2192	0.9555	0.9555	0.9555
11	3.0427	3.0427	0.3263	1.4221	1.4221	1.4221
12	3.5054	3.5054	0.3939	1.7168	1.7168	1.7168
13	4.7559	4.7559	0.4877	2.1259	2.1259	2.1259
14	3.6984	3.6984	0.7057	3.0760	3.0760	3.0760
15	2.7419	2.7419	0.5729	2.4974	2.4974	2.4974
16	2.1350	2.1350	0.4249	1.8522	1.8522	1.8522
17	4.6699	4.6699	0.4979	2.1702	2.1702	2.1702
18	4.9848	4.9848	0.6901	3.0081	3.0081	3.0081
19	2.4684	2.4684	0.2685	1.1702	1.1702	1.1702
20	2.0489	2.0489	0.3687	1.6073	1.6073	1.6073

This means that for the same number of selected PCs (r), the MDs in the unnormalised and normalised PC space are equal. It also shows that when all n PCs are used, these MDs are, except for the constant factor $\sqrt{n-1}$, equal to the ED in the normalised PC space.

□ When all n PCs are used, $(\boldsymbol{x}_i - \bar{\boldsymbol{x}})$ is equal to $\boldsymbol{u}_i \boldsymbol{\Lambda} \mathbf{V}^T$, and the MD in the original space (see Eq. (11)) can be rewritten as

$$\begin{aligned} MD_i^o &= \sqrt{\left(\boldsymbol{u}_i \boldsymbol{\Lambda} \mathbf{V}^T\right)\left[\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T\right)^T\left(\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T\right)/(n-1)\right]^{-1}\left(\boldsymbol{u}_i \boldsymbol{\Lambda} \mathbf{V}^T\right)^T} \\ &= \sqrt{n-1}\sqrt{\left(\boldsymbol{u}_i \boldsymbol{\Lambda} \mathbf{V}^T\right)\left(\mathbf{V}\boldsymbol{\Lambda}^{-2}\mathbf{V}^T\right)\left(\mathbf{V}\boldsymbol{\Lambda}\,\boldsymbol{u}_i^T\right)} = \sqrt{n-1}\sqrt{\boldsymbol{u}_i \boldsymbol{u}_i^T} = \sqrt{n-1} * ED_i^{\boldsymbol{u}}, \end{aligned} \tag{25}$$

which means that the MD in the original space is equal to the MD in the normalised and unnormalised PC space when all PCs are used.

Table 4 shows the computation of the ED and MD in the original variable space (ED^o and MD^o), the (unnormalised) PC space ($ED^{\boldsymbol{t}}$ and $MD^{\boldsymbol{t}}$) and in the normalised PC space ($ED^{\boldsymbol{u}}$ and $MD^{\boldsymbol{u}}$) for the simulated data of Table 1 using all four variables. It can be seen that in the original variable space, the ED and MD are indeed clearly different. The MDs in all three spaces are the same when using all PCs in the computations. The EDs in the original space and the PC space are also the same when using all PCs. The ED in the normalised PC space is equal to the MDs, but for a constant factor.

5. Chemometrical methods based on the MD

Many multivariate techniques require that a model is built using a training set. The training set contains objects of which the characteristics under examination are known. In calibration, this means that, e.g., the concentration of the measured substances is known. In process control, the training set contains measurements that rep-

resent all variations that are encountered during normal practice (i.e., when the process is considered to be in-control). In (supervised) pattern recognition, it is known to which class the objects of the training set belong.

5.1. Multivariate calibration

In multivariate calibration, distance measures are applied in outlier detection [2,3], sample selection techniques [4] and methods studying the representativity between two data sets [5,6].

An important step in the calibration procedure is the detection of outliers. The regression line in Fig. 3 was built using the unlabelled points. The labelled points are added to demonstrate the different types of outliers. Two types of outliers can be investigated before building the regression model. The first type of outliers, so-called outliers in y, can be detected by using only the information on the y-axis. These outliers (e.g., points 1 and 3) can be detected using univariate outlier tests such as, e.g., Grubbs [19–21] or the Dixon [21,22]. Since they are univariate, there is no need for using the MD. The second type, so-called outliers in **X**, are identified using only the information on the x-axis. In the example, points 2 and 3 are outliers in **X**. A classical way to detect outliers in **X** (training set) is to compute the squared MD between each point and the mean of the training set in the original variable space, using Eq. (11), and to compare it to the tabulated χ^2-distribution with $(p-1)$ degrees of freedom [23]. Using the χ^2 distribution implies that the real mean (μ) and variance–covariance matrix $(\mathbf{\Sigma})$ are known, which is mostly not the case. Hotelling's T^2 test [7,8] is basically the same, but the squared MD values are called T^2-values which are compared with a critical value obtained by [24]:

$$T^2_{\mathrm{UCL}} \cong \frac{(n-1)^2}{n}\beta_{(\alpha;p/2,(n-p-1)/2)}. \tag{26}$$

The limits constructed using the β-distribution take in account that $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are unknown and that the detection of outliers is retrospective. Indeed, if an outlier is present, the estimated mean $\bar{\boldsymbol{x}}$ and variance–covariance matrix $\mathbf{C}_x$ for computing the MDs are already influenced by it. An alternative is therefore to apply the leave-one-out or leave-few-out method. One or more objects are systematically deleted from the measurements and the remaining objects are used to estimate $\bar{\boldsymbol{x}}$ and $\mathbf{C}_x$. It is then checked whether the deleted objects are under or above the control limit in Eq. (28) with, in case of leave-one-out, n is replaced by $n-1$. As explained in Section 1, when working with a larger number of variables than objects, the MD cannot be computed anymore due to the

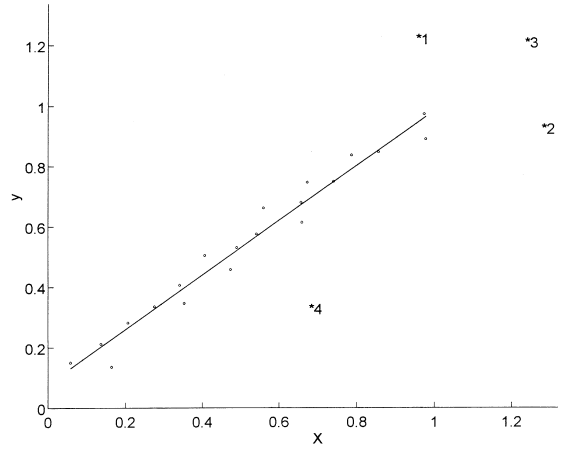


Fig. 3. Illustration of the different types of outliers.

singularity problem of the variance–covariance matrix. In that case, Hotelling's T^2 test can be applied in the PC space [25–27] (see further in Section 3).

The extreme behaviour of an object i in the **X**-space can also be measured by its so-called leverage value [2,28,29]. A point with a high leverage value (high leverage point) is a point that has a relatively large influence on the estimated regression parameters such as response, regression coefficient and standard error [30]. The leverage is closely related to the MD as shown in Eq. (27).

$$h_i = \frac{1}{n} + \frac{\left(\mathrm{MD}_i^o\right)^2}{n-1} \tag{27}$$

The computed leverage is compared to, e.g., a cut-off value of $2g/n$, with g the number of regression parameters to be estimated. The leverage can also be computed in the PC space using either all a or the first r significant PCs. In these methods, using the MD after PCA for the detection of high leverage points or outliers in **X**, one considers the deviation in each of the examined PCs equally important even when some PCs account only for a small part of the total variance within the data. Before building the regression model (e.g., using PC regression, PCR), one does not know whether these PCs are important or not for building the model. Therefore, it is important to take them all into account. Rousseeuw [31,32] introduced a robust version of the MD as a tool for the detection of outliers, the minimum volume ellipsoid (MVE) estimator. The variance–covariance is estimated using the at least $(n/2)+1$ objects of **X** which cover the smallest possible ellipsoid. In this way, the variance–covariance matrix is not influenced by multiple outliers.

A third type of outliers, so-called outliers towards the model, can be found only after building the regression model. They represent a different relationship between **X** and y than the majority of objects. Point 4 is an example of an outlier towards the model which is not an outlier in y, nor an outlier in **X**.

The MD is also applied for the selection of calibration samples. For the use with near infrared (NIR) spectra, two algorithms were proposed by Shenk and Westerhaus [4]. First, for each spectrum of a sample, the MD towards the mean spectrum is calculated. The measurements are then ordered according to this distance and represented in a histogram to detect and remove possible outliers. Then calibration samples are selected from a large set of measurements. It is based on the elimination of similar spectra. The MD between each of the spectra is calculated. The spectrum with the most neighbours at a distance smaller than a defined threshold is detected and all the neighbours are eliminated. Among the remaining spectra, this procedure is repeated until no points are detected anymore which have neighbours at a distance smaller than the threshold.

Representativity between two data sets was recently described as the simultaneous occurrence of three properties, namely: they should have similar directions in space, a similar variance–covariance matrix and the same position of the centroids [5,6]. In order to check this latter property, one can measure the MD between the centroids of each set, and compare it to a critical value computed by means of Hotelling's T^2 test. Among others, it is useful to check the representativity between the training and test sets.

5.2. Process control

The MD is used in Hotelling's T^2 test [7,8,33–37], a powerful tool to build multivariate process control charts using the original or latent variables (PCs). We will discuss Hotelling's T^2 test after performing PCA. The confidence limits are built using the training set which should contain the measurements representing the normal (in-control) situation. Therefore, in a first step, the training set itself has to be cleared from outliers. The MD of each object towards the center of the PC space is computed as in the third expression of Eq. (22). One usually applies only the first r important PCs as, e.g., determined by leave-one-out cross-validation [38], for the computation of MD. The computed MD (in this case, the so-called T^2-value) of each object is compared to the critical T^2-value as defined in Eq. (26). After clearing the training set from outliers, the MDs in Eq. (22) are recom-

puted and the upper control limit, which is used for predicting whether new objects are under control or not, is determined as [39]:

$$T^2_{\mathrm{UCL}} = \frac{p(n-1)(n+1)}{n(n-p)} F_{(\alpha;p,n-p)}. \tag{28}$$

The T^2-value of each measurement can be plotted together with this critical level to obtain a control chart. In this application of the MD, it is an advantage that the important PCs are equally weighted, since one wants to see all changes that occur in the measurements even when it occurs in a PC that explains a small amount of the total variance within the training set data.

5.3. Pattern recognition

Pattern recognition techniques can be divided into different categories according to certain properties. A first group of pattern recognition methods can be labelled as clustering techniques (unsupervised pattern recognition). The aim is to group similar objects (measured samples) together. In hierarchical clustering, one starts with all objects and links one by one the most similar objects together. As a measure of similarity, the ED is used. The MD can be used in the same way to link similar populations together by computing the MD between the population means [40]. A second group of pattern recognition techniques is called supervised because one possesses a training set of objects which are known to belong to a certain class. A mathematical model can be constructed to predict to which of the classes new measurements belong. Supervised pattern recognition methods can be divided in discrimination and class modelling techniques.

For discrimination methods, one global model is built to discriminate the classes. This means that each new object will always belong to (only) one of the classes. LDA [41], QDA and RDA [42,43] are discrimination techniques which can be explained in different manners [9]. In case of LDA, a new measurement $\boldsymbol{x}_i$ is predicted to be a member of the class K towards which it has the smallest so-called classification score:

$$\mathrm{cf}_K(\boldsymbol{x}_i) = (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_K)\mathbf{C}^{-1}_{\mathrm{pooled}}(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_K)^T - 2\ln \pi_K, \tag{29}$$

with $\mathrm{cf}_K(\boldsymbol{x}_i)$ the classification score, $\mathbf{C}^{-1}_{\mathrm{pooled}}$ is the pooled variance–covariance matrix of the different classes. The first term is in fact the squared MD computed between the new measurement and the centroid of class K. The second term $2\ln \pi_K$ is the prior probability that has to be taken in account when the number of objects in each class is not the same. In Fig. 4, the center points of two classes A and B are shown with around each class the lines of equal probability (or equal classification scores) towards the center point. Also the ellipse of equal

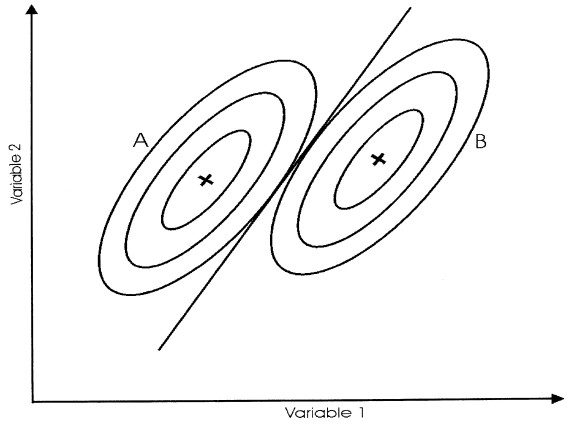


Fig. 4. The center points (+) of class A and B together with their ellipses of equal classification scores towards, respectively, the center point and the line that separates the classes according to LDA.

probability of class A which is in contact with the ellipse of equal probability of class B by exactly one point is drawn. In the contact point, the tangent on the ellipses can be constructed. This straight line defines the border between the two classes. Depending on the side of the border the new point is located, it is decided that it belongs to one of the two classes. In QDA, the variance–covariance matrices of all classes are considered to be unequal so that the respective variance–covariance matrix of each class is used in Eq. (29) instead of the pooled variance–covariance matrix in LDA. In this case, the borders between the classes can be represented by quadratic curves instead of straight lines in LDA. RDA is a combination of both LDA and QDA with borders graphically represented by combinations of straight and quadratic lines.

UNEQ, EQ and SIMCA are supervised pattern recognition techniques of the class modeling type, since confidence limits are calculated for each class separately. They are parametric methods, since it is assumed that the classes are multivariately normal distributed. In UNEQ [11] for each new object, the MD is computed between the object and the centroid of each class using the original variables. Similar to what is done in QDA, the particular variance–covariance matrix of each class is used for the computation of the MD. In UNEQ, however, a separate decision is made for each particular class whether the new object belongs to it or not. Similar as for the Hotelling's T^2 test in process control using the original variables, the decision is made according to a critical value (limit) which is based on the training set of the particular class. The critical value is the value obtained from a χ^2-distribution for p degrees of freedom [12], which is similar to the limit used for the MD as an outlier test described in Section 5.1. UNEQ, and the class modeling pattern recognition methods, in general, can indeed be considered as outlier tests for each separate class [44]. EQ is similar to UNEQ, but for all the classes in the training set, the same pooled variance–covariance matrix is used. In SIMCA [13,44], first, PCA is performed on the objects belonging to the class under examination (class training set). Second, the number of important PCs that describe the class variation is determined, e.g., by leave-one-out cross-validation [38]. The variance contained in the remaining (residual) PCs is very small and considered to be normally distributed. To test whether a new object belongs to the class under examination or not, the object is first projected in the PC space of the class. If the object belongs to the class, the selected (important) PCs of the class training set describe the object well so that the residual variance is small and comparable to the residual variance of the training set objects. If the object does not belong to the class, the important PCs of the class training set do not describe the object well so that the residual variance of the object is significantly larger than for the objects belonging to the class. As a measure of this residual variance, the squared ED between the point and the space spanned by the significant PCs of the particular class is computed. This is done by using only the scores on the remaining (residual) PCs in Eq. (10). The residual variance of the training set and the new object can be compared using an F-test for certain degrees of freedom as shown in Eq. (30).

$$F_j^{\text{new}} = \frac{s_j^2}{s_0^2} = \frac{\sum_{i=r+1}^{a} \left(t_{ji}^{\text{new}}\right)^2/(a-r)}{\sum_{k=1}^{m}\sum_{i=r+1}^{a} \left(t_{ki}^{K}\right)^2/(a-r)(m-r-1)}, \tag{30}$$

where r is the number of important PCs, a is the total number of computable PCs, m is the number of objects in the training set. Some variants of SIMCA use the squared MD distance between a new point and the space spanned by the significant PCs of the particular class instead of the ED for the computation of the residual variances s_j^2 and s_0^2 [45,46].

6. Some applications of the MD in chemistry

Smith et al. [47] and Caudill et al. [48] used Hotelling's T^2 test for controlling a laboratory process with 40 variables. Kourti and MacGregor [49], Nomikos and MacGregor [50] and MacGregor and Kourti [51] worked on

the monitoring of continuous and batch polymerisation processes in a petroleum refinery. In Fig. 5, an example from the literature [39] is shown. A continuous polymerisation process is monitored using a Hotelling's T^2 chart for five original variables such as, e.g., the average molecular weights and the branching of the polymer chains. The constructed $\alpha = 0.01$ and 0.05 limits are shown together with the new measurements of the process. From measurement 52 onwards, it is known that fouling in the reactor occurs, which is confirmed by the measurements being out of control. Hotelling's T^2 test based on PCA was applied to monitor the homogeneity of a powder mixture on-line using NIR spectroscopy [52]. In this case, the training set contains measurements of a homogeneous mixture that one wants to obtain. The blending process starts with an out of control situation (an inhomogeneous blend) and checks when the measurements become in control. Nijhuis et al. [53] used Hotelling's T^2 charts based on PCA, for checking the response of a gas chromatograph analysing the fatty acid composition in soya–maize oil. Cho and Gemperline [54] adapted the Hotelling's T^2 test to be more robust by estimating the variance–covariance matrix by the MVE to examine NIR spectra of sulfamethoxazole.

UNEQ was used for the classification of patients according to five laboratory tests for the functional state of the thyroid [55] and for classifying olive oils due to their region of origin [56]. Mark [57] used the MD to evaluate which sample preparation method is the best for analysing meat samples by NIR spectroscopy. Gemperline and Boyer [58] and Mark and Tunnell [59] used the MD for the classification of raw materials (pharmaceutical excipients) by NIR spectroscopy. Aldridge et al. [60] successfully applied the MD for the discrimination between the desired polymorph and other crystalline forms. Different pattern recognition methods such as, e.g., LDA were used to discriminate roasted coffees [61]. Discrimination methods were also applied on electronic nose data [62] and fatty acids composition measured by gas chromatography [63] of vegetable oils. Ramadan et al. [64] used LDA for classifying detectors for ion chromatography. SIMCA, adapted by using the MD, was used for the classification of tablets of clinical study lots [46].

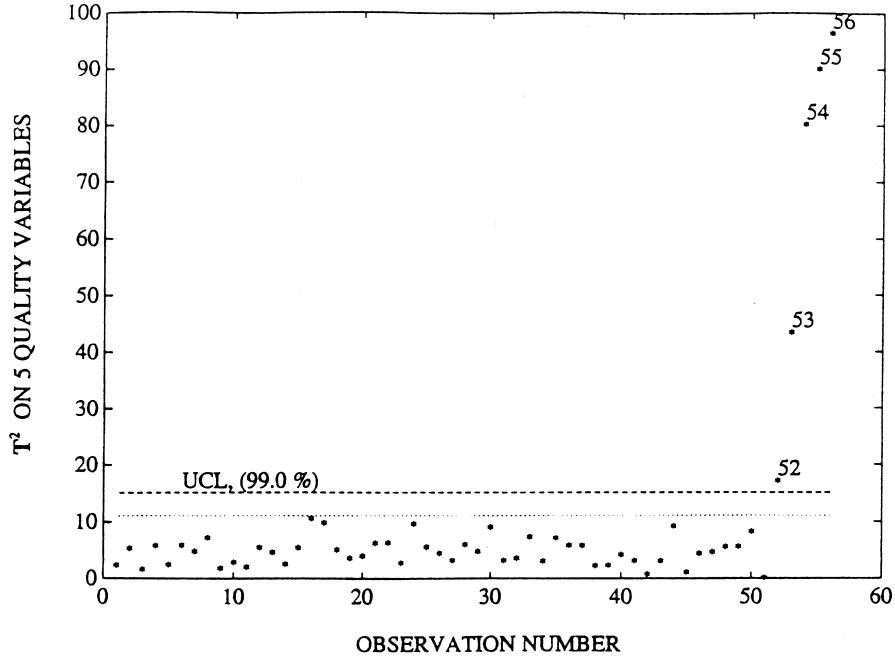


Fig. 5. T^2 chart on five product properties of polyethylene reprinted from Ref. [39].

References

[1] P.C. Mahalanobis, On the generalised distance in statistics, Proceedings of the National Institute of Science of India 12 (1936) 49–55.
[2] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chichester, 1991.
[3] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.
[4] J.S. Shenk, M.O. Westerhaus, Population definition, sample selection, and calibration procedures for near-infrared reflectance spectroscopy, Crop Science 31 (1991) 469–474.
[5] D. Jouan-Rimbaud, D.L. Massart, C.A. Saby, C. Puel, Characterisation of the representativity of selected sets of samples in multivariate calibration and pattern recognition, Anal. Chim. Acta 350 (1997) 149–161.
[6] D. Jouan-Rimbaud, D.L. Massart, C.A. Saby, C. Puel, Determination of the representativity between two multidimensional data sets by a comparison of their structure, Chemom. Intell. Lab. Syst. 40 (1998) 129–144.
[7] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (1933) 417-441, 498–520.
[8] H. Hotelling, Multivariate quality control, in: C. Eisenhart, M.W. Hastay, W.A. Wallis (Eds.), Techniques of Statistical Analysis, McGraw-Hill, New York, 1947, pp. 111–184.
[9] B.M.G. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier, Amsterdam, 1998.
[10] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heurding, F. Erini, Comparison of regularised discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data, Anal. Chim. Acta 329 (1996) 257–265.
[11] M.P. Derde, D.L. Massart, UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution, Anal. Chim. Acta 184 (1986) 33–51.
[12] D. Coomans, I. Broeckaert, M.P. Derde, A. Tassin, D.L. Massart, S. Wold, Use of a microcomputer for the definition of multivariate confidence regions in medical diagnosis bases on clinical laboratory profiles, Comp. Biomed. Res. 17 (1984) 1–14.
[13] S. Wold, M. Sjostrom, SIMCA: a method for analysing chemical data in terms of similarity and analogy, ACS Symposium Series 52 (1977) .
[14] G. Box, A. Luceno, Statistical Control by Monitoring and Feedback Adjustment, Wiley, New York, 1997.
[15] P. Dagnelie, Statistique Théoretique et Appliquée Tome 1, Les Presses agronomiques de Gembloux, 1992.
[16] A.C. Rencher, Multivariate Statistical Interference and Applications, Wiley, New York, 1998.
[17] J.E. Jackson, A User's Guide To Principal Components, Wiley, New York, 1991.
[18] F. Malinowski, D. Howery, Factor Analysis in Chemistry, Wiley, New York, 1980.
[19] F.E. Grubbs, G. Beck, Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics 14 (1972) 847–854.
[20] P.C. Kelly, Outlier detection in collaborative studies, J. Assoc. Off. Anal. Chem. 73 (1990) 58–64.
[21] V. Centner, D.L. Massart, O.E. de Noord, Detection of inhomogeneities in sets of NIR spectra, Anal. Chim. Acta 330 (1996) 1–17.
[22] J.C. Miller, J.N. Miller, Statistics for Analytical Chemistry, Ellis Horwood, Chichester, 1988.
[23] P. Dagnelie, Analyse Statistique a Plusieurs Variables, Vander, Paris, 1975.
[24] N.D. Tracy, J.C. Young, R.L. Mason, Multivariate control charts for individual observations, J. Qual. Technol. 24 (1992) 88–95.
[25] J.E. Jackson, Principal components and factor analysis: Part I. Principal components, J. Qual. Technol. 12 (1980) 201–213.
[26] J.E. Jackson, Principal components and factor analysis: Part II. Additional topics related to principal components, J. Qual. Technol. 13 (1981) 46–58.
[27] J.F. MacGregor, T. Kourti, Statistical process control of multivariate processes, Control Eng. Practice 3 (1995) 403–414.
[28] T. Næs, Leverage and influence measures for principal component regression, Chemom. Intell. Lab. Syst. 5 (1989) 155–168.
[29] D.L. Massart, B.M.G. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997.
[30] I.E. Frank, R. Todeschini, The Data Analysis Handbook, Elsevier, Amsterdam, 1994.
[31] P.J. Rousseeuw, Multivariate estimation with high breakdown point, Inst. Math. Stat. Bull. 12 (1983) 234.
[32] P.J. Rousseeuw, Introduction to positive breakdown methods, Handbook of Statistics 15 (1997) 101–121.
[33] J.E. Jackson, Quality control methods for several related variables, Technometrics 1 (1959) 359–377.
[34] D.C. Montgomery, P.J. Klatt, Economic design of T^2 control charts to maintain current control of a process, Management Science 19 (1972) 76–89.
[35] F.B. Alt, Multivariate quality control, in: S. Kotz, N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences, Vol. 6, Wiley, New York, 1985, pp. 110–122.
[36] T.P. Ryan, Statistical Methods for Quality Improvement, Wiley, New York, 1988.
[37] C. Fuchs, Y. Benjamini, Multivariate profile charts for statistical process control, Technometrics 36 (1994) 182–195.
[38] S. Wold, Cross-validatory estimation of the number of components in factor and principal components models, Technometrics 20 (1978) 397–405.
[39] T. Kourti, J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemom. Intell. Lab. Syst. 28 (1995) 3–21.

[40] P. Dagnelie, A. Merckx, Using generalised distances in classification of groups, Biom. J. 33 (1991) 683–695.
[41] P.A. Lachenbruch, Discriminant Analysis, Hafner Press, New York, 1975.
[42] J. Friedman, Regularized discriminant analysis, J. Am. Stat. Assoc. 84 (1989) 165–175.
[43] S. Aeberhard, D. Coomans, O. De Vel, Improvements to the classification performance of RDA, J. Chemom. 7 (1993) 99–115.
[44] B. Mertens, M. Thompson, T. Fearn, Principal component outlier detection and SIMCA: a synthesis, Analyst 119 (1994) 2777–2784.
[45] S. Wold, M. Sjostrom, Letter to the editor — Comments on a recent evaluation of the SIMCA method, J. Chemom. 1 (1987) 243–245.
[46] R. De Maesschalck, A. Candolfi, D.L. Massart, S. Heuerding, Decision criteria for SIMCA applied to near infrared data, Chemom. Intell. Lab. Syst. 47 (1999) 63–75.
[47] S.J. Smith, S.P. Caudill, J.L. Pirkle, D.L. Ashley, Composite multivariate quality control using a system of univariate, bivariate, and multivariate quality control rules, Anal. Chem. 63 (1991) 1419–1425.
[48] S.P. Caudill, S.J. Smith, J.L. Pirkle, D.L. Ashley, Performance characteristics of a composite multivariate quality control system, Anal. Chem. 64 (1992) 1390–1395.
[49] T. Kourti, J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemom. Intell. Lab. Syst. 28 (1995) 3–21.
[50] P. Nomikos, J.F. MacGregor, Multivariate SPC charts for monitoring Bach processes, Technometrics 37 (1995) 41–59.
[51] J.F. MacGregor, T. Kourti, Statistical process control of multivariate processes, Control Eng. Practice 3 (1995) 403–414.
[52] R. De Maesschalck, F.C. Sanchez, D.L. Massart, P. Doherty, P. Hailey, On-line monitoring of powder blending with near-infrared spectroscopy, Appl. Spectrosc. 52 (1998) 725–731.
[53] A. Nijhuis, S. de Jong, B.G.M. Vandeginste, Multivariate statistical process control in chromatography, Chemom. Intell. Lab. Syst. 38 (1997) 51–62.
[54] J. Cho, P.J. Gemperline, Pattern recognition analysis of near-infrared spectra by Robust distance method, J. Chemom. 9 (1995) 169–178.
[55] Coomans, I. Broeckaert, M.P. Derde, A. Tassin, D.L. Massart, S. Wold, Use of a microcomputer for the definition of multivariate confidence regions in medical diagnosis bases on clinical laboratory profiles, Comp. Biomed. Res. 17 (1984) 1–14.
[56] M.P. Derde, D.L. Massart, UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution, Anal. Chim. Acta 184 (1986) 33–51.
[57] H. Mark, Use of Mahalanobis distances to evaluate sample preparation methods for near-infrared reflectance analysis, Anal. Chem. 59 (1987) 790–795.
[58] P.J. Gemperline, N.R. Boyer, Classification of near-infrared spectra using wavelength distances: comparison to the Mahalanobis distance and residual variance methods, Anal. Chem. 67 (1995) 160–166.
[59] H.L. Mark, D. Tunnell, Qualitative near-infrared reflectance analysis using Mahalanobis distances, Anal. Chem. 57 (1985) 1449–1456.
[60] P.K. Aldridge, C.L. Evans, H.W. Ward II, S.T. Colgan, N. Boyer, P.J. Gemperline, Near-IR detection of polymorphism and process-related substances, Anal. Chem. 68 (1996) 997–1002.
[61] M.J. Martin, F. Pablos, A.G. Gonzalez, Application of pattern recognition to the discrimination of roasted coffees, Anal. Chim. Acta 320 (1996) 191–197.
[62] Y.G. Martin, J.L.P. Pavon, B.M. Cordero, C.G. Pinto, Classification of vegetable oils by linear discriminant analysis of Electronic Nose data, Anal. Chim. Acta 384 (1999) 83–94.
[63] D.S. Lee, B.S. Noh, S.Y. Bae, K. Kim, Characterisation of fatty acids composition in vegetable oils by gas chromatography and chemometrics, Anal. Chim. Acta 358 (1998) 163–175.
[64] Z. Ramadan, M. Mulholland, D.B. Hibbert, Classification of detectors for ion chromatography using principal components regression and linear discriminant analysis, Chemom. Intell. Lab. Syst. 40 (1998) 165–174.