Limit of detection for second-order calibration methods

MARÍA JOSÉ RODRÍGUEZ CUESTA

Doctoral Thesis
ROVIRA I VIRGILI UNIVERSITY
Tarragona 2006

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Limit of detection for second-order calibration methods

Doctoral thesis

ROVIRA I VIRGILI UNIVERSITY

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Rovira i Virgili University

Department of Analytical Chemistry and Organic Chemistry

Limit of detection for second-order calibration methods

Dissertation presented by

María José Rodríguez Cuesta

to receive the degree

Doctor of the Rovira i Virgili University

European PhD

Tarragona, 2006

Supervisor

Dr. Ricard Boqué i Martí

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

ROVIRA I VIRGILI UNIVERSITY
Department of Analytical Chemistry
and Organic Chemistry

Dr RICARD BOQUÉ I MARTÍ, Associate professor of the Department of Analytical Chemistry and Organic Chemistry at the Rovira i Virgili University

CERTIFIES:

That the Doctoral Thesis entitled: ”LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS”, presented by MARÍA JOSÉ RODRÍGUEZ CUESTA to receive the degree of Doctor of the Rovira i Virgili University, European PhD, has been carried out under my supervision in the Department of Analytical Chemistry and Organic Chemistry at the Rovira i Virgili University, and that all the results presented in this thesis were obtained in experiments conducted by the above-mentioned student.

Tarragona, April 2006

Dr Ricard Boqué i Martí

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

While writing this thesis I was supervised by Dr. Ricard Boqué, and the work was undertaken in the Chemometrics and Qualimetrics research group, headed by Prof. F. Xavier Rius, of the Analytical Chemistry and Organic Chemistry Department at the Rovira i Virgili University in Tarragona, Spain.

The work included a four-month research stay in the winter of 2004 in the Spectroscopy and Chemometrics group of the Quality and Technology Department of Food Science at The Royal Veterinary and Agricultural University (KVL) in Copenhagen, Denmark, where I was supervised by Prof. Rasmus Bro.

It also included a scientific collaboration with Prof. J. L. Martínez Vidal, Prof. A. Garrido Frenich, Prof. D. Picón Zamora and Prof. M. Martínez Galera of the Department of Hydrogeology and Analytical Chemistry of the University of Almeria

Finally, I appreciate the financial support given by the 'Agència de Gestió d'Ajusts Universitaris i Recerca' (2000FI-00632) of the Catalan Government.

For those who have helped me to grow as a person and as a researcher, and to all those have helped me to follow my path, each one in their different way. I hope that all of you, friends, family and colleagues, feel identified in these lines.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table of contents

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 1

DESCRIPTION OF THE THESIS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

1.1. Motivation

My earliest scientific interest in the field of chemistry was probably induced by my local environment: Tarragona, a town in Catalonia with a long and distinguished history and significant economic importance in Spain, is an area in which industry and services have grown without losing sight of the need to support the traditional role of agriculture and to consolidate the role of tourism.

Tarragona has an important industrial and service activity, a large number of energy resources, one of the most important sea ports on the Mediterranean and excellent road and rail communications with the rest of Spain and Europe. Numerous multinational companies have set up their production centres in Tarragona.

The chemical industry is an important part of Tarragona's productive base. Other important areas of production are foodstuffs and electronic components. These industries are located near the large urban centres of the province and spread throughout the territory in a highly integrated network that employs more than 70,000 workers.

The historical and cultural traditions of Tarragona have also consolidated a young Rovira i Virgili University (URV), which has already become an important research centre and a significant factor in the training of the area's human resources.

In this context, I enrolled for the Degree in Chemistry at the URV. Every course I attended, every piece of knowledge I acquired and every practical lesson I took confirmed my love for Chemistry. In the last year of my undergraduate course, the Department of Physical and Inorganic Chemistry offered me the opportunity (remunerated) to work on a research project for young chemists, and to collaborate on educational tasks.

I obtained my degree in Chemistry in 1998. I was rewarded by the University with the Bachelor's Extraordinary Award, which recognises the best academic performance of the year. Moreover, in November 1999 the Port Authority of Tarragona awarded me the title of "Best Chemistry Student".

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

After finishing my degree, my relationship with the research group into Chemometrics and Qualimetrics began as I joined a group of top quality, highly productive scientists and an excellent human team. My early participation with my new colleagues in a one-week Chemometric School in Granada (southern Spain) reinforced my enthusiasm for research, especially into chemometrics and multi-way analysis, and was the beginning of this doctoral thesis.

On a personal note, my family have always encouraged me to study and I grew up with both the desire to reach the highest level of academic achievement and the conviction that studying all my life would bring me great personal and professional success.

1.2. Objectives

Introduction

Chemistry is the science that deals with the properties of organic and inorganic substances and their interactions with other organic and inorganic substances. Historically, the science of chemistry is a recent development that has its roots in alchemy (from the Arabic word *kimia*, alchemy, in which *al* is Arabic for *the*) which has been practiced throughout the world for thousands of years.

Chemistry is often called the central science because it connects other sciences. Its field of study is broad and often overlaps with that of physics, biology or geology. Chemistry encompasses many specialized sub-disciplines. Analytical chemistry is the analysis of material samples to gain an understanding of their chemical composition and structure. Analytical chemistry can be split into two main types, qualitative and quantitative.

Qualitative analysis seeks to establish the presence of a given element, inorganic compound, functional group or organic compound in a sample. Quantitative analysis seeks to establish the amount of a given element or compound in a sample.

Most modern analytical chemistry is quantitative. Popular sensitivity to health issues is aroused by the *mountains* of government regulations that use science

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

to, for instance, provide public health information to prevent disease caused by harmful exposure to toxic substances. The concept of the minimum amount of an analyte or compound that can be detected or analysed appears in many of these regulations (for example, to discard the presence of traces of toxic substances in foodstuffs) generally as a part of method validation aimed at reliably evaluating the validity of the measurements.

The lowest quantity of a substance that can be distinguished from the absence of that substance (a blank value) is called the detection limit or limit of detection (LOD). Traditionally, in the context of simple measurements where the instrumental signal only depends on the amount of analyte, a multiple of the blank value is taken to calculate the LOD (traditionally, the blank value plus three times the standard deviation of the measurement). However, the increasing complexity of the data that analytical instruments can provide for incoming samples leads to situations in which the LOD cannot be calculated as reliably as before.

General objectives

The first objective of my doctorate was to study in depth the properties of the types of data, instruments and models that are used for calibration in analytical methods. The central axis of the study was the validation and estimation of figures of merit especially focusing on the detection parameters.

The aim of this thesis is to develop theoretical and practical strategies for calculating the limit of detection for complex analytical situations. Specifically, I focus on second-order calibration methods, i.e. when a matrix of data is available for each sample. To do so, I deal with different types of data. First I select a proper calibration method and then propose a LOD calculation that is practical and consistent with the international guidelines and recommendations on detection concepts.

Descriptions of the analytical situations treated in this thesis are given in the introductions to the corresponding chapters. In all cases under study, and because of the complexity of the data, calculating the LOD was a challenge. This was mainly because many analytes with very similar responses overlapped or because there were unwanted or unknown compounds that interfered with the analyte's response.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Specific objectives

1. To make a critical revision of the existing procedures for estimating the LOD for second-order calibration methods.
2. To develop a LOD estimator for some widespread methods applied to second-order data: PARAFAC, ITTFA, MCR-ALS and N-PLS.
3. To evaluate the relation between the LOD and other figures of merit.
4. To establish guidelines for improving the detection capability of second-order calibration methods based on the effect of other figures of merit (e.g. sensitivity or selectivity) and on the experimental design.

1.3. Structure

This thesis is divided into two stages. In the first stage I will describe some of the chemometric concepts and terminology in multi-way analysis in order to later explain the analytical and theoretical problems and justify my proposals for solving them. The bibliographic references in the text are cited by (first) author and year of publication. The full list of references is in the corresponding section of the *Appendix*.

In the second stage I discuss some practical applications and present my original work (chapters 3 to 6). The thesis is based on papers published in international journals. These have been edited to provide uniform format and mathematical notation throughout the thesis. However, for practical reasons, the original organization of the references is maintained.

Formally, the thesis is divided into eight chapters:

- Chapter 1. *Description of the thesis* contains the justification of this thesis and provides a brief overview of its context. The objectives of the thesis are explained and the structure of its contents is described.

- Chapter 2. *Introduction to chemometrics* contains the fundamental concepts of multi-way analysis, from the types of data and instruments the analyst may

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

have to deal with to the calibration models to be used. Section 2.2.2 provides an overview of the calibration methods and, in particular, describes the second-order methods used in this thesis: PARAFAC, ITTFA, MCR-ALS and N-PLS. Section 2.2.3 introduces the figures of merit for method validation, particularly the analytical detection parameters and how to calculate them for data of different orders.

- Chapter 3. *Estimation of the LOD for PARAFAC* contains the paper entitled *Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis. M. J. Rodríguez-Cuesta, R. Boqué, F. X. Rius, D. Picón Zamora, M. Martínez Galera and A. Garrido Frenich* in *Anal. Chim. Acta 491 (2003) 47-57*, where the LOD is calculated from the second-order excitation-emission matrices (EEM) of a set of calibration samples which are resolved using PARAFAC.

- Chapter 4. *Estimation of the LOD for ITTFA* contains the paper entitled *Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data. M. J. Rodríguez-Cuesta, R. Boqué and F. X. Rius* in *Anal. Chim. Acta 476 (2003) 111-122*, where the quality of the LOD estimator is defined as a function of the method performance characteristics and advice is given on how to improve the LOD by modifying experimental variables.

- Chapter 5. *Estimation of the LOD for MCR-ALS* contains the paper entitled *Development and validation of a method for determining pesticides from complex overlapped HPLC signals and multivariate curve resolution. M. J. Rodríguez-Cuesta, R. Boqué, F. Xavier Rius, J. L. Martínez Vidal and A. Garrido Frenich* in *Chemom. Intell. Lab. Syst. 77 (2005) 251-260*, where the LOD is calculated from the second-order spectro-chromatographic response of a set of calibration samples which are resolved using the ALS multivariate curve resolution method.

- Chapter 6. *Estimation of the LOD for N-PLS* contains the paper entitled *Standard error of prediction at low content levels and limit of detection estimation for multivariate and multi-linear regression. M. J. Rodríguez-Cuesta, R. Bro, R. Boqué and N. M. Faber* in *Anal. Chim. Acta (In press),* which

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

describes how to calculate the LOD for a three-way prediction method and shows that the distribution of the responses at zero concentration is statistically different for samples containing different analyte/interferent ratios, thus leading to sample-specific LODs.

- Chapter 7. *Conclusions.* Contains the general conclusions of the thesis and outlines suggestions for further research.

- The *Appendix* contains the list of tables, the list of figures and the references cited in the thesis (excluding the papers) as well as a glossary with the abbreviations used and the definitions of some additional terms. The list of papers and meeting contributions collects the publications, the oral communications and poster presentations at international meetings during the development of the thesis.

1.4. Notation and list of symbols

General scheme for notation

The notation of the quantities is important for differentiating between types of data. In one dimension, a scalar is a single value; in two dimensions, a vector is a "list" of values and a matrix is simply a "table" of values in three dimensions, three-way data are a "cube" of data, i.e. a set of tables of data.

Throughout the thesis, the following conventions are used for these types of data:

- Scalars are denoted by lowercase italic letters, e.g. ***a*.**
- Vectors are denoted by lowercase bold letters, for instance **a.** All vectors are column vectors unless explicitly written as transposed. The transposition of a vector –or a matrix– is symbolized by a superscripted 'T', e.g. $\mathbf{a}^T$.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

- Matrices are denoted by bold capital letters, for instance **A**. Matrix elements are denoted by doubly indexed lowercase italic letters, for instance **A***ij* is the element of the matrix **A** located in row *i* and column *j*. For a given matrix **A**, matrices $\mathbf{A}^{-1}$ and $\mathbf{A}^{+}$ stand for its inverse and pseudoinverse, respectively. In full rank matrices, $\mathbf{A}^{+}=(\mathbf{A}^{T}\mathbf{A})^{-1}\mathbf{A}^{T}$.
- Three-way structured data are denoted as matrices with an underlined letter, for instance **A**.

List of frequently used symbols

r, **r**, **R**, **R**	response (scalar, vector, matrix or cube of responses)
x, **X**, **X**	a set of responses grouped in a vector, matrix or cube of data
c, **c, C, C**	content or concentration (scalar, vector, matrix or cube of responses)
y, Y	a set of concentrations grouped in a vector or matrix of data
e, **e**, **E**, **E**	error
b, **b**, **B**	regression parameters of the model
i=1,…I, j=1,…J, k=1,…K	modes
L, M, P, Q, R	modes
f=1,…,*F*	number of components / number of analytes
A, **B**, **C, T, U, W, Q**	loading (scores) or weight matrices
α	probability of false positive or first type error
β	probability of false negative or second type error
ν	degrees of freedom
σ	standard deviation

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

z	normal distribution
t	Student-t distribution

This is the standard notation used in this thesis. In some cases, however, the same symbol is used with different meanings. This is in order to follow the accepted nomenclature found in the bibliography. When this occurs, it is explained in the text in order to avoid confusion.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 2

INTRODUCTION TO CHEMOMETRICS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

2.1. Types of data and instruments used in analytical chemistry

2.1.1. Historical background

Basic analytical measurements date from ancient times. Standard weights were used by the Babylonians in 2600 BC and, as early as that time, they were considered so important that users were supervised by their priests. However, it was not until the 17th century that the term “analyst” was introduced by Robert Boyle in his book The Sceptical Chymist (1661). Boyle can therefore be considered the father of this branch of chemistry.

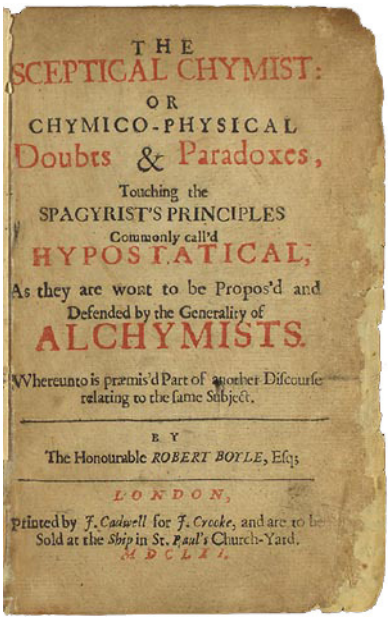

THE
SCEPTICAL CHYMIST:
OR
CHYMICO-PHYSICAL
Doubts & Paradoxes,
Touching the
SPAGYRIST'S PRINCIPLES
Commonly call'd
HYPOSTATICAL,
As they are wont to be Propos'd and
Defended by the Generality of
ALCHYMISTS.
Whereunto is præmis'd Part of another Discourse
relating to the same Subject.
BY
The Honourable ROBERT BOYLE, Esqs
LONDON,
Printed by J. Cadwell for J. Crooke, and are to be
Sold at the Ship in St. Paul's Church-Yard.
MDCLXI.

Figure 1. Title page of *The Sceptical Chymist* by Robert Boyle (1627-1691).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Most classical or wet-chemical methods were based on separating the analytes by precipitation, extraction or distillation procedures. After this separation step, qualitative analyses were carried out by reaction with a reagent that yielded products that were recognizable by their colours, boiling or melting points, solubilities in different solvents, optical activities or refractive indexes. For quantitative analysis, gravimetric or titrimetric measurements were used to determine the amount of analyte. Inventions and the development of analytical instruments followed one after the other between the 18th and early 20th century, and instrumental methods gradually supplanted the classical ones (Rubinson and Rubinson, 2000; Skoog *et al*., 1998). The main advances have been made since the 1950s when chemical structures and composition information were first converted into electrical and optical phenomena such as spectra, voltammograms and chromatograms. Modern instrumental methods grew parallel to the development of the electronics and computer industries. The introduction of computers into analytical chemistry laboratories has made data storage easier and enabled data to be reconstructed, manipulated or transformed by, for instance, subtracting a background, smoothing the noise in the signals or calculating a ratio of responses.

Nowadays, many analytical laboratories have computers and instruments that can generate huge amounts of (complex) data. These instruments can be classified according to the type of data they produce as response. Tensorial theory[1] provides a unified language that is useful for describing the chemical measurements, analytical instruments and calibration methods.

2.1.2. Tensor terminology

In the classical approach, tensors are mathematical objects that can be represented by arrays of components such as an *n*-dimensional generalization of scalars, 1-dimensional vectors and 2-dimensional matrices. The *order* of the tensor is the number of the modes or spaces spanned by the tensor (in the literature we find 'order', 'modes', 'spaces' or 'ways' used as synonyms). The

[1] The notation was developed around 1890 by Gregorio Ricci-Curbastro (1853-1925) and spread by his pupil Tullio Levi-Civita (1873-1941) with the publication of "The Absolute Differential Calculus" in 1900. Tensor Calculus was widely accepted with the introduction of Einstein's theory of general relativity, which was completely formulated in the language of tensors. Tensor theory is also used in areas such as elasticity, fluid mechanics, etc.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

number of elements in each order is called *dimensionality*. For each order of the tensor, the *rank* is defined as the number of varying independent factors and cannot be greater than the dimensionality of that order.

In tensorial language, a scalar is a tensor of order 0 with rank 0, a vector is a tensor of order 1 with rank 1, and a matrix is a tensor of order 2 with rank equivalent to the rank of its matrix[2].

Often we have to analyze several samples. When the measured responses for these samples are scalars (zeroth-order tensors) they can be arranged in a vector, thus giving rise to a one-way structure. If the responses are first-order tensors, e.g. spectra, they give rise to a two-way structure. Working with a set of second-order responses may involve handling three-dimensional arrays of data, i.e. three-way structures. To generalize, a collection of *n*th-order data from each of many samples creates an (*n*+1)-way structure. A graphical representation can help to understand these concepts (see Figure 2).

By analogy with the tensorial nomenclature for the data, instruments that generate a single datum per sample are called zeroth-order instruments. Examples of these instruments are the thermometer, which provides a specific temperature, and ion-selective electrodes such as the pH meter, which measures the pH of a solution. First-order instruments include all types of spectrometers and chromatographs. These produce multiple measurements per sample that are ordered in a data vector (or first-order tensor), such as the elution profile obtained with a liquid chromatograph or the spectrum obtained by a UV-Vis spectrophotometer.

Instruments that generate two-dimensional arrays of data are second-order instruments. A typical example is a spectrofluorometer, which provides a set of emission spectra obtained at different excitation wavelengths. Many of the so-called 'hyphenated methods' also provide second-order data generated by coupling two first-order instruments, e.g. a gas chromatograph coupled to an infrared spectrometer (GC-IR) or to a mass spectrometer (GC-MS), a liquid chromatograph coupled to a UV-Vis detector (HPLC-DAD), and a tandem mass

[2] Note that second-order tensors and matrices are not identical. A matrix is any arrangement of elements written as an array (Budianski 1974). See "Tensor" in the *Glossary* of terms.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

spectrometer (MS-MS) or a thermogravimetric analyzer coupled to an infrared spectrometer (TGA-IR). Table 1 summarizes these examples.

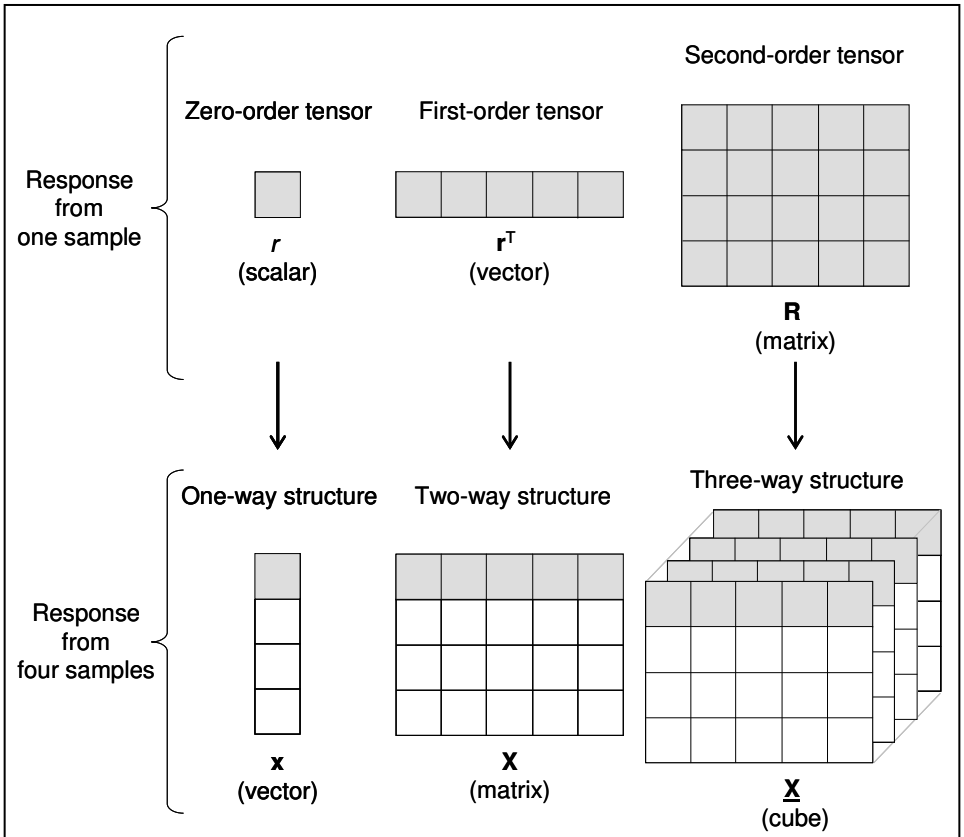


Figure 2. Graphical representation of the response from one sample (a tensor of zero, first or second-order), and the resulting array of responses from four samples (a structure of one, two or three ways).

Higher-order tensors can also be directly generated, and the increasing complexity of the instrumentation is generally compensated by the gain in analyte selectivity. Going from methods that use zeroth-order data to methods that use n*th*-order data requires a deeper knowledge of the mathematics and statistics involved in chemometrics methods than in the more intuitive and widely

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

disseminated univariate models. However, the use of higher-order data is justified for mainly two reasons: first, undesirable or unexpected compounds (interferents) that contribute to the measured signal can be detected and second, the analyte of interest can be quantified in the presence of these interferents. Section 2.2 describes these features in detail.

Table 1. Examples of data and instruments of different orders.

Data order	Data type	Instrument
Zeroth-order tensor	▪ Temperature ▪ pH value	▪ Thermometer ▪ pH-meter
First-order tensor	▪ Near infrared (NIR) spectrum ▪ Chromatogram at a single wavelength	▪ NIR spectrophotometer ▪ HPLC-UV/Vis
Second-order tensor	▪ Spectrochromatogram ▪ Excitation-emission matrix (EEM)	▪ HPLC-DAD ▪ EEM spectrofluorometer

2.1.3. The concepts of bilinearity and trilinearity

In the absence of noise, a second-order datum (i.e. the matrix of instrumental responses) corresponding to a *pure compound*, can ideally be written as the outer product of two vectors (or first-order tensors), also called a *dyad*, a *bilinear component* or, in the Principal Component Analysis (PCA) model (Wold *et al*., 1997), a *principal component*[3]. In HPLC-DAD data, for example, the column vector describes the concentration profile and the row vector describes the UV spectrum. The outer product concisely shoes that, ideally, the UV spectrum does not depend on elution time and that the elution profile is identical for all wavelengths.

In mixtures, the *bilinearity* of a data matrix **R** means that the contributions of the compounds in the two orders of measurement are additive and, therefore, that the response matrix can be decomposed as $\mathbf{R} = \mathbf{C}\mathbf{S}^T$. Following the HPLC-DAD example, the columns of **C** and **S** designate the intrinsic factors responsible for

[3] Principal components are also called, in a more general sense, *latent variables* in contrast to the original variables, which are called *manifest variables*.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

the observations, i.e. the pure chromatograms and spectra (up to a scalar that accounts for intensity) (see Figure 3).

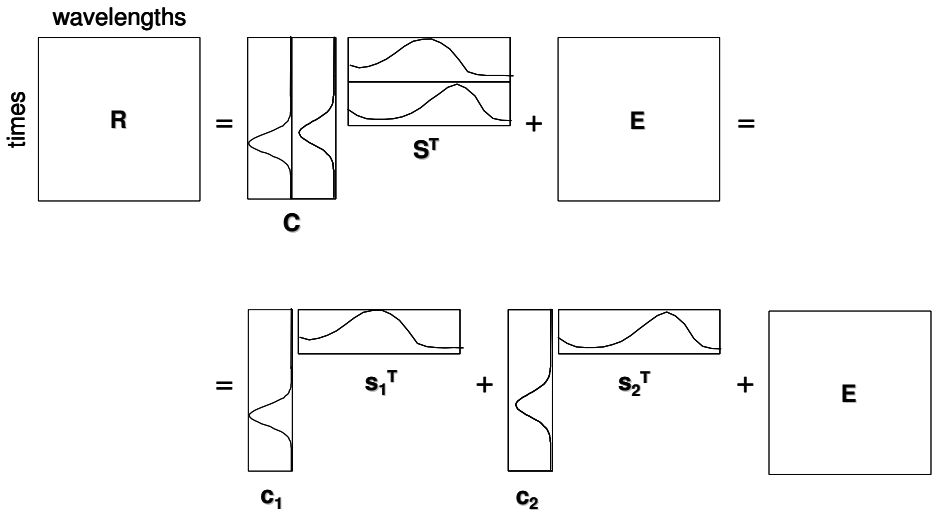


Figure 3. Schematic decomposition of a second-order matrix **R** into the individual pure profiles of each compound in the mixture.

In bilinear experimental data, the additional matrix of residuals **E** accounts for the experimental error and their elements are normally distributed with mean zero and standard deviation σ. Otherwise, i.e. if the residuals show a systematic variation, more components can be extracted due, for example, to the presence of an unexpected compound or to an instrumental baseline in one or more of the modes. In this latter case, however, data would not follow the bilinear model.

Three-way structured data can be obtained from a set of second-order matrices (see Figures 2 and 4), when, for instance, several standards with different concentration of analyte/s (and/or interferents) are measured. Among other important enhancements (see Section 2.2), by using this matrix augmentation one can limit the mathematical results to only those that have a chemical meaning.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

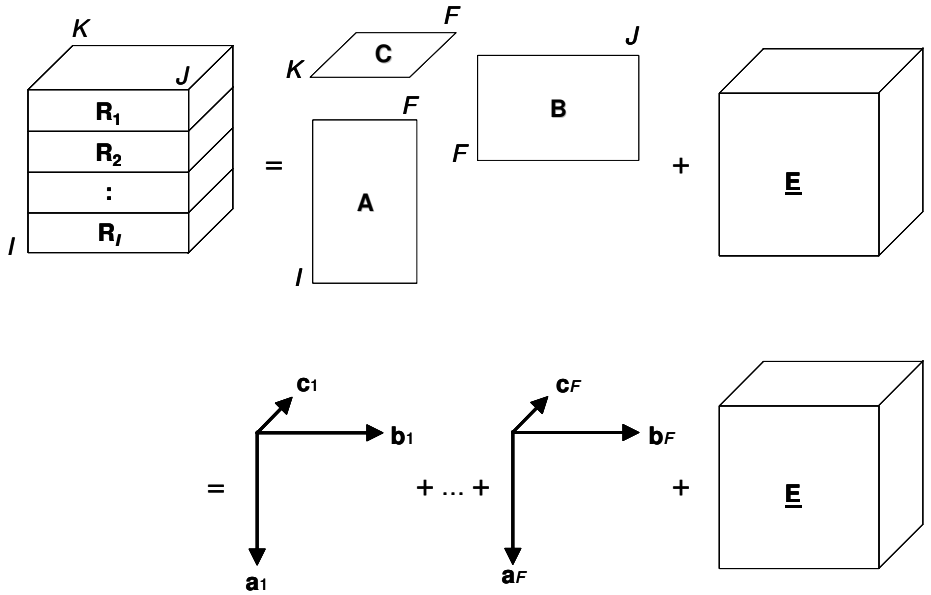


Figure 4. Decomposition of a set of second-order tensors in a three-dimensional structure into the product of **A**, **B** and **C** matrices that contain the first-order profiles in modes *I*, *J* and *K*, respectively. *F* is the number of underlying trilinear components and **E** is the residual matrix of non-modelled information.

When these second-order matrices can be decomposed into first-order tensors, and these first-order tensors are reproducible along the samples, the concept of bilinearity extends to the third mode (*trilinearity*) and the original data can be decomposed into trilinear components, or *triads*. Thus, trilinear models further assume that the underlying phenomena are parallel proportional profiles[4], i.e. the profiles are reproducible, have identical shape and differ only in intensity from experiment to experiment.

The trilinearity of a data cube can in practice be assessed as follows. The three-

[4] This is the PARAFAC model, which is described in Section 2.2.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

way structured data sized ($I \times J \times K$) can be matricized[5] in the three different directions, giving a row-wise augmented matrix $\mathbf{R}_r$ ($I \times JK$), a column-wise augmented matrix $\mathbf{R}_c$ ($J \times IK$), and a tube-wise augmented matrix $\mathbf{R}_t$ ($K \times IJ$). When the rank of these three matrices is the same, i.e. they all have the same number of components related to the chemical variations, the three-way structure follows a trilinear model.

In real analyses we do not always handle ideal scenarios. Deviations from linearity, if not considered, can lead to misinterpretation of the results from the mathematical decomposition. Fortunately, there are suitable methods for approximating many of these deviations.

2.1.4. Deviations from linearity

Real experimental data do not always behave ideally, and high-order data are not an exception. Linearity is assumed in many (principal components-based) models in order to decompose the data into a sum of dyads or triads. Actually, the Direct Trilinear Decomposition (Sánchez and Kowalski, 1990), DTD, schematized in Figure 4, is only feasible when data are intrinsically trilinear. For the example of the three-way analysis of HPLC-DAD data, this means that the shapes of the spectrum and the elution profile of each chemical compound remain invariant in all samples. In practice, there are some applications in which this is not achieved. Some situations can be regarded as a violation of the strict trilinearity, such as changes in spectra shape due to matrix effects or changes in the shape or shifts of the retention time in the elution profiles.

Another common example in which the linearity among the three modes breaks down appears in EEM fluorescence, when Rayleigh and Raman scattering are recorded. These diagonal patterns across the spectra are not multi-linear in its nature and therefore inefficiently modelled by trilinear calibration methods. To prevent the model from trying to fit these scatters, several useful procedures have been reported (Bro, 1997; JiJi and Booksh, 2000; Wentzell *et al.*, 2001; Bro *et al.*, 2002). Raman interference due to the solvent can often be almost

[5] See "Matricizing" in the *Glossary* of terms. Among the different equivalent terms that define this reorganization, the commonly used term "unfolding" should be avoided since it has a specific meaning outside the chemometric community.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

completely removed by subtracting the solvent spectra from the sample spectra. Rayleigh scattering, which occurs when the excitation wavelength is equal to the emission wavelength, can be disregarded by weighing the data or by restricting the analysis to the region in which the scatter does not appear.

Trilinearity also cannot be assumed when the variables of the different slabs are not comparable, for example in size. Let us imagine the sensory analysis of 10 attributes, e.g. colour, taste, etc., from 20 different wines samples carried out by 5 experimented or trained panellists (20 samples in the rows, 10 attributes in the columns, 5 panellists in the slabs). If any of the judges does not assess all attributes, its corresponding slab will have a lower dimension. In such a situation, a trilinear model cannot fit the data.

Generally, in order to adequately select the family of chemometric methods to use, any sampling problem, physical property or chemical interaction that involves a change in the underlying phenomena must be taken into account in case it affects the linearity of the data.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

2.2. Calibration methods

2.2.1. From zeroth- to second-order calibration

All instruments should be calibrated to check that the instrument provides the proper response according to the employed standard. As well as instrumental calibration, there is analytical or methodological calibration. All analytical methods that use instruments that do not directly provide the property the analyst is interested in should incorporate the analytical calibration, i.e. the establishment of the relationship between the instrumental response (the signal, e.g. an absorbance value) and the property of interest (usually a concentration). This mathematical relation is called the calibration model and can be classified following the tensorial terminology depending on the type of data collected from the samples.

The classical calibration curve built with a single-value response, e.g. an absorbance, from a set of calibration standards of known concentration, is an example of a zeroth-order calibration model. When a vector of absorbances, i.e. a spectrum, is collected from every standard, first-order calibration can be applied to extract information from the system. Similarly, second-order calibration concerns second-order data and so on.

In many cases, zeroth-order calibration models are based on theoretical models such as Lambert-Beer's law for spectral measurements or Nernst's equation for potentiometric ones. The zeroth-order response, r, is defined as a function (which is not necessarily linear) of the analyte concentration c (r=f(c), i.e. the "classical model"). In the simplest linear case, the univariate model is

$$r = b_0 + b_1 c \qquad \text{Eq. 1}$$

where b_0 and b_1 are the parameters of the model. The model can be defined from a single standard, as long as the straight line is forced to pass through the (0, 0) point. It is apparent that two points, (0, 0) and (c, r), define the univariate calibration line. However, it is usual to have several samples in order to average the influence of the random instrumental error and take into account the constant

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

offset in the response function. This function can be estimated by *ordinary least squares regression*, OLS.

The signal from any other active compound in the samples is assumed to give a constant contribution to the response and is therefore removed mathematically as an offset. If the intensity contribution of the interference is not constant, it is not possible to distinguish between the signal due to the analyte and the signal due to the interference, so the estimated analyte concentration is biased (see Figure 5).

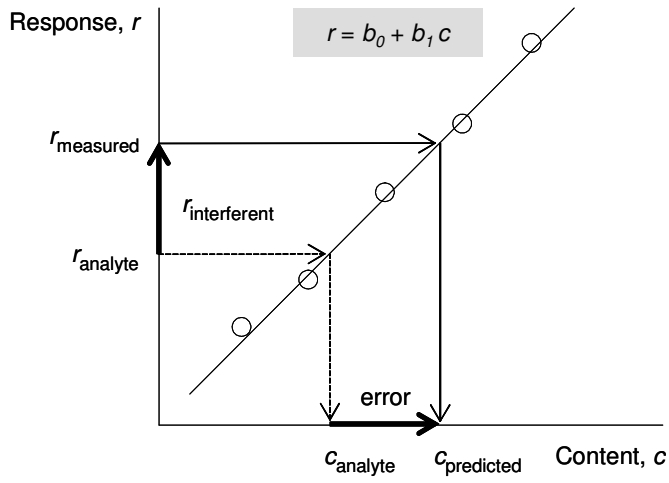


Figure 5. Zeroth-order linear calibration: unknown interferences contribute to the zeroth-order response and the predicted analyte concentration is biased.

In conclusion, a requirement of zeroth-order calibration is that the collected signal is due only to the analyte of interest, either using pure standards and samples or using a fully selective instrument.

When complete physical resolution or the required selectivity is not fulfilled, interferents can be dealt with by modelling them, i.e. by including their contribution in the mathematical model. For this, the samples used to establish the calibration model should contain the interfering compounds and be measured

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

with a multichannel instrument (i.e. a first-order instrument). First-order calibration methods are based on a *direct* or on an *inverse* multivariate model. The direct multivariate model (Equation 2) is a generalisation of the classical univariate model (Equation 1) and can be estimated by *classical least squares regression*, CLS. The first-order response measured at j=1, …, J variables (sensors) is expressed as a function of the concentration of every varying compound in the samples (f=1, …, F analytes), according to the classical model formulation. This means that the signal at unit concentration, e.g. unity spectrum, of all the active compounds that contribute to the signal must be available (matrix **S** in Equation 2), which unfortunately is not the case in many applications where little is known about the composition of the sample.

In the inverse multivariate model, the concentration of the analyte of interest (or generally speaking, the measurand) in the prediction sample is defined as a function of the response, as shown in Equation 3. The vector of regression coefficients is previously estimated from a set of calibration samples, where only the concentration (and not the unit signal) of the analyte of interest must be known. This is why, unlike zeroth-order calibration methods, the inverse model is often preferred for multivariate calibration.

$$\underset{(J\times 1)}{\mathbf{r}} = \underset{(J\times F)}{\mathbf{S}}\,\underset{(F\times 1)}{\mathbf{c}} \qquad \text{Eq. 2}$$

$$\underset{(1\times 1)}{c} = \underset{(1\times J)}{\mathbf{r}^{\mathrm{T}}}\,\underset{(J\times 1)}{\mathbf{b}} \qquad \text{Eq. 3}$$

The vector of regression coefficients in the inverse model represents the part of the response in **r** that is unique to the analyte of interest and orthogonal to the other active compounds in the sample. It is calculated from a set of calibration samples. When the multivariate model is used to predict the analyte concentration in a new sample (using Equation 3), its response is projected onto the regression vector. Every active compound should have been included in the calibration model. Otherwise, i.e. if the new sample contains non-calibrated interferences, abnormal behaviour can be detected (this is known as the *first-order advantage*, see Figure 6) but an unbiased analysis is not possible.

The most popular first-order methods among the chemometric community are *inverse least squares* (ILS) *regression* (also called *multiple linear regression*,

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

MLR) and, of the methods based on the reduction of variables, *principal component regression* (PCR) and *partial least squares regression* (PLSR).

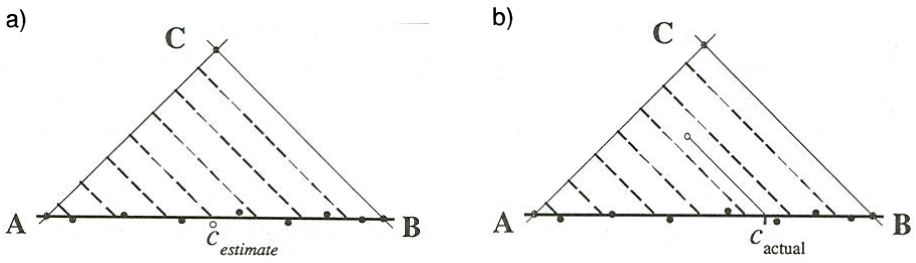


Figure 6. Figure 4 from Sanchez and Kowalski [Ref. 1988]: Multivariate linear calibration. Projection from a multidimensional space to a two-dimensional plane for a three-constituent system (A, B and C). The black dots represent the calibration samples and the white dot is the response for an unknown sample. The calibration set only contains mixtures of analytes A and B, and only the concentration of A is known. (a) Modelled interference: if the unknown sample also contains A and B, then it is possible to estimate the concentration of A. (b) New interference: if the unknown samples also contain C, it is not possible to estimate the concentration of A, but the distance to the A-B model reveals that a problem exists and alerts the analyst to the unsuitability of using that model to predict this sample.

Only with second-order data, is prediction in the presence of new interferences possible. This is known as the *second-order advantage* and can be achieved even with just one calibration sample.

In second-order calibration the data are arranged as three-way arrays. The elements in the three-way array result from the contribution of every analyte, *f*, in the three modes. This array of data is decomposed into a set of basis vectors so that the three-dimensional **R** is defined in terms of three two-way loading matrices (one for each mode) **A**, **B** and **C**, as shown in Figure 4. This decomposition is referred to as the *trilinear* or the *PARAFAC model* (Bro, 1997).

The additional complexity of three-dimensional organization is sometimes reduced by using multivariate models to deal with a collection of second-order

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

data. This is achieved by matricizing the matrix of responses from each sample into a long vector. However, this vectorization not only loses the original relationship in the rows, but also creates limitations of first-order calibration: a larger number of calibration samples are needed in order to span the variations in future samples, and prediction of the analyte concentration will not be possible in samples with new interferences that are not accounted for in the calibration step (Faber *et al*., 2002).

Table 2 is a schematic summary of the capabilities of the different types of calibration, including the commonest type of calibration. Useful references for dealing with this issue are Booksh and Kowalski 1994, Bro 2003, and Boqué and Ferré 2004.

Table 2. Capabilities of calibration of different orders.

Calibration type		Zeroth-order	First-order	Second-order
Detection of interferences		×	✓	✓
Quantification in presence of:	**calibrated Interferences**	×	✓	✓
	new Interferences	×	×	✓
Commonest type of calibration		Univariate linear calibration	Multivariate calibration	Second-order calibration

Note that the increasing complexity of models and methods involved in multi-way analysis, compared to those involved in zeroth- or first-order analysis, is even more evident in written language. Equations to transcribe such models and methods can be expressed graphically, e.g. Figure 4. These figures are easy to understand but they are limited to a maximum of three dimensions. A stricter notation is acquired by summations of indexed quantities, as in Equation 4. However, they lead to complex equations that are difficult to understand when many indices are involved. Actually, the conventional matrix and vector notation is normally used (Kiers, 2000), as in Equations 2 and 3, even though there is no absolutely agreed notation in multi-way analysis.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

2.2.2. Resolution and calibration methods for second-order data

The simplest three-way situation is to determine the concentration of the compounds in a mixture with unknown composition by directly comparing the matrix of the test sample and the matrix of a standard. This principle was introduced by Ho, Christian and Davidson in 1978 with the Rank Annihilation (RA) analysis (Ho *et al.*, 1978; Ho *et al.*, 1980). Their formulation was based on the assumption of bilinear data, i.e. second-order measurement from a sample is a linear combination of their respective contributions. Both the test sample $\mathbf{R}_u$ and the standard $\mathbf{R}_0$ therefore have one pair of vectors in common (first-order profiles of the analyte of interest; see Figure 3),

$$\mathbf{R}_u = \sum_{f=1}^{F} \mathbf{m}_f c_u \mathbf{s}_f^T \qquad \text{Eq. 4}$$

$$\mathbf{R}_0 = \sum_{f=1}^{F} \mathbf{m}_f c_0 \mathbf{s}_f^T \qquad \text{Eq. 5}$$

where c_u and c_0 are the amounts of analyte in the test sample and standard, respectively, and one can find the ratio r of concentrations of the common analyte f in both samples by considering the expression $\mathbf{R}_u - r_f \mathbf{R}_0$: the rank of the resulting matrix is one lower than that of $\mathbf{R}_u$ for the proper value of r_f giving the *actual* relation c_u/c_0 in the samples.

The original method consisted of trying different values of r_f until the effect of removing (*annihilating*) the contribution of the analyte led to the reduction of the rank in one (up to the measurement noise). This method, also called rank annihilation factor analysis (RAFA) (McCue and Malinowski, 1983), relies on only one calibration standard and there is no way of detecting an error in the standard. As in lower-order calibration, using many calibration samples can control this situation (Appellof and Davidson, 1983).

The quantification of several analytes at a time was also later dealt with (Ho *et al.*, 1981; Lorber, 1984), culminating in the popular generalized rank annihilation method (GRAM) developed by Sanchez and Kowalski in 1986.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Rank annihilation can also be explained as the most simplified decomposition of the *fundamental model* proposed by Tucker [6]. The Tucker3 models (Kroonenberg and Leeuw, 1980) reduce the three-way **X** array ($I \times J \times K$) sized into three loading matrices: **A**, **B** and **C** (see Equation 6 and Figure 7). As well as the remaining residual matrix **E**, a fourth matrix –the core matrix **G**– is also obtained, which gives the relations between the three modes.

$$x_{ijk} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R} a_{ip}b_{jq}c_{kr}g_{pqr} + e_{ijk} \qquad \text{Eq. 7}$$

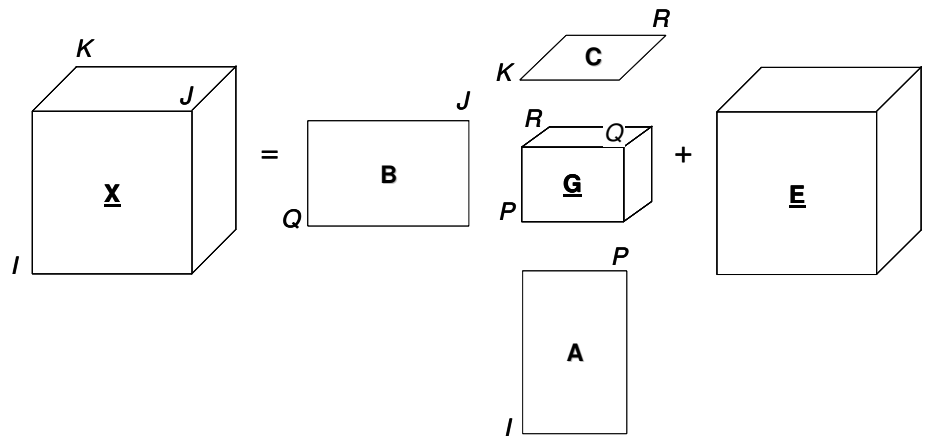


Figure 7. The general Tucker decomposition of **X** into the sum of a structured part (**A**, **B** and **C** loading matrices) and a noise part **E**. The core matrix **G** gives the relations between the three modes.

An important property of the Tucker models is that there is no requirement in the core matrix that the number of factors in **A**, **B** and **C** must be the same, i.e. this is

[6] Though the earlier presentation of the general decomposition given by Tucker in 1963 used a somewhat confusing notation, it was the same decomposition later published by Levin in 1965. One can find an analogue in the analysis of two-way arrays: the singular value decomposition (SVD). In the formula given here, it is not mandatory that C denotes concentrations or chromatographic profiles, the notation simply follows the standard convention (Kiers 2000).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

a general decomposition and *P*, *Q* and *R* do not have to be equal. The loading vectors can interact; this is the explicit meaning of a nondiagonal matrix, although a physical interpretation of these interactions may not always be evident.

When **G** is the identity matrix, i.e. a cube of size $R \times R \times R$ with ones in the super-diagonal and zeros in all the other positions, is equivalent to the PARAFAC model. PARAFAC is used to model trilinear data such as ideally, spectro-chromatographic or the fluorescence data. Specifically, the intensity of fluorophore emissions in dilute solutions or suspensions, despite several environmental factors that influence fluorescence properties, is linearly proportional to their concentration. This principle was used in Chapter 3 to determine a set of pesticides by excitation–emission matrix (EEM) fluorescence in combination with parallel factor analysis (PARAFAC). A more detailed description of the model is given in Section 2.2.2.1.

The methods that have been mentioned so far (RA based models, Tucker and PARAFAC) are *three-way component models*, i.e. models for three-way one-block data analysis. It is sometimes desirable to find a model for predicting a measurable property from an independent block of data. *Three-way regression models* find the connection between a three-way array of data and a vector, a matrix or a three-way array of measured quantities, i.e. they are used for three-way two-block data analysis, as schematized in Table 3.

Table 3. Some component and regression models.

	Two-way data	**Three-way data**
One-block Component models	PCA	TUCKER models PARAFAC
Two-blocks Regression models	PCR PLS	*N*-PLS

The same underlying theory of partial least squares regression (PLS) prevails in the extension to three-way data, the multi-linear PLS or *N*-PLS (Bro, 1996; Bro, 1998). Both an **X**-block of independent data, e.g. spectra, and a **Y**-block of dependent data, e.g. analyte concentrations, are decomposed simultaneously into two PARAFAC-like trilinear models in such a way that the score vectors from

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

these models have maximal covariance. The regression model relating the two decomposition methods maximizes both the variance of the scores (thus ensuring that the model is real and not due to small random variation) and the correlation between scores (thus optimizing the predictive ability of the model). The predictive ability of the *N*-PLS model is further analysed in Chapter 6. Further details of this model are given in section 2.2.2.4.

The analysis of multi-way data responds to the need for calibration (e.g. to calculate concentrations of overlapping constituents for every sample) but also to the need for resolution (e.g. to find pure spectra and chromatograms). As one may deduce from the summary here on calibration methods, some methods combine the ability to calibrate with the ability to resolve, while others specialize in one or the other.

As the name suggests, multivariate curve resolution methods focus on extracting pure profiles typically from two-way structured data. These methods can be classified according to several criteria. For example:

Space:
Some methods are carried out in the space of the original variables, i.e. they work with the raw data, while others transform the original space into a new one defined by the principal components that only account for the relevant information, i.e. compress the original variables to a lower number of latent variables and disregard the noise contribution.

Matrix analysis:
Depending on the volume of data considered at one time for the analysis, the methods can be classified as full-rank, which use the whole matrix, or as evolutionary or local-rank, which analyse successive sub-matrices or windows from the original matrix.

Procedure:
The most common classification differentiates iterative methods from non-iterative (or direct) methods. Iterative methods use an alternating least squares optimization to improve initial estimates in one of the orders and handle the data as a whole. They are fast and simple and require little user interaction or expertise. Non-iterative methods, on the other hand, use the concepts of local

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

rank and rank annihilation, so a deeper knowledge of the properties of the system and more user interaction are required. The main drawback with non-iterative methods is that data sets with non-sequential profiles (e.g. data sets with embedded peaks) cannot be solved.

The curve resolution methods applied in this thesis were the iterative approaches SIMPLISMA (SIMPle-to-use Interactive Self-Modelling Analysis), which extracts the most selective variables from the raw data, and EFA (Evolving Factor Analysis), which performs several local principal components analyses. Because of the information they provide, these methods can be performed as a previous step to calibration. The methods are explained in the Section 2.2.3. Iterative Target Transformation Factor Analysis (ITTFA) and Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) also belong to the group of iterative (and full-rank) methods. These methods were applied in two chapters of this thesis and are described in Sections 2.2.2.2 and 2.2.2.3, respectively. ITTFA was used in Chapter 4 to resolve, *one at a time*, second-order data matrices from simulated and real HPLC-DAD data to evaluate how the resolution is influenced by the chromatographic overlap, the spectral correlation and the relative sensitivity of the compounds. In Chapter 5, MCR-ALS was *simultaneously* applied to *a set of second-order data matrices* to identify and quantify several pesticides in multicomponent groundwater samples analysed by HPLC-DAD. In this application, second-order data were arranged into a two-way structure by setting one matrix below the other (i.e. matrix augmentation).

2.2.2.1. Parallel factor analysis (PARAFAC)

PARAFAC (PARAllel FACtor analysis) is a decomposition method that can be understood as a generalization of bilinear PCA. One of the main advantages of this model is the uniqueness of its solution, which means that the best parameters for the model are given in a least squares sense[7] (up to permutation, sign and scaling indeterminacy). In most circumstances the model is uniquely identified from the structure and therefore no post-processing is necessary since

[7] For comparative purposes, note that in first-order calibration, the decomposition of the calibration matrix is only unique when the basis vectors are constrained to be orthogonal. Generally in such case, the basis vectors do not represent the real first-order profiles, and a regression vector –the part of the analyte signal that is orthogonal to all other compounds– is required to estimate the concentration of the analyte of interest.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

the model is calculated in a least squares optimization[8]. Further, when (1) the data are strictly or *approximately* trilinear, (2) the right number of components is used and (3) the signal-to-noise ratio is appropriate, three-way structured data can be fitted to a trilinear model using a PARAFAC model and lead to reliable interpretations.

As above, a PARAFAC model of a three-way array **X** is given by three loading matrices, **A**, **B** and **C**, following the structural model written in Equation 7 and graphically represented in Figure 8. Detailed explanations of the model, the algorithm and the applications can be found in the literature (Bro, 1997; Bro, 1998).

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \qquad \text{Eq. 7}$$

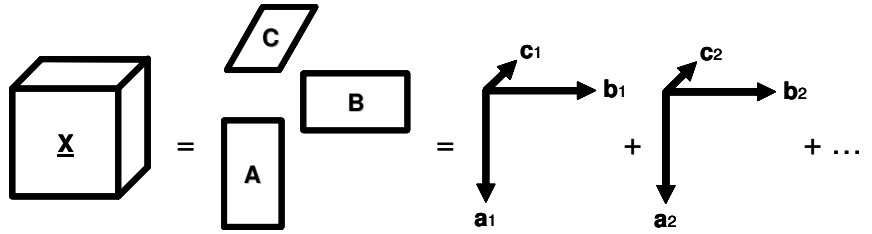


Figure 8. Scheme of a PARAFAC model of the three-way array **X** (residuals and indexes are omitted for brevity).

Across all the frontal slices of the cube of data, the components $\mathbf{a}_f$ and $\mathbf{b}_f$ remain the same and only their weights c_{k1}, …, c_{kf} are different. In practice this means that, for a set of *k* samples, the cube of data is decomposed into a nonorthogonal set of basis vectors (i.e. not independent) that directly represent the pure second-order profiles of every linearly independent compound present in the samples as

[8] Although no additional constraints are needed to identify the model, the algorithm can be modified according to external knowledge, such as non-negativity of the underlying parameters to ensure that the model parameters make sense.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

well as their concentration. In the case of excitation–emission matrix fluorometry, the *k*th slice of the trilinear cube $\underline{\underline{\mathbf{X}}}$ is the matrix of excitation and emission profiles of the fluorescent compounds for the *k*th sample, the elements in the loading matrices **A**, **B** are the relative intensity in the emission and in the excitation domain, respectively, and the elements in the loading matrix **C** are the relative concentrations in the samples of each component.

Using an alternating least squares (ALS) procedure, the trilinear model is found to minimise the sum of the squares of the residuals. In matrix notation, using the Khatri–Rao product[9], the PARAFAC model can be formulated in terms of the matricized array as shown in Equation 8:

$$\underbrace{\mathbf{X}_k}_{(IJ\times K)} = (\underbrace{\mathbf{B}}_{(J\times F)} \left|\otimes\right| \underbrace{\mathbf{A}}_{(I\times F)}) \cdot \underbrace{\mathbf{C}}_{(K\times F)}^{\mathbf{T}} = \underbrace{\mathbf{Z}}_{(IJ\times F)} \cdot \mathbf{C}^{\mathbf{T}} \qquad \text{Eq. 8}$$

Due to the symmetry of the model, the problem can be turned into a two-way problem of finding the least squares optimal **C** in the model $\mathbf{X}=\mathbf{Z}\mathbf{C}^{T}+\mathbf{E}$. The solution to this problem is

$$\mathbf{C}^{\mathbf{T}} = (\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}} \cdot \mathbf{X} = \mathbf{Z}^{+}\mathbf{X} \qquad \text{Eq. 9}$$

After giving starting values for **A** and **B**, the first estimation of **Z** is obtained (Equation 8) and the first estimation of **C** can then be calculated using Equation 9[10]. An update of **B** given **A** and **C**, an update of **A** given **B** and **C** and so on give rise to the optimization process that ends when the relative change in fit is small. Following the example of EEM fluorometry, the two loading matrices **A** and **B** contain the excitation and emission profiles of the active compounds and are used for qualitative identification of the fluorescent species in the samples (resolution purposes). Quantitative outputs are obtained from the **C**-loadings, which are the relative concentrations of the compounds. Absolute concentrations can be obtained from a regression of the **C**-loadings against the known concentration of the analyte of interest in the calibration samples.

[9] See “Khatri-Rao product” in the *Glossary* of terms.

[10] In Equation 16, the superscript + denotes the pseudoinverse. For full rank matrices, the pseudoinverse is $\mathbf{A}^{+}=(\mathbf{A}^{T}\mathbf{A})^{-1}\mathbf{A}^{T}$

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

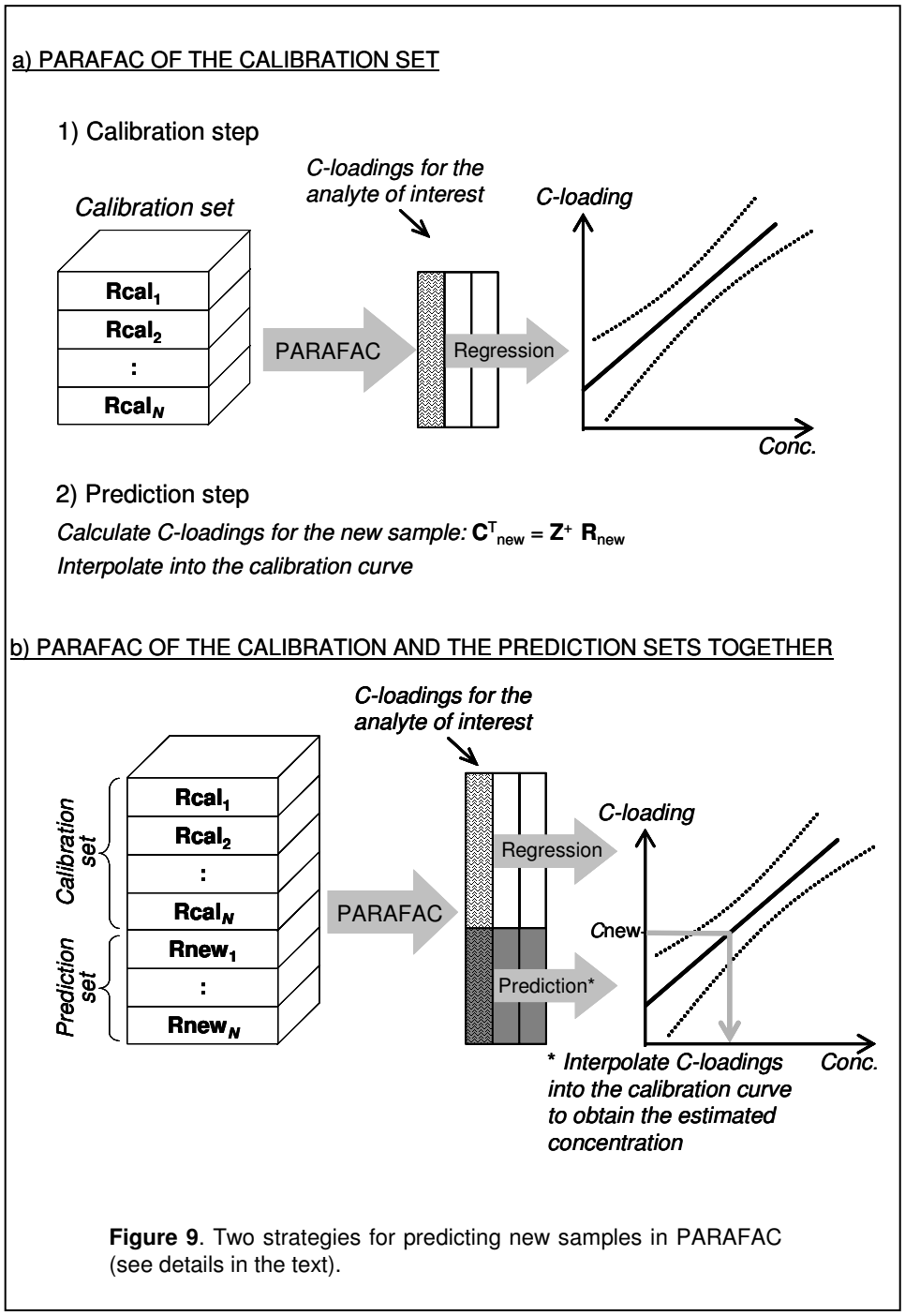


Figure 9. Two strategies for predicting new samples in PARAFAC (see details in the text).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

When the model has been fitted, it can be used to predict the concentration of the compounds in new samples. The **C**-loading of these new samples are simply calculated using Equation 9 with the **Z** matrix of the model and the new sample data, $\mathbf{R}_{new}$ (and converted on absolute concentration from the above-mentioned regression line).

Another strategy for predicting future samples is to include them in the initial PARAFAC model. In this way, the loading matrices for both the calibration and the prediction sets are recovered.

All the samples are considered simultaneously to calculate the model parameters. This forces the model to find the solution that best explains all the variations. This procedure may disguise outlying samples and lead to a poorer global fit, but in the absence of anomalous samples, and with a saving in calculation steps, the results may not differ significantly from those obtained by the first procedure. In figure 9 both procedures are schematized.

2.2.2.2. Iterative target transformation factor analysis (ITTFA)

Iterative Target Transformation Factor Analysis, ITTFA (Hopke *et al.*, 1983; Vandeginste *et al.*, 1985; Toft and Kvalheim, 1994), is a powerful multivariate curve resolution method whose aim is to determine the true factors of a second-order matrix. In the case of HPLC-DAD analyses, ITTFA determines the number of components in an overlapped chromatographic peak and mathematically resolves the spectro-chromatographic data into the elution profiles and spectra of all co-eluting compounds.

ITTFA tests whether a candidate factor or target is a true factor. The approach tries to improve targets in an iterative way using external knowledge such as non-negativity or unimodality to bring them close enough to one of the true factors.

The algorithm starts with the singular value decomposition of the peak under study, $\mathbf{R=UWV}^T$, where **U** is related to the chromatographic space and **V** is related to the spectral space. A normalized target vector, $\mathbf{x}_{target}$, with the shape of a chromatographic profile is proposed (possible shapes are Gaussian peaks of different size, needle peaks and triangular peaks).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

$\mathbf{x}_{target}$ is projected into the space described by the first column of **U** (Equation 10) and the norm of the difference between the target and the projected vectors is calculated (Equation 11).

$$\mathbf{x}_{projected} = \mathbf{u}_1\mathbf{u}_1^T \cdot \mathbf{x}_{target} \qquad \text{Eq. 10}$$

$$d = \left\| \mathbf{x}_{projected} - \mathbf{x}_{target} \right\| \qquad \text{Eq. 11}$$

The target vector is tested along the time axis, so that *d* can be represented as a function of the position of the maximum of $\mathbf{x}_{target}$. When $\mathbf{x}_{target}$ is described by the column vector $\mathbf{u}_1$, $\mathbf{x}_{projected}$ is similar to $\mathbf{x}_{target}$ and the difference *d* is small, which means that such target is representative of the peak of one of the compounds in **R** (Figure 10b). The further $\mathbf{x}_{target}$ is from the real one, the larger *d* is (Figure 10a).

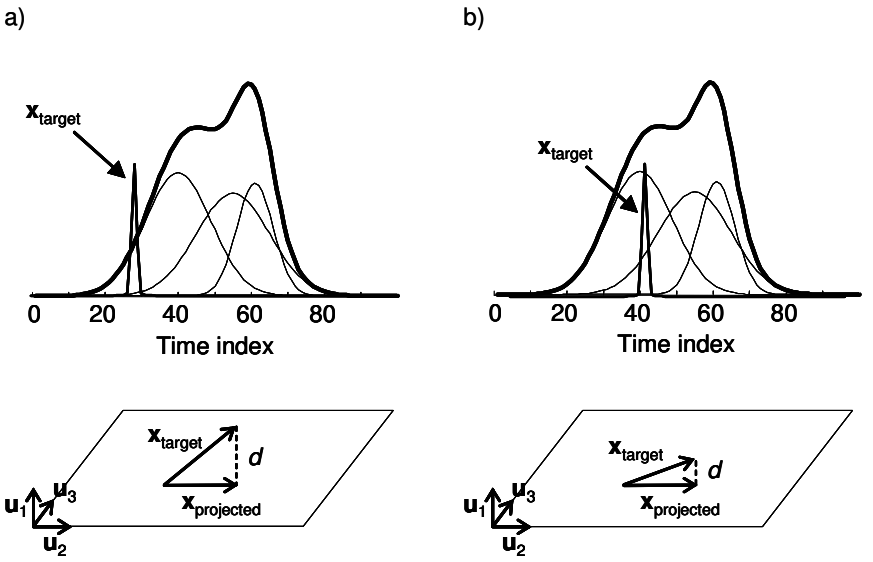


Figure 10. A target vector at different positions is projected into the chromatographic space. When it is close to an underlying factor, *d* becomes minimal.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

The same procedure is repeated by considering two, three, etc. columns of **U**. For each number of columns considered, i.e. the number of factors, *d* is plotted against the position of $\mathbf{x}_{target}$ where each minimum of these plots suggests the location of one peak. The optimal number of factors, *F*, is the one such that the number of minima does not increase when the number of factors is increased by one. Then, the number of factors corresponds to the number of compounds in the peak, and the positions of the maxima indicate the approximate situation of the maxima of the chromatographic profiles (see Figure 11).

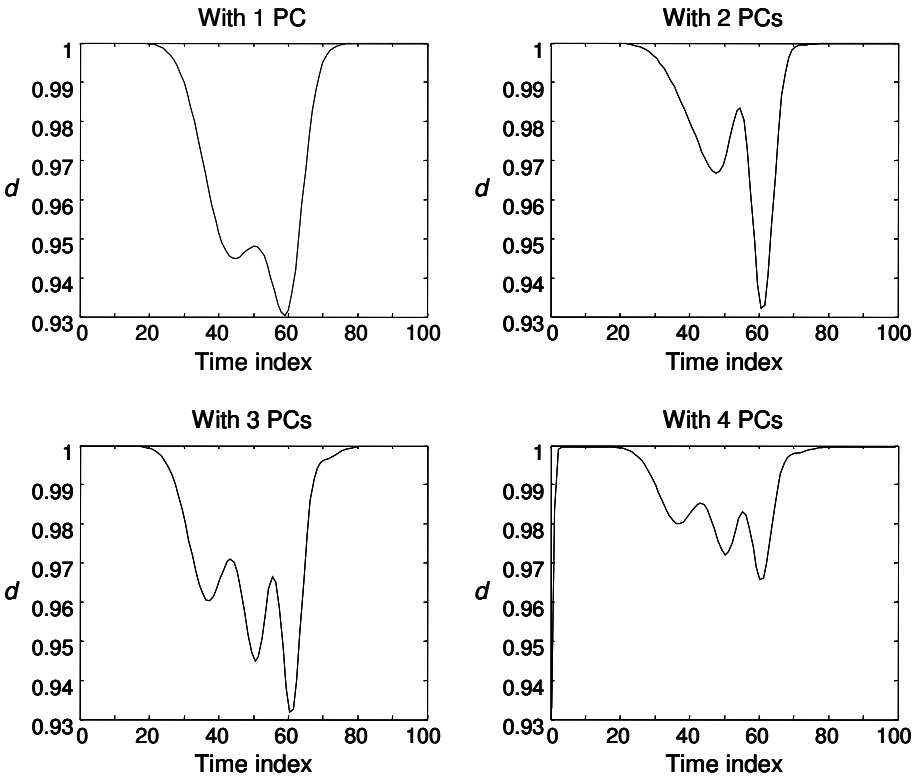


Figure 11. The number of minima that appear in the graphs and their position can be related to the number of compounds present in the sample and the time at which the elution peak of each compound appears, respectively. In this example, the proper number of factors is 3, because adding a new factor does not lead to a new minimum.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Once the number of factors is determined, for each factor a target vector whose maximum is at the estimated position from the graph is projected onto the chromatographic space. The projected vector can be improved using chemical knowledge, such as the adoption of non-negativity and unimodality according to the expected shape for a chromatographic profile. An iterative process, resubmitting the projected vector as the new input vector until convergence of the difference, leads to the estimation of the pure chromatogram. The process is repeated for each analyte whose position has previously been determined so that a chromatographic profile for each analyte is found. All the profiles are arranged in a new matrix of chromatographic profiles, **C**.

Estimates for the pure spectra, $\mathbf{S}=[\mathbf{s}_1,\ldots,\mathbf{s}_F]$, are then calculated using the generalised inverse of the matrix of chromatograms, as

$$\mathbf{S} = \left(\mathbf{C}^{+}\mathbf{R}\right)^{T} \qquad \text{Eq. 12}$$

Resolution is achieved because, for each spectro-chromatographic data matrix, the number of eluting compounds, as well as their pure profiles is obtained.

Once we have estimated the chromatograms of standards, we can regress the height or the area of the peak corresponding to the analyte of interest in each standard against the analyte concentration by fitting a univariate calibration model that can later be used to predict of new samples with overlapped signals.

2.2.2.3. Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS)

The versatile MCR-ALS (Tauler, 1995; Johnson *et al*., 1999; Tauler and de Juan) is a curve resolution method that is easily extended to the simultaneous analysis of several data matrices. Moreover, MCR-ALS shows the potential of handling second-order non-bilinear data, i.e. trilinear and non-trilinear structures, in contrast to PARAFAC-based models or generalized eigenvalue decomposition methods such as GRAM.

This method tries to explicitly recover the true underlying first-order profiles on each order of the measurement as constrained least squares optimal estimates. The usual initial assumption in multivariate resolution methods is that the

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

experimental data follow a linear model similar to that in Figure 3:

$$\mathbf{R} = \mathbf{C}\mathbf{S}^{\mathrm{T}} + \mathbf{E} \qquad \text{Eq. 13}$$

where **R** is, for example, the data matrix of spectra acquired at different times during the elution in an HPLD-DAD analysis, and **C** and **S** are the chromatographic and spectral profiles, respectively, of the eluting compounds. The number of chemical compounds is obtained directly from the chemical rank associated with the data matrix **R**. Starting estimates of **C** and **S** can be provided by techniques based on the detection of "purest" variables or from techniques based on Evolving Factor Analysis (an example of both is given in section 2.2.3). These initial estimations are optimized by solving Equation 13 iteratively by alternating least-squares optimization. At each iteration of the optimization, a new estimation of matrices **C** and **S** is obtained:

$$\mathbf{S} = \mathbf{C}^{+}\mathbf{R} = \mathbf{C}^{+}\mathbf{C}\mathbf{S}^{\mathrm{T}} \quad \text{and} \quad \mathbf{C} = \mathbf{R}^{*}\left(\mathbf{S}^{\mathrm{T}}\right)^{+} = \mathbf{C}\left(\mathbf{S}^{\mathrm{T}}\right)\left(\mathbf{S}^{\mathrm{T}}\right)^{+} \qquad \text{Eqs. 14}$$

where the matrix **R*** is the PCA reproduced data matrix for the selected number of components. At each iterative cycle, convenient constraints can be applied (some of these are explained in section 2.2.3) until convergence is achieved or a pre-selected number of cycles is reached. At this point, resolution is achieved for the individual compounds in the samples.

Extending this multivariate curve resolution method to the simultaneous analysis of several data matrices is easily performed by setting each of the individual data matrices $\mathbf{R}_k$ in a different reorganization that leads to a new augmented data matrix (see Figure 12).

Usually, this organization sets one matrix on top of the others, with the columns in common. This new augmented data matrix, shown in Figure 12b, is decomposed into an augmented matrix that describes the elution profiles of each compound in the different data matrices $\mathbf{R}_k$ (row space) times a smaller non-augmented matrix that describes the common spectral profile (column space). This data arrangement assumes that the spectra of the common active compounds are equal in the individual data matrices $\mathbf{R}_k$ but makes no assumption about the elution profiles in the different samples.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

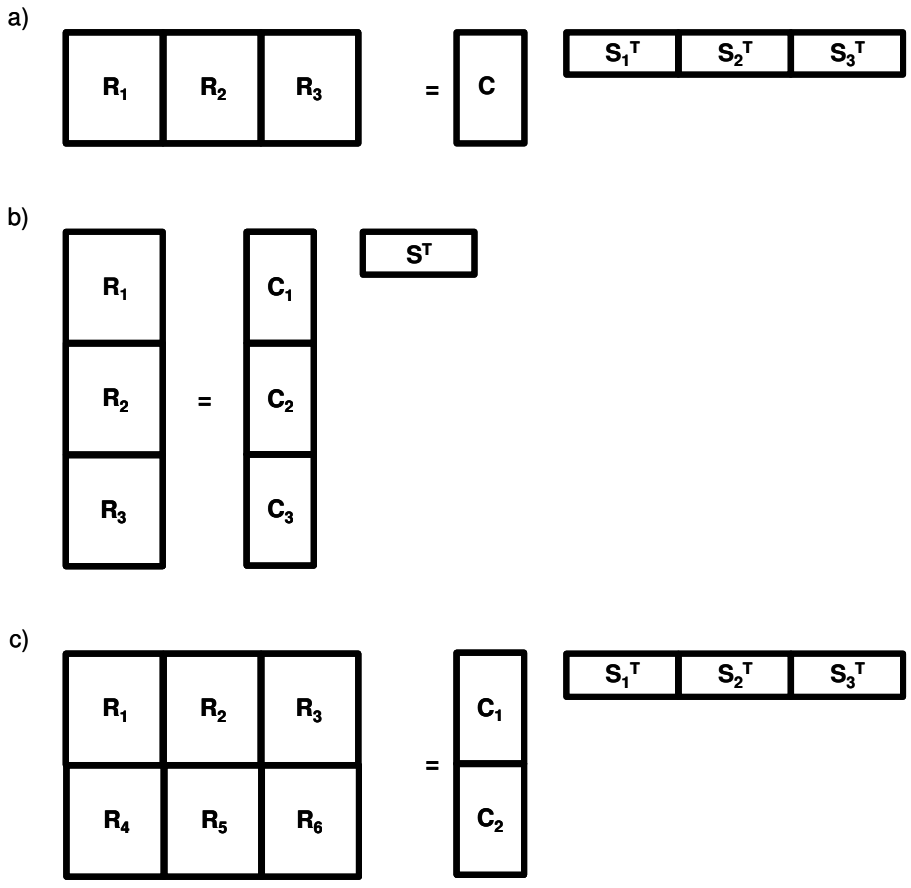


Figure 12. Augmentation of individual data matrices in the simultaneous analysis of several related data matrices. a) Row-wise augmented matrix, [$\mathbf{R}_1$ $\mathbf{R}_2$ $\mathbf{R}_3$]. b) Column-wise augmented data matrix, [$\mathbf{R}_1$;$\mathbf{R}_2$;$\mathbf{R}_3$]. c) Row- and column-wise augmented data matrix, [$\mathbf{R}_1$ $\mathbf{R}_2$ $\mathbf{R}_3$; $\mathbf{R}_4$ $\mathbf{R}_5$ $\mathbf{R}_6$].

Once resolution has been achieved for a particular compound following the iterative procedure described above, calibration for that particular compound is possible. Since the spectra for all the $\mathbf{R}_k$ must be the same, the appropriate column in the corresponding resolved $\mathbf{C}_k$ matrix contains the relative contribution of this compound in this matrix $\mathbf{C}_k$ in relation to the other matrices included in the augmented **C** matrix. The relative concentration of a particular compound

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

($c_{unknown}$) can be simply obtained from the quotient between the area of its resolved elution profile ($A_{unknown}$) and the area of the resolved elution profile of the same compound in another data matrix (of calibration) included in the same simultaneous analysis ($A_{standard}$), i.e.:

$$c_{unknown} = (A_{unknown} / A_{standard})\, c_{standard} \qquad \text{Eq. 15}$$

When several standards are employed in the decomposition, we can model the relative areas —with respect to the same reference standard— of the calibration samples against the known concentration of the analyte of interest. The relative areas of the prediction samples, which are also relative with respect to the reference standard, are then interpolated into the calibration line to obtain the estimated concentrations. Note that as prediction samples are used, together with the calibration samples, to build the calibration model, they have as much influence as the calibration set in the iterative decomposition procedure. On the other hand, the *second-order advantage* is fulfilled, i.e. the quantification of the analyte of interest in the prediction samples is possible even in the presence of new interfering compounds.

2.2.2.4. Multi-linear Partial Least Squares (*N*-PLS)

Partial least squares regression is a method for building regression models between independent (called *x*) and dependent (called *y*) variables. For multi-way calibration, multi-linear PLS or simply *N*-PLS (Bro, 1996) is an extension of the ordinary regression model PLS, where the independent data are modelled in a way that emphasizes variation that is especially relevant for predicting the dependent variables.

The general terminology of *N*-PLS depends on the order of the data: a Greek prefix indicates the order of **X** and an Arabic suffix after PLS indicates the order of **Y**. In tri-PLS1, therefore, each calibration sample is characterized by a $J \times K$ matrix $\mathbf{X}_i$, to build a cube of data $\underline{\mathbf{X}}$ of dimensions $I \times J \times K$, and for each sample there is a known measurand to be predicted by the independent variables, the values of which are collected in a general $\underline{\mathbf{Y}}$ matrix of dimensions $I \times L \times M$. When the measurands are concentrations, they are collected in a $I \times 1$ vector called **y**. If more than one variable is sought, these are collected in a matrix **Y** of dimensions $I \times L$, where L is the number of different analytes (tri-PLS2, tri-PLS3, etc.). When

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

several dependent variables are present, all y_i can be modelled simultaneously (as in the PLS2) and each dependent variable can be modelled with the PLS1 algorithm.

The main feature of the *N*-PLS algorithm is that it produces score vectors that, in a trilinear sense, have maximum covariance with the unexplained part of the dependent variable. For tri-PLS(*L*) (with $L \geq 2$), a trilinear decomposition of **X** and **Y** is sought, as in Figure 13. Dependent and independent data are both assumed to be centred in the *I* mode, i.e. column-wise, and sometimes even scaled.

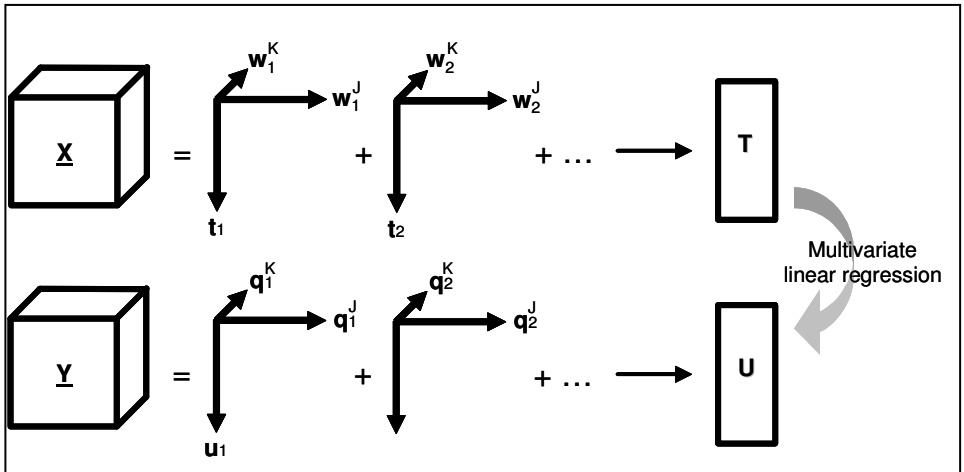


Figure 13. General scheme of an *N*-PLS decomposition.

The aim of the algorithm is to decompose the array **X** into triads. A triad consists of one score vector (**t**) and two weight vectors, one in the second mode called $\mathbf{w}^J$ ($J \times 1$) and one in the third mode called $\mathbf{w}^K$ ($K \times 1$).

Let **X** be the $I \times JK$ matricized array of independent data and **Y** be the $I \times LM$ matricized array of dependent data. The *N*-PLS models decomposes these matrices as

$$\underset{(I \times JK)}{\mathbf{X}} = \underset{(I \times F)}{\mathbf{T}} \cdot \underset{(JK \times F)}{\mathbf{W}}^{\mathsf{T}} = \mathbf{T}\left(\mathbf{W}^K \middle| \otimes \middle| \mathbf{W}^J\right)^{\mathsf{T}} + \mathbf{E}_X \qquad \text{Eq. 16}$$

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

$$\underset{(I\times LM)}{\mathbf{Y}} = \underset{(I\times F)}{\mathbf{U}} \cdot \underset{(LM\times F)}{\mathbf{Q}}^{\mathbf{T}} = \mathbf{U}\left(\mathbf{Q}^{\mathrm{L}} \middle| \otimes \middle| \mathbf{Q}^{\mathrm{M}}\right)^{\mathrm{T}} + \mathbf{E}_{\mathrm{Y}} \qquad \text{Eq. 17}$$

where the score vectors of **X** and **Y** are called **T** and **U**, respectively, and the weight vectors are called **W** and **Q**. The superscripts *J* or *K* and *L* or *M*, respectively, are used to specify which mode the vectors refer to. The model finds the set of vectors $\mathbf{w}^{\mathrm{J}}$, $\mathbf{w}^{\mathrm{K}}$, $\mathbf{q}^{\mathrm{L}}$ and $\mathbf{q}^{\mathrm{M}}$ such that the least-squares score vectors **t** and **u** have maximal covariance. The scores are successively determined following this criterion until enough components are calculated. Cross-validation can be used as a criterion for the number of components to use in the final model. The scores are then related by setting the regression model for the so-called inner relation:

$$\underset{(I\times F)}{\mathbf{U}} = \underset{(I\times F)}{\mathbf{T}} \cdot \underset{(F\times F)}{\mathbf{B}} + \mathbf{E}_{\mathrm{U}} \qquad \text{Eq. 18}$$

The regression matrix **B** must be calculated with the current and all previous **t**-vectors since **T** has no orthogonality properties (nor does **W)**.

The *N*-PLS model is unique in the sense that it consist of successively estimated one-component models. However, uniqueness in this case will seldom infer that real underlying phenomena such as pure-analyte spectra can be recovered, because the model assumptions do not reflect any fundamental or theoretical model. Here the focus is on predicting Y.

The dependent variable for a new sample $\mathbf{r}^{\mathrm{T}}_{\mathrm{unknown}}$ (1×*JK*) is predicted as follows (this can also be applied to a set of new samples, $\mathbf{X}_{\mathrm{unknown}}$.) From the model of **X** (Equation 16), **t** can be determined. Through Equation 18 scores in the **Y**-space can be predicted and through the model of **Y** (Equation 17) the prediction of $\mathbf{y}^{\mathrm{T}}_{\mathrm{pre}}$ is obtained:

$$\mathbf{r}^{\mathrm{T}}_{\mathrm{unknown}} = \mathbf{t}^{\mathrm{T}}_{\mathrm{unknown}} \cdot \mathbf{W}^{\mathrm{T}} \quad \rightarrow \quad \mathbf{t}^{\mathrm{T}}_{\mathrm{unknown}} = \mathbf{r}^{\mathrm{T}}_{\mathrm{unknown}} \cdot \left(\mathbf{W}^{\mathrm{T}}\right)^{+} \qquad \text{Eq. 19}$$

$$\mathbf{u}^{\mathrm{T}}_{\mathrm{unknown}} = \mathbf{t}^{\mathrm{T}}_{\mathrm{unknown}} \cdot \mathbf{B} \qquad \text{Eq. 20}$$

$$\mathbf{y}^{\mathrm{T}}_{\mathrm{pre}} = \mathbf{u}^{\mathrm{T}}_{\mathrm{unknown}} \cdot \mathbf{Q}^{\mathrm{T}} \qquad \text{Eq. 21}$$

If only the prediction of **Y** is wanted, it is possible to obtain a set of regression

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

coefficients that directly relates the centred/scaled **X** with the centred/scaled **Y**:

$$\mathbf{B}_{PLS} = \left(\mathbf{W}^T\right)^{+} \mathbf{B}\, \mathbf{Q}^T \qquad \text{Eq. 22}$$

Therefore, for a new sample $\mathbf{r}^T_{unknown}$ the predicted Y variable is

$$\mathbf{y}^T_{pre} = \mathbf{r}^T_{unknown} \cdot \mathbf{B}_{PLS} \qquad \text{Eq. 23}$$

2.2.3. Uniqueness, indeterminacies and constraints

Some mathematical models do not have a *unique* solution. For the decomposition of second-order data, which is expressed as the outer product of two vectors, there are infinite pairs of vectors that provide the exact same fit to the data due to the rotational freedom of the model. To find a unique solution in such decompositions, one can force the resolution to satisfy several *constraints*. For example, the decomposition of a data matrix by PCA is unique because the extracted factors are defined to explain the orthogonal variation of the data. Uniqueness is achieved when no additional constraints are needed to identify the model; hence any change in the model parameters results in a change of fit. However, even in unique multi-linear models there remain two indeterminacies. First, let us imagine the spectrochromatogram for a pure compound. The decomposition of this data matrix into a dyad will provide an estimation, of the pure spectrum and the elution profile, scaled by some unknown factor: the amplitude of one of these can be increased with a corresponding decrease in the other, and the product of the two vectors remains the same. This is known as a scale or intensity ambiguity and it can be solved using external knowledge that is introduced into the decomposition as constraints. Usually, spectra are scaled to norm one so that samples with increasing concentration show proportionally increasing amplitude in their elution profile, and a univariate regression between the areas or the heights and the reference concentrations can be established. Other strategies are a closure constraint (the fulfilment of a mass balance condition) or an equality constraint (e.g. known spectra). In the analysis of three-way structures, intensity/scale ambiguity is solved directly because relative terms are obtained and the absolute terms can be found by including known scale standards in the analysis.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

The other ambiguity is the order in which the factors are arranged. This permutation freedom, called rotational ambiguity, is solved in PCA by ordering the factors according to a decreasing explained variance. In multi-linear models permutation ambiguity can also be solved using external knowledge about the system, e.g. by identifying some selective variables that have intensity contribution from only one compound or the known rank of reduced data regions (local rank). Natural constraints such as non-negativity (for example in the estimated concentrations) or unimodality (in signals known to have only one maximum) may also help to diminish the model's rotational ambiguity.

A restricted model reduces the number of feasible solutions and always leads to poorer fits than unconstrained models. However, the gain in interpretability normally justifies this loss. At this point, several constraints have already been cited. A further description of these constraints and their uses, and of other constraints, can be found in the literature (Tauler *et al.*, 1995; de Juan *et al.*, 1997; Bro, 1998). Here, I will provide a brief overview of those used in this thesis.

Orthogonality

In the decomposition of an *n*-way structure into a set of basis vectors, these vectors can be defined as orthogonal[11], i.e. linearly independent. This mathematical property is useful for modelling algorithms because it enables different sets of variables to be treated independently. Forcing the loading vectors in certain modes to be orthogonal leads to non-correlated components, may speed up the algorithms and avoids local minima and degeneracy[12]. However, orthogonal factors seldom represent real underlying phenomena. Chromatograms or spectra, for example, are often quite similar when different compounds from the same chemical family are analysed. This constraint is therefore only used for calibration and quantification purposes and is not used when curve resolution or visual interpretation is the aim of the decomposition.

[11] See "Orthogonality" in the *Glossary* of terms.

[12] Degenerate solutions are found, for example, when a trilinear model is not appropriate for describing the data.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Normalization

Some loadings are usually scaled to norm one in order to solve the intensity ambiguities and ensure that the estimated profiles differ only in shape, not in intensity. When talking about the norm of a vector or, even, of a matrix, we usually refer to the Euclidean (or Frobenius) norm, i.e. the square root of the sum of the squared elements in the vector or matrix[13].

Correlation and equality

During the decomposition of an *n*-way structure, loadings, e.g. UV-Vis spectra, can be directly updated using the corresponding known values (equality) or according to their correlation with these previously known values. The known information can have different origins. We may, for example, know the pure profile of a certain compound (e.g. concentration profile or spectrum). In a local rank analysis (see below) of the data set, we may have detected selective regions of the data set in the concentration and/or in the spectral direction: in this case, we know that the rest of the species are absent, so the values of their related elements should be set to zero.

Selective variables

One variable, or a range of variables, are selective when the intensity contribution comes from only one compound of the mixture, e.g. wavelength ranges where only one compound absorbs. These variables, which are the most dissimilar or linearly independent variables in the system, can be detected using, for example, Orthogonal Projection Analysis (OPA) (Cuesta-Sánchez *et al.*, 1996) or SIMPLe-to-use Interactive Self-Modelling Analysis (SIMPLISMA) (Windig and Guilmet, 1991; Windig *et al.*, 2002). SIMPLISMA was used in Chapter 5. This procedure is based on the relative standard deviation of the *j* columns of the original data matrix **R** corrected with a small off-set value or noise level, *E*, usually calculated as a percentage (often 1 to 5%) of the maximum of the mean sample. Following the HPLD-DAD example, the larger the relative standard deviation, the higher the wavelength's selectivity or purity

[13] Note that there are other ways to calculate a norm, though all *p-norms* are equivalent for finite dimensional spaces. See "Norm" in the *Glossary* of terms.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

$p_j = s_j/(\bar{r} - E)$. The chromatograms at the purest wavelengths, i.e. the chromatograms that are least contaminated by the other compounds, are arranged in a matrix **P** and resolved into the pure spectra **S** by solving the equation $\mathbf{R}=\mathbf{P}\mathbf{S}^T$, i.e. $\mathbf{S}^T=(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{R}$. The estimated spectra in **S** are the pure variables (pure spectra) except for a normalization factor.

Local rank

To analyse the (global) rank in multicomponent mixtures, a second- (or higher-order[14]) tensor is needed and the number of rows, *I*, or the number of columns, *J*, of the data matrix must be equal to or larger than the number of pure compounds, *F*. For a matrix, the mathematical rank is the maximum number of linearly independent vectors. However, in chemical experiments the chemical rank, understood as the number of analytes that contribute to the measured variance, is generally lower than the mathematical rank. The number of principal components or latent variables into which the data matrix can be decomposed is itself an estimation of the (global) chemical rank in the sample. However, the local rank, i.e. the rank of reduced data regions, varies within the structure when some analytes are present only in some windows, depending, for example, on their elution pattern or absorptivity spectra. Local rank information can be obtained by the evolutionary rank analysis based methods (Toft, 1995). These methods analyze the two-way data structure piece-way, i.e. they define submatrices or windows and decompose each of them into latent variables by PCA or singular value decomposition (SVD)[15] to locally reveal the presence of analytes. Their power has been proven for multidetectional chromatographic structures in the retention time direction and in the wave numbers direction, separately, as well as in both directions simultaneously.

In Evolving Factor Analysis (EFA), which was used in Chapter 5, the first time window is defined by the spectra at the two first retention times, the second time

[14] The rank of a three-way array is equal to the minimal number of triads needed to describe the array. However, estimating the rank in multi-way arrays is not trivial and only some guidelines or results for particular cases are reported. For example, the maximal rank for a 2×2×2 array is 3, and 5 for a 3×3×3 array. For *i*×*j*×2 structures, Ten Berge [Ref. 2000] and Ten Berge *et al.* [Ref. 1999] offer some concluding inequalities, such as rank=2*j* for *i*≥*j* and rank=*i* for *j*<*i*>2*i*.

[15] See “SVD” in the *Glossary* of terms.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

window contains the first three spectra, and so on until the spectrum of the last retention time has been included. In each window, the number of latent variables is an estimate of the number of analytes present. This *forward* procedure reveals the appearance of the analytes (in the chromatogram) and the inverse procedure or *backward* EFA (starting from the last retention time) reveals their disappearance. A tutorial about EFA is cited in the bibliography (Keller and Massart, 1992) (see also Maeder [Ref. 1987] and Gemperline and Hamilton, [Ref. 1989]). Particularly useful for detecting minor components is the *fixed-size moving window* EFA (FSMW-EFA) (Keller and Massart, 1991), where the time window has a constant size (it usually contains 7 spectra), starts from the first retention time and moves stepwise until the spectrum at the last retention time is included.

Non-negativity

This constraint is a particular case of the inequality constraints in which the model parameters are forced to be equal or greater than zero. This measure can diminish rotational ambiguity and is actually based on the non-negativity of most physical properties such concentration and absorptivitiy. Non-negativity therefore reduces the number of feasible solutions and provides them with physical meaning. This was used in Chapters 3, 4 and 5.

Unimodality

Chromatographic peaks (i.e. elution profiles), concentration profiles in some chemical reactions and some peak-shaped instrumental responses can be assumed to show only one maximum. Unimodality is normally implemented by suppressing all local maxima to the right and to the left of the largest maximum. The concepts of *vertical* and *horizontal* unimodality refer to the way in which these local minima are suppressed, i.e. by setting the non-unimodal values to zero or to the nearest element maintaining the unimodal condition, respectively (de Juan *et al.*, 1997). Some algorithms, e.g. Unimodal Least Squares Regression (ULSR) (Bro and Sidiropoulos, 1998) have been specially developed to implement unimodality in a least squares sense.

Trilinearity

The three-way data structure constraint allows a trilinear model to be selected to

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

describe the data set. From an operational point of view, this means that a trilinear structure implies a common shape for all profiles related to the same compound in the different samples, e.g. in the submatrices of an augmented direction in MCR-ALS. In this case, in a column-wise augmented matrix, the elution profile from a given compound would have the same shape in all $\mathbf{C}_i$ submatrices.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

2.3. Validation of analytical methods

2.3.1. What is method validation?

In modern society, many decisions in science and technology depend crucially on the quality of the chemical measurements. Evidence shows that national and international regulations with commercial, medical or environmental impact implement limits and verify the implementation of these limits by setting out guidelines to reliably assess the quality of the measurements. Since the origin of such social and scientific necessity, the acceptance of quality has been based on an accreditation system (with more or less uniform and universal criteria). Concepts, terminology and approaches were not always as clear and universal as desired, however.

In July 1993 two international organizations, the International Union of Pure and Applied Chemistry (IUPAC) and the International Organization for Standardization (ISO), convened the "Harmonization Meeting" with the aim of developing common concepts and terminology for the international chemical and metrological communities[16].

According to current international recommendations (Inczédy *et al.*, 1998, EURACHEM 1998), analytical methods should be well characterized in order to define their application area and reliability, by meeting the specifications related to the intended use of the analytical results. This process, by which the method proves its "fitness for purpose", is known as validation[17].

The validation process typically starts by defining the scope of the analytical procedure, which includes the matrix, target analyte(s), analytical technique and intended purpose. The next step is to identify the performance characteristics that must be validated, which may depend on the purpose of the procedure, and

[16] IUPAC Recommendations 1995 (Currie 1995) and ISO Standard (11843.1-1997, "Capability of detection")

[17] "*Validation is the confirmation by examination and provision of the objective evidences that the particular requirements of a specific intended use are fulfilled*" (ISO 17025-1999, 5.4.4.1)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

the experiments for determining them. Finally, validation results should be documented, reviewed and maintained (if not, the procedure should be revalidated) as long as the procedure is applied in routine work.

2.3.2. Figures of merit. International recommendations

Method validation implies the confirmation that performance capabilities are consistent with what the application requires. The method performance parameters that need to be characterised can depend on customer requirements, prior experience of the method or the current guidelines in a particular sector of measurement, e.g. food analysis. When these performance parameters are used to prove that the procedure is under control or suitable for a certain analytical problem, they should be determined not under optimal instrumental conditions but in a real matrix under routine conditions.

The figures of merit of a chemical analytical process are 'those quantifiable terms which may indicate the extent of quality of the process. They include those terms that are closely related to the method and to the analyte (sensitivity, selectivity, limit of detection, limit of quantification, ...) and those which are concerned with the final results (traceability, uncertainty and representativity) (Inczédy *et al.*, 1998).

Typical figures of merit for the chemical measurement process, revised from the literature[18], are: accuracy, working/linear range, precision (repeatability and reproducibility), sensitivity, selectivity, robustness, detection limit and quantification limit. Some of these have been used in this thesis and are defined below.

- Accuracy

The accuracy of an analytical method is defined as 'the closeness of agreement between a test result and the accepted reference value'. Note that the term accuracy, when applied to a set of test results, involves a combination of random

[18] Danzer and Currie 1998; Danzer *et al*. 2004; Olivieri *et al*. 2006; Danzer 1996; Wegscheider 1996; IUPAC 'Orange' Book (Inczédy *et al*. 1998); Thomson *et al*. 2002; among others.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

components and a common systematic error or bias component (ISO 3534-1:1993).

The degree to which the determined value of an analyte in a sample corresponds to the true value may be determined by analysing a reference material of known concentration (this is not available for every determination), by the spiked-placebo[19] recovery method or by the standard addition[20] method. Conceptually, inaccuracy (lack of accuracy) comprises *imprecision* (or standard deviation) and *bias* (lack of trueness or existence of systematic errors).

- Working/linear range

The relationship between signal and concentration (or content) is numerically determined by means of regression. The functional relation can often be described by linear functions. Linearity is the ability of an analytical procedure to produce test results that are proportional to the concentration of analyte in samples within a given concentration range. The working range of an analytical procedure is the interval of analyte concentrations in the sample for which it has been demonstrated that the analytical method has suitable levels of precision, accuracy and linearity.

- Repeatability and reproducibility (precision)

Repeatability and reproducibility refer to the closeness of the agreement between the results of successive measurements of the same measurand carried out in certain conditions of measurement (ISO 3534-1/1993):

- Repeatability conditions, i.e. conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time. Thus, repeatability reflects the best achievable internal precision.
- Reproducibility conditions, i.e. conditions where test results are obtained with

[19] In the spiked-placebo method, sometimes also named standard addition method, a known amount of pure active compounds is added to a sample that contains all other ingredients except the active(s). The resulting mixture is assayed, and the results obtained are compared with the expected result.

[20] In the standard addition method, a sample is assayed, a known amount of pure active compound is added, and the sample is again assayed. The difference between the results of the two assays is compared with the expected value.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

the same method on identical test items in different laboratories with different operators using different equipment.

- Intermediate conditions, i.e. the test results are obtained within the same laboratory by changing the operator, the equipment and/or the day.

- Sensitivity

Sensitivity is the change in the response of a measuring instrument divided by the corresponding change in the stimulus[21] (EURACHEM Guide). In practice, in univariate calibration the sensitivity for a given analyte is defined as the slope of the analytical calibration curve which is usually constant at low concentration values (the calibration curve becomes a straight line). Analogously, for a classical multivariate model, the sensitivity is the slope of the pseudo-univariate calibration graph, and the inverse of the slope of this graph for an inverse model. Sensitivity can be defined as the net analyte signal NAS[22] (or part of the spectrum which is orthogonal to the spectra of the others compounds), generated by an analyte concentration equal to unity (Lorber, 1986). It can therefore be calculated as:

$$SEN_N = \mathrm{NAS}_N = \left\| \left(\mathbf{I} - \mathbf{S}_{-N} \mathbf{S}_{-N}^{+} \right) \mathbf{s}_N \right\| \qquad \text{Eq. 24}$$

I is the identity matrix and $\mathbf{S}_{-N}$ is the response of the calibration sample without the N column ($\mathbf{s}_N$).

For higher-order data, sensitivity is consistently generalized by Olivieri and Faber [Ref. 2004] for PARAFAC analyses.

- Selectivity

According to the latest IUPAC recommendation, 'selectivity refers to the extent to which the method can be used to determine particular analytes in mixtures or matrices without interferences from other compounds of similar behaviour' (Vessman *et al.*, 2001). In other words, selectivity is the ability of a method to

[21] '*Stimulus may for example be the amount of the measurand present. Sensitivity may depend on the value of the stimulus. Although this definition is clearly applied to a measuring instrument, it can also be applied to the analytical method as a whole, taking into account other factors such as the effect of concentration steps*' (VIM 1984 and IUPAC 'Orange' Book).

[22] See "NAS" in the *Glossary* of terms.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

accurately and specifically determine the analyte of interest in the presence of other compounds in a sample matrix under the stated conditions of the test. To quantitatively measure the extent to which a substance interferes with the determination of the analyte according to a given procedure, a pair-wise selectivity index has been defined as the ratio of the slopes of the calibration lines of the analyte of interest and the particular interference (Thomson *et al.*, 2002). In multivariate analysis, however, the interferences can be suitably modelled, which is why calculating the multivariate calibration selectivity has been approached in several ways. In any case, selectivity typically accounts for all the interferences in the mixture, though there is some traces of research on individual (pair-wise) multivariate selectivity coefficients (C.D. Brown, in preparation). In this thesis, selectivity is calculated, as defined by Lorber (Lorber, 1986 and Lorber *et al.*, 1997) as the net analyte signal divided by the length of the spectrum of the corresponding compound:

$$SEL_f = \frac{\mathrm{NAS}_f}{\left\| \mathbf{s}_f \right\|} = \frac{\left\| \left(\mathbf{I} - \mathbf{S}_{-f} \mathbf{S}_{-f}^{+} \right) \mathbf{s}_f \right\|}{\left\| \mathbf{s}_f \right\|} \qquad \text{Eq. 26}$$

The range of Lorber's selectivity is between 1 (fully selective or specific), when the spectra of the *f*th compound is completely free from overlap, and 0, when the spectra is the same as the other compounds' spectra or may be described as a linear combination of them. This selectivity measure indicates which part of the total signal is lost due to spectral overlap and, according to the recommendations of IUPAC, it can be used in multivariate calibrations.

For higher-order data, selectivity is also consistently generalized by Messick *et al*. [Ref. 1996]. These authors demonstrated that the Frobenius norm ($\|\cdot\|_F$) of the orthogonal projection of the matrix of responses of the *f* compound, $\|\mathbf{P}_f(\mathbf{R}_f)\|_F$ divided by $\|\mathbf{R}_f\|_F$ works effectively as a local second-order selectivity measure (Messick *et al. op. cit*. also provided a straightforward approach for computing the projections).

- Detection capability

The limit of detection (LOD, or detection limit) is, in broad terms, the smallest amount or concentration of analyte in the test sample that can be reliably

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

distinguished from zero. Although detection capability is widely considered to be a fundamental performance characteristic of the chemical measurement process, the detection limit as estimated in method development (or in the brochures of instruments) may not be identical in concept or numerical value to the one used to characterize an analytical method. For most practical purposes in method validation, it is usual to choose a simple definition that enables a quickly implemented estimation that is used only as a guide to the utility of the method. Analytical detection involves several conceptual approaches that have led the chemical community to controversy, which justifies the deeper development in the next section.

2.3.3. Analytical detection

2.3.3.1. Hypothesis testing

When an analyte-free sample is analyzed, the response produced by the measuring instrument may be nonzero. The signal observed for this blank is subtracted from the sample signal to obtain the *net signal*. However, if the blank measurement is repeated, the signal varies somewhat due to random errors, i.e. there is a blank signal distribution. To determine how large the instrument signal for a sample must be to reliably say that it contains the analyte, one calculates a threshold value for the net signal. This is called the critical value and is sometimes denoted by r_C. If the observed net signal for a sample exceeds the critical value, the analyte is considered to be "detected"; otherwise, it is considered to be "not detected".

The methods most often used for making detection decisions are based on statistical hypothesis testing (EPA 2004) and involve a choice between two hypotheses about the sample. The first hypothesis is the "null hypothesis" (H_0): the sample is analyte-free. The second hypothesis is the "alternative hypothesis" (H_1): the sample is not analyte-free.

In the hypothesis test there are two possible types of decision errors. An error of the first type, or Type I error, occurs when the signal for an analyte-free sample exceeds the critical value, leading one to conclude incorrectly that the sample contains a positive amount of the analyte. This type of error is sometimes called

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

a "false positive" and the probability of its being committed is often denoted by α.

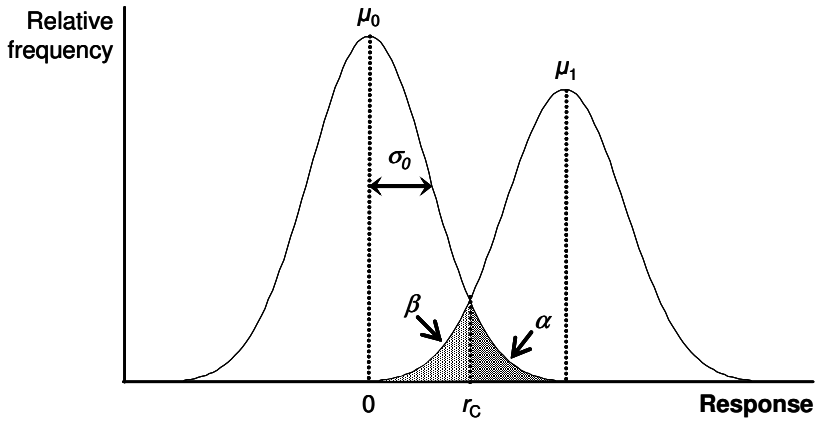


Figure 14. Distribution of the observed signal for a blank (with mean μ_0) and for a test sample (with mean μ_1). Type I errors (false positives) are represented by the value of α and type II errors (false negatives) are represented by the value of β.

An error of the second type, or Type II error, occurs if one concludes that a sample does not contain the analyte when it actually does. The probability of a Type II error, also known as a "false negative", is usually denoted by β.

To calculate the critical value, r_C, one must choose a significance level for the test. The significance level is a specified upper bound for the probability of a Type I error and is usually chosen to be α=0.05. This means that when an analyte-free sample is analyzed, there is a confidence level of $1-\alpha=0.95$, i.e. there should at most be a 5% probability of incorrectly deciding that the analyte is present. See Figure 14 for a graphical representation of these concepts.

2.3.3.2. Critical level and detection limit

Progresses in food science and technology in the 1940s led to a growth in knowledge and public awareness of food safety. Long before international organizations recommended chemical societies assess the performance

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

characteristics of the measurement processes, the need for analytical detection arose in order to provide end users (customers) with quality results, such as ensuring safe levels of hazardous materials.

Early mentions of chemical limits of detection were published by H. Kaiser and L.A. Currie (Kaiser, 1947; Kaiser, 1967; Currie, 1968). Currie defined three levels: the "critical level", L_C, the "limit of detection (LOD)", L_D; and the "limit of determination", L_Q, at which quantification was achieved with a given precision. Mathematically, these limits were expressed as

$$L_C = z_\alpha \cdot \sigma_0 \qquad \text{Eq. 27}$$

$$L_D = L_C + z_\beta \cdot \sigma_D = z_\alpha \cdot \sigma_0 + z_\beta \cdot \sigma_D \qquad \text{Eq. 28}$$

$$L_Q = k_Q \cdot \sigma_Q \qquad \text{Eq. 29}$$

where the factors z_α and z_β were abscissas of the standardized normal distribution for $1-\alpha$ and $1-\beta$ probability levels, σ_0, σ_D and σ_Q were the standard deviations of the (net) signal when the true signal was 0, L_D and L_Q, respectively, and $1/k_Q$ was the requisite relative standard deviation, which for many problems is set at 10%.

At zero concentration level, the standard deviation of the net concentration is expressed in a general way as:

$$\sigma_0 = \sigma_B \phi \qquad \text{Eq. 30}$$

σ_B is the standard deviation of the blank and $\phi = \sqrt{1/N + 1/n}$, where N and n are the number of replicates on the analysed sample and on the blank sample, respectively. $\phi = 1$ (and therefore, $\sigma_0 = \sigma_B$) in the special case when N=1 and n is high. When the net concentration is calculated in a paired experiment (i.e. as the concentration in the sample minus the blank), then $\phi = \sqrt{2}$. If σ_B is not known, it must be replaced by its corresponding estimate, s_B, in Equation 26.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

2.3.3.3. Limit of detection from linear calibration curves

Hubaux and Vos (1970) were pioneers in the evaluation of decision and detection limits from the calibration curve of an analytical procedure. Their implementation was based on external standards, a very common practice in calibration, and on some prior hypothesis to be accomplished. First, the standards were assumed to be independent (separately prepared) and to have accurately known concentrations. Second, the observed signals were assumed to have a Gaussian distribution around their expectation, with constant variance throughout the range of calibration. From these starting features, they described a new approach to define these lower limits, not from experimental repetitions of blanks, but from the confidence limits of the regression line.

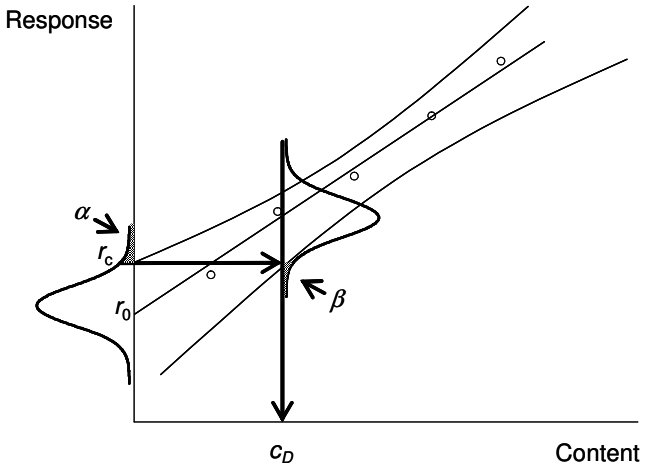


Figure 15. Hubaux and Vos' description of the critical level (r_C) and detection limit (c_D) from the upper and lower confidence limits of a linear calibration.

As Figure 15 shows, r_C is the lowest measurable signal and any signal under it will be disregarded unless the risk of stating that the analyte is present when it is absent is greater than α. The abscissa corresponding to r_C on the lower confidence limit, c_D, is the lowest content that can be distinguished from zero. This means that any sample with a content of less than c_D has a risk β of leading

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

to a signal lower than r_C and therefore stating that it is a blank. Using Currie's terminology, r_C is (an estimation of) the "critical level", L_C, in the signal domain and c_D is the "detection limit".

The critical level r_C was defined in the signal domain, which meant taking an *a posteriori* decision (after measuring the signal). As the detection limit, c_D, on the other hand, was related to content, it was inherent to the method and therefore computable *a priori*. Equations 31 and 32 below reflect the mathematical relationship between these limits and the confidence intervals of the regression line.

$$r_C = b_0 + t_{1-\alpha,n-2} \cdot s_0 = \; b_0 + t_{1-\alpha,n-2} \cdot s \sqrt{\frac{1}{N} + \frac{1}{n} + \frac{\bar{c}^2}{\sum (c_i - \bar{c})^2}} \qquad \text{Eq. 31}$$

The statistical parameters of the linear regression used were the intercept (b_0), the slope (b_1), the standard deviation of the response at zero level content (s_0), the estimate of the residual standard deviation (s) and the mean content of the n standards ($\bar{c}$).

$t_{1-\alpha,n-2}$ was the statistical value for a one-sided Student-t distribution with n–2 degrees of freedom and a probability α, and N were the number of repetitions of the future sample. Analogously, the limit of detection was expressed as

$$c_D = \frac{r_C - b_0}{b_1} + t_{1-\beta,n-2} \cdot \frac{s}{b_1} \cdot \sqrt{\frac{1}{N} + \frac{1}{n} + \frac{(c_D - \bar{c})^2}{\sum (c_i - \bar{c})^2}} =$$

$$= t_{1-\alpha,n-2} \cdot \frac{s}{b_1} \cdot \sqrt{\frac{1}{N} + \frac{1}{n} + \frac{\bar{c}^2}{\sum (c_i - \bar{c})^2}} + t_{1-\beta,n-2} \cdot \frac{s}{b_1} \cdot \sqrt{\frac{1}{N} + \frac{1}{n} + \frac{(c_D - \bar{c})^2}{\sum (c_i - \bar{c})^2}}$$

Eq. 32

where $t_{1-\beta,n-2}$ was the statistical value for a Student-t distribution with n–2 degrees of freedom and a one-sided probability β. This expression can be resolved by an iterative procedure or by directly using the formulas developed by Garner and Robertson in 1988:

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

$$c_D = u^{-1} \cdot \left[b_1(r_C - b_0) - \bar{c}V^2 + V\sqrt{(\bar{r} - r_C)^2 + u(SSX)\left(\frac{1}{N} + \frac{1}{n}\right)} \right] \quad \text{Eq. 33}$$

with $u = b_1^2 - V^2$, $V = t_{1-\beta,n-2} \cdot s/\sqrt{SSX}$, and $SSX = \sum (c_i - \bar{c})^2$.

Hubaux and Vos also indicated several ways of improving the decision and detection limits since they were directly related to the precision of the method (*s*). Of course, the number of standards, *n*, and the number of replication on the unknown, *N*, have strong influence (see Equations 31 and 32). However, they also studied the effect of the disposition of standards and the range of their contents. For example, in order to enhance sensitivity they found that the best arrangement of standards was for some of them to have the smallest admissible content and others to have the largest permissible content.

Table 4. International recommendations and nomenclature on analytical detection and quantification concepts.

RECOMMENDED TERM (and alternative)	Concept	Defining relations
CRITICAL VALUE, L_C (Critical Limit)	To distinguish a chemical signal from background noise.	$\Pr(\hat{L} > L_C \mid L = 0) \leq \alpha$ $L_C = z_\alpha \cdot \sigma_0$ $L_C = t_{1-\alpha,n-2} \cdot s_0$ (1)
MINIMUM DETECTABLE (TRUE) VALUE, L_D (Detection Limit)	Measure of the inherent detection capability of a chemical measurement process	$\Pr(\hat{L} \leq L_C \mid L = L_D) = \beta$ $L_D = z_\alpha \cdot \sigma_0 + z_\beta \cdot \sigma_D$ $L_D = \Delta_{(\alpha,\beta)} \cdot s_0$ (2)
MINIMUM QUANTIFICABLE (TRUE) VALUE, L_Q (Quantification Limit)	Measure of the inherent quantification capability of a chemical measurement process	$L_Q = k_Q \sigma_Q$ where $k_Q = 1/RSD_Q$

(1) When σ_0 is estimated by s_0, based on ν degrees of freedom, *z* must be replaced by a Student's-*t*.
(2) When $\sigma_0 = \sigma_D$ and it is estimated by s_0, based on ν degrees of freedom, $(z_\alpha + z_\beta)$ must be replaced by $\Delta_{(\alpha,\beta)}$, the non-centrality parameter of the non-central *t* distribution. For $\alpha = \beta$ and large ν, this parameter can be safely approached by 2*t*.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

This way of calculating the limit of detection from the regression line and its confidence intervals was (and still is) a widely accepted method. However, for decades, the evolution of the concepts of detection and quantification limits was characterized by differences in criteria and terminology. In 1990, the World Health Organization[23] formally expressed the need for homogeneous and international guidelines. ISO and IUPAC produced them in 1995. These organizations published separate documents (see above) based on the same formulation and nomenclature. These documents considered detection and quantification capabilities as fundamental performance characteristics of the chemical measurement process (CMP) derived from the statistical theory of hypothesis testing and the probabilities of Type I and Type II errors. These concepts and formulas are summarised in Table 4.

2.3.3.4. Limit of detection in multivariate calibration

The formulas developed for estimating the detection and quantification limits for zeroth-order data and univariate calibration essentially assume that the signal is highly selective for the analyte of interest. In complex samples, where selectivity is not achieved by improving the method either by experimental or instrumental means, multivariate calibration can be a powerful mathematical solution.

An early discussion of the need for an accurate multivariate limit of detection estimator was carried out by Garner and Robertson [Ref. 1998]. The theoretical development of the concept of detection limit in multicomponent systems, as well as the theoretical approaches and the hypothesis on which they are based, were critically reviewed by Boqué and Rius [Ref. 1996], van der Voet [Ref. 2002] and, more recently, by Olivieri *et al*. [Ref. 2006]. The most illustrative types of estimators are discussed below.

As explained above, multivariate calibration methods aim to construct a valid predictive model on the basis of (possibly highly) unselective multiple predictor variables (signal, response, etc.). The concept of net analyte signal (NAS) in multivariate calibration arises from the fact that a prediction sample response,

[23] *Codex Alimentarius Commision*, committee on Methods of Analysis and Sampling, Food and Agricultural Organization (FAO) of The United Nations and World Health Organization (WHO), URL: http://www.codexalimentarius.net

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

e.g. a spectrum, may have different contributions from other sample components. The spectrum can be decomposed into two orthogonal parts: one that can be uniquely assigned to the analyte of interest (the NAS) and one that contains the (possibly different) contribution from other components. The NAS is therefore proportional to the concentration of the analyte of interest and it is more important than the total signal (Booksh and Kowalski, 1994).

Many of the figures of merit that characterize multivariate methods are closely related to the concept of multivariate NAS as introduced by Lorber [Ref. 1986]. By considering the NAS, one can define error propagation[24], the multivariate detection limit and other figures of merit for each component.

Lorber (*op. cit.*) was one of the first researchers to calculate a multivariate LOD starting from the definition of the net analyte signal. Subsequently, Lorber *et al.* ([Ref. 1997] and Lorber and Kowalski, 1998) defined the multivariate detection limit estimator as a function of the confidence intervals associated with the predicted concentration by checking whether the predicted concentration and its confidence intervals for a given test sample included the zero concentration value. The expression developed for the multivariate LOD on the signal domain took into account only the uncertainty in the signal measurements, which meant that its real application was rather limited.

Several approaches to estimating the limit of detection apply an error-propagation-based formula for standard error of prediction to a zero concentration level. This was the approach used by Bauer *et al.* [Ref. 1991] and Faber and Kowalski [Ref. 1997] to calculate the multivariate LOD in the concentration domain for the calibration using the classical model, and by Boqué *et al.* [Ref. 1999] for the commonly used inverse calibration models. The work reviewed by Boqué (*op. cit.*) demonstrated the utility of error-propagation for estimating concentration uncertainties and limits of detection. The formula selected accounted for all sources of errors in the data (signals and

[24] The error propagation for a given component is the ratio *f* of the precision in concentration to the precision in the instrumental response. This value is an additional figure of merit that is concentration-independent and accounts for the possibility of dismissing this component with others because it has either the same instrumental response, e.g. spectrum, as another component or is a linear combination of the spectra of the others. This figure of merit is identified as the inverse of selectivity.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

concentrations) of calibration and prediction samples, and the resulting multivariate LOD was able to quantify the probabilities of committing type I and type II errors and was sample-specific (i.e. dependent on the level of interferences in the sample). This approach is used in Chapter 6 to calculate the LOD for multi-linear regression (N-PLS).

Typical method-specific approaches for higher-order data can be found in the literature. One example was developed by Boqué *et al*. [Ref. 2002] for estimating the limit of detection in second-order bilinear calibration with the generalized rank annihilation method (GRAM). A generally applicable non-parametric approach is to train a neural classifier to optimize the false positive and the false negative error rates for a fixed threshold concentration (Sarabia *et al*., 2002).

Another strategy for estimating multivariate limits of detection is to transform the multivariate model into a univariate one. Several authors have suggested substituting the multivariate signal by a variable that is directly related to the concentration of the analyte, e.g. the height or the area of a chromatographic peak for the analyte of interest in the calibration samples. A univariate regression line (signal *vs* analyte concentration) can then be built and figures of merit such as accuracy, sensitivity and limit of detection can be derived. This approach can be extended to higher-order data once the first-order profiles have been resolved. For instance, when spectro-chromatographic (second-order) data are obtained for a set of calibration samples, they can be decomposed into the pure first-order profiles by a proper second-order method (MCR-ALS, PARAFAC, etc.). For each calibration sample, the chromatographic peak of the analyte of interest is obtained and the height (or the area) of these peaks can be regressed against the corresponding reference concentrations (see Figure 16 for a graphical example).

The literature contains a few examples of the application of a “surrogate” signal to deal with the calculation of multivariate LODs. In the context of multivariate curve resolution, Saurina *et al*. [Ref. 2001] proposed using the areas of the recovered pure signal profiles of the analyte, which do not depend on the signal of interferent compounds. Ortiz *et al*. [Ref. 2003] applied a similar approach for multivariate and multi-linear PLS calibration. In this case, the surrogate variable is the concentration of the analyte in the calibration samples predicted by the optimal PLS model.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

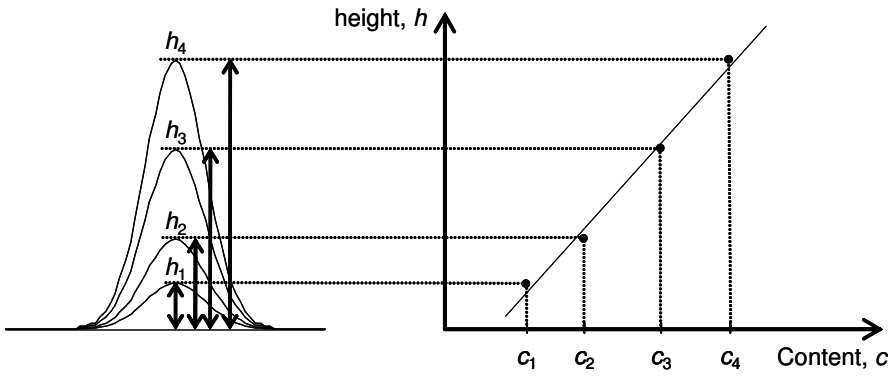


Figure 16. A univariate regression line built from the height of the peak of the analyte of interest in several calibration samples. From this regression line, figures of merit can be calculated using univariate statistics.

In this thesis, we used the transformation approach for several applications:

- In Chapter 3 we determined some pesticides by excitation-emission matrix (EEM) fluorescence in combination with PARAFAC. Decomposition of the three-way data by PARAFAC produced three loading matrices, one of which corresponded to the sample mode, i.e. the relative concentrations of the pesticides in the mixtures. In the calibration step, these loadings were regressed against the real concentrations of each pesticide in the mixtures to obtain the linear calibration line.

- In Chapter 4 we decomposed second-order data matrices of a set of standards into estimated chromatograms (and spectra) with the ITTFA algorithm. The height of the peaks in these chromatograms was plotted against concentration and classic estimators of LOD for univariate linear calibration were used.

- In Chapter 5 the resolved the second-order data matrices obtained by HPLC-DAD for a set of multicomponent groundwater samples by MCR-ALS, which provided the pure spectrum and chromatogram of each pesticide present in the samples. The relative areas estimated by the curve resolution method

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

(the areas under the estimated chromatograms) were used as the surrogate variable to perform the calibration line of areas *vs* concentration.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 3

ESTIMATION OF THE LIMIT OF DETECTION FOR PARALLEL FACTOR ANALYSIS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

3.1. Introduction

To put into practice the knowledge we have acquired on the estimation of detection limits, we began looking for an analytical problem involving complex samples. We consolidated our relationship with researchers in the Department of Hydrogeology and Analytical Chemistry at the University of Almería, who analysed organic microcontaminants (pesticides) in air, water, vegetables and biological samples (blood, urine, fat tissues, milk, etc). They also produced multicomponent data sets for which quantification and even identification was a challenge. The analytical measurements were taken with an HPLC-DAD (Chapter 5) in one of these joint applications and with an EEM fluorometer (described in this chapter) in the other. Both types of measurements were (or were assumed to be) bilinear, which conditioned the chemometric procedure used.

The linearity assumed in many models simplifies the decomposition of the data. Even when the data are not strictly linear, this assumption can be used for practical reasons. When we deal with high-order data, linearity should be assessed in each order since, depending on the nature of the data, it can be maintained in some data but lost in others. Three-way excitation-emission matrices from EEM fluorescence analyses can be assumed to be trilinear if we do not record (or reject after they have been recorded) the scattered data due to the Raman and the Rayleigh interferences.

In this chapter we present a published paper entitled "*Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis*" in which we estimate the limit of detection for type of data, which are *approximately* trilinear.

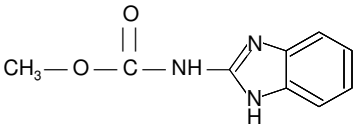
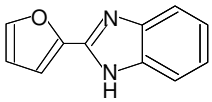
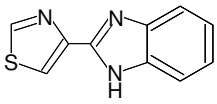
The samples under study were complex mixtures of three polluting compounds (carbendazim, fuberidazole and thiabendazole[49]) whose concentrations are regulated, for instance, in drinking waters. The molecular structure of these

[49] These pesticides are used as fungicides. Thiabendazole is also a parasiticide that is primarily used to control mold, blight and other fungally caused diseases in fruits and vegetables.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

pesticides is shown in Table 5:

Table 5. Molecular structure of the pesticides under study.

Commercial name	Chemical name	Structure
Carbendazim	methyl 2-benzimidazolecarbamate CAS*: 10605-21-7	
Fuberidazole	2-(2-furyl)benzimidazole CAS*: 3878-19-1	
Thiabendazole	2-(4'-thiazolyl)benzimidazole CAS*: 148-79-8	

* See footnote[50]

In the calibration step, we paid special attention to the criteria for selecting the calibration standards. The three-way calibration data was decomposed using PARAFAC and the LOD and other figures of merit were estimated by the transformation approach, i.e. from the univariate calibration line of model loadings versus the concentration of the standards.

Two prediction strategies were used to test the accuracy of the method. In the first one, the PARAFAC model was performed with the standards of known concentration and the univariate calibration line was formulated with their loadings. The loadings of the new samples were mathematically calculated from the parameters of the model and then interpolated or extrapolated in the regression line.

[50] CAS registry numbers are unique numerical identifiers for chemical compounds, polymers, biological sequences, mixtures and alloys. The *Chemical Abstracts Service* (CAS), a division of the *American Chemical Society*, assigns these identifiers to every chemical that has been described in the literature. As of February 2, 2006, there were 27281529 substances in the CAS registry. About 4000 new numbers are added each day. The current number is published at: http://www.cas.org/cgi-bin/regreport.pl

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

In the second strategy, both calibration and prediction samples were used to perform the model. Only the loadings of the calibration samples were regressed against the known concentrations and then the loadings of the prediction samples were used to predict their concentration.

The main conclusions of this paper are as follows:

- The model had to cover the experimental domain, so standards have preferably to be taken at the ends of the domain to avoid subsequent extrapolation.

- Information about the selectivity and sensitivity of the compounds may be useful for selecting the calibration samples. Including standards of the most sensitive or least selective compound of the mixtures may improve the model.

- Two prediction strategies are feasible for PARAFAC. When the model is only performed with the calibration set, further calculation is needed to obtain the loadings of the prediction samples. However, we can detect samples of a different nature or outliers because their loadings will be outside the experimental domain. When the prediction set is included in the model, the recovered profiles may correlate worse with the reference profiles, since the model is forced to explain all the variations in the calibration and the prediction samples. Also, outlying samples cannot be detected by this strategy.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

3.2. Paper. Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis.

Analytica Chimica Acta 491: 47-57 (2003)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation-emission matrix fluorescence and parallel factor analysis

M.J. Rodríguez-Cuesta, Ricard Boqué, F. Xavier Rius

Department of Analytical and Organic Chemistry, Rovira I Virgili University

Pça. Imperial Tàrraco,1. 43005-Tarragona (Catalunya). SPAIN

D. Picón Zamora, M. Martínez Galera, A. Garrido Frenich

Department of Hydrogeology and Analytical Chemistry, University of Almeria

04071-Almería (Andalucía). SPAIN

Abstract

We simultaneously determined carbendazim, fuberidazole and thiabendazole by excitation–emission matrix (EEM) fluorescence in combination with parallel factor analysis (PARAFAC). Three-way deconvolution provided the pure analyte spectra from which we estimated the selectivity and sensitivity of the pesticides, and the relative concentration in the mixtures from which we established a linear calibration. Special attention was given calculating such figures of merit as precision, sensitivity and limit of detection (LOD), derived from the univariate calibration curve. The method, which had a relative precision of around 2–3% for the three pesticides, provided limits of detection of 20 ng ml^{-1} for carbendazim, 4.7 ng ml^{-1} for thiabendazole and 0.15 ng ml^{-1} for fuberidazole. The accuracy of the method, evaluated through the root mean square error of prediction (RMSEP), was 27.5, 1.4, and 0.03 ng ml^{-1}, respectively, for each of the pesticides.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

1. Introduction

Pesticides are polluting compounds whose concentration is regulated by the European Commission in many samples such as drinking waters. Traditionally, instrumental techniques to determine these compounds involve gas (EPA method 515.1) or liquid chromatography (EPA method 531.1, 632 and 8318). Fluorimetric techniques can also be used to analyse pesticides in mixtures since many pesticides, including the ones studied in this paper, are intrinsically fluorescent. Fluorescence spectroscopy is a versatile analytical technique, which provides high sensitivity of detection. However, in multicomponent mixtures the fluorescence signal is normally overlapped and chemical procedures (in few cases) or chemometrical techniques of resolution have to be applied if it is to be quantitatively analysed. One of the mathematical resolution strategies is the collection of an entire excitation–emission matrix (EEM) fluorescence spectrum combined with multi-way deconvolution and calibration algorithms. In this way, Picón Zamora et al. [1] determined three pesticides by taking linear trajectories across the EEM and applying principal component regression (PCR) and partial least squares (PLS1 and PL2) algorithms; Saurina et al. [2] resolved the EEM of triphenyltin in synthetic and natural sea water samples with multivariate curve resolution (MCR); and JiJi et al. [3] determined carbamate pesticides by parallel factor analysis (PARAFAC) of the excitation–emission matrix. We applied a procedure similar to this latter one, and paid particular attention to the internal validation of the method and to the calculation of the figures of merit.

We determined carbendazim, fuberidazole and thiabendazole in mixtures of the three pesticides by PARAFAC deconvolution of the three-dimensional excitation–emission data. Internal validation was assessed from the correlation between the pure excitation and emission spectra of each compound, and the reference spectra, as a measure of the reliability of the model. The selectivities of the three pesticides were calculated from the recovered spectra. Next, a univariate regression was performed for each pesticide by relating the loadings of the PARAFAC decomposition with the known concentrations of the pesticides in the calibration samples. From these univariate calibration lines, figures of merit such as precision, sensitivity and limit of detection (LOD) were calculated. The effect of the spectral selectivities on these figures of merit is also discussed. Finally, the accuracy of the method was estimated by predicting a new set of samples, which were not included in the calibration step.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

2. Theoretical background

2.1. Three-way decomposition of fluorescence data

Fluorescence three-way data can be decomposed by parallel factor analysis because each analyte in the sample can ideally be described by one PARAFAC component [4]. This means that each fluorophore's contribution to the emission is independent of the contribution of the remaining fluorophores and identical for different samples (only varying in proportions). Hence, the PARAFAC model for a three-way array (r_{ijk}) can be denoted as:

$$r_{ijk} = \sum_{f=1}^{F} a_{if} \cdot b_{jf} \cdot c_{kf} + e_{ijk} \qquad \text{Eq. 34 (as Eq. 7)}$$

In the case of excitation–emission matrix fluorometry, the kth slice of the trilinear cube $\underline{\mathbf{R}}$ is the (I×J) matrix of excitation and emission profiles of the fluorescent components for the kth sample. Thus, a_{if}, b_{jf} and c_{kf} are the typical elements of the loading matrices **A**, **B** and **C** (emission wavelength, excitation wavelength and relative concentrations in the samples, respectively) for a given number of components F. Using an alternate least squares (ALS) procedure, the trilinear model is found to minimise the sum of squares of the residuals e_{ijk}. In matrix notation, and using the Khatri–Rao product [5], the PARAFAC model can be formulated in terms of the unfolded array as shown in Equation 35:

$$\underbrace{\mathbf{R}_k}_{(IJ\times K)} = (\underbrace{\mathbf{B}}_{(J\times F)} \otimes \underbrace{\mathbf{A}}_{(I\times F)}) \cdot \underbrace{\mathbf{C}}_{(K\times F)}^{\mathbf{T}} = \underbrace{\mathbf{Z}}_{(IJ\times F)} \cdot \mathbf{C}^{\mathbf{T}} \qquad \text{Eq. 35}$$

An important characteristic of the PARAFAC model is the uniqueness of its solution. This means that additional constraints, such as orthogonality or external information to solve rotational freedom are not needed to identify the model [6]. This property is an extension of the second-order advantage and so trilinear data (here fluorescence data) can be calibrated when there are unknown interferences in the samples.

In dilute solutions or suspensions, fluorescence intensity is linearly proportional to the solute concentration, and fluorescent excitation–emission measurements follow a trilinear model, such as the PARAFAC [3, 5, 7, 8]. However, spectral

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

properties are affected by the local environment. The most common environmental factors that influence fluorescence properties are solvent polarity, pH and fluorescence quenching (*quenching* is any process which decreases the fluorescence intensity of a sample, e.g. excited-state reactions, molecular rearrangements, energy transfer, ground–state complex formation and collision quenching, such as the one produced by molecular oxygen) [9].

Despite this environmental influence, fluorescence measurements can still fulfil the trilinear model if we keep the conditions constant throughout the experiments. However, other problems cannot be so effectively handled by PARAFAC based models. One of these is the emission region below excitation, where the intensity is approximately zero, i.e. the fluorophore shows no fluorescence. Whether these data are recorded or not, they should be treated as missing values and cannot be replaced with zeros to prevent the PARAFAC model from trying to fit them [5,10]. Other potential problems are Rayleigh and Raman scattering. Raman interference due to the solvent can often be almost completely removed by subtracting the solvent spectra from the sample spectra. Rayleigh scattering occurs in the EEM when the excitation wavelength is equal to the emission wavelength and there are no intrinsic profiles in either the *X*- or *Y*-order to extract. Several strategies for solving this problem have been described in the literature (e.g. data analysis can be restricted to regions were the scattering does not appear, a blank spectra can be subtracted if available or the data points can be weighed) [3,11].

2.2. PARAFAC calibration and prediction

The decomposition of the three-way data by PARAFAC gives rise to three loading matrices, one of which, **C**, corresponds to the sample mode. The **C**-loadings are the relative concentrations of the pesticides in the mixtures. In the calibration step, these loadings are regressed against the real concentrations of each pesticide in the mixtures to get a linear calibration line [2]. In the prediction step, this regression line can then be used to predict (if any new interferent is present) the concentration of each pesticide in future test samples, $\mathbf{R}_{un}$ that are not in the initial calibration dataset, by interpolating their loadings of relative concentration, $\mathbf{C}^T$. These loadings can be previously calculated from the Equation 35 multiplying the pseudoinverse of the **Z** matrix by the test sample data, as shows Equation 36:

$$\mathbf{C}^{\mathbf{T}} = (\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}} \cdot \mathbf{R}_{un} \qquad \text{Eq. 36}$$

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Another way to predict future test samples is to include them in the initial PARAFAC model. In this way, the loading matrices for both the calibration and the prediction sets are recovered. All the samples are considered to calculate the model parameters although the regression fit is only performed with the calibration samples. Finally, the loadings of the PARAFAC model for the prediction samples are interpolated into the corresponding regression line to obtain the predicted concentration of each analyte.

3. Experimental

3.1. Samples and standards

To determine the pesticides in synthetic samples, a set of 12 mixtures was prepared in methanol between 0 and 100 ng ml^{-1} for carbendazim, 0–0.7 ng ml^{-1} for fuberidazole and 0–40 ng ml^{-1} for thiabendazole. A standard of each pure pesticide was also prepared (Table 6).

Excitation–emission fluorescence matrices were recorded for all the standards. In each experiment, a methanol blank was subtracted to remove the interfering Raman effect of the solvent.

Table 6. Concentration of the pesticides in the synthetic mixtures and individual standards.

Sample	Carbendazim (ng ml^{-1})	Thiabendazole (ng ml^{-1})	Fuberidazole (ng ml^{-1})
M1	0	20	0.4
M2	50	20	0.0
M3	50	0	0.4
M4	30	30	0.5
M5	60	35	0.6
M6	20	15	0.1
M7	100	25	0.2
M8	90	35	0.4
M9	40	25	0.4
M10	60	20	0.5
M11	80	40	0.1
M12	90	35	0.3
Carbendazim	75	0	0.0
Fuberidazole	0	0	0.7
Thiabendazole	0	35	0.0

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

3.2. Instrumentation and data analysis

Measurements were performed with an Aminco–Bowman Series 2 luminescence spectrometer equipped with a 150W continuous xenon lamp. The EEM were defined so that they would collect the variation in the signal caused by the pesticides, and not record either excitation signals below emission or Rayleigh scattering. The dimension of the data matrices was 50×38, from 310 to 370 nm in the emission domain and from 260 to 306 nm in the excitation domain. The excitation and emission slits were both maintained at 4 nm and the scanning rate was 7 nm s^{-1}. All measurements were performed in a 10 mm quartz cell at 750 V. An AB2 software version 1.40, running under OS/2 2.0 was used for spectral acquisition and MATLAB 6.0 (The MathWorks Inc., 2000) was used for data analysis. In the Matlab environment, commercial [12] and home-made algorithms were used to process the data.

4. Results and discussion

4.1. Preliminary study of the data

A preliminary principal component analysis (PCA) can provide information about the degree of correlation of the data and the presence of outliers or influential samples. So we unfolded column-wise the EEM of each sample to a row vector of 1900 elements, and use them to build a new data matrix of 12 rows (samples) and 1900 columns (wavelengths). The dimensionality of the data was reduced with the unfold-PCA to two factors, which explained 97% of the variation in the data. In spite of there being three components, the high correlation between the fluorescence spectra explains the presence of only two significant principal components. Figure 17 shows the PC1–PC2 scores plot (explained variance is 86 and 11%, respectively) and the influence plot (squared residuals on *X* versus leverage). We can observe that sample M12 is far from the centre in the scores plot and has a high leverage value, which indicates that it is an influential sample.
A PARAFAC model with three factors was also performed with the whole set of mixtures to get an idea of the distribution of the samples. We represented the loadings of the sample mode in two dimensions and observed that M12 was again the most different sample, according to the U-PCA results.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

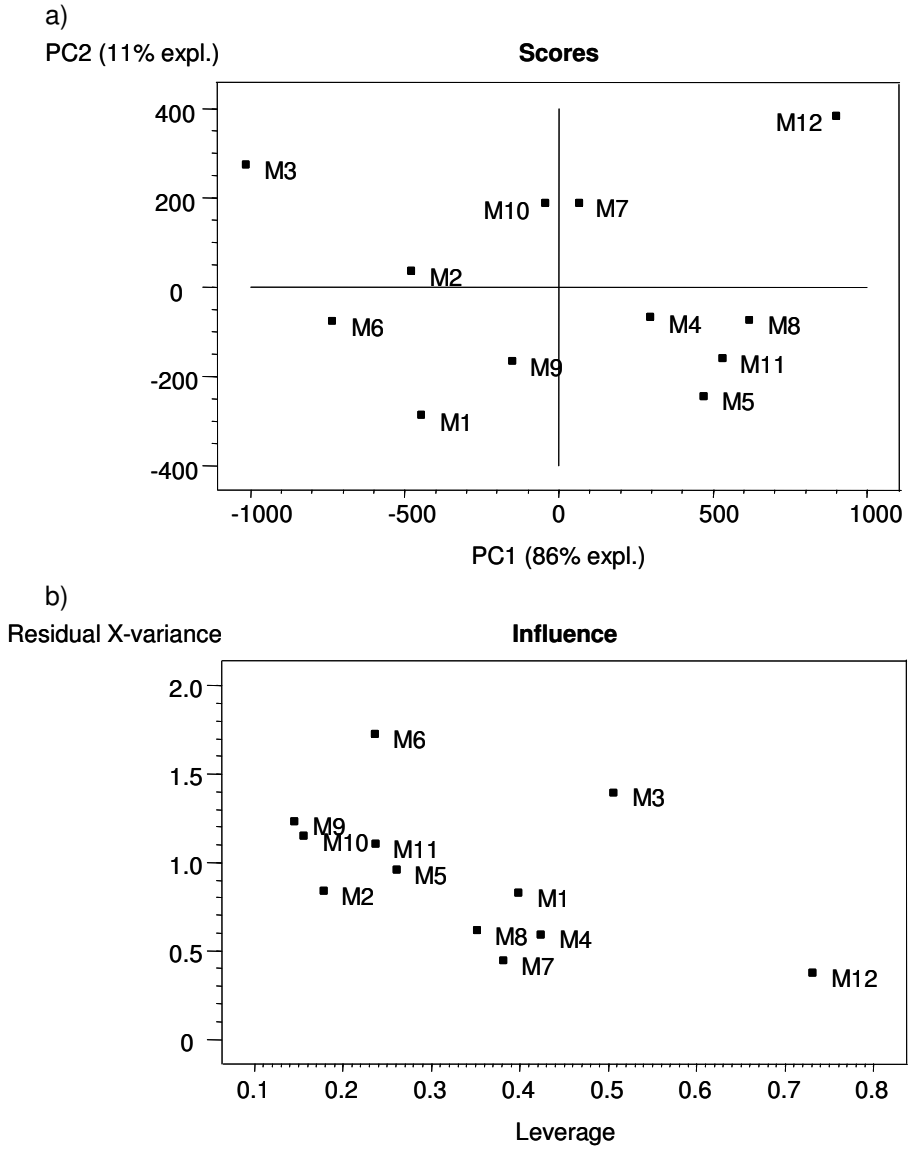


Figure 17. Scores (a) and influence (b) plot from the PCA of the columnwise unfolded matrices corresponding to the 12 synthetic mixtures (explained variance 97%).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

4.2. Individual standards

Figure 18 shows the excitation and emission spectra obtained for each pesticide. They were obtained by applying PARAFAC with one factor and non-negativity constraints to the three-dimensional matrices for the individual fluorescence spectra of each pesticide. We regarded these normalised profiles as the spectra of the *pure* pesticides and used them as reference spectra to evaluate the reliability of the models in the calibration.

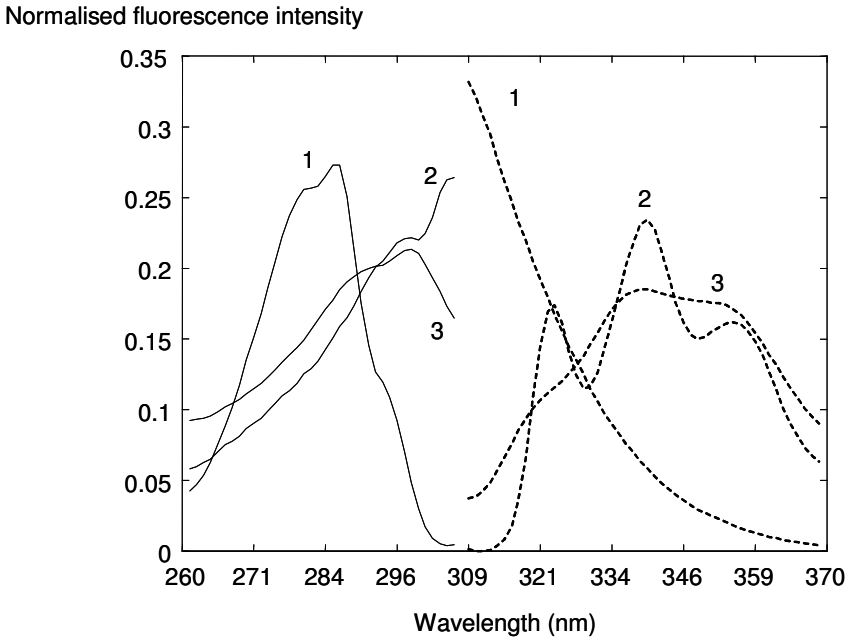


Figure 18. Excitation (solid lines) and emission (broken lines) spectra of (1) carbendazim (50 ng ml^{-1}), (2) fuberidazole (0.7 ng ml^{-1}) and (3) thiabendazole (35 ng ml^{-1}).

In a preliminary study [1], some of the authors recorded the pure excitation and emission spectra at the wavelengths of maximum emission and excitation, respectively. Correlation between those spectra and the ones in this paper, resolved by PARAFAC on the EEM of the individual pesticides, was higher than 0.998.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

We calculated the sensitivity and selectivity of the pesticides from their first-order profiles in both the excitation and the emission region (see Table 7). All figures of merit were calculated as described by Faber et al. [13]. In terms of first-order profiles, fuberidazole was the most sensitive compound, both in the excitation and the emission region, but one of the least selective, together with thiabendazole. The most selective pesticide was carbendazim, as its spectra are the most different in shape (see Figure 18) and therefore the least correlated. This preliminary information suggests that fuberidazole will be predicted at lower concentrations (it is the most sensitive) but with a higher error of prediction because it is highly correlated with thiabendazole. It also suggests that carbendazim will be the most accurately predicted because of its selectivity.

Table 7. First-order sensitivities and selectivities of carbendazim, thiabendazole and fuberidazole, estimated from the standards of the pure pesticides[a].

	Compound	Sensitivity	Selectivity
X (50×3) Emission spectra	Carbendazim	15.45	0.8068
	Thiabendazole	19.19	0.1976
	Fuberidazole	356.5	0.2073
Y (38×3) Excitation spectra	Carbendazim	7.635	0.3985
	Thiabendazole	11.23	0.1157
	Fuberidazole	240.9	0.1401

[a] Figures of merit for first-order spectral profiles

4.3. Calibration step

In the calibration step, we performed several PARAFAC models, which consisted of different calibration samples, and not necessarily the individual standards of each pesticide.

The models that included the individual standard of carbendazim in the calibration set recovered the profiles slightly better than the models that did not. The same occurred with the standard of thiabendazole. However, the presence of the standard of fuberidazole considerably improved the recovery of the profiles and, therefore, the correlation between them and the reference spectra. In the models which did not contain fuberidazole, on the other hand, the profiles of each pesticide could not be unequivocally identified. This is due to the sensitivity and

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

selectivity values of the compounds. When mixtures of three analytes were analysed by fluorescence, and the data were modelled by PARAFAC, the standard of the most selective and sensitive analytes did not have to be included, because the model was able to recover their profiles in the presence of other analytes. This was the case of carbendazim. However, thiabendazole and fuberidazole are less selective and highly correlated each other, so the model needs extra information (such as the individual standard) about the most sensitive of them (fuberidazole, in this case), so that the profiles can be reliably recovered.

The best calibration set consisted of the calibration samples located at the extremes of the domain, i.e. M2, M5, M7 and M11 (extrapolation out of the linear range is avoided) together with the individual standard of fuberidazole (see above). This PARAFAC model was built with three factors and non-negativity constraints in all the modes. The estimated profiles matched the reference spectra (see Figure 19), with correlation coefficients of 0.996 for carbendazim, 0.998 for thiabendazole and 1.000 for fuberidazole in the excitation region and 0.9995, 0.996 and 1.000, respectively, in the emission region.

Table 8 shows the figures of merit of this model. They were calculated, using univariate statistics, from the calibration line fitted with the loadings in the sample mode obtained with PARAFAC, as explained above.

Sensitivity was defined as the slope of the calibration curve. The values of this sensitivity measure correlate well with those calculated from the net analyte signal of the first-order spectra of each pesticide. Other authors have suggested a single measure of sensitivity for PARAFAC models based on net analyte signal calculations [14]. From the results of Table 8 we can conclude that fuberidazole is the most sensitive compound in the mixture, followed by thiabendazole and carbendazim, in agreement with the spectroscopic data from Table 7.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

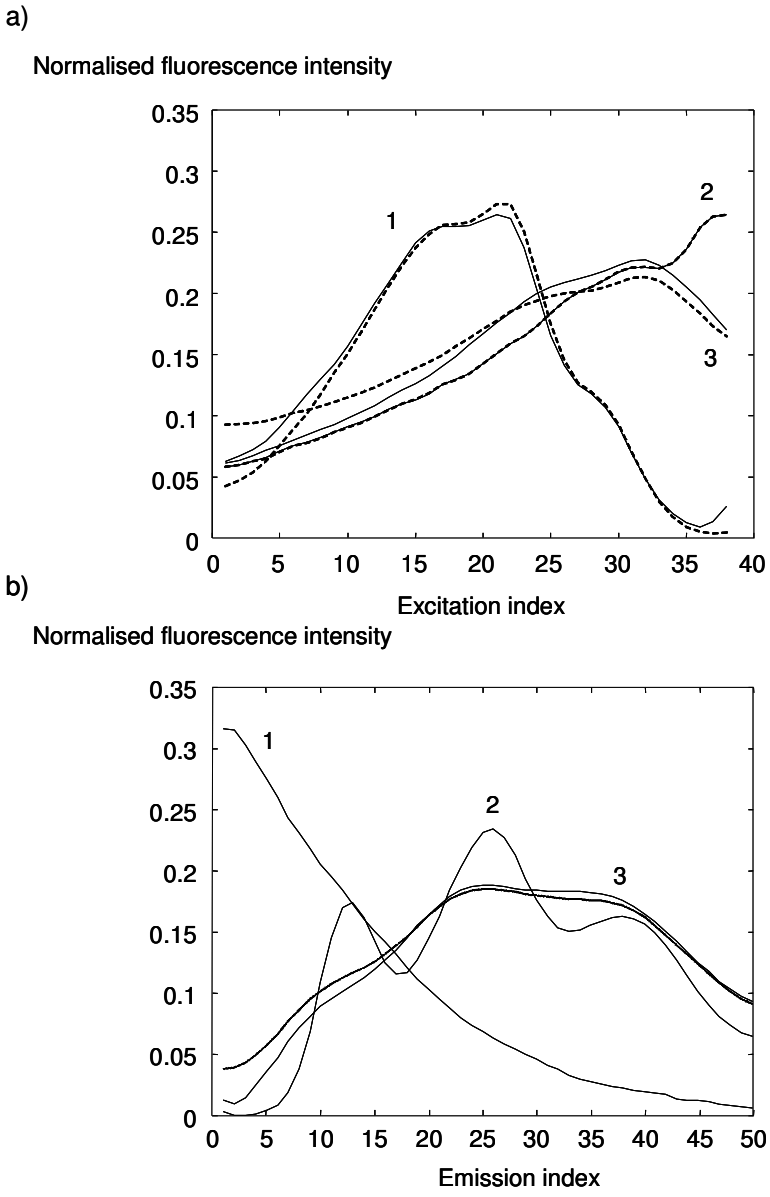


Figure 19. (a) Excitation and (b) emission spectra for carbendazim (1), fuberidazole (2) and thiabendazole (3). Broken lines: reference spectra; solid lines: spectra recovered by the PARAFAC model.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 8. Statistical parameters and figures of merit of the linear relationship between the proportion loadings calculated by PARAFAC and the true concentration of each pesticide.

	Carbendazim	**Thiabendazole**	**Fuberidazole**
Number of data points	5	5	4
Intercept	9.34×10^{-3}	-9.37×10^{-3}	-3.78×10^{-3}
Sensitivity (slope)	6.54×10^{-3}	16.4×10^{-3}	349.3×10^{-3}
Standard deviation of intercept	19.9×10^{-3}	11.9×10^{-3}	5.80×10^{-3}
Standard deviation of slope	2.96×10^{-4}	4.27×10^{-4}	181.3×10^{-4}
Standard error	22.3×10^{-3}	13.3×10^{-3}	8.26×10^{-3}
Correlation coefficient (r)	0.9969	0.9990	0.9973
Precision (%)	3.4	2.0	3.4
Limit of detection ($ng \cdot ml^{-1}$)	20	4.7	0.15

The precision for each pesticide was calculated in terms of concentration as the standard deviation of the *C*-loading residuals for all standards divided by the sensitivity, *SEN*, i.e. the slope of the calibration line as:

$$Precision = \frac{s_{res}}{SEN} = SEN^{-1}\sqrt{\frac{\sum (c_i - \hat{c}_i)^2}{n-2}} \qquad \text{Eq. 37}$$

c_i are the *C*-loadings for the given pesticide estimated from the PARAFAC model and $\hat{c}_i$ are the loadings estimated from the calibration line *C*-loadings versus pesticide concentration. *n* is the number of calibration standards.

Finally, the LOD for each pesticide was estimated from Equation 38, which takes into account the uncertainty of the calibration line and considers α and β probabilities of error, following the IUPAC recommendations [15]:

$$LOD = \delta(\alpha,\beta)\frac{s_{res}}{SEN}\sqrt{1+\frac{1}{n}+\frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} \qquad \text{Eq. 38}$$

x_i is the concentration of the given pesticide in each of the *n* calibration standard and $\bar{x}$ is the average concentration.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

The values of the correlation coefficients indicate the quality of the linear fits and the estimated precision shows that the calibration results are in close agreement. The detection limits of this method are in the order of magnitude of nanograms milliliters^{-1} (ng ml^{-1}). Expressed as a percentage of the higher analyte concentration in samples, they were around 20% for carbendazim and fuberidazole and below 12% for thiabendazole.

In order to test the second calibration strategy described in Section 2, we also performed the PARAFAC model with the chosen samples M2, M5, M7, M11, the standard of fuberidazole and the remaining samples to be predicted. We obtained the loadings of relative concentration for each sample, although the calibration line was fitted only with the calibration set. Table 9 shows the statistical parameters obtained for this regression. The estimated profiles were correctly recovered but the correlation coefficients with the reference spectra were slightly lower than those obtained with the model that did not include the prediction samples. This second model finds the solution that best explains all the variations, so we could not detect any outlying samples. Precision and limits of detection were of the same order for thiabendazole and higher for carbendazim and fuberidazole than with the previous model.

Table 9. Statistical parameters and figures of merit of the linear relationship between the loadings of calibration samples obtained from a PARAFAC that includes the prediction samples (see details in text).

	Carbendazim	Thiabendazole	Fuberidazole
Number of data points	5	5	4
Intercept	$2.01 \cdot 10^{-3}$	$-4.45 \cdot 10^{-3}$	$-7.37 \cdot 10^{-4}$
Sensitivity (slope)	$42{,}8 \cdot 10^{-3}$	$10.45 \cdot 10^{-3}$	$326.9 \cdot 10^{-3}$
Standard deviation of intercept	$14.1 \cdot 10^{-3}$	$70.7 \cdot 10^{-3}$	$103.1 \cdot 10^{-3}$
Standard deviation of slope	$2.10 \cdot 10^{-4}$	$2.55 \cdot 10^{-3}$	$32.2 \cdot 10^{-3}$
Standard error	$15.9 \cdot 10^{-3}$	$7.94 \cdot 10^{-3}$	$14.7 \cdot 10^{-3}$
Correlation coefficient (r)	0.9964	0.9991	0.9905
Precision (%)	3.7	1.9	6.4
Limit of detection (ng·ml^{-1})	22	4.4	0.29

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

4.4. Prediction step

We used the PARAFAC model built with the calibration set [M2, M5, M7, M11, Fub] to predict the concentration of carbendazim, thiabendazole and fuberidazole in the prediction set consisting of the remaining samples in Table 6. For each pesticide, we interpolated the estimated PARAFAC *C*-loadings, calculated from Equation 36, in the corresponding regression line (see Table 8) and calculated the predicted concentration.

Figure 20 shows the calibration model with a straight line, and the prediction samples indicated with crosses. The prediction results for thiabendazole and fuberidazole are very good.

The accuracy of the models was calculated by the root mean square error of prediction (RMSEP):

$$RMSEP = \sqrt{\frac{\sum (x_i - \hat{x}_i)^2}{M}} \qquad \text{Eq. 39}$$

where x_i and $\hat{x}_i$ are the measured and predicted concentrations of the given pesticide in each prediction sample, and M is the total number of prediction samples. RMSEP was 3.4% for thiabendazole and 3.9% for fuberidazole. However, in the prediction of carbendazim, samples M4, M10 and M12 behaved very differently from the model and the rest of the samples and increased the RMSEP from 5.6 to 27.5%. Work is in progress to efficiently detect and handle outliers in prediction for PARAFAC models.

We also expressed these values in terms of recovery (as the percentage ratio between the predicted and the true concentration) so that they could be compared with those obtained by the procedure described in a previous study using multivariate calibration methods [1] (Table 8, Model C). The average recoveries for the PARAFAC procedure were 103.6% for thiabendazole, 99.9% for fuberidazole and 111.1% for carbendazim. For the multivariate calibration procedure they were 97.7% for thiabendazole, 93.3% for fuberidazole and 102.3% for carbendazim. So, in all the cases, recoveries were around the ideal 100% for both methods, although the dispersion of the results was lower for the multivariate calibration procedure than for PARAFAC.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

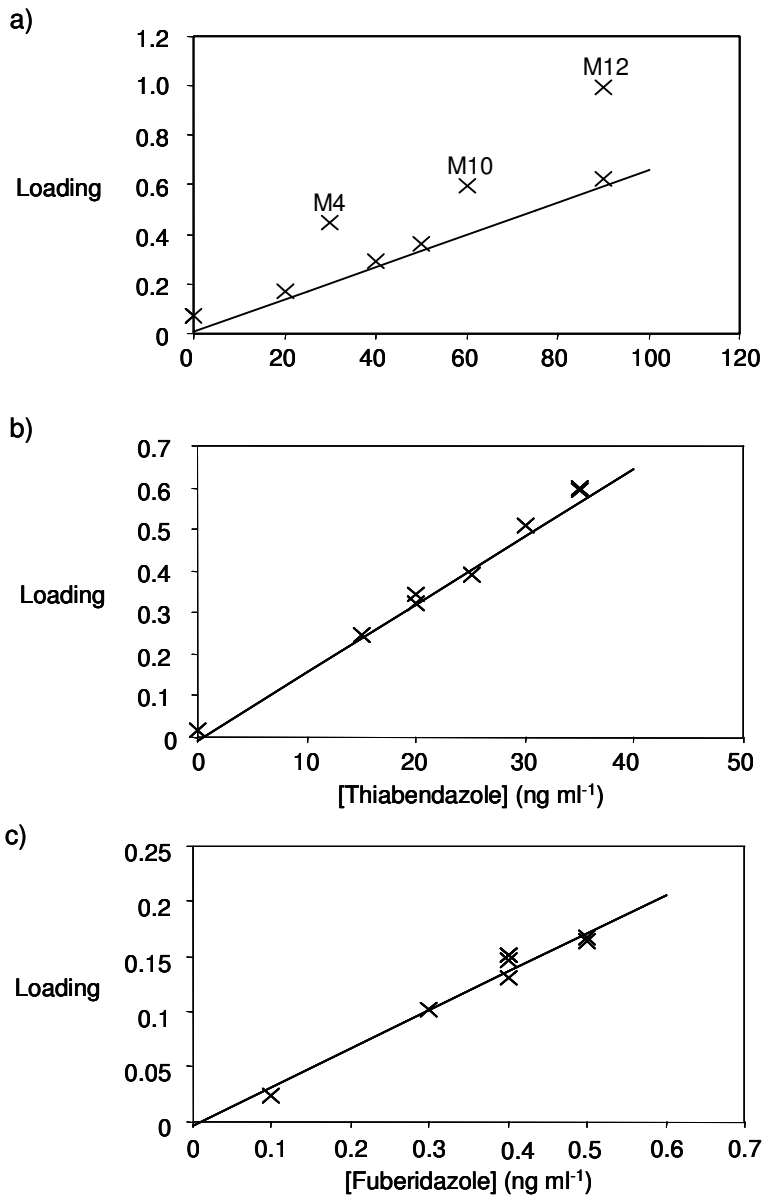


Figure 20. Calibration graphs (straight line) for carbendazim (a), thiabendazole (b) and fuberidazole (c). The crosses (×) represent the prediction samples.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

However, the methodology involving PARAFAC did not require as many calibration samples as the PLS models do and, what is more, would allow the determination of any of the three pesticides in the presence of unknown interferences (second-order advantage) even if they were not included in the model.

We also predicted the samples in the prediction set by including them in the PARAFAC model. The loadings estimated from the PARAFAC model were interpolated into the calibration line that was fitted only with the loadings of the calibration samples. In this way, anomalous samples went unnoticed and the RMSEP values were higher than those obtained by prediction from Equation 36 (see Table 10). What is more, we would have to rebuild the model for each new set of prediction samples, which is less practical in a real laboratory situation.

Table 10. Prediction errors of the two PARAFAC calibration procedures (see details in text).

RMSEP (%)	Carbendazim	Thiabendazole	Fuberidazole
PARAFAC (calibration set)	5.6	3.4	3.9
PARAFAC (calibration+prediction sets)	8.7	4.8	6.4

5. Conclusions

We determined carbendazim, fuberidazole and thiabendazole in mixtures of the three pesticides, by EEM fluorescence and three-way PARAFAC calibration. We used two main criteria for selecting the calibration standards. Firstly, the model had to cover the experimental domain, so standards were preferably taken at the extremes of the domain to avoid subsequent extrapolation. Secondly, we used the selectivity and sensitivity information about the compounds to select the calibration samples, and discussed whether it was necessary to include extra information (i.e. individual standards) in the model for the more sensitive or the less selective analytes in the mixtures.

In the light of the results, we selected for further study a calibration set with six samples: five combinations of extreme concentrations for each pesticide and an individual standard of fuberidazole. The prediction ability of this model compared favourably with the prediction ability of previous PLS models. PARAFAC

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

calibration also required fewer samples and would make quantification possible even in the presence of uncalibrated interferents.

Finally, the PARAFAC models were validated by calculating the univariate figures of merit from the calibration curves that had been built by regressing the PARAFAC sample loadings to the pesticide concentrations. These figures of merit were also used to compare the quality of the different models.

Acknowledgements

The authors would like to thank the MCyT (Project No. BQU2000-1256) for financial support and the CIRIT of the Catalan Government for providing M.J. Rodríguez's doctoral fellowship.

References

[1] D. Picón Zamora, M. Martínez Galera, A. Garrido Frenich, J.L. Martínez Vidal, Analyst 125: 1167-1174 (2000)

[2] J. Saurina, C. Leal, R. Campañó, M. Granados, M.D. Prats, R. Tauler, Anal. Chim. Acta 432: 241-251 (2001)

[3] R.D. JiJi, G.A. Cooper, K.S. Booksh, Anal. Chim. Acta 397: 61-79 (1999)

[4] R. Bro, The N-way online course on PARAFAC and PLS, 1998
URL: http://www.models.kvl.dk/courses

[5] R. Bro, Multiway analysis in the food industry: models, algorithms, and applications, Ph.D. Thesis, University of Amsterdam, 1998
URL: http://www.mli.kvl.dk/staff/foodtech/brothesis.pdf

[6] R. Bro, Chemom. Intell. Lab. Syst. 38: 149-171 (1997)

[7] P.D. Wentzell, S.S. Nair, R.D. Guy, Anal. Chem. 73: 1408-1415 (2001)

[8] D.K. Pedersen, L. Munck, S.B. Engelsen, J. Chemom. 16: 451-460 (2002)

[9] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, second ed., Kluwer Academic Publishers/Plenum Press, New York, 1999

[10] R. Bro, N.D. Sidiropoulos, A.K. Smilde, J. Chemom. 16: 387-400 (2002)

[11] R.D. JiJi, K.S. Booksh, Anal. Chem. 72: 718-725 (2000)

[12] C.A. Andersson, R. Bro, The N-way toolbox for MATLAB ver. 2.00, Chemom. Intell. Lab. Syst. 52: 1–4 (2000). URL: http://www.models.kvl.dk

[13] N.M. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11: 419-461 (1997)

[14] A.C. Olivieri, N.M. Faber, Chemom. Intell. Lab. Syst. 70: 75-82 (2004)

[15] L.A. Currie, Pure Appl. Chem. 67: 1699-1723 (1995)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

-

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 4

ESTIMATION OF THE LIMIT OF DETECTION FOR ITERATIVE TARGET TRANSFORMATION FACTOR ANALYSIS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

4.1. Introduction

The objects of study in this chapter are the practical implications of estimated detection limits. The fact that some experimental factors affect the efficiency of the resolution of second- or higher-order data led to the hypothesis that they would directly affect the estimation of the LOD and, therefore, that the quality of the LOD estimator depends on the performance characteristics of the determination.

In the paper in Section 4.2, we used HPLC-DAD data resolved by ITTFA to evaluate the influence of some significant parameters. Previous considerations for this work were as follows. The data from HPLC-DAD analyses lose their ideal linear structure when there are changes in the retention time between different runs and samples. This problem can be solved aligning the peaks[27] because, if it is not corrected, the lack of reproducibility leads to biased predictions and subsequently unacceptable LODs.

An alternative strategy for dealing with data that are not strictly bilinear is to treat each sample individually, i.e. to solve one second-order data matrix at a time. From the multivariate curve resolution methods that can be used for this purpose, for this paper we chose the iterative approach TTFA[28], which we used to resolve the spectro-chromatographic data of a calibration set.

The title of this paper was "*Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data*". Its main aims were to describe the influence of some performance characteristics on the ability to estimate spectra and elution profiles by the curve resolution method and analyse how this affects LOD estimation. Specifically, the three experimental factors chosen were:

[27] Some examples are in the following references: Gemperline *et al*. 1999, Bylund *et al*. 2002, Bogomolov and McBrien 2003 and Torgrip *et al*. 2003.

[28] Specifically for HPLC-DAD, Comas et al. (Ref. 2002) developed a method based on ITTFA for correcting the time shift.

- the chromatographic selectivity,
- the spectral selectivity, and
- the sensitivity of the analyte of interest (the relative intensity of the chromatographic peak).

To meet these objectives, we reproduced HPLC-DAD measurements of ternary mixtures and performed the experimentation according to a factorial design in order to consider the three factors simultaneously.

These samples were resolved by ITTFA and the LOD was estimated using the transformation approach, i.e. with univariate statistics from the calibration line obtained by the regression of the height of the estimated elution profile against the concentration.

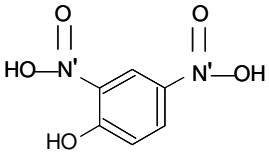
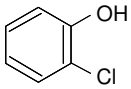
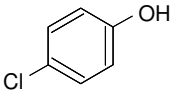
Our results showed that the three experimental factors (the sensitivity of the analyte and the selectivity of spectra and chromatograms) significantly affected the value of the LOD, so they can be modified in order to improve the detection threshold. We also indicated how the end user can improve the limit of detection by modifying these experimental variables.

To calculate the limits of detection for second-order experimental data and apply the method proposed in the paper, we analysed real samples containing mixtures of 2,4-dinitrophenol (2,4-DNP), 2-chlorophenol (2-CP) and 4-chlorophenol (4-CP) by HPLC-DAD. The molecular structure of these compounds is shown in Table 11.

The LOD was calculated for 2-CP, a commercially produced chemical that is toxic to aquatic organisms and is used as an intermediate in the production of other chemicals. The other two compounds present in the mixtures were considered as interferences.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 11. Molecular structure of the compounds under study.

Chemical name	Structure
2,4-dinitrophenol CAS: 51-28-5	
2-chlorophenol CAS: 95-57-8	
4-chlorophenol CAS: 106-48-9	

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

4.2. Paper. Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data.

Analytica Chimica Acta 476: 111- 122 (2003)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data

M.J. Rodríguez-Cuesta, Ricard Boqué, F. Xavier Rius

Department of Analytical and Organic Chemistry, Rovira I Virgili University.
Pça. Imperial Tàrraco,1. 43005-Tarragona (Catalunya). SPAIN

Abstract

We have established a procedure for calculating limits of detection for second-order data. One of the steps involves curve resolution by ITTFA and we have checked some experimental factors that affect the efficiency of resolution by ITTFA. Therefore, they directly affect the estimation of the LOD. In this paper we describe the quality of the LOD estimator as a function of the performance characteristics of a determination with HPLC-DAD (sensitivity and selectivity of spectra and chromatograms) and advise the end user about how he can improve it by modifying these experimental variables.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

1. Introduction

Analytical methods must be validated before they can be used routinely. Several performance characteristics, such as accuracy, precision, sensitivity, etc. must be tested to determine whether previously established requirements are met. This assesses the suitability of a particular method for a particular purpose. One common requirement of the validation process is to evaluate whether an analytical method can detect the presence of an analyte in a sample. This is particularly important for methods that work at low concentration levels.

Analytical performance characteristics are well-defined in univariate [1–7] and multivariate calibration [8–10] (in which a scalar and a vector of data, respectively, is measured for each chemical sample). Second order data (where a matrix of instrumental data is obtained for each sample) are usually produced by second-order instruments such as an excitation-emission spectrofluorometer or by instruments used in so-called hyphenated techniques, i.e. high performance liquid chromatography (HPLC)-diode array detection (DAD), CE-DAD, GC-MS GC-FTIR, LC-MS, HPLC-FTIR or 2D NMR. Although fundamental research in this field is very active, in the literature there are few applications of performance characteristics applied to second-order (or higher) calibration.

The second-order advantage, i.e. determining the analyte concentrations in the presence of unwanted and/or unknown components in the test samples, is possible when the data matrix corresponding to the pure standards or reference mixtures are available and second-order calibration methods such as rank-annihilation factor analysis (RAFA) [11] or the generalised rank-annihilation method (GRAM) [12,13] are used. In this way, Boqué *et al*. have recently presented a new estimator for calculating the limit of detection in bilinear second-order calibration with GRAM [14] yielding promising results.

However, second-order calibration methods are significantly limited when applied to data that deviate somewhat from trilinearity, which is one of the fundamental assumptions of these methods [15]. In HPLC-DAD analysis, these deviations are basically due to shifts in the peak retention times or changes in peak shapes. When such deviations are present we can use other methods, such as those based on curve resolution [16].

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

In this paper, we provide a procedure for calculating limits of detection (LOD) when one is working with experimental second-order data. Specifically, we have developed a strategy that involves a curve resolution method (iterative target transformation factor analysis, ITTFA) to obtain the LOD estimator from the estimated pure chromatograms and spectra of each component of a mixture. With estimated chromatographic profiles we can establish a relationship between the area or height of the peak and a concentration value (quantitative analysis). Classical univariate LOD estimators can therefore be used satisfactorily because we can represent resolved signal versus concentration.

The idea behind this procedure has already been devised by other authors [17]. Our aim in this work, however, is not only to provide a procedure for calculating the limit of detection in second-order data, but also to advise the end user about the quality of the LOD as a function of the characteristics of the problem at hand (sensitivity and chromatographic and spectral selectivities) in determinations that use a typical analytical technique such as HPLC with DAD. As well, Gemperline and Hamilton [18] calculated the limit of detection of minor components when overlapped with major components in hyphenated chromatographic techniques. They used principal component analysis for estimating the net signal due to the minor component in a binary mixture, and described the effect of relative concentration, chromatographic resolution and spectral similarity. Our approach is similar, but we performed the experimentation according to a factorial design to consider the effect of these three factors simultaneously, instead of considering them independently as they did. Moreover, the empirical basis of the procedure we described makes it more robust to face deviations from bilinearity in the data.

2. Theoretical background

In HPLC analysis, compounds may co-elute if they have similar interaction with the stationary phase. If this happens, the experimental conditions can be modified to achieve sufficient peak resolution. However, this can be a time-consuming and expensive process. Another solution is to use multivariate curve resolution (MCR) methods. This is a group of methods intended to determine the number of eluting compounds and to recover the response profiles (e.g. spectra, pH profiles, time profiles and elution profiles) of the components in an unresolved mixture obtained in evolutionary processes when no previous information about the nature or composition of the mixtures is available. MCR methods can be

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

extended to analyse many types of experimental data, including multiway data and non-evolutionary processes [19].

Mathematically, multivariate curve resolution breaks down a two-way data matrix into the product of two smaller matrices, each of which is related to one of the two orders of the original data matrix. It looks for decompositions with physical and chemical meaning and finds the true underlying causes of data variation [20]. The ambiguities inherent to factor analysis decompositions of a single data matrix can be partly overcome if we extract the information from submatrices of reduced rank or if we apply constraints such as linear additivity, non-negativity, unimodality or closure, to the full rank matrix.

Curve resolution methods can be divided into two groups: iterative methods and non-iterative or direct methods. Common examples of iterative approaches are ITTFA [21] and alternating regression (AR) [16]. Iterative methods are fast and simple to use and require less user interaction and expertise, but they do not use local rank information. Instead, they treat the data as a whole and do not provide unique results.

In this study, the MCR used was ITTFA, which was introduced at the beginning of the eighties by Hopke *et al*. [22] in environmetrics and by Gemperline [23,24] and Vandeginste *et al*. [21] in chromatography. The advantages of the method are that it is simpler to use than earlier curve resolution methods and it is more general because the number of co-eluting compounds is not limited. Vandeginste *et al*. [21] showed that ITTFA is suitable for mathematically resolving a data matrix from HPLC with DAD into the pure elution profiles and spectra of all co-eluting compounds.

TTFA tests whether a candidate factor or target is a true factor. The principle of ITTFA is to try to improve targets to bring them close to one of the true factors. The target is modified (using chemical knowledge) and resubmitted in an iterative way until the tested target is considered to match one of the true factors satisfactorily. Spectro-chromatographic data are particularly suited for this type of procedure because the chromatograms are non-negative and unimodal.

Details on how works ITTFA can be found elsewhere. TTFA starts decomposing the second-order data matrix using the singular value decomposition (SVD)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

algorithm. The iteration is started projecting a trial solution onto the chromatographic space and calculating the norm of the difference between initial and projected vector. The graph of this difference as a function of the time position is plotted considering a determined number of significant principal components. The number of minima that appear in the graphs and its position can be related to the number of compounds present in the sample and the time at which the elution peak of each compound appears, respectively. For each compound, a target vector whose maximum is at the estimated position from the graph, is projected onto the chromatographic space. The projected vector can be improved correcting it for shoulders and negative regions (chemical knowledge). An iterative process, resubmitting the projected vector as the new input vector until convergence of the difference, leads to the estimation of the pure chromatogram. Then estimates for the pure spectra are calculated using the generalised inverse of the matrix of chromatograms.
Finally, scale ambiguities inherent to the MCR method should be solved. This can be done using normalisation constraints. We divided the spectra by its norm and multiplied the chromatograms by the same factor (the norm of spectra), so the result of multiplying spectra by chromatograms did not change.

Once we have estimated the chromatograms of standards, we can regress the height of the peak corresponding to the analyte of interest in each standard against the analyte concentration. From the univariate regression line, we can calculate the limit of detection LOD by applying classical univariate LOD estimators such as that recommended by Hubaux and Vos [3]:

$$LOD = t_{1-\alpha,n-2} \cdot \frac{s}{b_1} \sqrt{\frac{1}{N} + \frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} + t_{1-\beta,n-2} \cdot \frac{s}{b_1} \cdot \sqrt{\frac{1}{N} + \frac{1}{n} + \frac{(x_D - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

Eq. 40 (as Eq. 32)

where s is the standard deviation of the residuals of the regression line, b_1 the slope of the regression line, N the number of replications performed on the unknown sample, n the number of standards, $\bar{x}$ the mean of the concentrations of the standards of the calibration curve and $t_{1-\alpha,n-2}$ and $t_{1-\beta,n-2}$ are the $(1-\alpha)$ and $(1-\beta)$ quantiles of the Student's t-distribution with $n-2$ degrees of freedom. In this calculation, α and β incorporate the probability of false positive (or type I error) and false negative (or type II error) as recommended by IUPAC [2].

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

3. Procedure

To prove that the procedure for calculating limits of detection when working with second-order data is valid, we first simulated second-order data by reproducing HPLC-DAD measurements in approximate real conditions. Our simulations covered several analytical situations, so we were able to approximately determine the quality of the estimated LOD when working with real data. We then applied the method to calculate the limit of detection of 2-chlorophenol in synthetic mixtures of three phenolic compounds.

3.1. Limit of detection for second-order simulated data

We simulated the response matrices of three components in unresolved mixtures. One of these components was the analyte of interest (analyte 2), for which we calculated the limit of detection. We obtained the second-order responses by multiplying a matrix of chromatographic profiles by the transposed matrix of normalised spectral profiles. We added homoscedastic noise to the spectral profiles to approximate to real conditions. The strength of this noise was about 2% of the maximum intensity of the spectral profiles (e.g. see Figure 21). Second-order data matrices had 100 rows and 80 columns. In other words, simulated chromatograms represented measurements collected at 100 retention times, at each of which absorbance was measured at 80 wavelengths.

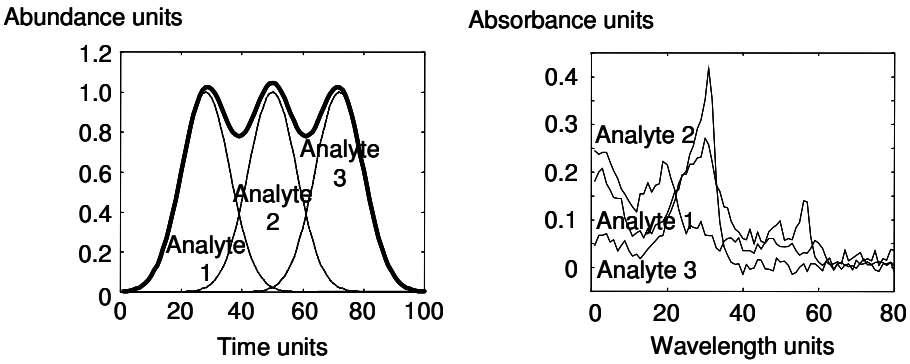


Figure 21. Example of simulated chromatogram and noisy spectra used in this study. The second-order response is obtained by multiplying the matrix of chromatograms by the transpose of the matrix of spectra.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

As we wished to establish a linear regression, we needed standards with decreasing concentration levels. To obtain simulated responses at different concentration levels, we set the elution profile of the analyte of interest at different heights. The heights of the elution profiles of the standards were h, 0.75, 0.50, 0.25, and 0.10 (see example in Figure 22).

We resolved each of the simulated second-order matrices by ITTFA, and calculated the limit of detection as described in Section 2 of this work. The method we have used in this study is summarised graphically in Figure 23.

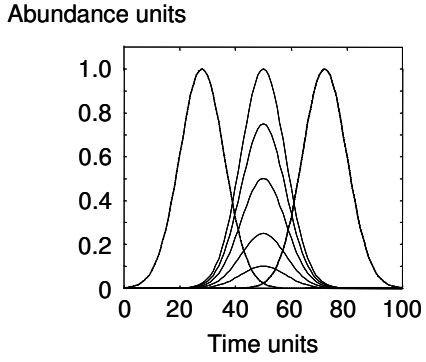


Figure 22. Example of simulated elution profiles of standards at different concentration levels of analyte 2.

3.2. Quality of the LOD estimator

In general terms, and assuming a fixed (rather high) level of noise, the ability to estimate spectra and elution profiles by MCR depends on the similarity of the spectra, the overlap of the elution profiles and the relative intensity of the signal [25–27]. As the limits of detection calculated by this procedure are affected by the quality of the resolved profiles, we performed simulations according to a full 3^3 factorial design of these three factors investigated at three levels.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

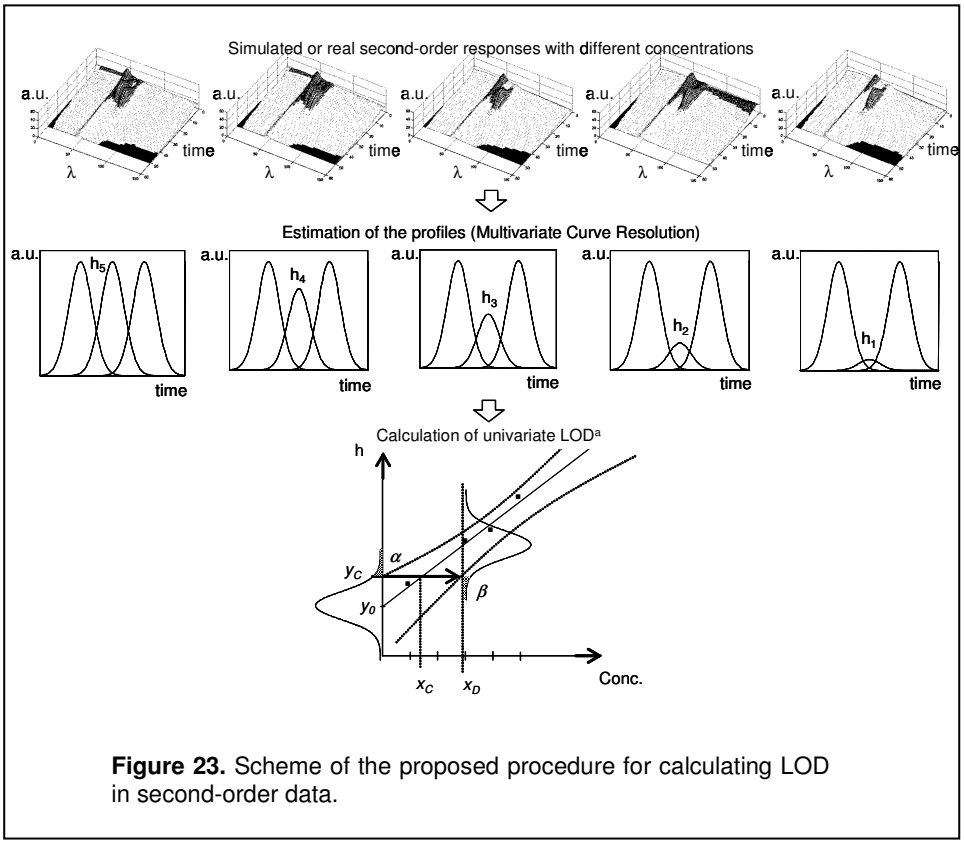


Figure 23. Scheme of the proposed procedure for calculating LOD in second-order data.

The next step was to test how significant the effect of each factor was. We did this by comparing the variance of the response accounted for by the effect with the residual variance, which summarises the experimental error. If the variance due to the effect is smaller than the variance due to the error, the effect can be considered negligible. If it is significantly larger than the error, it is regarded as significant. Numerically this can be expressed with the *F*-ratios, which compare the variance due to the effect with the residual variance. The higher the ratio, the greater the effect. A complementary measure is the *P*-value, i.e. the probability that effect=0. The smaller the *P*-value, the more likely it is that the effect is not due to chance. Usually an effect is declared significant if *P*-value>0.05 (confidence level of 95%).

We discuss these three factors studied below: chromatographic and spectral selectivities (that measure how unique the profile of the analyte of interest is

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

compared with the other species) and the sensitivity (understood as the relative intensity of the instrument response for the species present in the mixture), in certain conditions. Note that we worked with coded variables to denote the factor's low, intermediate and high levels.

3.2.1. Chromatographic selectivity

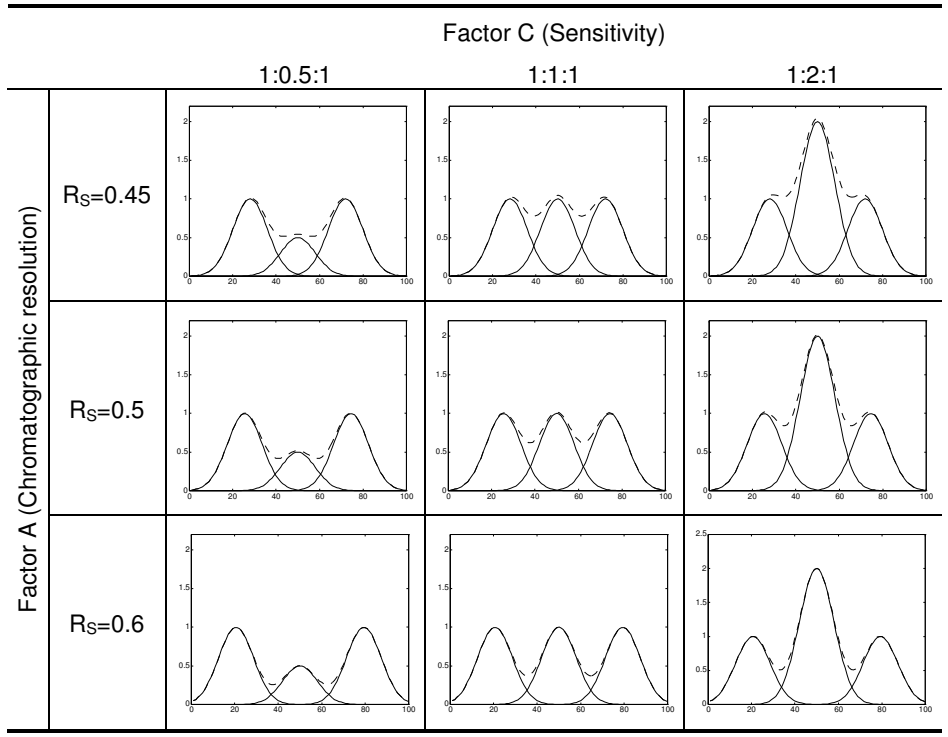


Figure 24. Levels of the factors A (selectivity of chromatograms) and C (sensitivity). The individual profiles are displayed as full curves and the synthetic chromatograms are displayed as broken curves. The analyte of interest is the second peak in all cases. Units of *x*-axis: units of absorbance. Units of *y*-axis: units of time.

Factor A was the chromatographic selectivity expressed as the resolution of the elution profiles. In the simulated mixtures of three components, the analyte of interest eluted the second and overlapped with the other two, which were considered the interferents. As recommended by IUPAC, we calculated

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

chromatographic resolution as $R_S=2(t_{R2}-t_{R1})/(W_{b1}+W_{b2})$, where t_R is the retention time ($t_{R2}>t_{R1}$) and W_b is the Gaussian peak-width at base. As shown in the literature [25–27], spectra of good quality estimated by some MCR methods are obtained for a chromatographic resolution $R_S>0.45$. We then chose the lowest level of this factor at this value, and the intermediate and high levels were 0.5 and 0.6, respectively (see Figure 24). The coded (proportional) values of the levels were –1, –0.33, +1.

3.2.2. Spectral selectivity

Factor B was the spectral selectivity and was expressed as the correlation coefficient of the normalised spectra (norm=1, when dividing the spectra by its Euclidean norm). The higher the selectivity, the lower the correlation coefficient (C_f), and vice-versa. As we had a mixture of three components, we fixed the correlation between the spectra of the analyte of interest and the spectra of the interferent that eluted at higher retention time (C_f=0.16). In this way, factor B implied a variation in the correlation between the spectra of the analyte of interest and the spectra of the interferent that elutes first.

The low, intermediate, and high levels of spectral similarity corresponded to C_f=0.86, 0.49, and 0.20, respectively (see Figure 25). After proportional coding, they corresponded to –1, 0.12, +1.These values can also be expressed in terms of Lorber et al.'s selectivity [28], i.e. the net analyte signal (or part of the spectrum which is orthogonal to the spectra of the others components) divided by the length of the spectrum of the sample.

This selectivity indicate what part of the total signal is lost due to spectral overlap and, according to IUPAC recommendations [29], it can be used in multivariate calibrations with some degree of success. Lorber *et al.*'s selectivity for the analyte of interest at the low, intermediate and high level was 0.36, 0.45, 0.51.

3.2.3. Relative intensity of the signal

Factor C was the sensitivity. This was expressed as the relative height of the elution profiles. The three levels chosen were 1:0.5:1, 1:1:1, and 1:2:1 (see Figure 24). These were coded as –1, 0, +1.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

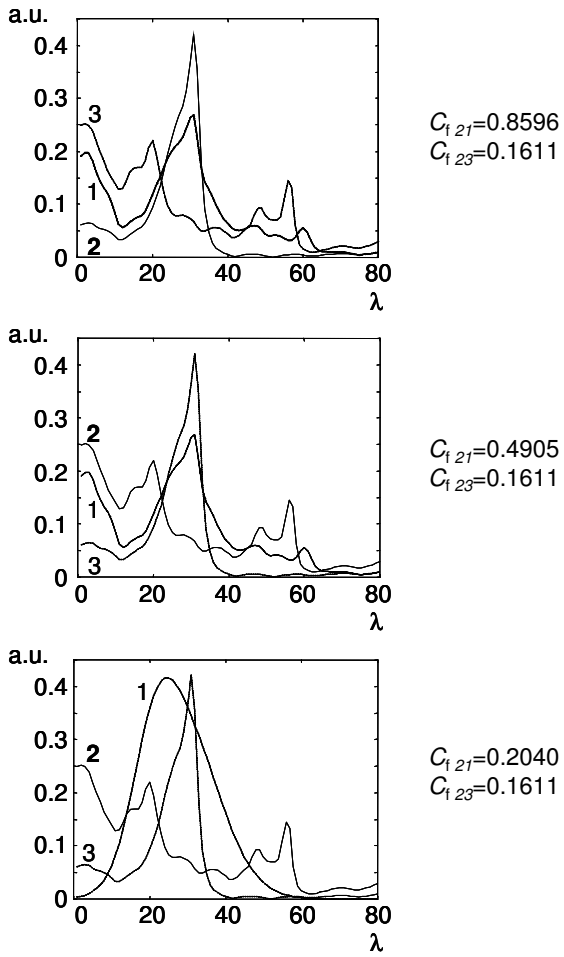


Figure 25. Levels of factor B (selectivity of spectra). Spectral selectivity is expressed as the correlation coefficient (C_f) of the normalised spectra. Spectra of the analyte of interest are displayed as full curves (number 2) and spectra of the interferents are displayed as broken curves.

To estimate how the selected factors affected the accuracy of a quantitative analysis, and particularly how they affected the calculation of the limits of detection, we performed simulations according to this design (see Table 12).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 12. Simulations according to a full 3^3 factorial design. Factor A is the selectivity of chromatograms, factor B is the selectivity of spectra and factor C is the sensitivity. The coding of the levels is explained in the text.

	Factors		
point	**A**	**B**	**C**
1	-1	-1	-1
2	-0.33	-1	-1
3	1	-1	-1
4	-1	0.12	-1
5	-0.33	0.12	-1
6	1	0.12	-1
7	-1	1	-1
8	-0.33	1	-1
9	1	1	-1
10	-1	-1	0
11	-0.33	-1	0
12	1	-1	0
13	-1	0.12	0
14	-0.33	0.12	0
15	1	0.12	0
16	-1	1	0
17	-0.33	1	0
18	1	1	0
19	-1	-1	1
20	-0.33	-1	1
21	1	-1	1
22	-1	0.12	1
23	-0.33	0.12	1
24	1	0.12	1
25	-1	1	1
26	-0.33	1	1
27	1	1	1

We processed our data according to the scheme in Figure 23. We generated five simulated samples with decreasing concentrations of the analyte of interest (as explained above), and resolved these second-order matrices using the ITTFA algorithm. After obtaining the estimated elution profile of each compound in the standards, we plotted the height of the analyte of interest against the concentration and calculated the limit of detection from the calibration curve using Equation 40 (we will call this value the *true* LOD or x_D).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

To evaluate the quality of the estimated *true* LOD, x_D, we generated the second-order response of *M* samples (*M*=100) with the concentration of the analyte of interest at the level of the *true* LOD (we set the height of the elution profile proportional to the LOD). We resolved these M samples by ITTFA to estimate their elution profiles and interpolated in the calibration curve the heights of the analyte of interest in the *M* samples. We could then compare the predicted *M* concentrations (from the *M* samples simulated at the LOD level) with the true LOD that we had previously calculated.

The quality of the LOD can then be expressed in terms of the percentage of bias calculated from the difference between the mean of the *M* predicted concentrations and the true value of the limit of detection, x_D, calculated from Equation 40 (see Equation 41).

$$\text{Percent bias} = \frac{\left(\sum \hat{c}_i / M\right) - x_D}{X_D} \times 100 \qquad \text{Eq. 41}$$

With this procedure we obtained a bias estimation of the calculated LOD. Moreover, for every simulated condition of spectral selectivity, chromatographic selectivity and sensitivity, we repeated the procedure 50 times, i.e. we estimated 50 values of bias for every specific condition. We expressed the quality of the LOD in robust average terms, by calculating the median value of the 50 estimated biases.

3.3. Material and methods

To calculate the limits of detection for second-order experimental data, we analysed real samples containing mixtures of 2,4-dinitrophenol (2,4-DNP), 2-chlorophenol (2-CP) and 4-chlorophenol (4-CP) by HPLC-DAD. We used a 1050 Hewlett-Packard high-pressure liquid chromatograph equipped with a 1050 HP automatic injector and a HP 1050M diode array UV-Vis detector (interface: Hewlett-Packard 35900 model) coupled to an HP ChemStation, v.A.05.02. Analytes were eluted using a Spherisorb reversed-phase column (250mm×45mm i.d., 5 ml) as the stationary phase (partially cross-linked phase) at 60ºC. The injection volume was 20 ml and the flow rate was 1.0 ml/min. The mobile phase included methanol for gradient HPLC, and Millipore water (pH=2.6, acidified with

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

HAc for HPLC). Determinations were performed with an HPLC elution program as follows: an initial gradient elution from 55% of solvent A (MeOH) to 95% of the same solvent in 0.6 min, followed by an isocratic elution until minute 5. Finally, the program reached the initial conditions in 0.5 additional minutes to stabilise the corresponding mobile phase. Spectra were recorded from 250 to 350 nm in steps of 1.96 nm at a rate of 150 spectra per minute.

The LOD was calculated for 2-CP, a commercially produced chemical that is used as an intermediate in the production of other chemicals and is toxic to aquatic organisms [30,31]. The other two compounds present in the mixtures were considered interferents. The initial mixture contained the three compounds at the following concentrations: [2,4-DNP]=0.6 mg/l, [2-CP]=6.5 mg/l and [4-CP]=9.7 mg/l. To calculate the limit of detection of 2-CP, the synthetic samples contained decreasing concentrations of 2-CP (6.5, 4.9, 3.3, 1.6, and 0.7 mg/l), while the concentration of the two interferents was kept constant. The second-order data recorded for each sample was transferred from the HP software to the Matlab [32] environment to be resolved using a home-made ITTFA algorithm.

4. Results

4.1. Simulated data

We processed the median values of the calculated (true) LOD and their percentage of bias, calculated from Equation 41 obtained for all experiments in the experimental design (see Table 13 and Figure 26), by an analysis of variance (ANOVA) (Table 14) to evaluate how the factors affected the calculation of the LOD by this procedure.

From the *F*-test we found that all the factors were significant and this was confirmed by the low values of statistical *P*. Of all the factors, factor B (similarity of spectra) had the weakest effect. Interactions between the effects were also significant in all combinations, but again the highest interaction involved factors A and C since, at the high level of A, the accuracy of the LOD was best when factor C was low, and vice-versa.

We, therefore, assessed how important are the chromatographic and spectral selectivities, the sensitivity and their interactions (in the range studied) for calculating the limit of detection reliably. Also, the quality map indicates how we

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

can modify one or more of the factors to reduce, when necessary, the bias of the LOD calculated with the procedure discussed in this paper.

Table 13. Median (*M*) of the calculated LOD and its accuracy.

Point	$M(x_D)$	*M* (percent bias)	Point	$M(x_D)$	*M* (percent bias)
1	0.01686	16,931	15	0.01629	0,260
2	0.01843	6,756	16	0.01456	18,867
3	0.01399	1,315	17	0.01815	4,903
4	0.01587	18,665	18	0.01557	0,823
5	0.01483	10,587	19	0.01728	2,201
6	0.01922	0,653	20	0.01562	1,160
7	0.01779	36,004	21	0.01743	0,093
8	0.01638	19,587	22	0.02183	1,386
9	0.01634	3,533	23	0.01436	0,912
10	0.01467	10,247	24	0.01637	-0,296
11	0.01648	2,911	25	0.01522	6,285
12	0.01798	-0,004	26	0.01637	1,890
13	0.01529	7,698	27	0.01599	0,199
14	0.01490	3,788			

Table 14. Results of ANOVA.

Effects	*F*-ratio	*p*-value	Interactions	*F*-ratio	*p*-value
A	90.6	0.0000	A × B	6.5	0.0124
B	23.5	0.0005	A × C	17.4	0.0005
C	74.1	0.0000	B × C	5.6	0.0193

Tabulated *F*-value for 95% significance: $F_{0.05,1,27}$ =4.21

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

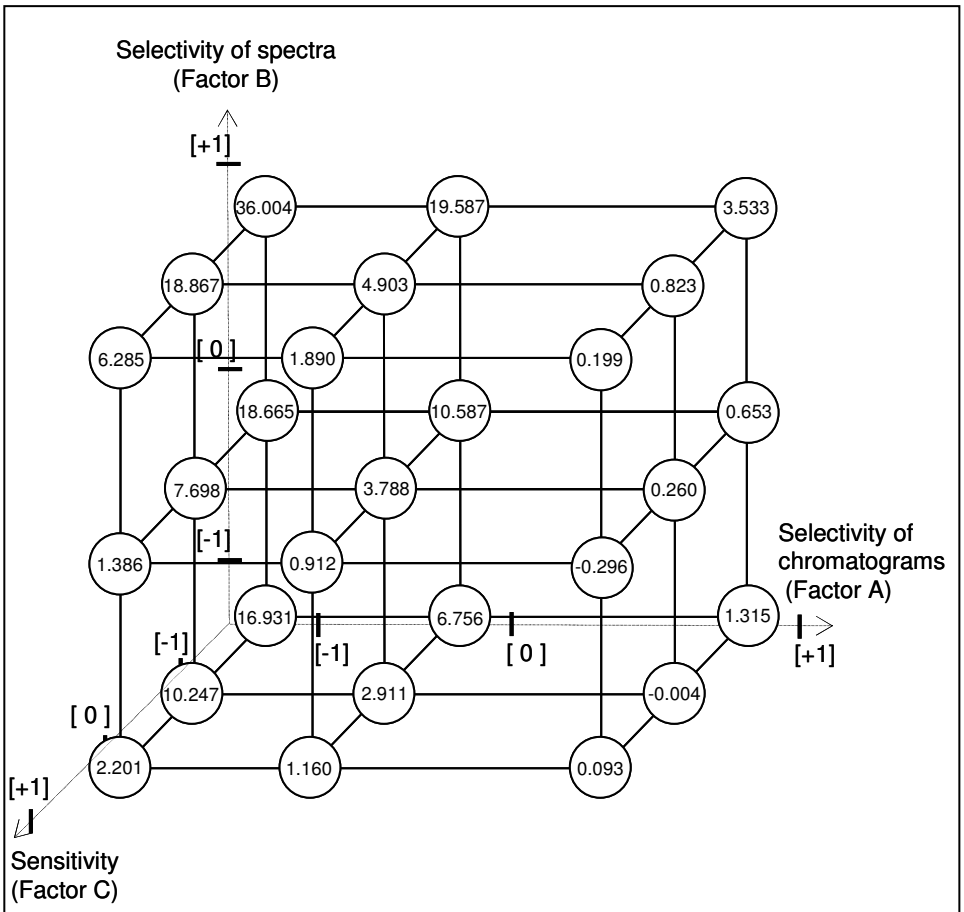


Figure 26. Quality of the LODs calculated. Graphical representation of the results expressed in percent bias, as explained in text.

4.2. Real data

To show the applicability of this procedure, i.e. to calculate the limit of detection and indicate how to improve it, we analysed real samples of mixtures of 2,4-DNP, 2-CP and 4-CP and determined the limit of detection of the 2-CP. We resolved the second-order data recorded for the samples by the ITTFA algorithm. From the estimated spectra (Figure 27a), we unambiguously identified the two interferents and the analyte of interest by comparing them with a library of

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

spectra (the correlation between the estimated and the real spectra was always above 0.99).

As the heights of the peaks are proportional to the concentration, the estimated chromatograms (Figure 27b) provided quantitative information, from which we calculated the limit of detection.

As we can see in Figure 27a, the spectrum of the 2-CP is highly correlated with the spectrum of one of the interferents: C_i(2-CP–2,4-DNP)=0.87 and C_i(2-CP–4-CP)=0.63, which we interpreted as a low level of spectral selectivity (Lorber *et al.*'s selectivity is 0.30), i.e. low level of factor B in this study. We calculated chromatographic selectivity from the estimated chromatographic profiles of the standard with the higher concentration of 2-CP. This selectivity, expressed in terms of resolution, was 0.60 between the first interferent and 2-CP and 0.61 between the second interferent and 2-CP. This means that the level of factor A is high in the terms of this study. Also, the estimated chromatograms show us that the signal intensity ratio was approximately 1:1:1, which is interpreted as an intermediate level of factor C.

We also compared the estimated chromatogram of each compound with the chromatogram obtained with the first-order instrument, e.g. HPLC with a UV-Vis detector. Figure 28 shows the chromatogram at 280 nm under the gradient conditions described in Section 3 of this work.

At this wavelength, the three compounds overlapped so it was difficult to identify or quantify them. In this Figure 28, the broken lines are for the chromatograms estimated by ITTFA, with which we calculated the chromatographic resolution between the peaks and determined the relative intensity of the signal, two of the parameters discussed in this paper.

We used the proposed procedure to calculate the limit of detection for 2-chlorophenol (Table 15). In the previously described experimental conditions, this was 0.9 mg/l.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

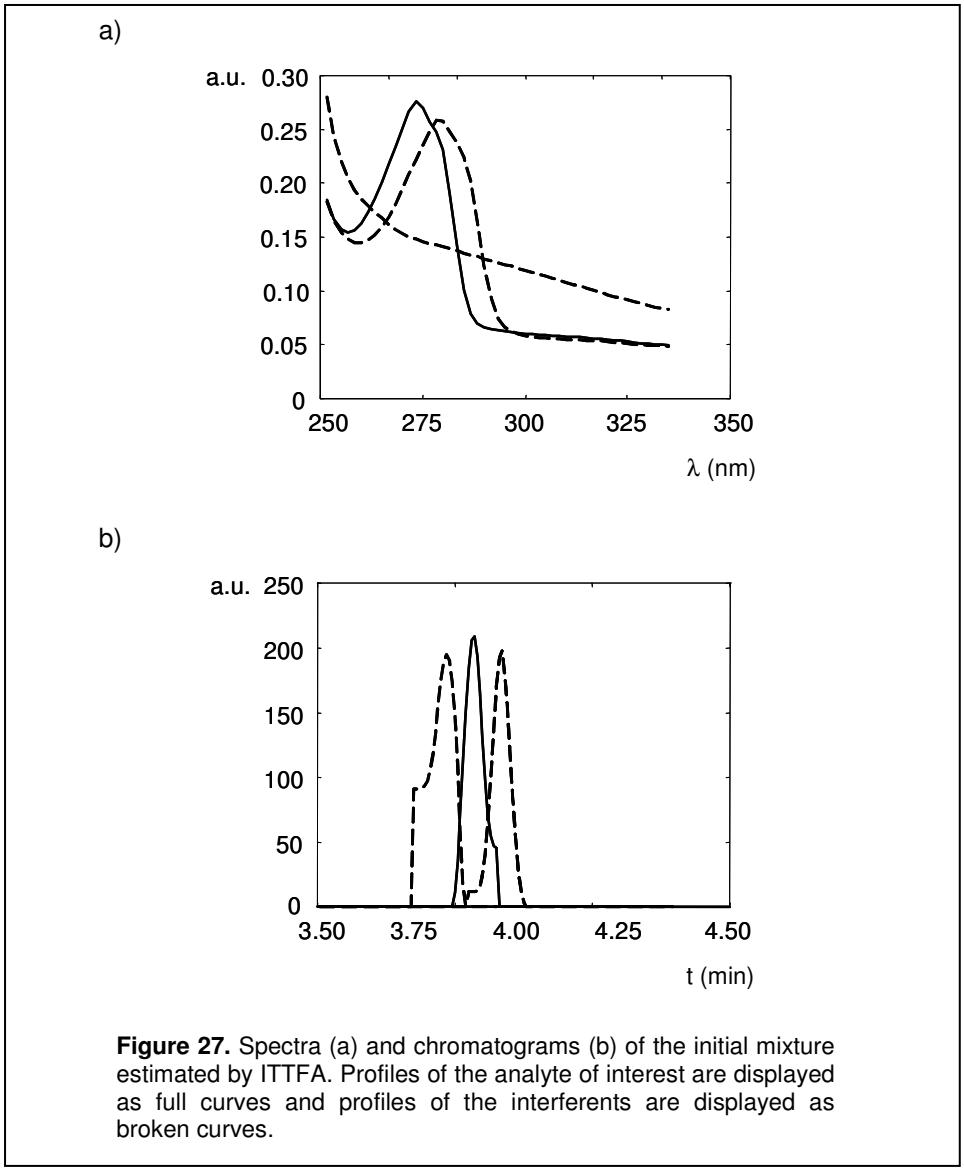


Figure 27. Spectra (a) and chromatograms (b) of the initial mixture estimated by ITTFA. Profiles of the analyte of interest are displayed as full curves and profiles of the interferents are displayed as broken curves.

Table 15. Concentration of 2-CP in the standards and height of the estimated peak.

[2-CP] (mg l-1)	6.5	4.9	3.3	1.6	0.7
Height (a.u.)	181.0	147.5	105.2	52.7	40.9

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

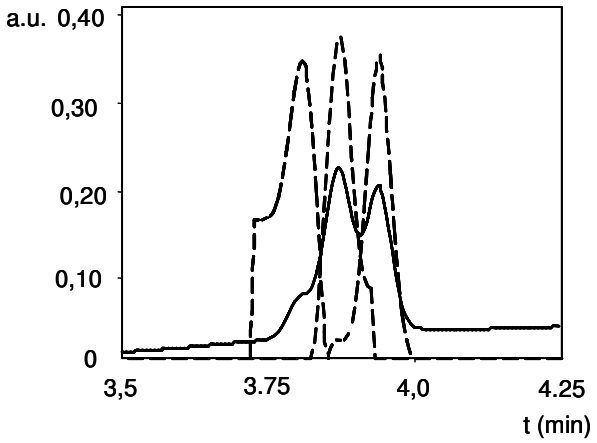


Figure 28. Chromatogram at 280 nm (full curve), and estimated chromatograms by ITTFA (broken curves).

According to the assays of quality performed with the simulated data (Figure 26), the LOD estimated for the experimental samples was biased at an order of magnitude of 0.004% (high level factor A; low level factor B; intermediate level factor C; as in simulation number 12). Depending on the problem at hand, the level of accuracy estimated may or may not be acceptable for the end user. The results of the simulations indicate how we can modify one or more of the factors to reduce this bias if necessary.

For example, we can increase the selectivity of the analyte of interest removing chemically or physically some interferent, if possible (for example by previous preconcentration or separation). Thus, the recorded signal is more selective, resolution improves and it yields to a better quantification. To show this, we have recalculated the Lorber *et al.*'s selectivity of the analyte of interest used in the simulations. The three levels of spectral selectivity in a simulated mixture of three components were 0.36, 0.45, and 0.51. If the second interferent is removed, the three levels of spectral selectivity of the analyte of interest, in the binary mixture, would go up to 0.49, 0.60, 0.80.

We can also improve the LOD by using more standards in the linear model or by increasing the number of replications of the unknown sample (n and N, respectively, in Equation 40).

In cases of severe overlap, MCR may not provide satisfactory results, but we can still increase selectivity by unfolding the matrices and performing partial least squares (PLS) [33].

5. Conclusions

This procedure for calculating limits of detection for second-order data involves one step of curve resolution by ITTFA. The efficiency of the resolution by ITTFA depends on the characteristics of the data, e.g. deviations from the bilinearity due, for instance, to shifts in the peak retention times. In this paper, we have discussed the quality of the LOD estimator as a function of the performance characteristics of a determination with HPLC-DAD (sensitivity, spectral selectivity and chromatographic selectivity). We have shown that these sensitivity and selectivity parameters significantly affect the calculation of LOD, so we also show the end user how to improve this LOD by modifying the experimental variables. For example, we can modify the chromatographic resolution by changing the length of the column or the polarity of the mobile phase, etc. or we can modify the sensitivity by changing the ratio of signal intensity by removing some interferent.
As the accuracy of the LOD estimation strongly depends on the quality of the curve resolution results, we are focusing our research on comparing the results obtained in this paper with the ones obtained with other MCR methods.

Acknowledgements
The authors would like to thank the MCyT (Project no. BQU2000-1256) for financial support and the CIRIT from the Catalan Government for providing M.J. Rodr´ıguez's doctoral fellowship.

References

[1] L.A. Currie, Anal. Chem. 40: 586-593 (1968)

[2] L.A. Currie, Pure Appl. Chem. 67: 1699-1723 (1995)

[3] A. Hubaux, G. Vos, Anal. Chem. 42: 849-855 (1970)

[4] J.D. Winefordner, G.L. Long, Anal. Chem. 55: 712A-724A (1983)

[5] C.A. Clayton, J.W. Hines, P.D. Elkins, Anal. Chem. 59: 2506-2514 (1987)

[6] ISO 11843-1:1997, Capability of detection—Part 1: terms and definitions.

[7] ISO 11843-2:2000, Capability of detection—Part 2: methodology in the linear calibration case.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

[8] G. Bauer, W. Wegscheider, H.M. Ortner, Fresenius J. Anal. Chem. 340: 135-139 (1991)

[9] R. Boqué, F.X. Rius, Chemom. Intell. Lab. Syst. 32: 11-23 (1996)

[10] R. Boqué, N.M. Faber, F.X. Rius, Anal. Chim. Acta 423: 41-49 (2000)

[11] C.N. Ho, G.D. Christian, E.R. Davidson, Anal. Chem. 50: 1108-1113 (1978)

[12] E. Sanchez, B.R. Kowalski, Anal. Chem. 58: 496-499 (1986)

[13] E. Sanchez, B.R. Kowalski, J. Chemom. 2: 265-280 (1988)

[14] R. Boqué, J. Ferré, N.M. Faber, F.X. Rius, Anal. Chim. Acta 451: 313-321 (2002)

[15] P.J. Gemperline, J.H. Cho, B. Archer, J. Chemom. 13: 153-164 (1999)

[16] R. Tauler, Chemom. Intell. Lab. Syst. 30: 133-146 (1985)

[17] J. Saurina, C. Leal, R. Compañó, M. Granados, M.D. Prat, R. Tauler, Anal. Chim. Acta 432: 241-251 (2001)

[18] P.J. Gemperline, J.C. Hamilton, Anal. Chem. 61: 2240-2243 (1989)

[19] Group of Solution Equilibria and Chemometrics, Analytical Chemistry Department, University of Barcelona. URL: http://www.ub.es/gesq/mcr/mcr.htm

[20] R. Tauler, A. Smilde, B. Kowalski, J. Chemom. 9: 31-58 (1995)

[21] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173: 253-264 (1985)

[22] P.K. Hopke, D.J. Alpert, B.A. Roscoe, Comput. Chem. 7: 149-156 (1983)

[23] P.J. Gemperline, J. Chem. Inf. Comput. Sci. 24: 206-212 (1984)

[24] P.J. Gemperline, Anal. Chem. 58: 2656-2663 (1986)

[25] B.G.M. 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, p. 271

[26] L.K. Strasters, H.A.H. Billiet, L. de Galan, B.G.M. Vandeginste, G. Kateman, Anal. Chem. 60: 2745-2751 (1988)

[27] B.G.M. Vandeginste, F. Leyten, M. Gerritsen, J.W. Noor, G. Kateman, J. Frank, J. Chemom. 1: 57-71 (1987)

[28] A. Lorber, K. Faber, B.R. Kowalski, Anal. Chem. 69: 1620-1626 (1997)

[29] Selectivity in analytical chemistry (IUPAC Recommendations 2001), Pure Appl. Chem. 73: 1381-1386 (2001)

[30] Ambient Water Quality Criteria for 2-Chlorohenol, United States Environmental Protection Agency, EPA 440/5-80-034

[31] International Chemical Safety Cards (ICSC:0849), National Institute for Occupational Safety and Health (NIOSH)

[32] MATLAB 5.3, The MathWorks Inc., Natick, MA, 1999

[33] N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivos, Chemom. Intell. Lab. Syst. 63: 107-116 (2002)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 5

ESTIMATION OF THE LIMIT OF DETECTION FOR MULTIVARIATE CURVE RESOLUTION - ALTERNATING LEAST SQUARES

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

5.1. Introduction

The aim of the application in this chapter was to estimate the LOD for second-order data that are or can be treated as approximately linear.

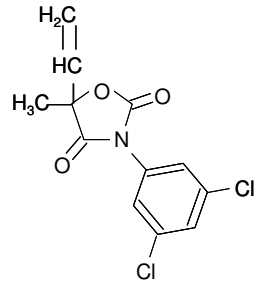
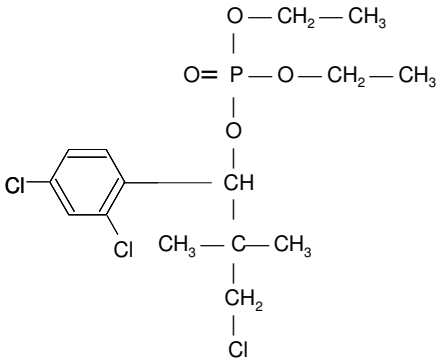
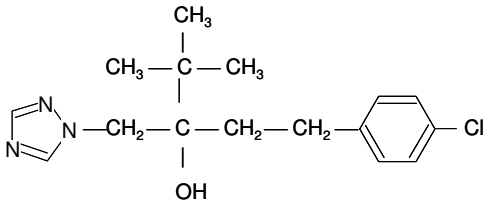
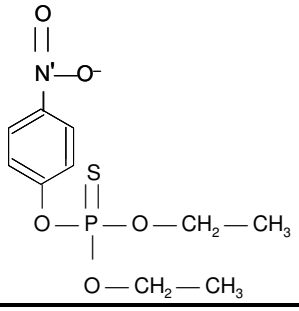
Our colleagues in the Department of Hydrogeology and Analytical Chemistry at the University of Almería provided us with a set of highly complex natural groundwater samples. These contained several pesticides and were analysed by HPLC-DAD. These data were used for this application because three-way data from HPLC-DAD analyses are bilinear under ideal conditions (the circumstances that prompt violation of linearity are changes in the shape of the spectra due, for example, to matrix effects, or changes in the shape or shifts of the retention time). The chemometric technique selected to elucidate and quantify the analyte of interest was the versatile multivariate curve resolution–alternating least squares (MCR-ALS) method.

In Section 5.2 I present the paper entitled "*Development and validation of a method for determining pesticides from complex overlapped HPLC signals and multivariate curve resolution*". As I have mentioned, we aimed to elucidate highly complex natural groundwater samples that contained several pesticides. The analytical problem was that the samples showed highly overlapped spectra, which made quantification especially difficult. Some of the co-authors had been working on these samples using several methods for a long time, so the results of the paper were a challenge in themselves. The procedure was established using synthetic samples. A set of natural groundwater samples was also used to validate the model and assess the fitness of the method for quantifying the pesticides in natural samples. We focused on determining four pesticides whose spectra were located in the region with the most severe overlapping. The molecular structures of these pesticides are shown in Table 16[29].

[29] Vinclozolin is a common fungicide used in vineyards and is a known endocrine disruptor. Clorfenvinphos is a colorless or amber liquid with insecticide activity that has harmful effects on the human nervous system. Tebuconazole and parathion-ethyl are pesticides with fungicide action.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 16. Molecular structure of the pesticides under study.

Commercial name	Chemical name	Structure
Vinclozolin	3-(3,5-dichlorophenyl)-5-ethenyl-5-methyl-2,4-oxazolidinedione CAS: 50471-44-8	
Chlorfenvinphos	2-chloro-1-(2,4-dichlorophenyl) ethenyl diethyl phosphate CAS: 470-90-6	
Tebuconazole	1-(4-chlorophenyl)-4,4-dimethyl-3-(1,2,4-trizole-1-yl-methyl)pentane-3-ol CAS: 80443-41-0	
Parathion-ethyl	diethyl 4-nitrophenyl phosphorothionate CAS:56-38-2	

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

HPLC-DAD data were resolved by MCR-ALS. A main feature of this study was to maintain the three-way structure of the data when decomposing several samples simultaneously and to impose trilinearity only for some compounds while preserving bilinearity for others. Prediction samples were used together with the calibration samples to build the calibration model. We then modelled the relative areas–with respect to the same reference standard–estimated by the MCR-ALS model for the calibration samples against the known concentration of the pesticide of interest. The relative areas of the prediction samples, which were also relative with respect to the reference standard, were then interpolated into the calibration line to obtain the estimated concentrations. Figures of merit were calculated by the transformation approach, i.e. from the fitted line and using univariate statistics.

One proof of the complexity of these mixtures of pesticides, apart from the obvious signal overlap, was the presence of systematic data variation that could not be attributed to the expected four analytes. Further assays were performed to find out the correct number of species in the mixtures, the nature of these species, and what causes this behaviour e.g. whether there is a contamination or decomposition process for some compounds. However, none of the strategies we tried explained the abnormal number of sources of variance.

In summary:

- We use different calibration sets to test the proper number of standards from which the model would be fitted and the composition of these samples. Special attention was paid to the chromatographic and spectral selectivities and to the sensitivities of the compounds.

- We used the prediction samples, together with the calibration samples, to build the calibration model so they had as much influence as the calibration set in the iterative decomposition procedure.

- We used a multivariate curve resolution alternating least squares (MCR-ALS) mixed bilinear-trilinear model to determine the pesticides in groundwater. In this model trilinearity was imposed only for some compounds while bilinerarity was preserved for others.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

- We validated the method by assessing its figures of merit not only to assess the reliability of the results but also to apply the method to future situations in a consistent way.

- We used a representative suite of test data and applied the method to determine these pesticides in natural samples. We obtained good results for three of the pesticides.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

5.2. Paper. Development and validation of a method for determining pesticides from complex overlapped HPLC signals and multivariate curve resolution.

Chemometrics and Intelligent Laboratory Systems 77(1-2): 251-260 (2005)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Development and validation of a method for determining pesticides from complex overlapped HPLC signals and multivariate curve resolution

M.J. Rodríguez-Cuesta, Ricard Boqué, F. Xavier Rius

Department of Analytical and Organic Chemistry, Rovira I Virgili University
Pça. Imperial Tàrraco,1. 43005-Tarragona (Catalunya). SPAIN

J. L. Martínez Vidal, A. Garrido Frenich

Department of Hydrogeology and Analytical Chemistry, University of Almeria
04071-Almería (Andalucía). SPAIN

Abstract

A method is presented to determine pesticides in groundwater by way of multivariate curve resolution alternating least squares (MCR-ALS) mixed bilinear–trilinear models of overlapped chromatographic second-order data (HPLC-DAD). Four of the pesticides in the samples showed highly overlapped spectra, which made the quantification especially difficult. Different calibration and validation sets were used to test the proper number and distribution of the samples. Performance characteristics (figures of merit) such as sensitivity, precision and limit of detection were assessed from the calibration mixtures. Accuracy in the predictions was estimated in terms of the root mean square of percentage deviation (RMSPD) and explained variance (Q^2). We obtained values of RMSPD of 7%, 10% and 8% and Q^2-values of 99%, 93% and 99% for the pesticides vinclozolin, chlorfenvinphos and parathion-ethyl, respectively. A set of natural groundwater samples was also used to validate the model and to assess the fitness of the method for quantification of the pesticides in natural samples. Three of the pesticides under study were satisfactorily resolved and quantified in the groundwater samples by the proposed procedure achieving 88%, 96% and 94% of the explained variance for the pesticides vinclozolin, chlorfenvinphos and parathion-ethyl, respectively.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

1. Introduction

The analysis of mixtures of pesticides using methods based on high-performance liquid chromatography (HPLC) sometimes results in complex separations and overlapped peaks. Nevertheless, complex multicomponent mixtures, as those coming from environmental analysis, can in many cases be qualitatively and quantitatively resolved by means of chemometrics. Depending on their nature, data can be arranged in a two-way structure (a table or a matrix), as in the case of collecting the absorbance spectra for many samples, or in a three-way structure, e.g. in HPLC-DAD, where spectra are recorded at several retention times for each sample (Figure 29). Such data arrangements in three- or higher way arrays can be handled using multi-way methods of analysis. These methods must also be validated following international ISO guidelines [1] in order to reliably use them as routine methods for the identification of the species present in the sample and the quantification of the analytes of interest.

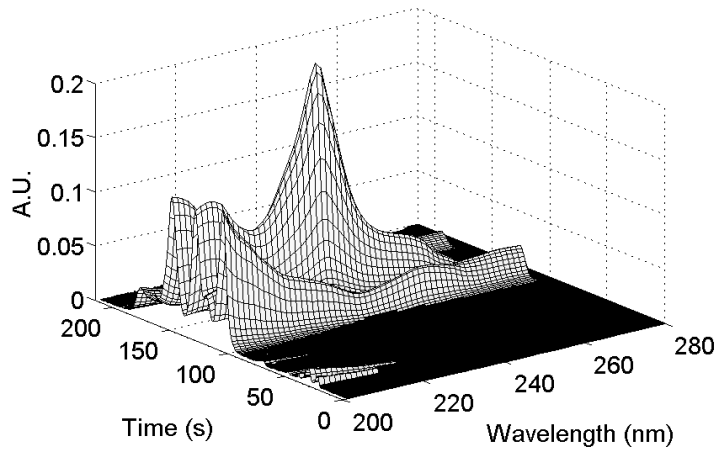


Figure 29. Three-dimensional plot of absorbance, wavelength and retention time corresponding to a multicomponent mixture of pesticides.

An important step forward to build the best model is the selection of the calibration set, which implies to decide the number of calibration samples as well

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

as their composition. The number of standards from which the model will be fitted must be high enough to assess the robustness of the model. Nevertheless, increasing the number of standards will also increase the number of analyses and the time of experimentation. In any case, the user must be aware of that the selected calibration samples will define the interval of applicability of the method and the associated precision.

Selectivity and sensitivity of the compounds present in the samples must be taken into account, too. These parameters influence the multi-way decomposition of the data and, therefore, the figures of merit of the analytical procedure [2]. Selectivity and sensitivity information about the system must be collected, if possible, beforehand and used to properly select the calibration samples. The authors have shown in a previous work [3] that this knowledge notably improves the fit of the model and its prediction ability. Therefore, special attention should be given to the chromatographic and spectral selectivities and sensitivities and to the selection of the calibration samples on the basis of these parameters.

We put into practice these ideas by establishing a procedure to determine the pesticides of a multicomponent mixture. From the 12 pesticides present originally in the mixtures [4] (chlorfenvinphos, chlorothalonil, fenamiphos, iprodione, malathion, parathion-ethyl, parathion-methyl, procymidone, tebuconazole, triadimefon, triazophos and vinclozolin), we focused on four of them, located in the region where severely overlapped signals appeared: vinclozolin (Vi), chlorfenvinphos (Cf), tebuconazole (Te) and parathion-ethyl (Pe).

These pesticides are usually determined by gas chromatography (GC) coupled to mass detection [5–7]. In addition, capillary electrophoresis has also been used for the determination of some of them [8]. To properly quantify the pesticides, these methods use a preliminary separation step, which is time-consuming and require expensive instrumentation and consumption of solvents. Application of chemometric approaches may avoid these disadvantages, since the separation step is eliminated and the resolution may be achieved mathematically. For example, partial least squares (PLS) models [9], cross-section (CS) approaches in combination with PLS and principal components regression (PCR) [10], peak purity assays such as SIMPLe to use Interactive Self-modelling Mixture Analysis (SIMPLISMA), orthogonal projection analysis (OPA), needle search (NS), positive matrix factorization (PMF2) and alternating least squares (ALS) [11,12]

were used to resolve that 12-analyte HPLC-DAD data into the individual concentration profiles and spectra of each pesticide.

Among the other resolution strategies that, to the best of our knowledge, have not been reported to resolve this complex mixture, e.g. Heuristic Evolving Latent Projections (HELP) [13, 14], Stepwise Key Spectrum Selections (SKSS) [15] or Iterative Target Transformation Factor Analysis (ITTFA) [16], we evaluated the ability of the multivariate curve resolution (MCR-ALS) [17] procedure to obtain reliable qualitative and quantitative results for these pesticides.

We established and validated a procedure for the determination of four pesticides using multivariate curve resolution (MCR-ALS) of the overlapped HPLC-DAD data. We resolved the mixtures and obtained the pure spectra of the pesticides. A main feature of this work is to keep the three-way structure of the data when decomposing several samples simultaneously, imposing the trilinearity property only for some compounds, while for others only bilinerarity is preserved. Validation of the method, through assessing its figures of merit, is a crucial step, not only for assessing the reliability of the results, but for consistently applying the method to future situations. Finally, we used a representative suite of test data and applied the proposed methodology to the determination of the pesticides in natural samples.

2. MCR-ALS method

The multivariate curve resolution method has been described elsewhere [17] and used for solving analytical problems similar to the one reported in this paper [18–21]. It is based on a linear model which assumes the additivity of the response of all the compounds in the samples. The spectrochromatographic data is usually arranged column-wise, i.e. an augmented single matrix **R** is built by setting one matrix in top of the other. These data are modelled with the equation:

$$\underset{(IK\times J)}{\mathbf{R}} = \underset{(KI\times F)}{\mathbf{C}} \cdot \underset{(F\times J)}{\mathbf{S}^{\mathbf{T}}} + \underset{(KI\times J)}{\mathbf{E}} \qquad \text{Eq. 42 (as Eq. 13)}$$

where **R** is the column-wise augmented matrix, built with *K* matrices of *I* spectra in the rows and *J* chromatograms in the columns, **C** is the augmented matrix of

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

chromatographic profiles for the *F* analytes, $\mathbf{S}^T$ is the transposed matrix of absorption profiles and **E** is the matrix of residuals (see Figure 30). To estimate the matrices **C** and $\mathbf{S}^T$, from the matrix **R** of measured data, an alternating least squares (ALS) procedure is used, starting with initial estimates of the chromatographic elution profiles.

During the iterative optimisation, several constraints, such as non-negativity, unimodality, closure and selectivity constraints [22, 23], may be applied to obtain chemically meaningful solutions. The optimisation ends when a convergence criterion is reached. The common minimization of the squared sum of the residuals was the criterion used in this work, despite others, like requiring that the change in $\det(\hat{\mathbf{S}}_f\hat{\mathbf{S}}_f^T)$ to be sufficiently small (~10^{-8}), have been reported to provide good indication of the completion of the resolution process [24].

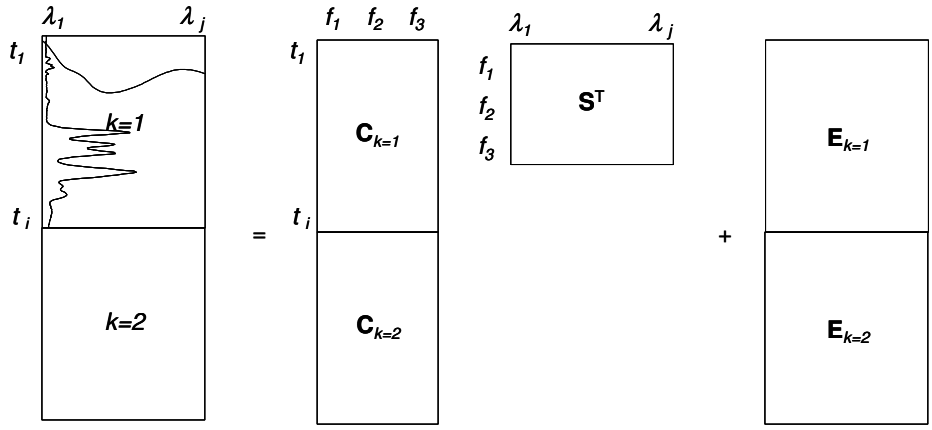


Figure 30. Graphical representation of the mathematical model. In this example of HPLC-DAD data, a column-wise augmented matrix, built with *K*=2 matrices of *I* spectra measured at *J* wavelengths, the resolution for *F*=3 components gives an augmented matrix **C** of chromatographic profiles and a matrix $\mathbf{S}^T$ of common absorption profiles.

The prediction step can be performed by comparing the area below the chromatographic profile of the analyte in a reference calibration standard and in the unknown sample as $x_{unkown}=(A_{unkown}/A_{standard})x_{standard}$, where x_{unkown} and $x_{standard}$

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

are the concentrations of the analyte of interest in the unknown and standard samples, respectively, and A_{unkown} and $A_{standard}$ are the areas below the chromatographic profiles in the unknown and in the standard samples, respectively. When several standards are employed in the decomposition, we can model the relative areas —with respect to the same reference standard— of the calibration samples against the known concentration of the analyte of interest. Then the relative areas of the prediction samples, which are also relative with respect to the reference standard, are interpolated into the calibration line to obtain the estimated concentrations.

One of the main features that differentiates MCR-ALS second-order calibration from uni- or multivariate calibration are: (1) that prediction samples are used, together with the calibration samples, to build the calibration model, and therefore they have as much influence as the calibration set in the iterative decomposition procedure, and (2) that quantification of the analyte of interest in the prediction samples is possible even in the presence of new interfering compounds.

3. Figures of merit

Figures of merit were calculated after resolving the overlapped peaks by using univariate statistics from the calibration line fitted with the relative areas of the calibration standards against the known concentration of analyte of interest, as explained above.

Sensitivity was defined as the slope of the calibration curve. Precision was calculated in terms of concentration as the standard deviation of the residuals relative areas for all standards divided by the sensitivity, SEN, as:

$$Precision = \frac{s_{res}}{SEN} = SEN^{-1}\sqrt{\frac{\sum\left(A_i - \hat{A}_i\right)^2}{n-2}} \qquad \text{Eq. 43 (as Eq. 37)}$$

A_i are the relative areas estimated by MCR-ALS, $\hat{A}_i$ are the areas estimated from the calibration line of areas versus concentration and n is the number of calibration standards.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

The limit of detection, LOD, was estimated from Equation 40, which takes into account the uncertainty of the calibration line and considers α and β probabilities of error, following the IUPAC recommendations [25]:

$$LOD = \delta(\alpha,\beta)\frac{s_{res}}{SEN}\sqrt{\frac{1}{N}+\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}} \qquad \text{Eq. 44 (as Eq. 38)}$$

x_i is the concentration of analyte in each of the n calibration standards and $\bar{x}$ is the average concentration. N is the number of replications performed on the unknown sample.

To estimate the accuracy of the prediction, we calculated the root mean square error of the percentage deviation [26], RMSPD, for each model validation as

$$RMSEPD = \sqrt{\frac{\sum(c_{ref,i}-\hat{c}_i)^2}{M}}\times 100 \qquad \text{Eq. 45}$$

where $c_{ref,i}$ and $\hat{c}_i$ are the reference and the predicted concentrations of analyte in the ith prediction sample, and M is the number of prediction samples. Another generalised way of measuring the prediction ability is by means of the explained variance in the prediction set. Called Q^2-value [27], this variance can be calculated using Equation 46:

$$Q^2 = \left(1-\frac{\sum_{m=1}^{M}(c_{ref,i}-\hat{c}_i)^2}{\sum_{m=1}^{M}(c_i-\bar{c})}\right)\times 100 \qquad \text{Eq. 46}$$

with $c_{ref,i}$, $\hat{c}_i$, c_i and $\bar{c}$ as defined above.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

4. Experimental

4.1. Reagents and standards

Pesticide standards (Pestanal quality) of chlorfenvinphos (Cf), chlorothalonil, fenamiphos, iprodione, malathion, parathion-ethyl (Pe), parathion-methyl, procymidone, tebuconazole (Te), triadimefon, triazophos and vinclozolin (Vi) were obtained from Riedel-de HaJn (Seelze, Germany); purity always >99%. Pesticide stock standards were prepared at 200 mg l^{-1} in acetonitrile, where they were stable for several months. Working solutions were prepared daily by diluting the stock standards. All standard solutions were stored at 4 ºC. Analytical-reagent grade solvents, acetonitrile, acetone and methylene chloride, obtained from Merck (Darmstadt, Germany), were also used. HPLC-grade water provided by a Millipore Milli-Q water filtration/purification system (Bedford, MA, USA) was used and anhydrous Na2SO4 was supplied by Panreac (Barcelona, Spain).

4.2. Samples

To determine the pesticides in synthetic samples, a data set of 34 mixtures was prepared in acetonitrile. The concentrations were randomly selected, covering the domain of concentrations and avoiding collinearity between mixtures. The linear ranges for the 12 pesticides were in the limits 0.3–12 µg ml^{-1} for all of them, except for malathion and tebuconazole, for which the values were between 2 and 15 µg ml^{-1}. Concentration data are shown in Table 17.

Groundwater samples were taken from different deep wells (>300 m) on the Campo de Dalías (Almería), in the south eastern coast of Spain, in order to avoid a possible contamination by pesticides. Samples were collected in 1 litre amber glass bottles capped with Teflon-lined screw caps. Replicated samples were taken. After filling with water, the bottles were shaken vigorously for 1 min, iced down in the field and kept refrigerated at 4 ºC away from light prior to extraction, which was done within 48 h. The extracts were analysed before two weeks of collection. To determine the pesticides in these natural groundwater samples, a set of eight samples were spiked with the pesticides and pre-treated as explained below. Concentration data are also shown in Table 17.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 17. Concentration data (μg ml^{-1}) for the synthetic mixtures (S#) and the spiked groundwater samples (AG#).

Sample	Vi	Cf	Te	Pe	Sample	Vi	Cf	Te	Pe
S1	2	8	8	3	S22	8	5	7	5
S2	2	7.5	8	3	S23	4	4	5	7
S3	10	2	3	8	S24	3	7	3	4
S4	9.5	0	3	8	S25	3	7	3	4
S5	5	2	5	5	S26	6	3	8	3
S6	5	4	5	5	S27	6	3	8	3
S7	4	7	4	0	S28	5	6	6	6
S8	8	3	7	8	S29	5	6	6	6
S9	6	3	7	8	S30	6	5	10	5
S10	0	5	4	4	S31	6	5	10	5
S11	7	2	6	6	S32	4	4	6	6
S12	7	2	6	6	S33	4	4	6	6
S13	3	6	8	7	S34	8	7	4	6
S14	3	6	8	7	AG1	3	4	4	6
S15	5	10	10	3	AG2	4	4	6	4
S16	5	10	10	3	AG3	8	8	8	5
S17	4	3	5	4	AG4	6	5	10	6
S18	4	3	5	4	AG5	4	6	8	8
S19	7	6	8	8	AG6	5	6	6	4
S20	7	6	8	8	AG7	8	3	8	6
S21	5.5	8	4	3	AG8	5	5	4	4

4.3. Apparatus

A Waters (Milford, MA, USA) model 990 liquid chromatographic system was used, equipped with a Model 600 E constant-flow pump, a Rheodyne six-port injection valve with a 20 μl sample loop; a Model 990 UV–visible photodiode-array detector, a printer/plotter and a microcomputer using the Waters 991 software package.

4.4. Experimental procedure

Synthetic mixtures were prepared by adequate dilution of working standard solutions and 20 μl volumes were injected into the HPLC system.

Spiked groundwater samples (500 ml) containing 50 ml of acetone were shaken with 50 ml of methylene chloride for 2 min. Three extractions with methylene chloride were carried out. The combined organic phases were dried, by passing them through anhydrous Na_2SO_4, and evaporated using a rotary vacuum

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

evaporator. The samples thus concentrated were eluted with 1 ml of acetonitrile and 20 µl volumes injected into the HPLC system.

HPLC separations were performed on a Hypersil C_{18} column (100×0.46 mm i.d., 5 µm particle size). The mobile phase, under isocratic conditions, was acetonitrile/water (60:40 v/v). This mobile phase composition was used to reduce the time of analysis and avoid too much dispersion of peaks.

The solvents were filtered daily through a 0.45 µm cellulose membrane filter before use, and degassed with helium before and during use. Samples (20 µl) were injected with the solvent flow rate maintained at 1 ml min^{-1}. Photometric detection was performed in the range 200–280 nm, with a spectral resolution of 1.4 nm. Data was obtained over an integration period of 1.4 s per spectrum. That is, in the original spectrochromatographic data absorbances were measured at 215 times and 60 wavelengths.

4.5. Chemometric procedure

In order to focus the problem on the four unresolved peaks, we defined a window in the chromatographic direction from minute 3.50 to 4.43, obtaining matrices of 41 rows (retention time) and 60 columns (wavelengths) (see Figure 31). The four analysed compounds were expected to elute in this region in the order Vi, Cf, Te and Pe.

We did principal component analysis on a new unfolded data (unfold PCA), to check the presence of outliers or influential samples. The new data was a (34×2460) matrix, where the rows were each of the 34 (41×60) matrices previously converted into a (1×2460) vector.

Evolving factor analysis (EFA) [28,29] and SIMPLISMA [30] were applied to collect information about the proper number of components and the variability sources in the system.

For the calibration, the number of samples should be sufficient to statistically define the relationship between the signal and the concentration (usually established as a multiple of the number of factors employed in the calibration model). In order to optimize the procedure and to find a compromise between the number of analysis and the proper description of the experimental domain, we

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

tested three different situations by combining different training and validation sets. The calibration set #1 contained six standards covering the expected range of concentrations and uniformly distributed over the total range. The calibration set #2 contained 12 samples in the same way. To prove the validity of the method, we tested two validation sets, covering also the working range of concentrations. The prediction sets #1 and #2 contained 6 and 22 validation samples, respectively. The effect of the number of calibration and validation samples on the resolution and on the prediction error was evaluated in cases A (calibration set #1 and prediction set #1), B (calibration set #2 and prediction set #1) and C (calibration set #1 and prediction set #2), as summarised in Table 18.

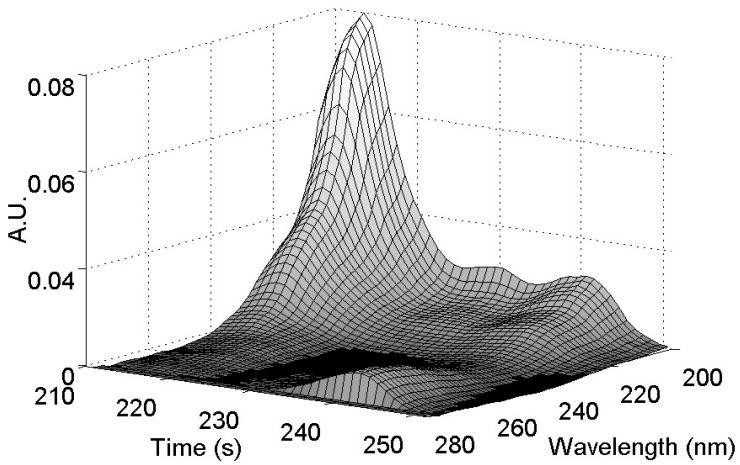


Figure 31. Three-dimensional plot of the elution window of tebuconazole, vinclozolin, clorfenvinphos and parathion-ethyl.

Alternating least squares (ALS) [17, 31] was used as multivariate curve resolution method to quantify the pesticides from the multicomponent HPLC-DAD data (selected region). MCR-ALS was applied to resolve the column-wise augmented matrices into individual concentration and spectral profiles, in an iterative least squares process in which constraints can be introduced in both concentration and spectral matrices. The constraints applied were non-negativity in all the modes

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

and equal shape and retention time in some chromatographic profiles (mixed bilinear–trilinear models).

Table 18. Definition of the calibration and prediction sets for cases A, B and C.

	Calibration samples	Prediction samples
Case A	Calibration ser #1: S7, S8, S10, S11, S15, S24	Prediction set #1: S1, S12, S16, S17, S25, S34
Case B	Calibration set #2: S7, S8, S10, S11, S15, S24, S2, S9, S21, S26, S32, S33	Prediction set #1: S1, S12, S16, S17, S25, S34
Case C	Calibration set #1: S7, S8, S10, S11, S15, S24	Prediction set #2: S1, S12, S16, S17, S25, S34,S3, S4, S5, S6, S13, S14, S18, S19, S20, S22, S23, S27, S28, S29, S30, S31

All the calculations were performed in MATLAB 6.5 (The MathWorks, 2002) using commercial [32, 33] and home-made routines.

5. Results and discussion

5.1. Rank analysis

After performing unfold PCA on the (34×2460) matrix, the dimensionality of the data was reduced from 2460 variables to four factors, which explained 99.5% of the variation in the data. Neither outliers nor different groups of data were detected in this preliminary evaluation.

The mathematical rank of each of the thirty-four 41×60 matrices was evaluated by singular value decomposition (SVD), as an estimation of the chemical rank or number of compounds in the mixtures. The number of statistically nonzero singular values (associated with the variance described by the eigenvector) was in all the cases between four and six depending on the sample. The matrices were also individually row-wise factor-analysed by means of EFA. From the

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

typical plot of eigenvalues as a function of the number of rows included in the window, we should have found four factors in each sample, theoretically corresponding to the four analytes. However, we found six factors in most of the samples. This behaviour pointed out systematic data variance other than the considered eluted components to be present. SIMPLISMA was also performed over an augmented matrix built up with the 34 samples. The pesticides' spectra were really correlated and seriously overlapped over the full spectral domain. To identify them, we had to allow a large deviation, which led on five factors with 23% of noise with respect to the maximum of the average spectrum. By selecting a smaller data set, SIMPLISMA could extract the expected four purest variables (with 15% of noise), which were in good agreement with those of reference, Vi, Cf, Te and Pe (see Figure 32), with correlations higher than 0.994 for Vi and Cf, 0.96 for Te and 0.925 for Pe.

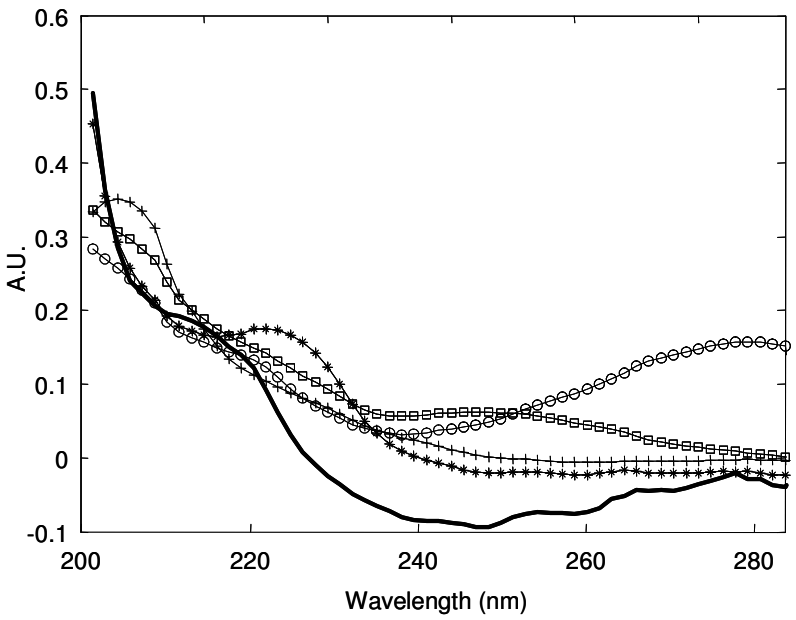


Figure 32. Spectra at purest time variables, corresponding to Vi (+), Cf (□), Te (∗), Pe (○) and the unexpected profile (solid line).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

The shape of the fifth UV–Vis spectra profile did not match to any of the pesticides. This might be due to different causes. One hypothesis is the contamination of the samples with an unexpected compound. Another hypothesis is the degradation of one of the pesticides from sample to sample. Since Vi, Cf and Pe were successfully determined and quantified in previous works [9,10], but not Te, we believed that tebuconazole possibly changed its spectra from sample to sample.

Selectivity of these pesticides was calculated in both the chromatographic and the spectral domains. Spectral selectivities, whose values range between 0 and 1, were estimated as described by Faber et al. [34] using the part of the signal due to the analyte, i.e. the net analyte signal (NAS), from the profiles recovered by MCR-ALS for one of the samples with known concentration (concretely, sample S34). As shown in Table 19, the spectral selectivities were low because the pesticides' spectra were quite overlapped.

Table 19. Spectral selectivity and chromatographical resolution of vinclozolin, clorfenvinphos, tebuconazole and parathion-ethyl, in the mixture S34.

Compound	Spectral selectivity	Chromatographical resolution
Vinclozolin	0.221	$Rs_{(Vi,\ Cf)}$=0.211
Clorfenvinphos	0.301	$Rs_{(Cf,\ Te)}$=0.682
Tebuconazole	0.279	$Rs_{(Te,\ Pe)}$=0.743
Parathion-ethyl	0.439	

Selectivities of the chromatograms (see Figure 33) were expressed in terms of chromatographic resolution, *Rs*. The first two elution profiles were quite overlapped (*Rs*=0.2) while the two last profiles showed larger resolution (*Rs*=0.74).

Despite of the bell-shaped chromatograms obtained by MCR-ALS for some samples such as those plotted in Figure 33, tebuconazole resolution was in most of the cases not successful. Its elution profile was wider than expected (as had been reported previously [12]) and irregular. In many resolutions, tebuconazole chromatogram was already present at the beginning of the time window, and it

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

was present all along the recorded range, meanwhile other compounds appeared and disappeared.

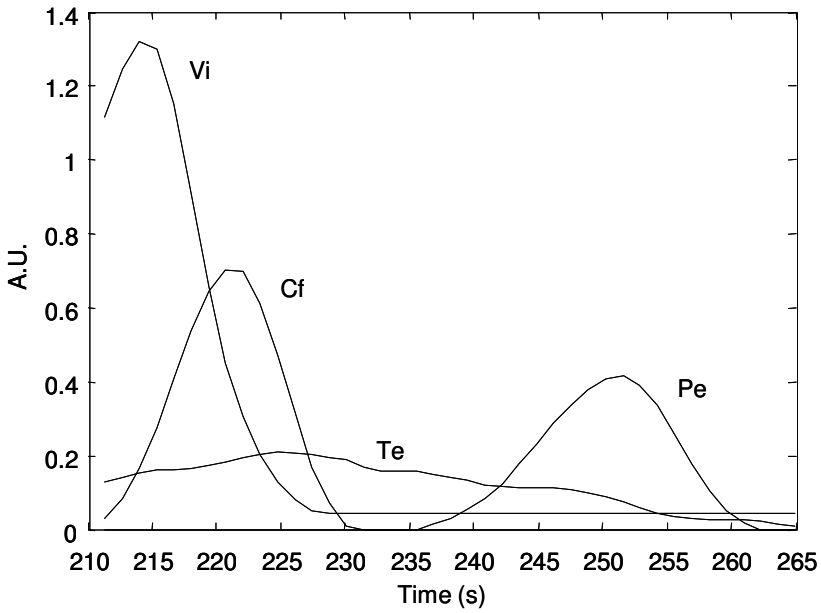


Figure 33. Pesticide's elution profiles estimated by MCR-ALS from one of the samples. Solution was constrained to non-negativity in both modes and, for Vi, Cf and Pe, also to trilinearity.

Although the main idea of this work was to give a practical example following the guidelines of a proper calibration and validation procedure (selection of the training and test sets, assessment of the figures of merit, etc.), further strategies were carried out to find out the proper number of species in the mixtures and their nature. For instance, we performed PARAllel FACtor (PARAFAC) [35] analysis with four and five factors. The fifth component described only noise in the data, since the corresponding loadings in the first mode (concentration) were approximately constant for all the mixtures and the loadings in the third mode (spectrum) were around the baseline. A model in a smaller time window using two factors (because only Te and Pe were expected to elute) and using three factors (in case one interfering compound was present) was also tested. We did a sample by sample meticulous analysis that showed the tebuconazole spectrum

evolved indeed from sample to sample. We finally measured a standard of Te along a day to check the hypothesis of degradation (by air or light, for example). None of the strategies could explain the abnormal number of factors, and work is in progress to find an answer.

5.2. MCR-ALS calibration and prediction

An augmented matrix of calibration and prediction samples was arranged by setting one matrix on top of the others and keeping the common absorption wavelengths in the same column. We evaluated three different combinations of calibration and prediction sets, as explained above, so that we had 12 samples in case A (n=6, m=6), 18 samples in case B (n=12, m=6) and 28 samples in case C (n=6, m=22) (see Table 18). We resolved each resulting augmented matrix with MCR-ALS using the chromatographic profiles obtained from the SIMPLISMA analysis as initial estimates to start the iterative algorithm. All the models were calculated considering four factors and restricting the solution to non-negative values (in the three modes). For each situation, the results of the curve resolution procedure were the particular elution profile in each of the samples and the pure spectra of the four pesticides underlying in all the mixtures. The recovery of these spectra was in the three models higher than 0.99 for Vi and Cf, 0.96 for Te and 0.98 for Pe.

We established a linear relationship between the area comprised under the estimated chromatograms and the reference concentration for the calibration samples. Figures of merit related with this calibration step are shown in Table 20. It can be seen that models A and C were very similar, since the calibration samples were the same. The differences arose when we included different prediction samples in the MCR-ALS procedure. In these two models, the goodness of the linear relation can be assessed from the small values of the residual deviation (or standard error) for Vi, Cf and Pe calibration, and from the correlation coefficients, higher than 0.995 for Cf and Pe, and 0.97 for Vi. Using more calibration samples, i.e. in model C, did not lead to a significant improvement, which indicates that the small calibration set sufficiently spanned the concentration range and the possible sources of variation. Tebuconazole was deficiently modelled in all cases, not achieving the validation requirements.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 20. Statistical parameters and figures of merit of the calibration curves for cases A, B and C when MCR-ALS solution was constrained to non-negativity in all the modes.

Model A	**Vi**	**Cf**	**Te**	**Pe**
Data points (*n*)	6	6	6	6
Intercept $\times 10^{-2}$	-0.3	-183	-26	-1.3
Sensitivity [a] $\times 10^{-2}$	12	96	31	12
S.D. Intercept $\times 10^{-2}$	7.2	28	63	3.0
S.D. slope $\times 10^{-2}$	1.4	4.5	10	0.61
Standard error $\times 10^{-2}$	8.9	30	60	3.7
Correlation coefficient (*r*)	0.97	0.996	0.8	0.995
Precision $\times 10^{-1}$ (%) [b]	7.5	3.1	25	3.0
LOD ($\mu g\ ml^{-1}$)	3.8	1.7	13	1.6
Prediction samples (*m*)	6	6	6	6
RMSPD (%)	20	13	46	16
Q^2 (%)	91	85	31	93

Model B	**Vi**	**Cf**	**Te**	**Pe**
Data points (*n*)	12	12	12	12
Intercept $\times 10^{-2}$	5.6	-152	-	-7.9
Sensitivity [a] $\times 10^{-2}$	10.8	88	-	13
S.D. Intercept $\times 10^{-2}$	6.1	33	-	4.5
S.D. slope $\times 10^{-2}$	1.2	5.6	-	0.89
Standard error $\times 10^{-2}$	8.9	47	-	6.9
Correlation coefficient (*r*)	0.94	0.98	<0.5	0.98
Precision $\times 10^{-1}$ (%) [b]	8.2	5.2	>40	5.3
LOD ($\mu g\ ml^{-1}$)	3.4	2.2	<0	2.2
Prediction samples (*m*)	6	6	6	6
RMSPD (%)	21	13	56	12
Q^2 (%)	93	89	<0	97

Model C	**Vi**	**Cf**	**Te**	**Pe**
Data points (*n*)	6	6	6	6
Intercept $\times 10^{-2}$	-2.0	-137	-36	-0.21
Sensitivity [a] $\times 10^{-2}$	12	78	33	12
S.D. Intercept $\times 10^{-2}$	7.3	20	69	4.1
S.D. slope $\times 10^{-2}$	1.4	3.3	11	0.84
Standard error $\times 10^{-2}$	9.0	22	65	5.1
Correlation coefficient (*r*)	0.97	0.997	0.8	0.990
Precision $\times 10^{-1}$ (%) [b]	7.5	2.8	19	4.2
LOD ($\mu g\ ml^{-1}$)	3.8	1.5	13	2.2
Prediction samples (*m*)	22	22	22	22
RMSPD (%)	17	23	79	17
Q^2 (%)	70	67	<0	59

[a] Defined as the slope of the calibration curve.
[b] Precision relative to the maximum concentration of the analyte in the calibration samples

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 21. Statistical parameters and figures of merit of the calibration curves for cases A, B and C when MCR-ALS solution was constrained to non-negativity in all the modes and, for Vi, Cf and Pe, also to trilinearity.

Model A	**Vi**	**Cf**	**Te**	**Pe**
Data points (*n*)	6	6	6	6
Intercept $\times 10^{-2}$	1.1	-14	-	-0.95
Sensitivity [a] $\times 10^{-2}$	9.4	8.6	-	3.5
S.D. Intercept $\times 10^{-2}$	4.2	1.3	-	0.68
S.D. slope $\times 10^{-2}$	0.80	0.21	-	0.14
Standard error $\times 10^{-2}$	5.1	1.4	-	0.85
Correlation coefficient (*r*)	0.990	0.999	<0.60	0.997
Precision $\times 10^{-1}$ (%) [b]	5.5	1.6	-	2.4
LOD (μg ml^{-1})	2.8	0.92	-	1.3
Prediction samples (*m*)	6	6	6	6
RMSPD (%)	16	11	-	13
Q^2 (%)	90	88	-	98

Model B	**Vi**	**Cf**	**Te**	**Pe**
Data points (*n*)	12	12	12	12
Intercept $\times 10^{-2}$	6.7	-9.8	-	-2.4
Sensitivity [a] $\times 10^{-2}$	8.2	6.8	-	3.7
S.D. Intercept $\times 10^{-2}$	2.5	2.0	-	1.0
S.D. slope $\times 10^{-2}$	0.50	0.33	-	0.20
Standard error $\times 10^{-2}$	3.7	2.8	-	1.6
Correlation coefficient (*r*)	0.98	0.990	<0.50	0.98
Precision $\times 10^{-1}$ (%) [b]	4.5	4.1	-	4.3
LOD (μg ml^{-1})	1.9	1.8	-	1.8
Prediction samples (*m*)	6	6	6	6
RMSPD (%)	7	10	-	8
Q^2 (%)	99	93	-	99

Model C	**Vi**	**Cf**	**Te**	**Pe**
Data points (*n*)	6	6	6	6
Intercept $\times 10^{-2}$	4.1	-10	-	-0.58
Sensitivity [a] $\times 10^{-2}$	8.5	7.1	-	3.5
S.D. Intercept $\times 10^{-2}$	3.1	1.7	-	0.52
S.D. slope $\times 10^{-2}$	0.60	0.27	-	0.11
Standard error $\times 10^{-2}$	3.8	1.8	-	0.66
Correlation coefficient (*r*)	0.990	0.997	<0.60	0.998
Precision $\times 10^{-1}$ (%) [b]	4.5	2.6	-	1.9
LOD (μg ml^{-1})	2.3	1.4	-	1.0
Prediction samples (*m*)	22	22	22	22
RMSPD (%)	8	24	-	13
Q^2 (%)	95	83	-	71

[a] Defined as the slope of the calibration curve.
[b] Precision relative to the maximum concentration of the analyte in the calibration samples

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Prediction step was carried out by interpolating into the corresponding curve the area under the chromatograms for the prediction samples, also estimated by the MCR-ALS procedure. For cases A and B, where the prediction set was the same and only the calibration set differed, prediction errors and the explained variance for the prediction samples were similar. Concretely in model B, the most favourable case, Q^2 was 93% for Vi, 89% for Cf and 97% for Pe. In the prediction of Te, the models only accounted for less than 50% of the variance.

When predicting a higher number of samples (case C), the prediction parameters were in general worst, explaining around the 70% of the variance in the predictions set for Vi and Cf and about the 60% for Pe. Regarding Te, prediction was again not possible because it was not properly calibrated. One possible explanation is that, in the mixtures we studied, tebuconazole seemed to contravene one of the necessary conditions leading to uniqueness (unique resolution), described as *resolution theorems* by Manne [36] (as mentioned above, the resolved elution profile for Te was in most of the cases wider than expected, so that, for every interfering pesticide, tebuconazole had a subwindow where the interference was absent, and then it is mathematically possible to calculate the spectrum of this analyte; but, on the other hand, Cf and Pe elution profiles, which appeared inside the concentration windows of tebuconazole, did not appear outside its window, and therefore, it is not possible to calculate without ambiguities the concentration profile of Te).

In case that the origin of the problem was a decomposition process or an external contamination of the mixtures, we imposed a trilinearity constraint to Vi, Cf and Pe, i.e. their elution profiles were supposed to be trilinear (same shape, only different intensity from sample to sample) and synchronized (same elution time, no significant time shifting), and allowed the fourth compound to change from sample to sample, i.e. bilinearity but no trilinearity was assumed. With this additional constraint applied only to some of the compounds, we performed the mixed bilinear–trilinear models following the same calibration and prediction sets definition as in cases A, B and C. Recovered spectra were well correlated with reference spectra in all the models, with correlation coefficients above 0.99 for Vi, Cf, above 0.98 for Pe and above 0.97 for Te. From the recovered elution profiles corresponding to the calibration samples, we established the linear relation with the known concentration. Figures of merit for these calibration curves are shown in Table 21. Vi, Cf and Pe calibration clearly improved, achieving better

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

correlation coefficients and lower detection limits in all the cases. Prediction of their content in the new samples also enhanced considerably: compared with the previous models, the RMSPD values were lower in all the cases and the explained variance (Q^2) increased up to 99% for Vi, up to 93% for Cf and up to 99.4% for Pe, e.g. in model B (see Table 21).

5.3. Determination of the pesticides in groundwater samples

To test the performance of the proposed methodology, eight groundwater samples were spiked at levels between 0 and 10 μg l^{-1} (Table 17) and pre-treated and analysed by HPLC-DAD as described in the literature [9]. Spectrochromatographic data were processed as explained above and two augmented matrices were built, by adding column-wise the groundwater samples (1) to the calibration set of six standards (case 1, likewise in model A) and (2) to the calibration set of 12 standards (case 2, likewise in model B). Both MCR-ALS models were performed with four factors and applying non-negativity in all the modes. The trilinearity constraint was applied to three of the compounds (since that improved significantly the prediction), so that we built mixed bilinear–trilinear models. The spectra recovered by the two models were very similar, showing good correlation with the reference ones: 0.995 for Vi, 0.993 for Cf, 0.97 for Te and 0.97 (in case1) or 0.98 (in case 2) for Pe. For each case, areas under elution profiles from the natural samples were interpolated in the previously calibration curve relating the areas and the known concentration corresponding to the calibration samples. The concentrations of the pesticides in the eight groundwater samples were estimated from these calibration curves. In both cases under study, tebuconazole responses did not show behaviour linear enough and therefore calibration was, as expected, not satisfactory. For the other three pesticides, in both cases the models were appropriate to predict the groundwater samples. Figure 34 shows the calibration curves for each pesticide and the prediction for the new samples. Case 2, where more calibration samples were used for the curve resolution procedure, provided better predictions, i.e. lower RMSPD and higher Q^2-values. The curve for Vi showed high residuals and the prediction samples with the higher concentration did not seem to follow the same linear tendency, and so the explained variance was only 88%.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

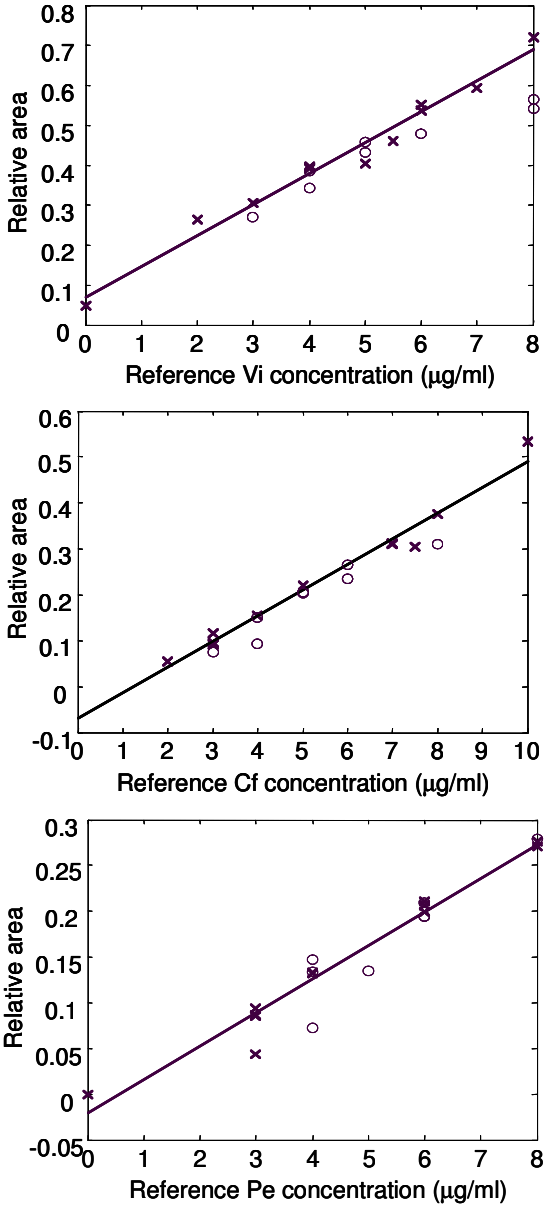


Figure 34. Calibration curves (in case 2, see details in text) for Vi, Cf and Pe. Crosses are the calibration samples and circles the natural samples to be predicted.

The calibration curve for Cf was appropriate for the prediction of the unknown concentrations. With that model, we explained the 96% of the variance in the prediction set. In the Pe model, we could detect one sample (AG2) with lower response than the others with the same Pe concentration (AG6 and AG8), pointing it as a probable prediction outlier. The explained variance for the full prediction set was 94%, but it could be increased up to 98% if we discard sample AG2. In general, the Q^2-values were lower and the RMSPD values higher for the prediction of natural samples than the values obtained for the prediction of synthetic samples, which can be explained from the increase in the matrix complexity.

6. Conclusion

In this work, we have faced the actual analytical problem of determining pesticides in complex mixtures. The complexity of these environmental samples was aforementioned in previous works. Actually, curve resolution assays showed some anomaly in the peak purities regarding one of the compounds. Calibration of the other pesticides under study was performed by bilinear–trilinear MCR-ALS, and the validity of these models was assessed by calculating the figures of merit. A data set of natural samples was also analysed and predicted with this procedure, giving good results for three of the pesticides. The quantitative results given in this paper could be improved including samples with more selective signals for each analyte, i.e. pure standards, if available, which is actually a very common practice.

Always, the user must decide whether the validation parameters fit the particular purpose. This work provides a reasonable practice of validation for a method generating second-order data and gives the tools to decide whether an analytical methodology can be used for routine analysis.

References

[1] ISO/IEC 17025:1999, “General Requirements for the Competence of Testing and Calibration Laboratories”, 1999

[2] M.J. Rodríguez-Cuesta, R. Boqué, F.X. Rius, Anal. Chim. Acta 476: 111-122 (2003)

[3] M.J. Rodríguez-Cuesta, R. Boqué, F.X. Rius, D. Picón Zamora, M. Martínez Galera, A. Garrido Frenich, Anal. Chim. Acta 491: 47-56 (2003)

[4] M. Martínez Galera, J.L. Martínez Vidal, A. Garrido Frenich, M.D. Gil García, J. Chromatogr., A 778: 139-149 (1997)

[5] S.M. Garland, R.C. Menary, N.W. Davies, J. Agric. Food Chem. 47: 294-298 (1999)

[6] J.W. Wong, M.G. Webster, C.A. Halverson, M.J. Hengel, K.K. Ngim, S.E. Ebeler, J. Agric. Food Chem. 51: 1148-1161 (2003)

[7] P. Cabras, A. Angioni, V.L. Garau, M. Melis, F.M. Pirisi, E.V. Minelli, F. Cabitza, M. Cubeddu, J. Agric. Food Chem. 45: 2708-2710 (1997)

[8] Y.S. Wu, H.K. Lee, S.F.Y. Li, J. Chromatogr., A 912: 171-179 (2001)

[9] M. Martínez Galera, A. Garrido Frenich, J.L. Martínez Vidal, J. Liq. Chromatogr. Relat. Technol. 23: 1187-1202 (2000)

[10] A. Garrido Frenich, J.L. Martínez Vidal, M. Martínez Galera, Anal. Chem. 71: 4844-4850 (1999)

[11] A. Garrido Frenich, J.R. Torres-Lapasió, K. De Braekeleer, D.L. Massart, J.L. Martínez Vidal, M. Martínez Galera, J. Chromatogr., A 855: 487-499 (1999)

[12] A. Garrido Frenich, M. Martínez Galera, J.L. Martínez Vidal, D.L. Massart, J.R. Torres-Lapasió, K. De Braekeleer, Ji-Hong Wang, P.K. Hopke, Anal. Chim. Acta 411: 145-155 (2000)

[13] O.M. Kvalheim, Y.Z. Liang, Anal. Chem. 64: 936-946 (1992)

[14] Y.Z. Liang, O.M. Kvalheim, Anal. Chem. 64: 953-956 (1992)

[15] F. Gan, Q.S. Xu, Y.Z. Liang, Analyst 126: 161-168 (2001)

[16] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173: 253-264 (1985)

[17] R. Tauler, Chemom. Intell. Lab. Syst. 30: 133-146 (1985)

[18] A. Espinosa-Mansilla, A.M. de la Pena, T.C. Goicoechea, A.C. Olivieri, Appl. Spectrosc. 58: 83-90 (2004)

[19] J.S. Salau, M. Honing, R. Tauler, D. Barceló, J. Chromatogr., A 795: 3-12 (1998)

[20] R. Tauler, S. Lacorte, D. Barceló, J. Chromatogr., A 730: 177-183 (1996)

[21] S. Lacorte, D. Barceló, R. Tauler, J. Chromatogr., A 697: 345-355 (1995)

[22] R. Tauler, A.K. Smilde, B.R. Kowalski, J. Chemom. 9: 31-58 (1995)

[23] A. de Juan, Y. Vander Hieden, R. Tauler, D.L. Massart, Anal. Chim. Acta 346: 307-318 (1997)

[24] F. Gan, P.K. Hopke, Anal. Chim. Acta 495: 195-203 (2003)

[25] L.A. Currie, Commission on analytical nomenclature, nomenclature in evaluation of analytical methods including detection and quantification capabilities, Pure Appl. Chem. 67: 1699-1723 (1995)

[26] N.J. Miller-Ihli, T.C. O'Haver, J.M. Harnly, Spectrochim. Acta, Part B: Atom. Spectrosc. 39: 1603-1614 (1984)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

[27] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Introduction to Multi- and Megavariate Data Analysis Using Projection Methods (PCA/PLS), Umetrics AB, Ume3, Sweden, 1999

[28] H.R. Keller, D.L. Massart, Chemom. Intell. Lab. Syst. 12: 209-224 (1992)

[29] M. Maeder, Anal. Chem. 59: 527-539 (1987)

[30] B.G.M. 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, Chapt. 34

[31] R. Tauler, D. Barceló, Trends Anal. Chem. 12: 319-327 (1993)

[32] Tauler, R., de Juan, A., MATLAB program MCR-ALS.
URL: http:// www.ub.es/gesq/mcr/mcr.htm

[33] C.A. Andersson, R. Bro, The N-way toolbox for MATLAB ver. 2.00, Chemom. Intell. Lab. Syst. 52(1): 1-4 (2000). URL: http://www. models.kvl.dk

[34] N.M. Faber, A. Lorber, B.R. Kowalski, J. Chemom. 11: 419-461 (1997)

[35] R. Bro, Multiway analysis in the food industry: models, algorithms, and applications, PhD thesis, University of Amsterdam, 1998
URL: http://www.mli.kvl.dk/staff/foodtech/brothesis.pdf, sections 3.3 and 4.3

[36] R. Manne, Chemom. Intell. Lab. Syst. 27: 89-94 (1995)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 6

ESTIMATION OF THE LIMIT OF DETECTION FOR MULTI-LINEAR PARTIAL LEAST SQUARES

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

6.1. Introduction

In winter 2004 I spent four months with the Spectroscopy and Chemometrics group of the Quality and Technology Department of Food Science at the Royal Veterinary and Agricultural University (KVL) in Copenhagen, Denmark. During that research stay, I was supervised by Prof. Rasmus Bro. What began as an opportunity to further investigate chemometric techniques also become the source of interesting discussions. Out of these discussions came the idea behind the paper presented in this chapter.

The aim of the paper entitled "*Standard error of prediction at low content levels and limit of detection estimation for multivariate and multi-linear regression*" was to develop a LOD estimator for a genuine three-way prediction method: multi-linear partial least squares (N-PLS). In N-PLS, resolution capacity is lost in favour of higher prediction ability.

A second objective was to study the influence of the predicted sample, through its leverage and uncertainty of prediction, on the LOD estimated. For samples at (or close to) zero concentration, the distribution of responses (net analyte signal) may depend on the amount and nature of interferences. LOD is usually considered to be a specific parameter of an analytical method but, if the standard deviation of prediction were statistically different for samples containing different analyte/interferences ratios, this would lead to a sample-specific LOD and mean that the whole concept of LOD would have to be re-evaluated.

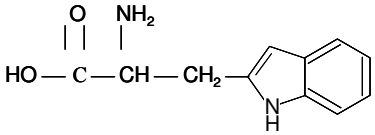
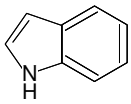
As in N-PLS the resolution capacity is lost in favour of higher prediction ability, and as our second objective was to study the influence of the predicted sample, through its leverage, on the LOD estimated, there was a different LOD for every sample.

Finally, we also studied how the interferences (degree of overlapping and concentration level) affect the performance of the calibration model. This would make the LOD model-dependent for calibration samples with the same concentration of the analyte of interest. To study this hypothesis, we evaluated the standard error of prediction at low content (zero) level in the multivariate and multi-way scenarios. In both cases, complex three-component systems were

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

reproduced using the excitation-emission matrix of tryptophan[30], indole[31] and hydroquinone[32]. The molecular structures of these compounds are shown in Table 22.

Table 22. Molecular structure of the compounds under study.

Commercial name	Chemical name	Structure
Tryptophan	(S)-2-amino-3-(1H-indol-e-yl)-propanoic acid CAS: 73-22-2	
Indole	2,3-benzopyrrole CAS: 120-72-9	
Hydroquinone	Benzene-1,4-diol CAS: 123-31-9	

One conclusion of this paper was that the main source of variability is the replicate, especially for low levels of interferences. However, as the level of interferences increases, so does the variability due to the calibration model. This may be due to the effect of the increasing leverage of the prediction samples, since the average SEP_0 also increases as the level of interferences increases.

The higher the level of interferences, the greater the influence of the fit of the model to the calibration data on the variability of the predicted concentrations. This is because small changes in fit, i.e. small changes in the calibration model,

[30] Tryptophan is one of the 20 amino acids in the genetic code and is essential in human nutrition. Its main function is as a building block in protein synthesis. Only the L-stereoisomer appears in mammalian protein.

[31] Indole is an aromatic heterocyclic organic compound. It is solid at room temperature and has an intense fecal smell. At very low concentrations, however, it has a flowery smel, and is a constituent of many flower scents and perfumes.

[32] Hydroquinone is an aromatic organic compound with several uses that are mainly associated with its action as a reducing agent that is soluble in water.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

have more effect on the ends of the domain than in the centre and the samples that are far from the centre therefore have a higher increase in their leverage. If the new prediction samples are similar to the calibration samples, they will have low leverages. Only when the new samples are far from the experimental domain of calibration will the prediction samples have high leverages and therefore high SEP. In our example, the calibration samples cover the experimental domain well. When the samples of the calibration set are not so homogeneously distributed, for instance with a distribution of interferences that does not cover the entire experimental domain, leverage has a greater effect on prediction.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

-

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

6.2. Paper. Standard error of prediction at low content levels and limit of detection estimation for multivariate and multi-linear regression.

Chemometrics and Intelligent Laboratory Systems. *In press*

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

-

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Standard error of prediction at low content levels and limit of detection estimation for multivariate and multi-linear regression

M.J. Rodríguez-Cuesta, Ricard Boqué

Department of Analytical and Organic Chemistry, Rovira I Virgili University
Campus Sescelades. 43005-Tarragona (Catalunya). SPAIN

Rasmus Bro

Food Technology, Royal Veterinary and Agricultural University
Rolighedsvej 30, 1958 Frederiksberg C. DENMARK

Nicolaas (Klaas) M. Faber

Chemometry Consultancy, Rubensstraat 7, 6717 VD Ede. THE NETHERLANDS

Abstract

The internationally accepted guidelines concerning analytical detection are intrinsically bound to the probability of committing false positive and false negative errors, as assessed from the probability density functions of the response of samples with concentrations at zero and at the limit of detection (LOD) levels, respectively. For the calculation of the LOD, accurate estimates of the standard deviation at these two low levels are needed. The distribution of the responses at zero concentration is here shown to be statistically different for samples containing different analyte/interferents ratios, and this leads to a sample-specific LOD in contrast to the traditional method-dependent LOD concept.

Calibration of an analyte of interest was performed by multivariate PLS (two-way PLS) and by multi-linear PLS (three-way PLS or N-PLS) in the neighbourhood of the presupposed detection level. The standard deviations for blanks, with and without interferents, were compared. LODs were calculated from these standard

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

errors of prediction and validated using samples with known content of the analyte, by calculating the percentage of false positive and false negative errors. The proposed LODs fulfil the validation requirement both in multivariate and in multi-way situations, in contrast to the traditional-based procedures, which were proved in most of the cases to overestimate the detection limits. Some practices for LOD estimation are suggested, based on a capability concept linked to the distribution of the predicted concentration for the new test set.

1. Introduction

According to the IUPAC Nomenclature rules [1], "*the performance characteristics of the chemical analytical processes are those quantifiable terms(…) which are closely related to the method and to the analyte, like: sensitivity, selectivity, limit of detection, limit of determination, etc. (…)*". The limit of detection (LOD), commonly defined as the minimum amount or concentration of substance that can be reliably detected [2], is a performance characteristic of a given analytical method and thus, an a priori defined value.

However, for samples at zero concentration it is possible that the distribution of responses (net analyte signal) and hence the distribution of the predicted concentrations may differ as a function of the amount and nature of the interferents. As introduced, LOD is usually considered as *one* specific parameter of an analytical method, but if the standard deviation of prediction would be statistically different for samples containing different analyte/interferents ratios, then it would lead to a sample-specific LOD, implying that the whole concept of LOD may have to be re-evaluated.

For the calculation of the LOD, accurate estimates of the uncertainty at zero and at the LOD levels are needed. Although the only generally accepted approach to prediction uncertainty in multivariate calibration is to use an overall measure such as the root mean square error of prediction (RMSEP), the necessity of a sample-specific prediction uncertainty to yield realistic prediction intervals is fully justified [3]. For univariate calibration, the standard error of prediction (uncertainty) for a given concentration level can be estimated from the confidence limits of the regression line [4, 5, 6]. Since an inherent assumption of such zeroth-order calibration methods is to have full selectivity for the analyte of interest -otherwise

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

estimations are biased-, this standard deviation is estimated in the exemption of interferents situation (or assuming the interferents' contribution to be constant so that they can be removed). When going beyond zeroth-order calibration, a full selectivity of the analyte of interest is not mandatory. The incorporation of the interferents in the multivariate or first-order models means that the distribution of the response (net analyte signal) and of the predicted concentration can differ as a function of the amount and nature of the interferents. This subject has been widely dealt with in multivariate calibration, resulting in several discussions and expressions for estimating sample-specific standard error of prediction for partial least squares regression (PLSR) [7-14], most incorporating an indirect dependence on the amount of interferents (usually through a leverage-expression).

The purpose of the current work is to assess whether the error of prediction depends on the level of interferents and to estimate a limit of detection on such basis. We work in two scenarios: in the multivariate scenario we use first-order data and models (excitation spectra from a set of samples with one analyte of interest and two interfering compounds at different proportions, which are modelled by multivariate partial least squares, PLS) and in the multi-linear scenario we used second-order excitation-emission matrices (EEMs) for samples with one analyte of interest and two interfering compounds at different proportions, which are modelled by multi-linear PLS or N-PLS.

We simulate samples at zero and LOD levels of the analyte of interest with different level of interferents and compare the calculated method-characteristic LOD with the proposed sample-specific LOD estimated from the error of prediction.

2. Theory

2.1 Preliminaries

Partial least squares regression is a widespread regression technique whose basis and algorithms can be easily found both for multivariate [15-16] and for multi-linear [17-18] PLS. Whatever the order of the data is, the model establishes a relation between the independent variables (**X**-block), e.g. spectra, and the dependent variables (**Y**-block), e.g. analyte concentrations, in such a way that the covariance between **X** and **Y** is maximized in the new reduced data space

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

defined by the new variables called PLS factors. Typically, the final goal of the calibration is to predict the concentration of a given analyte, $\hat{y}$, in an unknown sample from its instrumental response. Although the true concentration y is in practice unknown, the prediction error (PE) is defined as $PE \equiv \hat{y} - y$ and the standard error of prediction (SEP) is defined as the square root of the prediction error variance (prediction uncertainty).

2.2. Estimation of the sample-specific standard error of prediction

There are several expressions for standard error of prediction suitable for multivariate PLS calibration (see references [7-20] in [19]). For the case of negligible error in the independent variables, the simplest approach (ignoring the most complicated terms) was first proposed by Höskuldsson [20]. An equivalent expression can be found in ASTM (ASTM Standard E1655). On the basis of such approach, Faber and Bro [19] derived an expression for multi-linear partial least squares (N-PLS) to estimate sample-specific standard errors of prediction. It was recently applied to a fluorescence spectroscopic calibration problem where N-PLS regression was appropriate, highlighting the validity and the generality of the proposed expression [21].

In the absence of bias, the sample-specific root mean squared error of prediction (standard error of prediction or SEP, for shortness) as derived by Faber and Bro depends on the leverage for the sample and on the mean square error of calibration (MSEC):

$$SEP_i = \left[(1 + h_i)\, MSEC - V_{\Delta y}\right]^{1/2} \qquad \text{Eq. 47}$$

For zero-intercept models, the MSEC is obtained in the usual way from the squared fit errors as

$$MSEC = \frac{\sum_{i=1}^{I} (\hat{y}_i - y_i)^2}{\nu} \qquad \text{Eq. 48}$$

where I denotes the number of samples in the calibration set. The degrees of freedom in Equation 48 are calculated as $\nu = I - F$, where F are the number of

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

factors used in the model (mean-centring, if it is applied, is accounted for by subtracting one additional degree of freedom). General requirements on the appropriateness of the MSEC (no significant changes by the in- or exclusion of one sample, consistency with the cross-validated MSE, etc) should necessarily be assessed to guarantee a reliable estimate of SEP.

The leverage, denoted by h, quantifies the Mahalanobis distance from the prediction sample to the mean of the calibration set in the F-dimensional space. For zero-intercept models, the leverage is calculated as $h=\mathbf{t}^T(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{t}$ (mean-centring is accounted by adding an $1/I$ term), where $\mathbf{t}$ is the $F\times1$ score vector of the unknown sample and $\mathbf{T}$ the $I\times F$ matrix of scores for the calibration set.

The correction term $V_{\Delta y}$ is an estimate of the measurement error variance associated with the reference method; if unknown, it can be set to zero leading to a more pessimistic (higher) SEP.

The application of Equation 47 only requires the standard deviation of the measurement error in the reference values, in contrast to prior expressions (see Eq. (8) in [19]) where estimates of the error variances for both predictor and predictand variables were required. Moreover, this expression is designed to work not only in the homoscedastic case but also in heteroscedastic conditions.

2.3. SEP dependence on the level of interferents

From Equation 47 it follows that SEP may actually vary for zero-analyte samples if the leverage varies. Assuming the expression is valid, it therefore immediately follows that LOD can not be constant even for zero-analyte samples for a given model if the level of interferents vary such that the leverages differ. This though relies on the adequacy of the approximation in Equation 48 and also on the extent to which the leverages vary significantly in realistic settings.

To determine whether and to what extent the prediction uncertainty (SEP) depends on the level of interferents, it should be assessed whether the predictions of samples without analyte of interest, i.e. $\hat{y}_0$, and different amounts of interferent species, actually include the *true* (reference) concentration, i.e. $y_0 = 0$. For the theory of hypothesis testing, the risk one accepts of committing a false positive decision is given by the value of α, and the risk of committing a

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

false negative decision is given by the value of β. Theoretically, for a given two-tailed probability density function of e.g. predicted concentrations, 100(1–2(α/2))% of the prediction intervals should contain the zero, while 100(α/2)% of the predictions will be in the right side as false positives, i.e. stating that the analyte is present when in fact it is not.

The prediction intervals can not be set up from Equation 47 because it does not lead to *t*-statistics: subtracting $V_{\Delta y}$ not only may lead negative variance estimates for relatively large measurement error variances, but also induce prediction variance need not be approximately distributed proportional to a simple χ^2 (see [19, 22]). Alternatively, the prediction intervals (*PI*≡ predicted concentration ± associated uncertainty) are built with the endpoints indicated in Equation 45 using the $(1-2(\alpha/2))^{th}$ percentile of a Student's *t*-distribution with $\nu=I-F$ degrees of freedom, where *I* is the size of the population (number of samples in the prediction set) and SEP_0 is the standard error of prediction for the sample without the analyte of interest.

$$PI = \hat{y}_0 \pm t_{1-\alpha/2,\nu} SEP_0 \qquad \text{Eq. 49}$$

It is essential to construct a large number of independent (randomly generated) models to calculate an accurate average value of the percentage predictions that include zero and compare it to the nominal value, e.g. 100(1-2(α/2))=95% for α=0.05.

2.4. LOD estimation

The decision of whether a given analyte is present or not in a sample is based on a comparison of the estimated quantity with a critical level, y_C, such that the probability of exceeding y_C is no greater than α if analyte is absent [23]. This decision is then taken a posteriori, i.e. once the sample has been measured. However, LOD has been defined as an a priori parameter [23-24], characteristic of the analytical method at hand. ISO [24] defines the LOD (they use the term minimum detectable net concentration or amount) as the true net concentration or amount of the analyte in the material to be analysed which will lead, with probability (1-β) to the conclusion that the concentration or amount of the analyte in the analysed material is larger than that in the blank material. Analytical

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

detection guidelines are hence intrinsically bound to the probability of committing false positive and false negative errors. These come from the probability density functions of the response of samples with concentrations at zero and at the LOD levels, respectively, as shown in Figure 35.

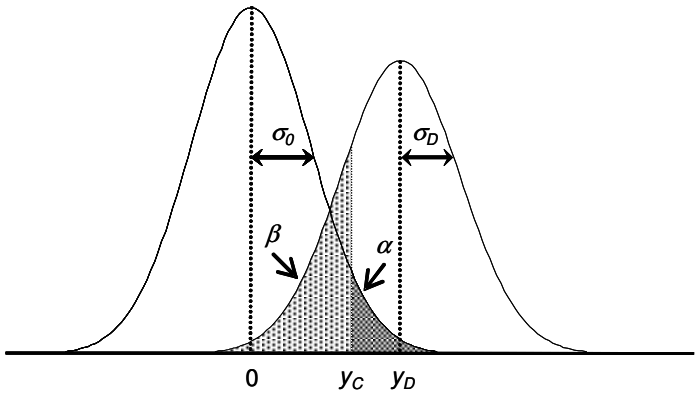


Figure 35. Representation of the density probability functions at zero and LOD levels, with the probability of committing false positives (α) and false negatives (β) (see details in text).

According to IUPAC Recommendations [23], the critical value y_C for making the detection decision can be estimated as $\hat{y}_C = z_{1-\alpha}\,\sigma_0$ and the minimum detectable value y_D or *inherent* detection capability of a chemical measurement process as $\hat{y}_D = \hat{y}_C + z_{1-\beta}\,\sigma_D$. Here $z_{1-\alpha}$ and $z_{1-\beta}$ represent respectively the (1-α)th and (1-β)th percentage point of the standard normal variable and σ is the standard deviation of the estimated quantity. Assuming that the variances of the predicted concentrations $\hat{y}_0$ and $\hat{y}_D$ are the same and taking the default values for α and β (i.e. 0.05), the previous expression reduces to $(z_{1-\alpha}+z_{1-\beta})\sigma_0$. If the standard deviation is replaced by an estimate s_0 based on ν degrees of freedom, then the quantile z must be replaced by Student's-t and $(z_{1-\alpha}+z_{1-\beta})$ must be replaced by the non-centrality parameter of the non-central-t distribution, $\delta_{\alpha,\,\beta,\,\nu}$.

On these statistical bases, Hubaux and Vos [4] introduced an estimator of the limit of detection from linear calibration curves. This LOD estimator was

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

connected to the concept of confidence limits and emphasized its dependence upon the calibration standards, as it is thereby explicit in the summary of expressions shown in the Appendix.

Moving to higher-order calibrations, however, it has been recently reported by Ortiz *et al*. [25] that the limit of detection (as defined by the ISO in the norm 11843 [23, 26]) was not suitable for soft multivariate or multi-way calibration methods since a vector or a matrix of numbers is available for each analyzed sample. Based on the fact that such LOD is invariant for linear transformations of the response variable, Ortiz *et al*. calculated the minimum detectable net concentration from the linear regression between the concentration of the analyte, y_i, and that calculated with the (two- or three-way) PLS model of calibration, $\hat{y}_i$ (see Appendix).

In the current work, the critical value $\hat{y}_C$ and the limit of detection $\hat{y}_D$ were estimated following the IUPAC recommended expressions based on the standard deviation of the estimated concentration at low content levels. The singular contribution lies in the use of the standard error of prediction of Equation 47, yielding the following expressions,

$$\hat{y}_C = t_{1-\alpha,\nu}\, SEP_0 = t_{1-\alpha,\nu} \sqrt{(1+h_0)\, MSEC} \qquad \text{Eq. 50}$$

$$\hat{y}_D = \delta_{\alpha,\beta,\nu}\, SEP_0 = \delta_{\alpha,\beta,\nu} \sqrt{(1+h_0)\, MSEC} \qquad \text{Eq. 51}$$

Note that Equation 51 is equivalent to Equation 49 (and to Eq. 5b in terms of concentration) of ref. 6, proposed by Currie, for the univariate case. Moreover, SEP_0 in these equations is analogous, except for the term $V_{\Delta y}$, to Equation 46 and paragraph above in ref. 6.

3. Description of the simulations

The standard error of prediction at zero content level was evaluated in the multivariate and in the multi-way scenarios. For both, three-component systems were simulated, in such a way that the response of the analyte of interest overlapped the response of the two interferent species, at different ratios. The

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

suitability of using simulated data is justified from the necessity of having as many observations as required to demonstrate a difference in performance, in addition to knowing their "true" concentration.

To build the simulated first-order data we used the excitation spectra of tryptophan, hydroquinone and indole because of their moderate/high overlap. Tryptophan was the analyte of interest. Its spectrum is highly correlated with that of hydroquinone (correlation coefficient, *r*, was 0.93) and moderately correlated with indole (*r*=0.49). The left box in Figure 36 shows the three normalized spectra, **S**. The multivariate data **X** for each sample were built multiplying the corresponding diagonal matrix of concentrations, **C** (generally speaking, the **Y**-block), by the normalized spectra transpose matrix, $\mathbf{S}^T$, and summing these weighted spectra.

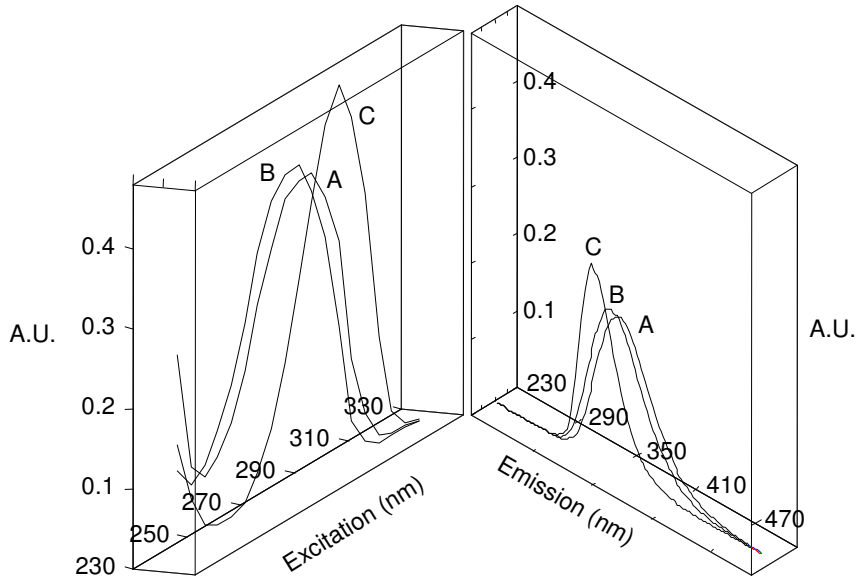


Figure 36. Normalized excitation (left box) and emission (right box) spectra of tryptophan (A), hydroquinone (B) and indole (C).

The simulated second-order data, excitation-emission matrices (EEMs), were built using the normalized excitation and emission spectra of tryptophan, hydroquinone and indole. Overlap in the excitation mode has already been

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

shown. In the emission mode there was also a strong overlap, with correlations of r=0.96 between tryptophan and hydroquinone and r=0.59 between tryptophan and indole (right box in Figure 36).

The second-order data for each sample were built by multiplying the normalized emission spectra, **A**, by the corresponding diagonal matrix of concentrations, **B**, and by the transpose of the normalized excitation spectra matrix, $\mathbf{C}^T$. These weighted analyte EEMs were summed to give the overall EEM, $\underline{\mathbf{X}}$.

Both to the multivariate and the multi-way data, white noise, **E**, was added such that the noise level was approximately 2% in terms of relative standard deviation at the maximum of the response.

Table 23. Concentration of the three components in the 20 calibration samples.

Sample	Tryptophan	Hydroquinone	Indole
1	0	0.1	0.1
2	0	0.1	2.0
3	0	2.0	0.1
4	0	2.0	2.0
5	0.5	0.1	0.1
6	0.5	0.1	2.0
7	0.5	2.0	0.1
8	0.5	2.0	2.0
9	1.0	0.1	0.1
10	1.0	0.1	2.0
11	1.0	2.0	0.1
12	1.0	2.0	2.0
13	1.5	0.1	0.1
14	1.5	0.1	2.0
15	1.5	2.0	0.1
16	1.5	2.0	2.0
17	2.0	0.1	0.1
18	2.0	0.1	2.0
19	2.0	2.0	0.1
20	2.0	2.0	2.0

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Twenty calibration samples were defined in order to relate their (multivariate or multi-way) responses with the concentration of the analyte of interest, tryptophan, in the range from 0 to 2.0 (arbitrary) concentration units. For each level of tryptophan, the interferent analyte concentrations were distributed following a full factorial design 2^2 (low level 0.10 c.u., high level 2.0 c.u.) in order to span the range of concentrations of interferents between 0.10 and 2.0 c.u.. These concentrations (**Y**-block) are shown in Table 23. No noise was added to the concentration and no pre-treatment was applied on the data.

The calibration models were built with two- or three-way PLS, respectively, using commercial [27] and home-made routines performed in MATLAB 6.5 (The MathWorks Inc., 2002). The models were all different due to the randomly distributed error added to the multivariate response, i.e. $\mathbf{X}=\mathbf{C}\mathbf{S}^T+\mathbf{E}$, or to the multi-way response. As an illustration, Figure 37 shows the multivariate data (rows in the two-dimensional **X**-block) used in one of the two-way PLS models and Figure 38 shows the EEM of one of the calibration samples (one slab in the three-dimensional **X**-block) used in one of the N-PLS models.

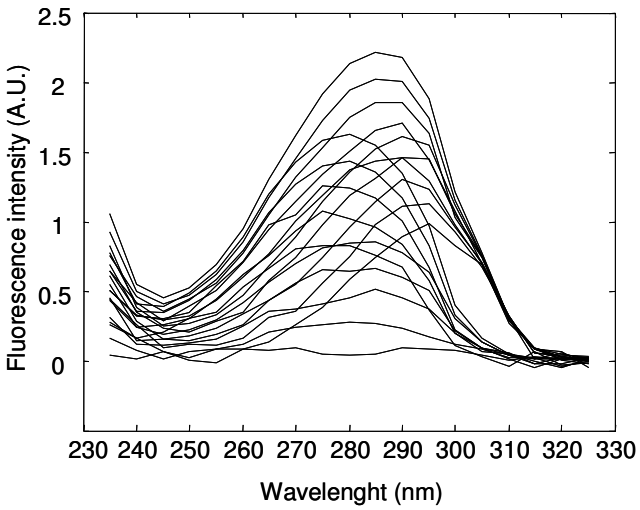


Figure 37. Multivariate response of the calibration samples. Each line represents the spectrum of one sample.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

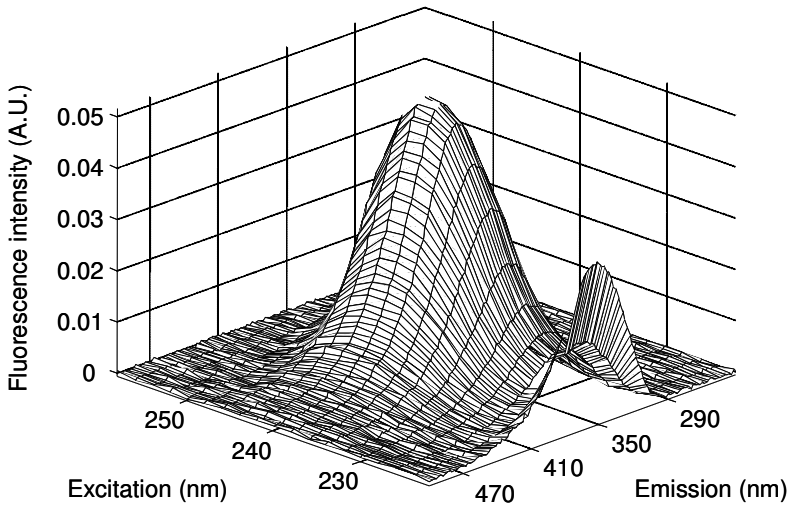


Figure 38. Calibration sample (sample 11) of one of the models. The composition of the mixture is 2.00 c.u. of tryptophan, 1.75 c.u. of hydroquinone and 0.89 c.u. of indole.

The prediction data sets were defined at two levels of tryptophan concentration, y_0 and y_D. To test whether the standard error of prediction was affected by the presence of interferents, four different situations were analyzed at the zero content level for the analyte of interest, in which the concentration of the interferent species were randomly distributed in a given range of concentration named *NONE*, *LOW*, *MEDIUM* and *HIGH*. Additionally, when dealing with second-order data and N-PLS, a fifth situation was checked named *EXTREME* where the concentration of interferents was higher than that one modelled in the calibration, i.e. above 2.0 c.u. (see Table 24).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 24. Ranges of concentration (in arbitrary concentration units) of the three components in the calibration and prediction sample sets.

Situation		Tryptophan	Hydroquinone	Indole
Calibration samples		**0-2.0**	**0-2.0**	**0-2.0**
Prediction samples	NONE -without interferences	0	0	0
	LOW level of interferences	0	0-0.4	0-0.4
	MEDIUM level of interferences	0	0.8-1.2	0.8-1.2
	HIGH level of interferences	0	1.6-2.0	1.6-2.0
	EXTREME level of interferences	0	2.4-2.8	2.4-2.8

3.1. Evaluation of the sample-specific SEP dependence on the level of interferents

One hundred two-way PLS and one hundred N-PLS models were built (all with 3 factors) to draw general conclusions in the end that come from averaging and not from a particular situation. In both scenarios and for each level of interferents analyzed (*NONE*, *LOW*, *MEDIUM* and *HIGH* for the multivariate analysis and the additional situation *EXTREME* for the multi-way analysis), the concentration of the analyte of interest, i.e. tryptophan, is zero. The prediction sets consisted of 200 prediction sample response vectors and 5 replicates of each one. These 1000 samples were simulated in the same way as the calibration samples, i.e. by multiplying the normalized spectra (for the multivariate data) or excitation-emission matrix (for the multi-way data) by the corresponding concentrations and adding ~2% of white noise in terms of relative standard deviation at the maximum of the response. The only difference from the calibration data was that the analyte concentration was zero.

The concentration of tryptophan in these 1000 samples was predicted using the (multivariate or multi-linear) PLS models, concretely, we predicted 10 samples (2 samples with 5 replicates) with each of the 100 models. The associated sample-specific standard error of prediction SEP_0 was estimated using Equation 47 and the RMSEP was also calculated for each sample from its 5 replicates. For each sample, we estimated the 95% prediction interval, PI, as shown in Equation 49, with α=0.05 and $\nu=I-F=20-3=17$ degrees of freedom (standard calculation of degrees of freedom was used but the pseudo-degrees of freedom by H. van der Voet [28] may be used for complex predictive models). Within the specified level

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

of interferents, we calculated the percentage of those prediction intervals that actually did not include the reference value, i.e. zero, and compared it with the nominal value of 5%.

3.2. LOD estimation

LOD was estimated from simulated data for multivariate and for multi-linear PLS following three different procedures. Two of them are based on traditional LOD concepts, i.e. as an inherent capability of the method, and the third one is the one introduced in the current work, which is based on the standard error of prediction at the individual sample level. These procedures are schematized in Figure 39 and described next.

3.2.1. Method-characteristic LODs

Both for first- and second-order data, we had already built 100 PLS models. From the linear regression between the concentration of the analyte, y_i, and that calculated with the (two- or three-way) PLS model of calibration, $\hat{y}_i$, we estimated the critical level and the limit of detection following (I) Ortiz et al. procedure and (II) Hubaux and Vos approximation. For the latter, the LOD equation (see Appendix) was resolved by means of algebraic manipulations and the quadratic formula using the Garner and Robertson estimator [29].

We simulated 4 (or 5, for second-order data) different levels of interferents sets of 100 prediction samples at those detection levels of concentration. So that for each of the 100 models, we estimated $\hat{y}_C$ and $\hat{y}_D$ (using approximation I or II) and predicted the four/five sets of 100 samples each one with the corresponding PLS model. To test the adequacy of the method-dependent LODs, we compared the predicted concentration with the associated $\hat{y}_D$ and obtained the experimental probability of false negatives, β_{exp}, which would have to be comparable to the pre-defined value of 5%.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

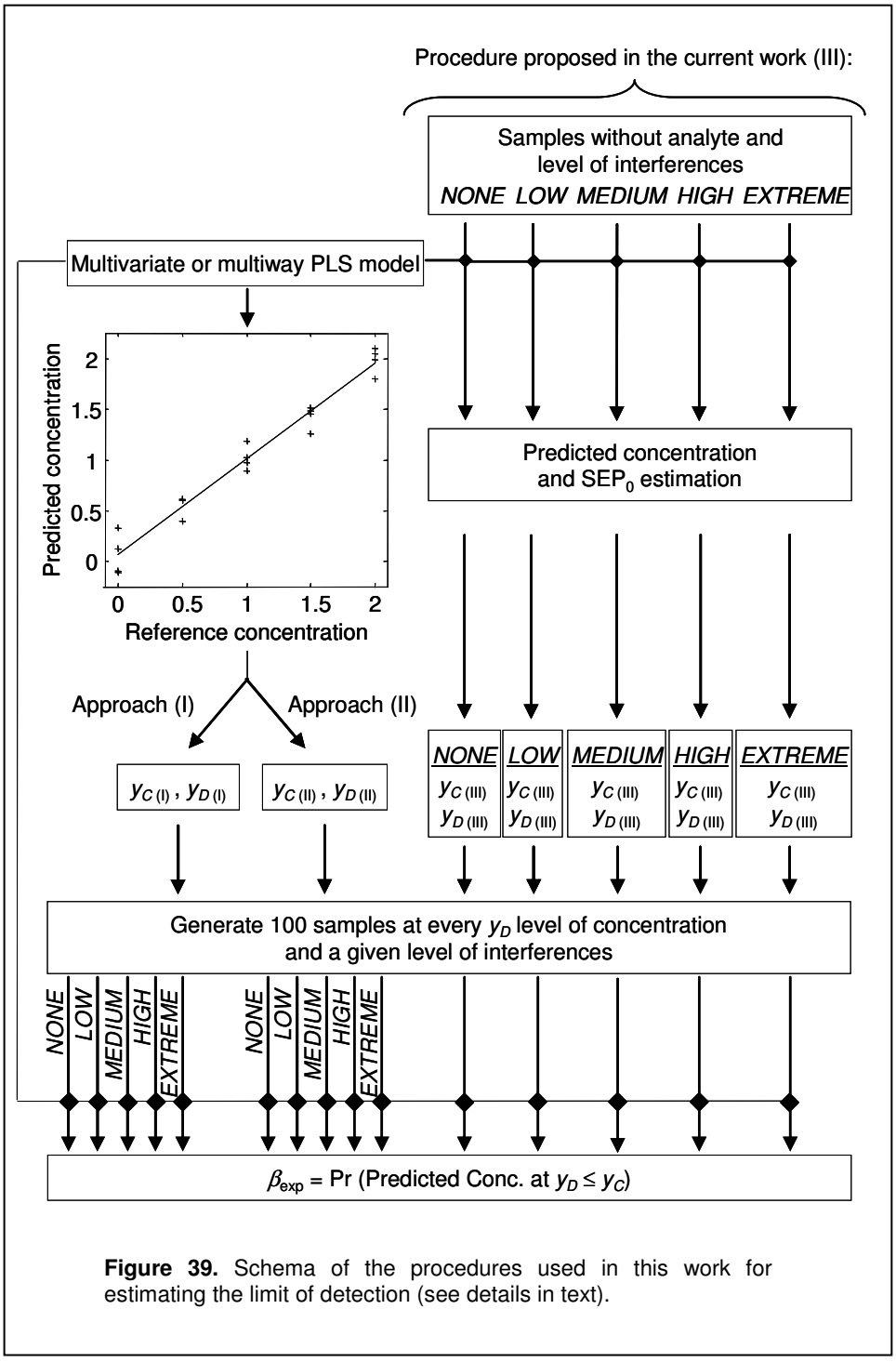


Figure 39. Schema of the procedures used in this work for estimating the limit of detection (see details in text).

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

3.2.2. Sample-specific LODs

Both for first- and second-order data, we had already samples at different level of interferents and no analyte of interest. We selected 100 of those samples from each interferent level data set, and predicted their concentration using different PLS models, as well as their leverage, $h_{0,i}$ and their $SEP_{0,l}$ (in this work, the error in the reference measurement was zero and thus we discarded the last term in Equation 47). Then, we calculated the sample-specific critical level and limit of detection for each of the 100 samples of each set using Equations 50 and 51. As for the method-dependent LODs, we simulated 4 (or 5, for second-order data) different level of interferents sets of 100 prediction samples each one at those detection levels of concentration, and predicted them using different PLS models. Finally, we tested the adequacy of the LODs for the individual sample levels comparing the predicted concentration with the associated $\hat{y}_C$ and obtained the experimental probability of false negatives, β_{exp}, which would have to be comparable to the pre-defined value of 5%. In addition, we compared the proposed LOD estimations with the results from the two method-dependent LOD estimators. The SEP was also analyzed at the LOD level to further evaluate its dependence on the level of interferents.

4. Results and discussion

4.1. SEP dependence on the level of interferents

For a fixed confidence level of 95% there should be 5% of the PI of samples without analyte of interest (tryptophan) that do not include the reference concentration, i.e. $y_0 = 0$). We used histograms to graphically summarize and display the distribution of the predicted concentrations in each situation analyzed. The range of the data (horizontal axis: predicted concentration) was split into nonoverlapping equal wide bars. The heights of the bars (vertical axis: frequency, i.e. counts for each bar) represent observed frequencies, i.e. what proportion of cases fall into each specified range. Among other information, these histograms show the location and scale of the variable, i.e. the mean and the standard deviation of the predicted concentration.

4.1.1. First-order situation

The mean estimated concentration of the first-order data prediction sets was found to be slightly biased from zero (Figure 40). This bias was negligible for the

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

prediction set without interferents (situation *NONE*), with mean of the predicted concentration $\bar{\hat{y}}$=–6·10^{-3}, and it was increasing when going to higher level of interferents, up to $\bar{\hat{y}}$=0.11.

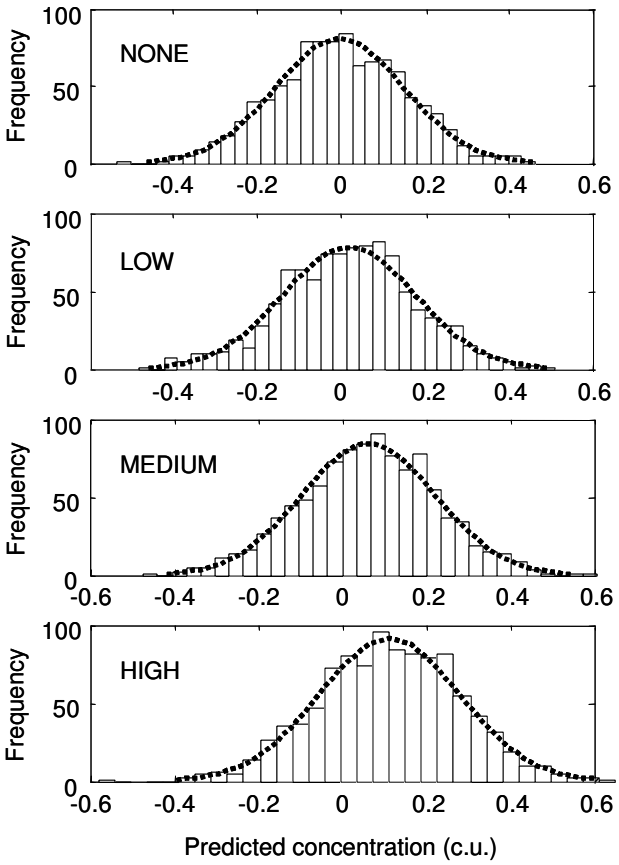


Figure 40. Distribution of the predicted concentration by multivariate PLS for samples with zero content level of analyte of interest at the different level of interferences analyzed.

After bias correction, e.g. subtracting the mean deviation from zero to all the predictions, we estimated the associated uncertainty (with the above equation) of each predicted concentration to estimate the PI and counted how many of those

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

intervals do not include the zero. The percentage of PI that fulfil that requirement (see Figure 41) was not exactly the expected probability because of the experimental error and the limited number of simulations but they agreed with the pre-defined value of β=5%, which gave an indication of the suitability of the expression used to estimate the prediction error. The prediction error was further evaluated. The mean SEP was approximately the same for all the situations analyzed. For each level of interferents, the RMSEP was calculated from the prediction of the replicates (same analyte concentration, i.e. y_0 with random noise). The mean RMSEP was of the same order of magnitude as the (mean) SEP for none or low level of interferents, while their values were prone to discrepancy for higher level of interferents (Table 25, upper rows) accordingly to the slight increase of the standard deviation of the predicted concentration.

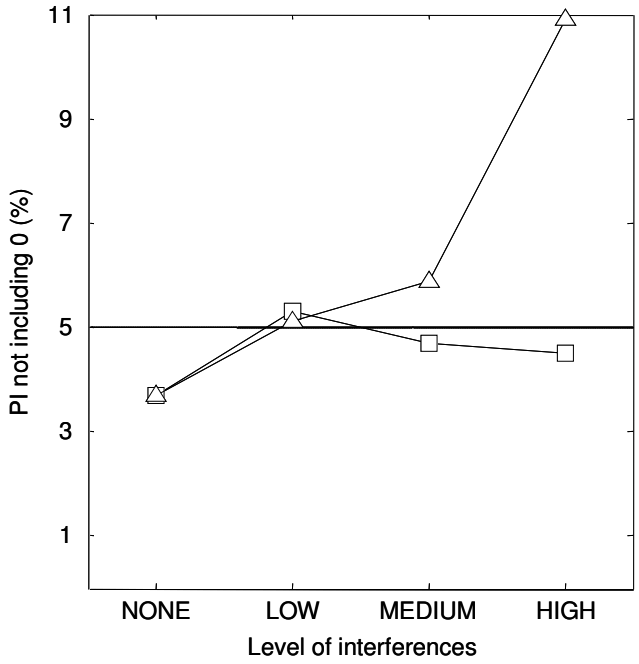


Figure 41. First-order data and multivariate PLS models. Percentage of prediction intervals (PI) for samples without analyte of interest that do not include 0, at the different levels of interferents analyzed. The symbol (△) correspond to the original results and (□) to the bias corrected PI. The expected 5% is indicated with a dotted line.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Table 25. Standard deviation of the predicted concentrations, mean SEP for prediction samples at y_0=0 and mean RMSEP calculated from the prediction of 5 replicates.

	Level of interferences:	*NONE*	*LOW*	*MEDIUM*	*HIGH*	*EXTREME*
First-order data and models	Std. deviation of $\hat{y}$	0.153	0.159	0.160	0.168	-
	$\overline{SEP}$ (a)	0.156	0.156	0.157	0.161	-
	$\overline{RMSEP}$ (b)	0.147	0.150	0.163	0.193	-
Second-order data and models	Std. deviation of $\hat{y}\times10^3$	0.794	0.835	0.831	0.900	0.946
	$\overline{SEP}\times10^3$ (a)	0.831	0.831	0.832	0.843	0.885
	$\overline{RMSEP}\times10^3$ (b)	0.751	0.801	0.799	0.854	0.907

(a) for 1000 prediction samples
(b) from 5 replicates

4.1.2. Second-order situation

Unlike first-order data, accurate concentrations were predicted for the second-order data from samples without analyte, regardless of the level of interferents (Figure 42). The PIs were small and they included the expected reference value around 95% of the times in all cases (Figure 43), in agreement with the assumed (1-β) probabilities. The SEP, for each individual sample, and the RMSEP calculated from 5 replicates, were of the same order of magnitude, and both slightly increased when the level of interferents increased (Table 25, lower rows), as well as the standard deviation of the predicted concentration.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

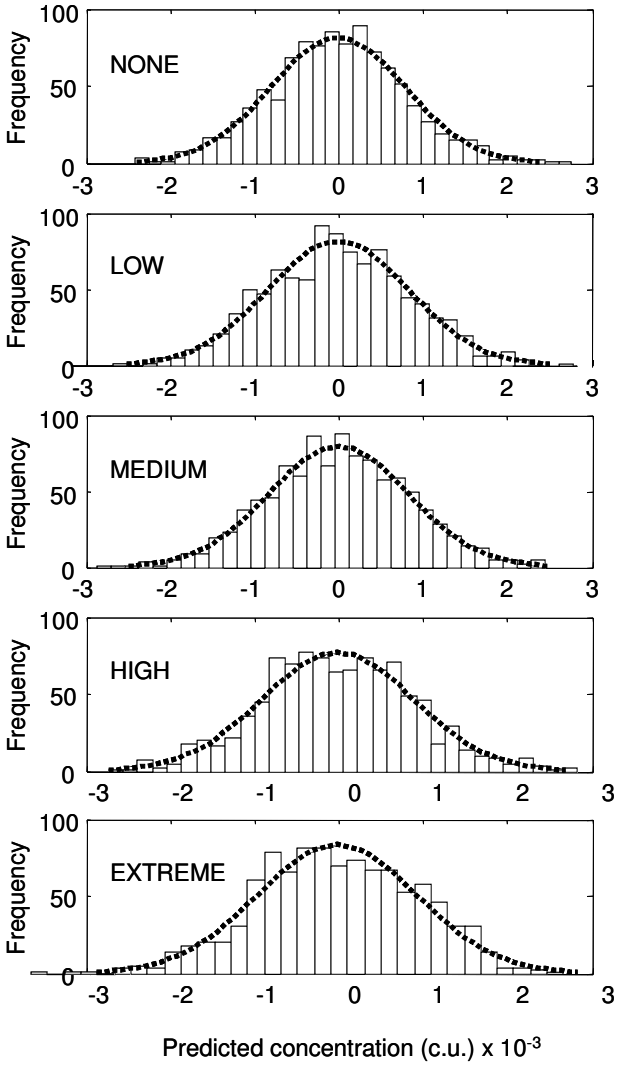


Figure 42. Distribution of the predicted concentration by N-PLS for samples with zero content level of analyte of interest at the different level of interferences analyzed.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

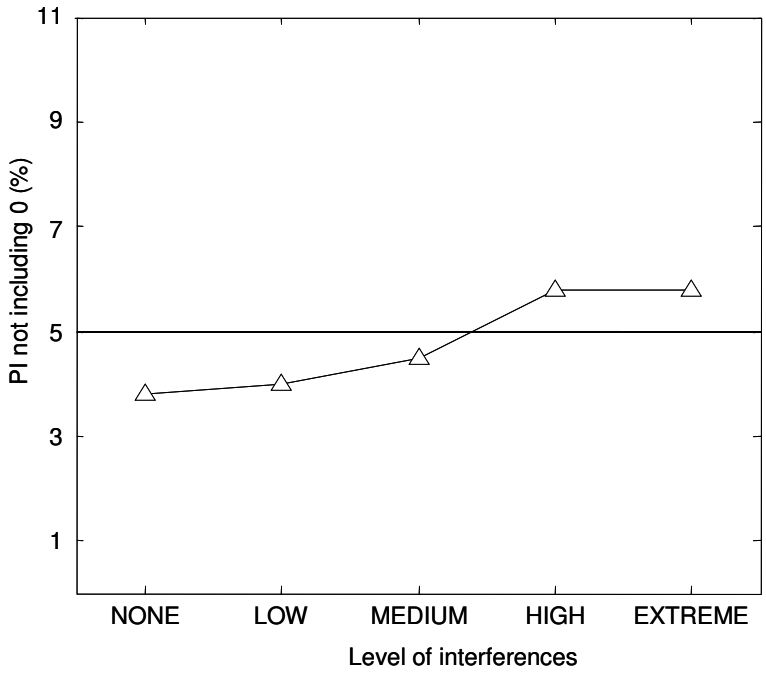


Figure 43. Second-order data and N-PLS models. Percentage of prediction intervals (PI) for samples without analyte of interest that do not include 0, at the different levels of interferents analyzed. The expected 5% is indicated with a dotted line.

4.2. LOD estimation

4.2.1. First-order situation

As seen in Figure 40, the predictions from multivariate data, also at the LOD level, needed to be bias-corrected. The LOD estimated for the traditional procedures and the mean of the proposed LODs at each level of interferents set are plotted in Figure 44a. The mean individual standard error of prediction was almost the same at the LOD concentration for none, low and medium level of interferents and increased for the higher level of interferents. Therefore, the mean LOD estimated from these SEPs remain almost constant for low level of interferents too and then increased with the amount of interferents. In any case, these LODs were lower than the ones estimated by the traditional-based procedures. Suitable conclusions could be drawn from the validation. The

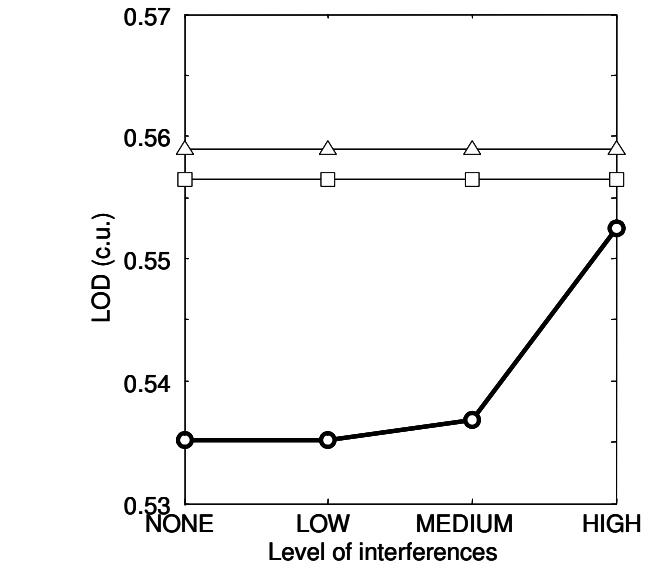
UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

proposed LODs were satisfactorily validated by comparison of the experimental probability of false negative with the pre-defined value, 5%, for all the levels of interferents (Figure 45a). However, both method-dependent estimations were prone to overestimate the LOD, especially for lower level of interferents, i.e. the distribution of the predicted concentrations is shifted to higher values, giving less prediction below the critical level than expected.

4.2.2. Second-order situation

Predicted concentrations from second-order data and models were accurate (predictions were not biased). The LOD estimated by the two traditional-based procedures and the mean of the estimated LODs as proposed in this work at each level of interferents set are plotted in Figure 44b. As observed before, the LOD for the individual sample level differed for different level of interferents. The remarkable observation is that the sample-dependent LOD from multi-way determinations of samples with none or low content of interferents were higher than the ones obtained by the traditional procedures but lower than those ones when the level of interferents increased and, in addition, almost unchangeable for medium, high or extreme level of interferents in the samples. Validation was again the filter to a proper discussion. The probability of committing false negatives was, in general terms, around the pre-defined 5% (Figure 45b). Focusing our attention on the situations with lower level of interferents (*NONE*, *LOW* and *MEDIUM*), we could assess that the traditional-based approaches provided overestimated LODs since the experimental β were lower than expected (as in the multivariate simulations), while the proposed LOD seemed to be more accurate. Moving to very higher level of interferents, all the LODs were somewhat underestimated, resulting in high β probabilities. However, the sample-dependent LOD provided again the nearest approximation to the expected value.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007



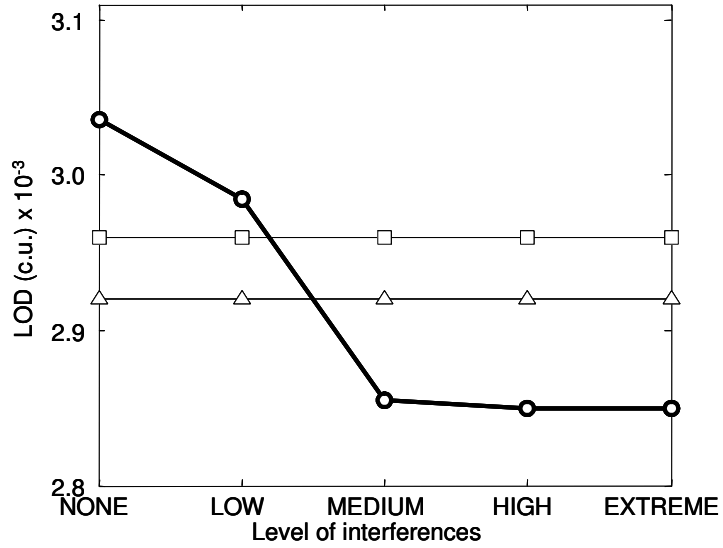


Figure 44. Both in multivariate (a) and in multi-way (b) simulations, the LOD estimated by procedure I (△) and by procedure II (□) were constant for all the prediction sets with different level of interferences, while the mean LOD estimated by the proposed procedure (○) varied from set to set.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

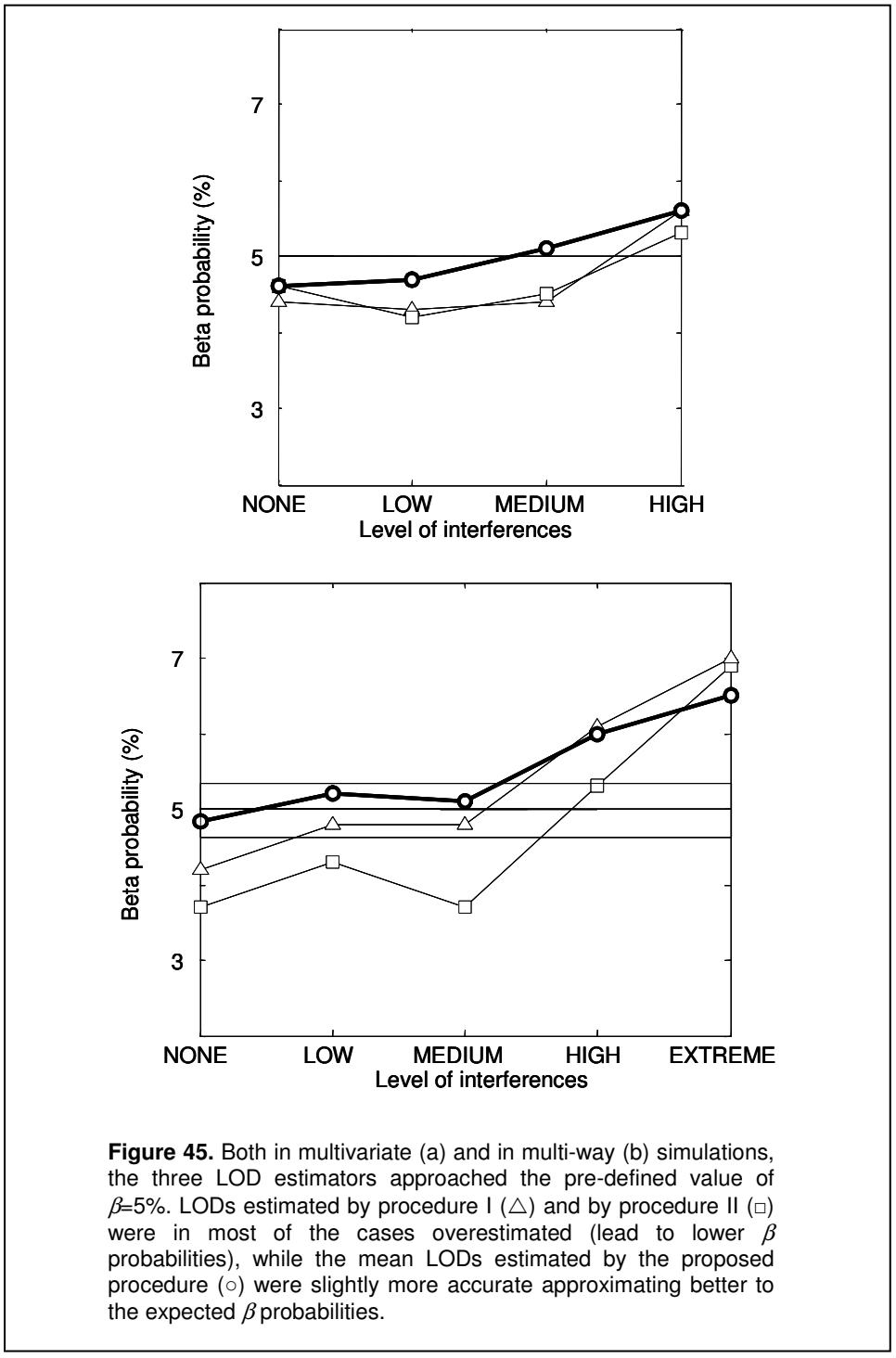


Figure 45. Both in multivariate (a) and in multi-way (b) simulations, the three LOD estimators approached the pre-defined value of β=5%. LODs estimated by procedure I (△) and by procedure II (□) were in most of the cases overestimated (lead to lower β probabilities), while the mean LODs estimated by the proposed procedure (○) were slightly more accurate approximating better to the expected β probabilities.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

5. Conclusions

The distribution of the predicted concentrations at low content levels of the analyte of interest is statistically different for samples containing different amount of interferents. The standard error of prediction defined at the individual analyte/interferent ratio is used in this work to estimate the limit of detection. Therefore, the LOD depends on the uncertainty of the analyte-specific part of the sample which depends on the sample matrix; hence it makes sense to take a detection decision for each particular sample. In this way, the LOD is sample-specific and it is calculated after the measurement on the new sample is made, contrary to the traditional concept of the LOD that defines this performance characteristic as an inherent detection capability of the chemical measurement procedure.

A sample-specific detection limit is the lowest level of an analyte that would have been reliably detected in the specific sample. This value will be a more realistic detection decision since it takes advantage of being a posteriori estimation. However, note that this approach is fully compatible with the official requirements because that sample-specific or matrix-specific detection limits can always be integrated to obtain a (traditional) detection limit, which pertains to a well-defined population of samples [30].

The sample-specific LODs were validated following a common practice: to assess that the percentage of samples generated at the LOD level whose predicted concentration is below the critical level agree, up to the experimental error, with the pre-defined probability β or type II error. In the simulations performed, the proposed sample-specific LOD fulfil the validation requirements both for multivariate and for multi-way data and PLS models. It was properly assessed for different pre-fixed β values (1%, 5% and 10%), although only results from β=5% were shown here.
We do not forget that, to calculate the sample-dependent LOD, we need an estimation of the SEP at zero content level and the same analyte/interferent ratio than in the test sample. However, this apparent limitation can be overcome easily. One possibility is to build a synthetic test set with the given range of interferents and no analyte of interest and consider the mean of their SEP_0 (or the highest one, to be warier). This solution would provide an approximate LOD valid for such range of analyte/interferent ratio. In actual situations, however, it

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

may not be so easy to know or to reproduce the matrix of the unknown sample. An alternative, subject to further research, would be to obtain the prediction error from the cross-validation results of the lower content standards.

In any case, the results of the current work open a new door in detection capability concepts. LOD is not only an inherent characteristic of the method, but the particular detection level that can be reached for any new sample. The use of one or the other notion will depend on the particular purpose of the matter, but once the sample dependence is realised, the (multivariate or multi-way) calibration sets should be designed in such a way that the calculated LOD is representative of the future samples, e.g. including several realistic levels of interferents at each level of concentration.

Appendix

The classical univariate LOD estimators recommended by Hubaux and Vos [4] are here given.

$$y_C = t_{1-\alpha,\nu}\,\frac{s}{b_1}\sqrt{\frac{1}{I}+\frac{1}{n}+\frac{\bar{y}^2}{\sum_{i=1}^{I}(y_i-\bar{y})^2}} \qquad \text{Eq. 52}$$

$$y_D = y_C + t_{1-\beta,\nu}\,\frac{s}{b_1}\sqrt{\frac{1}{I}+\frac{1}{n}+\frac{(L_D-\bar{y})^2}{\sum_{i=1}^{I}(y_i-\bar{y})^2}} \qquad \text{Eq. 53}$$

where $t_{1-\alpha,\nu}$ and $t_{1-\beta,\nu}$ are the quantiles of the one-sided Student's *t*-distribution with the corresponding α and β probability respectively, s is the standard deviation of the residuals of the regression line, b_1 the slope of the regression line, I the number of standards, n the number of replications performed on the unknown sample and $\bar{y}$ the mean of the concentrations of the standards of the calibration curve. The degrees of freedom are constricted by the regression line, i.e. $\nu=I-2$.

The critical level and the limit of detection as defined by the ISO 11843-2 norm [26], were used in reference [25] to calculate the minimum detectable net

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

concentration from the linear regression between the concentration of the analyte, y_i and that calculated with the (two- or multi-way) PLS model of calibration, $\hat{y}_i$. The nomenclature of these formulations is here conveniently adapted to simplify the presentation:

$$y_C = t_{1-\alpha,\nu}\,\frac{s}{b_1}\sqrt{\frac{1}{I}+\frac{1}{n}+\frac{\bar{y}^2}{\sum_{i=1}^{I}(y_i-\bar{y})^2}} \qquad \text{Eq. 54}$$

$$y_D = \delta_{\alpha,\beta,\nu}\,s_0 = \delta_{\alpha,\beta,\nu}\,\frac{s}{b_1}\sqrt{\frac{1}{I}+\frac{1}{n}+\frac{\bar{y}^2}{\sum_{i=1}^{I}(y_i-\bar{y})^2}} \qquad \text{Eq. 55}$$

where t is the quantile of the one-sided Student's t-distribution and δ is the non-centrality parameter of the non-central t-distribution, with the corresponding α and β probabilities and $\nu=I-2$ degrees of freedom (from the linear regression between predicted and reference concentration). The terms s, b_1, I, n, and $\bar{y}$ have been already defined.

Acknowledgements

The authors thank the MCyT (Project No. BQU2003-00500) for financial support.

References

[1] J. Inczédy, T. Lengyel, A.M. Ure, A. Gelencsér, A. Hulanicki (Eds.), IUPAC Compendium of Analytical Nomenclature (definitive rules 1997), Blackwell, Oxford, 1998

[2] L.A. Currie, Pure Appl. Chem., 67: 1699 (1995)

[3] A.C. Olivieri, N.M. Faber, J. Ferré, R. Boqué, J.H. Kalivas, H. Mark, Pure Appl. Chem., *accepted for publication.*

[4] A. Hubaux, G. Vos, Anal. Chem. 42: 849 (1970)

[5] C.A. Clayton, J.W. Hines, P.D. Elkins, Anal. Chem., 59: 2506 (1987)

[6] L.A. Currie, Chemom. Intell. Lab. Syst., 37: 151 (1997)

[7] A. Höskuldsson, J. Chemom., 2: 211 (1988)

[8] N.M. Faber, B.R. Kowalski, Chemom. Intell. Lab. Syst., 34: 283 (1996)

[9] N.M. Faber, B.R. Kowalski, J. Chemom., 11: 181 (1997)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

[10] M.C. Denham, J. Chemom., 11: 39 (1997)

[11] A. Lorber, B.R. Kowalski, J. Chemom., 2: 93 (1998)

[12] N.M. Faber, J. Chemom., 14: 363 (2000)

[13] N.M. Faber, Chemom. Intell. Lab. Syst., 52: 123 (2000)

[14] N.M. Faber, X.-H. Song, P.K. Hopke, Trends in Anal. Chem., 22: 330 (2003)

[15] E: Sanchez, B.R. Kowalski, J. Chemom., 2: 247 (1988)

[16] H. Martens, T. Naes, Multivariate Calibration. J Wiley & Sons, Chichester, 1989, p. 116

[17] R. Bro, J. Chemom., 10: 47 (1996)

[18] A.K. Smilde, J. Chemom., 11: 367 (1997)

[19] N.M. Faber, R. Bro, Chemom. Intell. Lab. Syst., 61: 133 (2002)

[20] A. Höskuldsson, J. Chemom., 2: 211 (1988)

[21] R. Bro, A. Rinnan, N.M. Faber, Chemom. Intell. Lab. Syst., *In press*

[22] S.R. Searle, G. Casella, C.E. McCullogh, Variance Components, Wiley, New York, 1992

[23] L A.Curie, Anal. Chim. Acta, 391: 105 (1999).

[24] ISO 11843-1: 1997. Capability of detection - Part 1: Terms and definitions, Geneva, Switzerland

[25] M.C. Ortiz, L.A. Sarabia, A. Herrero, M.S. Sánchez, M.B. Sanz, M.E. Rueda, D. Giménez, M.E. Meléndez, Chemom. Intell. Lab. Syst., 69: 21 (2003).

[26] ISO 11843-2: 2000, Capability of detection – Part 2: Methodology in the linear calibration case, Geneva, Switzerland

[27] C.A. Andersson, R. Bro, Chemom. Intell. Lab. Syst., 52: 1 (2000)

[28] H. van der Voet, J. Chemom., 13: 195 (1999)

[29] F.C.Garner, G.L. Robertson, Chemom. Intell. Lab. Syst., 3: 53 (1988)

[30] H. van der Voet, Detection limits, in: Encyclopedia of Environmetrics vol. 1, John Wiley & Sons, Chichester, 2002, p. 504

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Chapter 7

CONCLUSIONS OF THE THESIS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

7.1. Introduction

In this chapter I present an overview of the main conclusions of this thesis and make several suggestions for future work on second-order calibration methods with non-selective high-order data. Finally, I express some personal impressions about how my time spent working on this PhD has developed me as a scientist and as a person.

7.2. Conclusions

Conclusions of the thesis

In the last few years, second-order data and instruments have become usual in most laboratories of analysis. Rather than rejecting some of the measurements in order to simplify to first-order measurements, or even to single signals, and performing multivariate or univariate calibration models, second-order calibration methods are a practical solution.

The wide range of second-order calibration methods means it is possible to select the most suitable method for each analytical problem. I have found that there is no absolute or definitive method: it always depend primarily on the type of data we are dealing with, and on the characteristics of the laboratory or analyst concerned e.g. the material and human resources, the time available and the desired quality of results.

Validation of the method is an essential step. Assessing the figures of merit related to the procedure used is indispensable for ensuring the quality of the results.

The limit of detection is usually one of the figures of merit to be estimated. For univariate and multivariate calibration, LOD estimation is well defined. For second-order calibration, however, there is still certain confusion for many users.

In this thesis we have shown, with practical examples, that the limit of detection can be calculated for complex analytical data using second-order calibration methods and different strategies.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Our main conclusions are:

LOD for PARAFAC:

- Parallel Factor Analysis (PARAFAC) can be used to identify and quantify pesticides in non-selective EEM data that are approximately trilinear.

- Information about the selectivity and sensitivity of the compounds can be used to select the calibration samples. Including standards of the most sensitive or the least selective compound in the mixtures may improve the model.

- Two prediction strategies are feasible for PARAFAC. When the model is performed only with the calibration set, further calculation is needed to obtain the loadings of the prediction samples. However, we can detect outliers and the model is valid for future samples within the experimental domain. Including the prediction set in the model can save time, but outliers may not be detected. Also, if outliers exist, there is a loose fit of the model and of the correlation between the estimated and the reference profiles.

- The LOD (and other figures of merit) can be suitably estimated for PARAFAC using the transformation approach, i.e. with univariate statistics from the calibration line obtained by the regression of the model loadings against the concentration of the standards.

LOD for ITTFA:

- Iterative Target Transformation Factor Analysis (ITTFA) can be used to identify and quantify compounds in non-selective HPLC-DAD data that are not trilinear since it resolves one second-order (bilinear) data matrix at a time.

- The LOD can be suitably estimated for ITTFA using the transformation approach, i.e. with univariate statistics form the calibration line obtained by regression of the height of the estimated elution profile against the concentration.

- Chromatographic selectivity, spectral selectivity and the sensitivity of the analyte of interest (e.g. the relative intensity of the chromatographic peak) are experimental factors in HPLD-DAD determinations that significantly affect the resolution and, therefore, the estimated LOD.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

LOD for MCR-ALS:

- Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) can be used to identify and quantify pesticides in complex natural groundwater samples that are analysed by HPLC-DAD.

- The resolution of complex data by MCR-ALS can be improved by preserving the three-way structure of the data when decomposing several samples simultaneously, imposing trilinearity only for some compounds and preserving bilinearity for others.

- The LOD can be suitably estimated for MCR-ALS using the transformation approach, i.e. with univariate statistics form the calibration line obtained by regression of the relative areas estimated by MCR-ALS against the concentration.

LOD for N-PLS:

- The distribution of the predicted concentration at low levels of the analyte of interest is statistically different for samples containing different amounts of interferents.
- The LOD can also be suitably estimated for both multivariate and multi-linear PLS from the standard error of prediction defined at the individual analyte/interferent ratio. The LOD therefore depends on the uncertainty of the analyte-specific part of the sample, which depends on the sample matrix.
- We suggest taking a detection decision for each sample. This means that the LOD is sample-specific and that it is calculated after the measurement of the new sample has been made. This definition is contrary to the accepted concept of the LOD, which defines it as an inherent parameter of the chemical measurement procedure. In our case, LOD is understood as the lowest level of an analyte that would have been reliably detected in the specific sample. To be compatible with the IUPAC recommendations, it would be possible to define a LOD for a particular type of sample from the mean LOD of representative samples.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Suggestions for the future

- For N-PLS, as well as the suggested sample-dependence of the LOD, a model-dependence of the LOD could be studied.
- For the various calibration methods, guidelines on how to improve the LOD as a function of the experimental design of the calibration set of samples could be established.
- A LOD estimator for non-bilinear data, such as MS-MS data could be developed because it is not available.

Personal impressions

Doctoral programmes are considered a crucial source of a new generation of researchers and are meant to serve as the main bridge between the European Higher Education and Research Areas. Doctoral training is, however, markedly different from the first and second cycles of education. Its main characteristic, which makes it specific, is that its most predominant and essential component is research[33].

Certainly, PhD candidates do produce much of the research conducted in universities. Moreover, we often participate in teaching activities, tutoring and even outreach activities. By completing this level of studies and working full time, we can be considered as professionals, specifically as *early stage researchers* (ESRs). In the context of the Bologna Process, "*ESRs are professionals who are trained through research in the conception or creation of new knowledge, products, processes, methods and systems, and in the management of the projects concerned*". I therefore feel that, with the presentation of this thesis I have completed part of the early stage of a research career.

Doctoral candidates have to prove their ability to perform original and independent research in a scientific discipline. Scientific training in core research skills is usually mandatory in doctoral programmes. However, training in transferable professional and personal skills and competences is offered more on a voluntary basis. Individuality, originality and a certain autonomy are important features of a doctorate and I believe that during my PhD I have made good use

[33] Extracted and adapted from "Doctoral Programmes for the European Knowledge Society", Report on the EUA Doctoral Programmes Project 2004-2005 of the European University Association asbl (EUA). URL: http://www.eua.be

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

of my time to develop several skills and competences, most of which have been self-trained ("*learning by doing*"). Important personal improvements I have made in core and transferable skills include writing and communication skills such as scientific writing; networking and team-working; time management; problem solving; research methodology and techniques; research management; analysis and diffusion; and awareness of scientific ethics and intellectual property rights.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Appendix

REFERENCES

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Appellof *et al*. 1983

C. Appellof, E. Davidsoni. *Three-Dimensional Rank Annihilation for Multi-Component Determinations*, Anal. Chim. Acta 146: 9-14 (1983)

Bauer *et al*. 1991

G. Bauer, W. Wegscheider, H.M. Ortner. *Limits of Detection in Multivariate Calibration*, Fres. J. Anal. Chem. 340: 135-139 (1991).

Berge 2000

J. M. F. Ten Berge. *The Typical Rank of Tall Three-Way Arrays*, Psychometrika 65: 525-532 (2000)

Berge *et al*. 1999

J. M. F. Ten Berge, H. A. L. Kiers. *Simplicity of Core Arrays in Three-Way Principal Component Analysis and the Typical Rank of p×q×2 Arrays*, Linear Algebra and its Applications 294: 169-179 (1999)

Bogomolov and McBrien 2003

A. Bogomolov, M. McBrien. *Manual Peak Matching in a Series of HPLC-DAD Mixture analyses*, Anal. Chim. Acta 490: 41-58 (2003)

Booksh and Kowalski 1994

K. S. Booksh, B. R. Kowalski. *Theory of Analytical Chemistry*, Anal. Chem. 66: 782A-791A, 1994

Boqué and Ferré 2004

R. Boqué, J. Ferré. *Using Secod-Order Data in Chromatographic Analysis*, LC-GC Europe, 16: 402-407 (2004)

Boqué and Rius 1996

R. Boqué, F. X. Rius. *Multivariate Detection Limits Estimators (Review)*, Chemom. Intell. Lab. Syst. 32: 11-23 (1996)

Boqué *et al*. 1999

R. Boqué, M. S. Larrechi, F. X. Rius. *Multivariate Detection Limits with Fixed Probabilities of Error*, Chemom. Intell. Lab. Syst. 45: 397-408 (1999)

Boqué *et al*. 2002

R. Boqué, J. Ferré, N.M. Faber, F. X. Rius. *Limit of Detection Estimator for Second-Order Bilinear Calibration*, Anal. Chim. Acta 451: 313-321 (2002)

Bro 1996

R. Bro. *Multi-Way Calibration. Multi-Linear PLS*, J. Chemom. 10: 47-61 (1996)

Bro 1997

R. Bro. *PARAFAC. Tutorial and Applications*, Chemom. Intell. Lab. Syst. 38:149-171 (1997)

Bro 1998

R. Bro. *Multi-Way Analysis in the Food Industry. Models, Algorithms, and Applications*, Doctoral Thesis, University of Amsterdam, 1998.

Bro 2003

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

R. Bro. *Multivariate Calibration. What is in Chemometrics for the Analytical Chemist?*, Anal. Chim. Acta 500: 185-194 (2003)

Bro and Sidiropoulos 1998
R. Bro, N. Sidiropoulos. *Least Squares Algorithms under Unimodality and Non-Negativity Constraints*, J. Chemom. 12: 223-247 (1998)

Bro *et al*. 1999
R. Bro, C. A. Andersson, H. A. L. Kiers. *PARAFAC2-PART II. Modeling Chromatographic Data with Retention Time Shifts*, J. Chemom. 13: 295-309 (1999)

Bro *et al*. 2001
R. Bro, A. K. Smilde, S. de Jong. On the Difference between Low-Rank and Subspace Approximation: Improved Model for Multi-Linear PLS Regression, Chemom. Intell. Lab. Syst. 58: 3-13 (2001)

Bro *et al*. 2002
R. Bro, N. D. Sidiropoulos, A. K. Smilde. *Maximum Likelihood Fitting Using Ordinary Least Squares Algorithms*, J. Chemom. 16: 387-400 (2002)

Budianski 1974
B.Budiansky. *Tensors* in *Handbook of Applied Mathematics*, ch.4, p. 179, ed. by C.E. Pearson, Van Nostrand Reinhold Company: New York, 1974

Bylund *et al*. 2002
D. Bylund, R. Danielsson, G. Malmquist, K. E. Markides. *Chromatographic Alignment by Warping and Dynamic Programming as a Pre-Processing Tool for PARAFAC Modelling of Liquid Chromatography-Mass Spectrometry Data*, J. Chromatogr. A 961: 237-244 (2002)

Comas *et al*. 2002
E. Comas, R. A. Gimeno, J. Ferré, R. M. Marcé, F. Borrull, F. X. Rius. *Time Shift Correction in Second-Order Liquid Chromatographic Data with Iterative Target Transformation Factor Analysis*, Anal. Chim. Acta 470: 163-173 (2002)

Cuesta-Sánchez *et al*. 1996
F. Cuesta-Sánchez, J. Toft, B. van den Bogaert, D. L. Massart. *Orthogonal Projection Approach Applied to Peak Purity Assessment*, Anal. Chem. 68: 79-85 (1996)

Currie 1968
L.A. Currie. *Limits for Quantitative Detection and Quantitative Determination. Application to Radiochemistry,* Anal. Chem. 40: 586-593 (1968)

Currie 1995
L. Currie. *Nomenclature in Evaluation of Analytical Methods Including Detection and Quantification Capabilities (IUPAC Recommendations 1995)*, Pure Appl. Chem. 67: 1699-1723 (1995)

Danzer 1996
K. Danzer. *Significance of Statistics in Quality Assurance* in *Accreditation and Quality Assurance* in *Analytical Chemistry*, ch. 5, p. 105, ed. by H. Günzler, Springer Verlag: Berlin, 1996

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Danzer and Currie 1998

K. Danzer, L. A. Currie. *Guidelines for Calibration in Analytical Chemistry. Part 1. Fundamentals and Single Component Calibration (IUPAC Recommendations 1998).* Pure Appl. Chem. 70: 993-1014 (1998)

Danzer *et al*. 2004

K. Danzer, M. Otto, L. A. Currie. *Guidelines for Calibration in Analytical Chemistry. Part 2. Multispecies Calibration (IUPAC Technical Report)* Pure Appl. Chem. 76: 1215-1225 (2004)

EPA 2004

EPA 402-B-04-001C. *Detection and Quantification Capabilities* (vol. III, ch. 20) and *Key Analytical Planning Issues and Developing Analytical Protocol Specification* (vol. I, ch. 3) in *Multi-Agency Radiological Laboratory Analytical Protocols Manual*, United States, 2004
URL: http://www.epa.gov/radiation/marlap/manual.htm

EURACHEM 1998

EURACHEM Guide. *The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics*. London, 1998

Faber and Kowalski 1997

N. M. Faber, B.R. Kowalski, *Improved Estimation of the Limit of Detection in Multivariate Calibration*, Fres. J. Anal. Chem. 357: 789–795 (1997)

Faber *et al*. 2002

N. M. Faber, J. Ferré, R. Boqué, J. H. Kalivas. *Second-Order Bilinear Calibration: the Effects of Vectorising the Data Matrices of the Calibration Set*, Chemom. Intell. Lab. Syst. 63: 107-116 (2002)

Garner and Robertson 1988

F.C. Garner, G.L. Robertson. *Evaluation of Detection Limit Estimators*, Chemom. Intell. Lab. Syst. 3: 53-59 (1988)

Gemperline and Hamilton 1989

P. J. Gemperline, J. C. Hamilton. *Evolving Factor Analysis Applied to Flow Injection Analysis Data*, J. Chemom. 3: 455-491 (1989)

Gemperline *et al*. 1999

P. J. Gemperline, J. H. Cho, B. Archer. *Multivariate Background Correction for Hyphenated Chromatography Detectors*, J. Chemom. 13: 153-164 (1999)

Ho *et al*. 1978

C.-N. Ho, G. D. Christian, E. R. Davidson. *Application of the Method of Rank Annihilation to Quantitative Analyses of Multicomponent Fluorescence Data from the Video Fluorometer*, Anal. Chem. 50:1108-1113 (1978)

Ho *et al*. 1980

C.-N. Ho, G. D. Christian, E. R. Davidson. *Application of the Method of Rank Annihilation to Fluorescent Multicomponent Mixtures of Polynuclear Aromatic Hydrocarbons*, Anal. Chem. 52:1071-1079 (1980)

ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Ho *et al*. 1981
C.-N. Ho, G. D. Christian, E. R. Davidson. *Simultaneous Multicomponent Rank Annihilation and Applications to Multicomponent Fluorescent Data Acquired by the Video Flurometer*, Anal. Chem. 53:92-98 (1981)

Hopke *et al*. 1983
P. K. Hopke, D. J. Alpert, B. A. Roscoe, *Fantasia - A Program for Target Transformation Factor Analysis to Apportion Sources in Environmental Samples*.. Comput. Chem. 7:149-156 (1983)

Hubaux and Vos 1970
A. Hubaux, G. Vos. *Decision and Detection Limits for Linear Calibration Curves*, Anal. Chem. 42: 849-855 (1970)

Inczédy *et al*. 1998
J. Inczédy, T. Lengyel, A.M. Ure, A. Gelencsér, A. Hulanicki, IUPAC, *Compendium of Analytical Nomenclature*, ed. by Blackwell, Oxford, 1998

ISO 3534-1:1993
Statistics – Vocabulary and Symbols – Part 1: Probability and General Statistical Terms (1993) Geneva: International Organization for Standardization

JiJi and Booksh 2000
R. D. JiJi, K. S. Booksh. *Mitigation of Rayleigh and Raman Spectral Interferences in Multiway Calibration of Excitation-Emission Matrix Fluorescence Spectra*, Anal. Chem. 72: 718-725 (2000)

Johnson *et al*. 1999
K. Johnson, A. de Juan, S. C. Rutan. *Three-Way Data Analysis of Pollutant Degradation Profiles Monitored Using Liquid Chromatography-Diode Array Detection*, J. Chemom. 13: 331-341 (1999)

de Juan *et al*. 1997
A. de Juan, Y. Vander Hieden, R. Tauler, D. L. Massart. *Assessment of New Constraints Applied to the Alternating Least Squares Method*, Anal. Chim. Acta, 346: 307-318 (1997)

Kaiser 1947
H. Kaiser. *Die Berechnung der Nachweisempfindlichkeit,* Spectrochimica Acta 3: 40-67 (1847)

Kaiser 1965
H. Kaiser. *Zum Problem der Nachweisgrenze*, Fresenius Zeitschrift der Analytischen Chemie, 209: 1-18 (1965)

Keller and Massart 1991
H. R. Keller, D. L. Massart. *Peak Purity Control in Liquid-Chromatography with Photodiode Array Detection by Fixed Size Moving Window Evolving Factor Analysis*, Anal. Chim. Acta 246: 379-390 (1991)

Keller and Massart 1992
H. R. Keller, D. L. Massart. *Evolving Factor Analysis*, Chemom. Intell. Lab. Syst. 12: 209-224 (1992)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Kiers 2000
H. A. L. Kiers. *Towards a Standardized Notation and Terminology in Multiway Analysis*, J. Chemom. 14: 105-122 (2000)

Kiers *et al*. 1999
H. A. L. Kiers, J. M. F. Ten Berge, R. Bro. *PARAFAC2-PART I. A Direct Fitting Algorithm for the Parafac2 Model*, J. Chemom. 13: 275-294 (1999)

Kroonenberg and Leeuw 1980
P. M. Kroonenberg, J. de Leeuw. *Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithm*, Psychometrika 45: 69-97 (1980)

Levin 1965
J. Levin. *Three-Mode Factor Analysis*, Psychological Bulletin, 64: 442-452 (1965)

Lorber 1984
A. Lorber. *Quantifying Chemical Composition from Two-Dimensional Data Arrays*, Anal. Chim. Acta 164: 293-297 (1984)

Lorber 1986
A. Lorber, *Error Propagation and Figures of Merit for Quantification by Solving Matrix Equations*, Anal. Chem. 58: 1167-1172 (1986)

Lorber and Kowalski 1988
A. Lorber, B.R. Kowalski. *The Effect of Interferences and Calibration Design on Accuracy: Implications for Sensor and Sample Selection*, J. Chemom. 2: 67-79 (1988)

Lorber *et al*. 1997
A. Lorber, N. M. Faber, B.R. Kowalski. *Net Analyte Signal Calculation in Multivariate Calibration*, Anal. Chem. 69: 1620-1626 (1997)

Maeder 1987
M. Maeder. *Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks*, Anal. Chem. 59: 527-539 (1987)

Messick et al. 1996
N. J. Messick, J. H. Kalivas, P. M. Lang. *Selectivity and Related Measures for Nth-Order Data*. Anal. Chem. 68: 1572-1579 (1996)

McCue and Malinowski 1983
M. McCue, E. R. Malinowski. *Rank Annihilation Factor Analysis of Unresolved LC Peaks*, J. Chromatogr. Sci. 21: 229-234 (1983)

Öhman *et al*. 1990
J. Öhman, P. Geladi, S. Wold. *Residual Bilinearization. Part 1: Theory and Algorithms*, J. Chemom. 4: 79-90 (1990)

Olivieri and Faber 2004
A. C. Olivieri, N. M. Faber. *Standard Error of Prediction in Parallel Factor Analysis of Three-Way Data*. Chemom. Intell. Lab. Syst. 70: 75-82 (2004)

ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Olivieri *et al*. 2006

A. C. Olivieri, N. M. Faber, J. Ferré, R. Boqué, J. H. Kalivas, H. Mark. *Uncertainty Estimation and Figures of Merit for Multivariate Calibration*, Pure Appl. Chem., 78: 633-661 (2006)

Ortiz *et al*. 2003

M. C. Ortiz, L. A. Sarabia, A. Herrero, M. S. Sánchez, M. B. Sanz, M. E. Rueda, D. Giménez, M. E. Meléndez. *Capability of Detection of an Analytical Method Evaluating False Positive and False Negative (ISO 11843) with Partial Least Squares*, Chemom. Intell. Lab. Syst. 69: 21-33 (2003)

Rubinson and Rubinson 2000

K. A. Rubinson, J. F. Rubinson. *Contemporary Instrumental Analysis*, Prentice-Hall: New Jersey, 2000

Sánchez and Kowalski 1986

E. Sanchez, B. R. Kowalski. *Generalized Rank Annihilation Factor Analysis*, Anal. Chem. 58:496-499 (1986)

Sánchez and Kowalski 1990

E. Sánchez, B. R. Kowalski. *Tensorial Resolution: a Direct Trilinear Decomposition*, J. Chemom. 4: 29-45 (1990)

Sanchez *et al*. 1988

E. Sanchez, B.R. Kowalski. *Tensorial Calibration: II. Second-Order Calibration*, J. Chemom. 2: 265-280 (1988)

Sarabia *et al*. 2002

L.A. Sarabia, M. Cruz Ortiz, M. Julia Arcos, M. Sagrario Sánchez, A. Herrero, S. Sanllorente. *Multivariate Detection Capability Using a Neural Classifier for Nonselective Signals*, Chemom. Intell. Lab. Syst. 61: 89-104 (2002)

Saurina *et al*. 2001

J. Saurina, C. Leal, R. Campañó, M. Granados, M.-D. Prat, R. Tauler. *Estimation of Figures of Merit Using Univariate Statistics for Quantitative Second-Order Multivariate Curve Resolution*, Anal. Chim. Acta 432: 241-251 (2001)

Skoog *et al*. 1998

D. A. Skoog, F. J. Holler, T. A. Nieman. *Principles of Instrumental Analysis*, Fifth Edition. Saunders College Publishing: Fort Worth (USA), 1998

Smilde *et al*. 2004

A. Smilde, R. Bro, P. Geladi. *Multi-Way Analysis. Application in the Chemical Sciences*, John Wiley & Sons, Chichester, 2004.

Tauler 1995

R. Tauler. *Multivariate Curve Resolution Applied to Second Order Data*, Chemom. Intell. Lab. Syst. 30: 133-146 (1995)

Tauler and de Juan

R. Tauler, A. de Juan, *Multivariate Curve Resolution Homepage*
URL: http://www.ub.es/gesq/mcr/mcr.htm

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Tauler *et al*. 1995
R. Tauler, A. K. Smilde, B. R. Kowalski. *Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution*, J. Chemom. 9: 31-58 (1995)

Thomson *et al*. 2002
M. Thompson, S. L. R. Ellison, R. Wood. *Harmonized Guidelines for Single-Laboratory Validation of Methods of Analysis (IUPAC Technical Report)*. Pure Appl. Chem. 74: 835-855 (2002)

Toft 1995
J. Toft. *Evolutionary Rank Analysis Applied to Multidetectional Chromatographic Structures*, Chemom. Intell. Lab. Syst. 29:189-212 (1995)

Toft and Kvalheim 1994
J. Toft, O. M. Kvalheim. *Multi-array Resolution Parameter for Multidetection Chromatography. Performance of Alternating Regression, Iterative Target Transformation Factor Analysis and Heuristic Evolving Latent Projections*, Chemom. Intell. Lab. Syst. 25: 61-75 (1994)

Torgrip *et al*. 2003
R. J. O. Torgrip, M. Alberg, B. Karlberg, S. P. Jacobson. *Peak Alignment Using Reduced Set Mapping*, J Chemom. 17: 573-582 (2003)

Tucker 1963
L. Tucker. *Implications of Factor Analysis of Three-Way Matrices for Measurement of Change*, pp. 122-137 in C. Harris (Editor), *Problems of Measuring Change*, The University of Wisconsin Press, Madison, WI (1963)

Vandeginste *et al*. 1985
B. G. M. Vandeginste, W. Derks, G. Kateman. *Multicomponent Self-Modelling Curve Resolution in High-Performance Liquid Chromatography by Iterative Target Transformation Analysis*, Anal. Chim. Acta 173: 253-264 (1985)

van der Voet 2002
H. van der Voet, *Detection Limits*, pp. 504-515 in A. H. El-Shaarawi and W. W. Piegorsch (Editors), *Enciclopedia of Environmentrics*, Vol. 1, Wiley, Chichester (2002)

Vessman *et al*. 2001
J. Vessman, R. I. Stefan, J. F. van Staden, K. Danzer, W. Lindner, D. T. Burns, A. Fajgelj, H. Müller. *Selectivity In Analytical Chemistry (IUPAC Recommendations 2001),* Pure Appl. Chem. 73: 1381-1386 (2001)

Wang *et al*. 1993
Y. Wang, O. S. Borgen, B. R. Kowaslki, M. Gu, F. Turecek. *Advances in Second-Order Calibration*. J. Chemom. 7: 117-130 (1993)

Wegscheider 1996
W. Wegscheider. *Validation of Analytical Methods in Quality Assurance* in *Accreditation and Quality Assurance in Analytical Chemistry*, ch. 6, p. 135, ed. by H. Günzler, Springer Verlag: Berlin, 1996

ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Wentzell *et al*. 2001

P. D. Wentzell, S. S. Nair, R. D. Guy. *Three-Way Analysis of Fluorescence Spectra of Polycyclic Aromatic Hydrocarbons with Quenching by Nitromethane*, Anal. Chem. 73: 1408-1415 (2001)

Wilson *et al*. 1989

B. E. Wilson, W. Lindberg, B. R. Kowalski. *Multicomponent Quantitative Analysis Using Second-Order Nonbilinear Data: Theory and Simulations*, J. Am. Chem. Soc. 111: 3797-3804 (1989)

Windig and Guilmet 1991

W. Windig, J. Guilmet. *Interactive Self-Modeling Mixture Analysis*, Anal. Chem. 63: 1425-1432 (1991)

Windig *et al*. 2002

W. Windig, B. Antalek, J. L. Lippert, Y. Batonneau, C. Brémard. *Combined Use of Conventional and Second-Derivative Data in the SIMPLISMA Self-Modeling Mixture Analysis Approach*, Anal. Chem. 74: 1371-1379 (2002)

Wold *et al*. 1997

S. Wold, K. Esbensen, P. Geladi. *Principal Components Analysis*, Chemom. Intell. Lab. Syst. 2: 37-52 (1987)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Appendix

GLOSSARY AND ABBREVIATIONS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Bilinearity

A data matrix is bilinear when the contributions of the compounds in the two *orders* of measurement are additive and therefore it can be decompose into the outer product of two matrices containing the intrinsic factors of each compound in each mode.

Khatri-Rao Product

The Khatri-Rao product of two matrices**,** with the same number of columns, is denoted by the symbol |⊗| (or ⊙). For its definition, the Kronecker tensor, ⊗, should be firstly defined.
The Kronecker tensor product of **A** and **B** where **A** is of size $I \times J$ holds that

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} x_{11}\mathbf{B} & \cdots & x_{1J}\mathbf{B} \\ \vdots & & \vdots \\ x_{I1}\mathbf{B} & \cdots & x_{IJ}\mathbf{B} \end{bmatrix} \qquad \text{Eq. Ap1}$$

Then, the Khatri-Rao product between $\mathbf{A}=[\mathbf{A}_1 \ldots \mathbf{A}_J]$ and $\mathbf{B}=[\mathbf{B}_1 \ldots \mathbf{A}_J]$ is defined as

$$\mathbf{A}|\otimes|\mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B}_1 \ldots \mathbf{A}_K \otimes \mathbf{B}_K] \qquad \text{Eq. Ap2}$$

The Khatri-Rao product is a useful matrix product used in this thesis because it provides an efficient way to write a PARAFAC model, especially simple and transparent for higher-order PARAFAC models.

NAS (Net Analyte Signal)

The NAS concept arises in multivariate calibration from the fact that a prediction sample response, e.g. a spectrum, may have varying contributions from other sample compounds. The spectrum can be decomposed in two orthogonal parts: a part than can be uniquely assigned to the analyte of interest (the NAS), and the remaining part that contains the (possibly varying) contribution from other compounds. Hence the NAS is proportional to the concentration of the analyte of interest.

NAS for first order data is defined in Equation Ap1, in which **I** is the identity matrix, $\mathbf{R}_{-f}$ is the matrix of pure profiles (e.g. spectra) of all constituents except the *f*th analyte, and $\mathbf{r}_f$ is the profile (e.g. spectrum) of the analyte. The superscript + symbolizes the pseudoinverse. NAS is a vector, and it is related to the regression

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

vector **b** by Equation Ap2.

$$\mathrm{NAS}_f = \left(\mathbf{I} - \mathbf{R}_{-f}\mathbf{R}_{-f}^{+}\right)\mathbf{r}_f \qquad \text{Eq. Ap1}$$

$$\left\|\mathbf{b}\right\| = \frac{1}{\left\|\mathrm{NAS}_f\right\|} \qquad \text{Eq. Ap2}$$

Since the NAS is the only part of the spectrum employed for prediction, no information is lost when transforming the vector NAS into a scalar (typically, the Euclidean norm is taken, i.e., its length).

Norm

Mathematically, the norm is any scalar-valued function, denoted by the double bar notation $\|\cdot\|$, which satisfies three properties:
(1) The norm of every *tensor* is positive: $\|\mathbf{A}\| \geq 0$ ($\|\mathbf{A}\| = 0$ if $\mathbf{A} = 0$)
(2) Scaling a vector scales the norm by the same amount, $\|s\mathbf{A}\| = |s|\,\|\mathbf{A}\|$
(3) The "size" of the sum of two tensors is less than or equal to the sum of their "sizes", $\|\mathbf{A}+\mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.
Particular norms are distinguished by subscripts, such as $\|\cdot\|_V$, when referring to a norm in the space *V*. A class of vector and matrix norms, called *p*-norms and denoted by $\|\cdot\|_p$, is defined as:

$$\left\|\mathbf{a}\right\|_p = \left(\left|\mathbf{a}_1\right|^p + \left|\mathbf{a}_2\right|^p + \cdots + \left|\mathbf{a}_n\right|^p \right)^{1/p} \qquad \text{Eq. Ap3}$$

The most widely used are the 1-norm, 2-norm (sometimes also called the Euclidean or Frobenius norm) and ∞-norm, shown for vectors and for matrices in the next table:

	For a column vector **a** with *M* elements	For a matrix **A** with dimensions *N*×*M*
1-norm	$\left\|\mathbf{a}\right\|_1 = \left\|\mathbf{a}_1\right\| + \left\|\mathbf{a}_2\right\| + \cdots + \left\|\mathbf{a}_M\right\|$	$\left\|\mathbf{A}\right\|_1 = \max_{1\leq j\leq M} \sum_{i=1}^{N} \left\|a_{i,j}\right\|$ Maximum absolute column sum
2-norm	$\left\|\mathbf{a}\right\|_2 = \sqrt{\sum_{i=1}^{M} \left\|a_i\right\|^2} = \sqrt{\mathbf{a}^T\mathbf{a}}$	$\left\|\mathbf{A}\right\|_2 = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} \left\|a_{i,j}\right\|^2}$ or the square root of the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$
∞-norm	$\left\|\mathbf{a}\right\|_\infty = \max_{1\leq i\leq M} \left\|\mathbf{a}_i\right\|$	$\left\|\mathbf{A}\right\|_\infty = \max_{1\leq i\leq N} \sum_{j=1}^{M} \left\|a_{i,j}\right\|$ Maximum absolute row sum

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Orthogonality

Two vectors, $\mathbf{a}=(a_1,a_2,\ldots,a_i)'$ and $\mathbf{b}=(b_1,b_2,\ldots,b_i)'$, are orthogonal when their cross product sum is zero: $(a_1\, b_1 + a_2\, b_2 + \ldots + a_i\, b_i) = 0$. In a three-dimensional space that means that the angle between them is 90º. For higher dimensional space, this concept can be taken as the definition of orthogonality, which can be written in matrix algebra as $\mathbf{a}'\mathbf{b}=0$.

SVD (Singular Value Decomposition)

Any data matrix **A**, of dimension $I \times J$, can be decomposed according to the relationship: $\mathbf{A=UWV}^T$, where **U** ($I \times J$) is related to the scores of the objects, **V** ($J \times J$) is related to the loadings of the original (or manifest) variables and **W** ($J \times J$) is a diagonal matrix related to the variation explained by successive latent variables. The elements on the diagonal of **W** are arranged in descending order of magnitude and correspond to the singular values w_i, i.e. the square root of the eigenvalues λ_i ($\lambda_i = w_i^2$ or in matrix notation $\mathbf{\Lambda=W}^2$).

Tensor

Mathematical object defined as a generalization of scalars, vectors and matrices. It can be represented by an array of components that fulfil certain transformation rules (Budianski 1974).

Dimensionality of a tensor: Number of elements in each *order*. Figure Ap1 shows an example of third-order tensor with different dimensionality in each order.

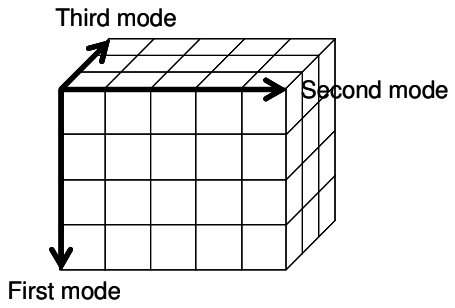


Figure Ap1. Third-order tensor with the following dimensionalities: 4 in the first order, 5 in the second mode and 3 in the third mode. Each cube represents a datum, i.e. a scalar.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Order of a tensor: Number of spaces that the *tensor* spans. Also named 'mode', 'way' and even 'dimension' (do no confuse this term with the *dimensionality* of the tensor).

Rank of a tensor: Number of varying independent factors for a given *order of the tensor*. It cannot be greater than the *dimensionality* of that order.

Trilinearity

Extension of the *bilinearity* concept to a third mode. That involves that the underlying profiles are reproducible, i. e. they have identical shape and only different intensity (proportional to concentration) from experiment to experiment.

Unfolding

In the chemometric community, to unfold a first- or higher-order tensor is to reorganize its elements for a particular convenience. Also called matricizing, concatenation, augmentation, etc., this reorganization can help to handle the data but, on the other hand, some information may be lost in the process, such as the interrelation between the slices.

The following figures exemplify the principle of unfolding:

Figure Ap2: The second-order tensor can be concatenated row-wise to a row vector (or column-wise to a column vector).

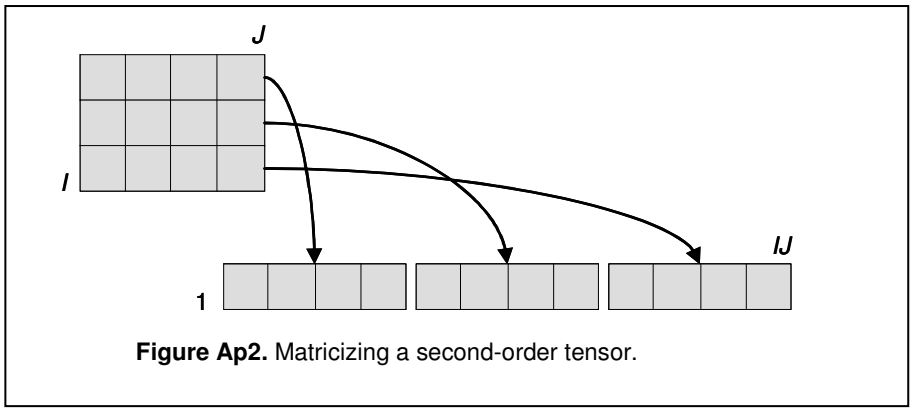


Figure Ap2. Matricizing a second-order tensor.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Figure Ap3: a third-order tensor of dimension $I \times J \times K$ can be unfolded preserving one mode and confounding the other two. When the first mode is preserved, the left-most matrix in the resulting array is the $I \times J$ matrix equal to the first frontal slab (K=1). The following frontal slices are concatenated to the right side of the previous one. The resulting array has dimension $I \times JK$. When the second mode is preserved, the frontal slices are arranged one on the bottom of the other building a $IK \times J$ array. When the preserved mode is the third, the resulting array has dimension $K \times IJ$.

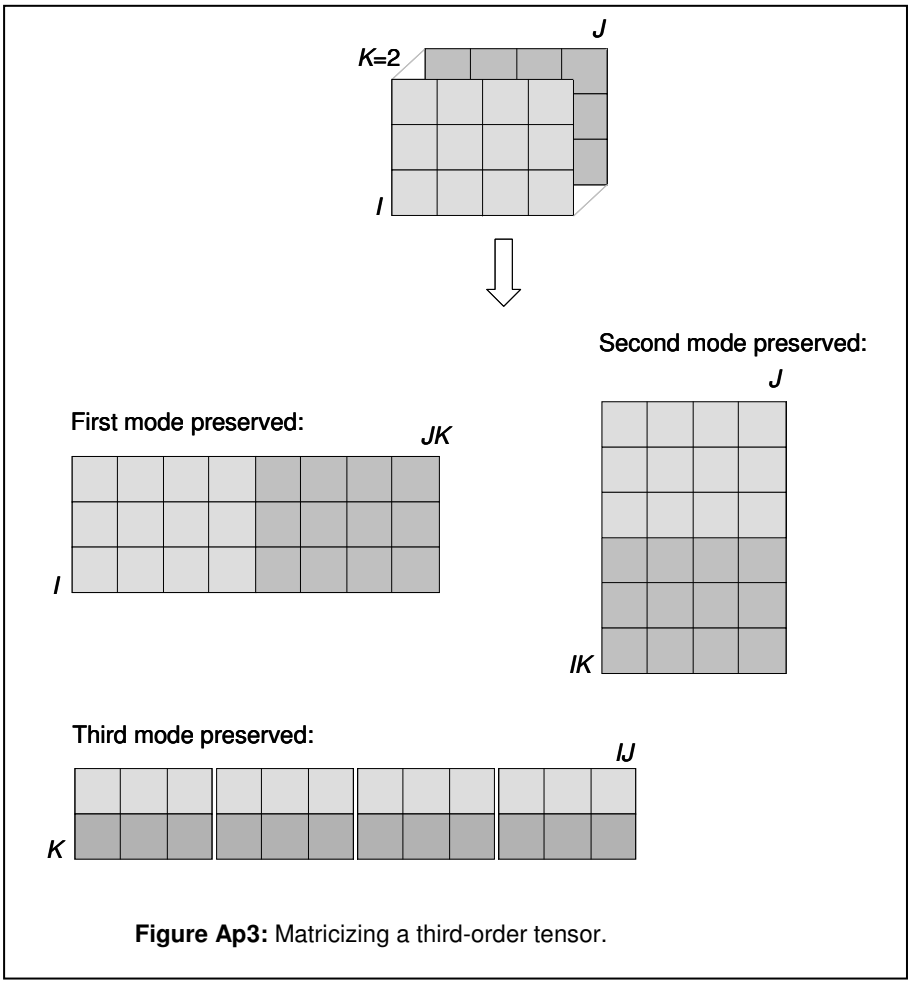


Figure Ap3: Matricizing a third-order tensor.

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Abbreviations

ANOVA	Analysis of Variance
ALS	Alternating Least Squares
CAS	Chemical Abstracts Service
CE	Capillary Electrophoresis
CLS	Classical Least Squares
CMP	Chemical Measurement Process
CS	Cross-Section
DAD	Diode Array Detector
DTD	Direct Trilinear Decomposition
EEM	Excitation-Emission Matrix
EFA	Evolving Factor Analysis
EPA	Environmental Protection Agency
FAO	Food and Agricultural Organization
FTIR	Fourier Transform Infrared Spectroscopy
GRAM	Generalized Rank Annihilation Method
GC	Gas Chromatography
HELP	Heuristic Evolving Latent Projections
HPLC	High Performance Liquid Chromatography
IR	Infrared
ILS	Inverse Least Squares
ISO	International Organization for Standardization
ITTFA	Iterative Target Transformation Factor Analysis
IUPAC	International Union of Pure and Applied Chemistry
LC	Liquid Chromatography
LOD	Limit of Detection
MCR	Multivariate Curve Resolution
MLR	Multiple Linear Regression
MS	Mass Spectrometry
NAS	Net Analyte Signal
NBRA	Non-Bilinear Rank Annihilation
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
N-PLS	Multi-linear Partial Least Squares
NS	Needle Search
OLS	Ordinary Least Squares
OPA	Orthogonal Projection Analysis
PARAFAC	Parallel Factor Analysis
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

PMF2	Positive Matrix Factorization
RA	Rank Annihilation
RAFA	Rank Annihilation Factor Analysis
RBL	Residual Bilinearization
RMSEP	Root Mean Square Error of Prediction
RMSPD	Root Mean Square of Percentage Deviation
SEL	Selectivity
SEN	Sensitivity
SIMPLISMA	Simple-to-use Interactive Self-modelling Analysis
SKSS	Stepwise Key Spectrum Selections
SVD	Singular Value Decomposition
TGA	Thermogravimetric Analysis
U-PCA	Unfold Principal Component Analysis
UV-Vis	Ultraviolet Visible
WHO	World Health Organization

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

Appendix

LIST OF PAPERS AND MEETING CONTRIBUTIONS

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

List of papers

List of papers by the author presented in this thesis (in chronological order):

1. M. J. Rodríguez Cuesta, R. Boqué, F. X. Rius
Influence of selectivity and sensitivity parameters on detection in multivariate curve resolution of chromatographic second-order data.
Analytica Chimica Acta 476: 111-122 (2003)
(Chapter 4)

2. M. J. Rodríguez Cuesta, R. Boqué, F. X. Rius, D. Picón Zamora, M. Martínez Galera, A. Garrido Frenich
Determination of carbendazim, fuberidazole and thiabendazole by three-dimensional excitation–emission matrix fluorescence and parallel factor analysis.
Analytica Chimica Acta 491: 47-57 (2003)
(Chapter 3)

3. M.J. Rodríguez Cuestą, R. Boqué, F. Xavier Rius, J. L. Martínez Vidal, A. Garrido Frenich
Development and validation of a method for determining pesticides in groundwater from complex overlapped HPLC signals and multivariate curve resolution.
Chemometrics and Intelligent Laboratory Systems 77(1-2): 251-260 (2005)
(Chapter 5)

4. M. J. Rodríguez Cuesta, R. Bro, R. Boqué and N.M. Faber
Standard error of prediction at low content levels and limit of detection estimation for multivariate and multi-linear regression.
Chemometrics and Intelligent Laboratory Systems (*In press*)
(Chapter 6)

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

List of meeting contributions

List of contributions to international meetings (in chronological order):

1. M. J. Rodríguez Cuesta, R. Boqué, F. X. Rius
Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data
CAC'02 – 8th Chemometrics in Analytical Chemistry
Seattle (EE.UU.) September, 2002
Poster communication

2. M. J. Rodríguez Cuesta, R. Boqué, F. X. Rius, A. Garrido Frenich, M. Martínez Galera, D. Picón Zamora
Determination of carbendazim, thiabendazole and fuberidazole by three-dimensional EEM fluorescence and PARAFAC
CCM5 – V Colloquium Chemometricum Mediterraneum
Ustica (Italy). June, 2003
Oral communication

3. M. J. Rodríguez Cuesta, R. Boqué, F. X. Rius, R. Bro
Standard error of prediction at low content level and limit of detection estimation for multivariate and multi-way PLS.
CAC'04 – 9th Chemometrics in Analytical Chemistry
Lisbon (Portugal). September, 2004
Poster communication

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007

UNIVERSITAT ROVIRA I VIRGILI
LIMIT OF DETECTION FOR SECOND-ORDER CALIBRATION METHODS.
M. José Rodríguez Cuesta
ISBN: 978-84-690-7787-0 / DL: T.1349-2007