



TECHNICAL REPORT



Chemometrics for process analytical technologies – Part 1: General provisions, and methods for univariate statistics and chemometric processing of data

INTERNATIONAL
ELECTROTECHNICAL
COMMISSION

ICS 25.040.40

ISBN 978-2-8322-7584-9

Warning! Make sure that you obtained this publication from an authorized distributor.

CONTENTS

CONTENTS	2
FOREWORD	4
INTRODUCTION	6
1 Scope	8
2 Normative references	8
3 Terms and definitions	8
4 Fields of application.....	8
4.1 Process control and process analytical technologies (PAT).....	8
4.2 Physical and chemical properties	9
4.3 PAT fields of application	10
4.3.1 Definition of chemometrics.....	10
4.3.2 Overview on PAT fields of applications	10
4.3.3 Chemometrics for sensors	10
4.3.4 Chemometrics for production units.....	11
4.3.5 Chemometrics along a production chain	11
5 Pre-requisites of chemometric data analysis	12
5.1 Data has to be adequate and reliable.....	12
5.2 Data representativeness	12
5.3 Data acquisition	13
5.4 Data management.....	13
5.5 Databases versus spreadsheets	13
5.6 Data quality	14
5.7 Data validation.....	14
5.8 Data corruption	14
5.9 Data security and fraudulent data detection	14
5.10 Data management for data mining	15
6 Pre-requisites of chemometric data analysis	15
6.1 Technical requirements of chemometric data analysis.....	15
6.2 Data dimensionality	15
6.3 Method classification	16
6.4 Data pre-processing.....	17
6.4.1 Filtering	17
6.4.2 Smoothing	17
6.4.3 Data reduction	17
7 Methods of chemometric data analysis	20
7.1 Univariate analysis.....	20
7.1.1 Descriptive statistics	20
7.1.2 Hypothesis testing	21
7.1.3 Analysis of variance (ANOVA)	23
7.1.4 General linear models.....	25
7.2 Bivariate analysis.....	25
7.2.1 Regression analysis.....	25
7.2.2 Time series analysis	28

Annex A (informative) Advice on software validation for process analytical applications.....	31
A.1 General.....	31
A.2 Basic recommendations	31
A.3 Software validation	33
Annex B (informative) Reference data sets available for software benchmarking	35
Bibliography.....	36
Figure 1 – Different levels of chemometric applications: (a) within an (intelligent or smart) sensor, (b) within a production unit, e.g., process control system, process control environment, or laboratory information management system (LIMS), (c) along a production chain, e.g., ERP system, data mining, etc.....	10
Figure 2 – influence of pre-processing techniques for classification of the geographical origin of wine	19
Figure A.1 – Different paths for the introduction of new software in a laboratory	33
Table 1 – Data analysis techniques and data formats	17
Table A.1 – Categories of software	32
Table A.2 – Software validation levels	32

INTERNATIONAL ELECTROTECHNICAL COMMISSION

CHEMOMETRICS FOR PROCESS ANALYTICAL TECHNOLOGIES –

Part 1: General provisions, and methods for univariate statistics and chemometric processing of data

FOREWORD

- 1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
- 2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
- 3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
- 4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
- 5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
- 6) All users should ensure that they have the latest edition of this publication.
- 7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
- 8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.
- 9) Attention is drawn to the possibility that some of the elements of this IEC Publication may be the subject of patent rights. IEC shall not be held responsible for identifying any or all such patent rights.

The main task of IEC technical committees is to prepare International Standards. However, a technical committee may propose the publication of a Technical Report when it has collected data of a different kind from that which is normally published as an International Standard, for example "state of the art".

IEC TR 62869-1, which is a Technical Report, has been prepared by subcommittee 65B: Measurement and control devices, of IEC technical committee 65: Industrial-process measurement, control and automation.

The text of this Technical Report is based on the following documents:

Enquiry draft	Report on voting
65B/1062/DTR	65B/1095B/RVDTR

Full information on the voting for the approval of this Technical Report can be found in the report on voting indicated in the above table.

This document has been drafted in accordance with the ISO/IEC Directives, Part 2.

A list of all parts in the IEC 62829 series, published under the general title *Chemometrics for process analytical technologies*, can be found on the IEC website.

The committee has decided that the contents of this document will remain unchanged until the stability date indicated on the IEC website under "<http://webstore.iec.ch>" in the data related to the specific document. At this date, the document will be

- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

A bilingual version of this publication may be issued at a later date.

IMPORTANT – The 'colour inside' logo on the cover page of this publication indicates that it contains colours which are considered to be useful for the correct understanding of its contents. Users should therefore print this document using a colour printer.

INTRODUCTION

Chemometrics is a rapidly developing subject. It was thus felt that a report offering guidance on its application to process analytical applications would both be helpful to all users of such technology and would stimulate specialists in chemometrics to work with users and developers of this technology.

This document does not seek to do other than provide a useful overview and a brief bibliography that enables interested parties to learn about and, hopefully, apply chemometrics in the most useful and appropriate ways for their circumstances. In that sense, it is definitely not prescriptive but constructively critical and seeks to encourage good practice and a wider appreciation.

It also aims at encouraging new research and development, as well as innovation, in applications of chemometrics for process analytical applications by highlighting areas to which such activities might usefully be directed.

Nowadays, the use of chemometric data analysis methods is widespread. Applications are in fields like

- design of statistical/chemometric sampling strategies, design of experiments, design of observational studies,
- design of data collection (including signal processing) protocols, data validation methods and database management (including metadata management),
- quality management, including quality assurance and quality control,
- data analysis and interpretation, not only in the use of multivariate (many variable) methods but also univariate (one variable) and bivariate (two variable) methods,
- process monitoring, optimization and control,
- chemical process and property modelling,
- guiding decision analysis and designing decision analysis methods/protocols in process control and optimization,
- method and instrumentation performance validation (Annex A) and calibration.

Because of the interdisciplinary and multidisciplinary nature of the discipline of chemometrics, it is often possible to be able to make unusual links and thereby solve problems taking cues from disciplines that are as diverse as medical diagnostics, decision sciences and quality assurance.

For example, in diagnosing the likely environmental impact of discharges of waste water from an industrial process, we might want to link toxicity assessment to chemical composition, the route and extent of discharge and the organisms likely to be affected. This might involve establishing a chemometric (mathematical) model of the impact of the discharge, bio-sensing the toxicity of the discharge on-line and relating both to the time, volume and concentration variations in chemical composition and physicochemical properties. This could then be used to assess the predictive reliability of the model and how this might be linked to process control and optimization of the discharge treatment and any associated risk assessment of the discharge process.

Conventionally, process control has involved using control charts for individual variables and this sometimes leads us to false impressions of process behaviour. Since 2010, techniques including both commercial and other software have become available to construct a wide variety of useful multivariate control charts that sometimes reveal "out-of-control" situations not apparent using conventional univariate control charts.

Due to the applicability of chemometric methods to a nearly unlimited number of cases in all fields of measurement and testing, but particularly due to need of using chemometric techniques in process analytical applications, it was felt a necessity to have guidance on the available methods and their appropriate choice.

CHEMOMETRICS FOR PROCESS ANALYTICAL TECHNOLOGIES –

Part 1: General provisions, and methods for univariate statistics and chemometric processing of data

1 Scope

This part of IEC 62829, which is a Technical Report, covers

- a study into the pre-requisites of chemometric (exploratory) data analysis,
- an overview of common data analysis procedures for univariate, bivariate and multivariate data analysis,
- explanations of the basic principles and major application areas of the different methods),
- some recommendations on the selection of an appropriate data analysis strategy.

These recommendations not covered earlier by other guidance documents on the topic are complemented by some advice on the validation of commercial (at the site of installation) and tailored software for process analytical purposes. Recommendations are given on available reference data sets (Annex B) for benchmarking of software implementing the data analysis methods covered (if available). An application example is given.

2 Normative references

There are no normative references in this document.