IEC TR 62829-1

Edition 1.0 2019-11

TECHNICAL REPORT

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

IEC TR 62829-1:2019-11(en)

IEC Central Office
3, rue de Varembé
CH-1211 Geneva 20
Switzerland

Tel.: +41 22 919 02 11
info@iec.ch
www.iec.ch

About the IEC
The International Electrotechnical Commission (IEC) is the leading global organization that prepares and publishes International Standards for all electrical, electronic and related technologies.

About IEC publications
The technical content of IEC publications is kept under constant review by the IEC. Please make sure that you have the latest edition, a corrigendum or an amendment might have been published.

IEC publications search - webstore.iec.ch/advsearchform
The advanced search enables to find IEC publications by a variety of criteria (reference number, text, technical committee,…). It also gives information on projects, replaced and withdrawn publications.

IEC Just Published - webstore.iec.ch/justpublished
Stay up to date on all new IEC publications. Just Published details all new publications released. Available online and once a month by email.

IEC Customer Service Centre - webstore.iec.ch/csc
If you wish to give us your feedback on this publication or need further assistance, please contact the Customer Service Centre: sales@iec.ch.

Electropedia - www.electropedia.org
The world's leading online dictionary on electrotechnology, containing more than 22 000 terminological entries in English and French, with equivalent terms in 16 additional languages. Also known as the International Electrotechnical Vocabulary (IEV) online.

IEC Glossary - std.iec.ch/glossary
67 000 electrotechnical terminology entries in English and French extracted from the Terms and Definitions clause of IEC publications issued since 2002. Some entries have been collected from earlier publications of IEC TC 37, 77, 86 and CISPR.

IEC

IEC TR 62829-1

Edition 1.0 2019-11

TECHNICAL REPORT

colour inside

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

CONTENTS

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.

3 Terms and definitions

No terms and definitions are listed in this document.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- IEC Electropedia: available at http://www.electropedia.org/
- ISO Online browsing platform: available at http://www.iso.org/obp

4 Fields of application

4.1 Process control and process analytical technologies (PAT)

There is currently a considerable trend to use process analytical technologies (PAT) for reaction monitoring and (direct loop) process control. Current developments in the field of process engineering are not imaginable without PAT, such as modern process design, integrated processes (e.g., reactive separation processes), and intensified processes along with requirements to process control, model-based control, and soft sensing – all involving chemometrics.

The process industry relies on the design, operation, control, and optimization of chemical, physical, or biological processes. This involves creating production facilities that translate raw materials into value-added products along the supply chain. Such conversions typically take place in repeated reaction and separation steps – either in batch or continuous processes. The end products of a chemical production facility are the result of several production steps that are connected not only in a sequential fashion, but also involve recycling of unused raw materials and by-products, as well as waste treatment stages. Production processes in the process industry are particularly disturbed by variations in feed-stocks and other influences that impact the product quality. An integrated process control approach enables constant product quality and prevents out-of-spec production by effectively compensating for such process variations. In a conventional approach, quality is determined by withdrawing samples from material streams and conducting offline analytics, which is called in-process control or at-line or off-line control. By applying quality by design (QbD, see ICH definition in 4.2) approaches, quality can significantly improve to generate less waste, reduce reprocessing of substandard material, and create products of superior quality.

Today's optimized process design relies heavily on computer aided tools, which account for, for example, mass transfer, thermodynamic, kinetic, and other physical properties of the treated materials. Typically, a sufficient understanding of such properties is available and implemented in dynamic numeric models. Dynamic models are in turn the essential basis for optimized process and plant design. Unfortunately, they are only sparsely used for process control. A definition from Lee (2008) brings this to a contemporary level:

"Cyber-physical systems (CPS) are integrations of computation with physical processes. Embedded computers and networks monitor and control the physical processes, usually with feedback loops where physical processes affect computations and vice versa."

4.2 Physical and chemical properties

Process analytical techniques are extremely useful tools for chemical production and manufacture and are of particular interest to the pharmaceutical, food and (petro-)chemical industries. It can be easily transferred to manufacturing for process control and for quality assurance of final products to meet required product specifications, since it provides dynamic information about product properties, material stream characteristics, and process conditions.

Normally, the quality of the final product is assessed after processing by adequate testing procedures. The rationale behind process analytical technologies (PAT) is to measure, and assess, physical and chemical properties over, and throughout, the production process in order to assure a product which is within the tolerance limits or regulatory restrictions.

The quality of any product or the properties of a material can be described by a complex functional relationship to physical and/or chemical properties of the constituents of the product/material and its temporal changes during processing.

According to ICH Q8(R2), quality is the suitability of either a drug substance or drug product for its intended use (ICH: International Conference on Harmonisation). This term includes attributes such as identity, strength, and purity. Before the launch of the PAT initiative, pharmaceutical production was confronted with challenges like drug shortages due to manufacturing difficulties, process deviations coupled with frequent inconclusive investigations, batch failures and rejections, in-process test debates (e.g., blend uniformity), slow and protracted cGMP (current good manufacturing practice) remediation, warning letters, and others.

The science-based regulatory guidances such as the FDA (US Federal Drug Administration) and ICH PAT guidance have recognized spectroscopic techniques as potentially useful tools on building quality into the product and manufacturing processes, as well as continuous process improvements. The goal of PAT is to enhance understanding and thereby control the manufacturing process.

The common future vision in pharmaceutical production is continuous manufacturing (CM), based on real-time release (RTR), i.e. a risk-based and integrated quality control in each process unit. This will allow for flexible hook-up of smaller production facilities, production transfer towards fully automated facilities (featuring less operator intervention and less down-time), and end-to-end process understanding over product life cycle, future knowledge, and faster product to market.

Both methods, i.e. in-process and final-product testing, intensively use statistical methods as described in this document.

4.3 PAT fields of application

4.3.1 Definition of chemometrics

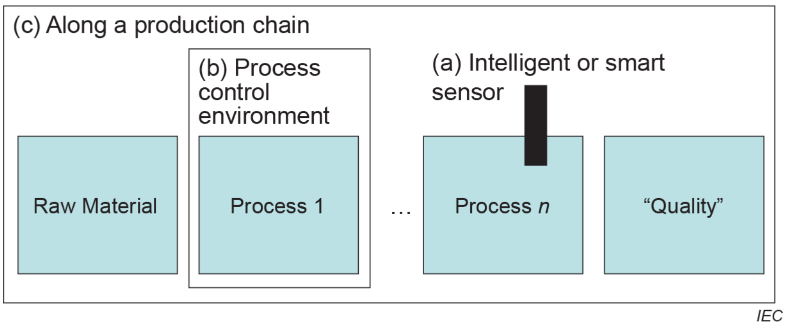
At the dawn of chemometrics, an appropriate definition of the term was that chemometrics is "… the use of statistical, mathematical and other logic-based techniques, together with chemical knowledge to solve chemical problems" (D1) with a clear emphasis on chemical problems, thus explaining the word "chemo" in the term. A definition taken from the Journal of Chemical Information and Computer Sciences (1975), Vol. 15, page 201, defines chemometrics as the "development/application of mathematical/statistical methods to extract useful chemical information from chemical measurements" (D2) thus detailing the aim, namely the extraction of (useful) information from measurement data.

"Chemometrics is a sub-discipline of metrology dealing with the application of mathematical, statistical and other methods employing formal logic to evaluate and interpret (chemical, analytical) data, optimize and model (chemical, analytical) processes and instrumentation, extract maximum (chemical, analytical) information from experimental and observational data" (D3).

To date, no internationally agreed and standardized definition of chemometrics exists. Although all three definitions reflect, in principle, both the intention of, and the instruments used for, chemometrics, definition D3 is the most general. The application of the principles and tools of chemometrics is explicitly not limited to the field of chemical/analytical measurement, so definition D3 may by read and used without the specification "chemical/analytical", deliberately put in parentheses here.

4.3.2 Overview on PAT fields of applications

Fields of application of PAT are described below and visualized in Figure 1.



(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

Figure 1 – Different levels of chemometric applications

4.3.3 Chemometrics for sensors

Sensors are the sense organs of process automation. At present there are serious changes in the areas of information and communication technology, which offer a great opportunity for optimized process control and value-added production with dedicated network communicating sensors. These kinds of smart or intelligent sensors (see Figure 1) provide services within a network and use information from the process information systems.

Intelligent field devices, digital field networks, Internet protocol (IP)-enabled connectivity and web services, historians, and advanced data analysis software are providing the basis for the future project “Industrie 4.0”, and industrial Internet of Things (IIoT). This is a prerequisite for the realization of cyber physical systems (CPS) within these future automation concepts for the process industry.

As a consequence, smart process sensors enable new business models for users, device manufacturers, and service providers.

4.3.4 Chemometrics for production units

With the introduction of advanced process analytical technology, the closeness of key process variables to their limits can be directly monitored and controlled and the processes can automatically be driven much closer to the optimal operating limit. Classical, non-model-based solutions reach their limits when sensor information from several sources has to be merged. In addition, their adaptation causes a high effort during the life cycle of the process. This calls for adaptive control strategies, which are based on dynamic process models as mentioned above. Model-based control concepts have also the potential to automatically cope with changes of the raw-materials as well as process conditions.

Chemical process control technology has advanced significantly during the last decades. For world-scale high-throughput continuous processing units such as crackers and separation trains, in most cases classically engineered control solutions (proportional–integral–derivative (PID) controllers, cascade and override structures) have been replaced by model-based techniques, most prominently model-predictive control (MPC) based on linear plant models. However, the engineering and implementation costs of such advanced controllers are still high. For smaller, flexible processes in which varying products or intermediates are manufactured, it is not economic to re-engineer the control concept or to re-model the process for all intended processes.

Advanced control strategies have to be built upon empirical – often data-based – models which describe cause-effect mappings between the degrees of freedom of the process and the product properties in a black-box fashion. Chemometric techniques for the derivation of empirical models – e.g. partial least squares (PLS), principal component analysis (PCA) – are available but currently mostly used for off-line data analysis to detect the causes of variations in the product quality. An automated application along the life cycle is still very limited. The development of such models requires significant experimental work, and the reduction of the effort needed for these experiments is the focus of ongoing research. When such stationary models are available and are combined with dynamic models that describe the times needed for the transition from one steady state to the other, feedback control and iterative optimization schemes can be built that make use of the novel sensors.

The departure from current automation measurement to smart sensor systems has already begun. Further development is based on the actual situation over several steps. Possible perspectives will be via additional communication channels to mobile devices, bidirectional communication, integration of the cloud and virtualization. The cost of connectivity is dropping dramatically, providing powerful potential to connect people, assets, and information across the industrial enterprise. While only providing add-on information, the first cloud services may not require a high availability or real-time capabilities. But when available in the future, even process control tasks will be possible using cloud services, e.g. when complex computing algorithms are needed, which require computing performance and availability.

4.3.5 Chemometrics along a production chain

Current focuses of research are closed-loop adaptive control concepts for plant-wide process control, which make use of specific or non-specific sensors along with conventional plant instruments. Such advanced control solutions could give more information than only control information, such as sensor failure detection, control performance monitoring, and improve simulation-based engineering. At present, such data is typically collected and analysed in an enterprise resource management system (ERP).