analytical chemistry

pubs.acs.org/ac

Feature

How Can Chemometrics Support the Development of Point of Need Devices?

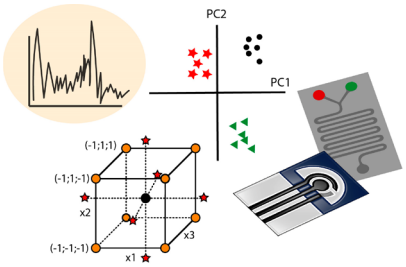
The necessity to establish novel solutions for decentralized monitoring is attracting attention in all fields of analytical chemistry, i.e., clinical, pharmaceutical, environmental, agri-food. The research around the terms "point-of-need", "point-of-care", "lab-on-chip", "biosensor", "microfluidics", etc. is/has been always aimed at the possibility to produce easy-to-use and fast-response devices to be used by nonspecialists. However, the routes to produce the optimal device might be time-consuming and costly. In this Feature, we would like to highlight the role of chemometric-based approaches that are useful in the conceptualization, production, and data analysis in developing reliable portable devices and also decrease the amount of experiments (thus, costs) at the same time. Readers will be provided a concise overview regarding the most employed chemometric tools used for target identification, design of experiments, data analysis, and digitalization of results applied to the development of diverse portable analytical platforms. This Feature provides a tutorial perspective regarding all the major methods and applications that have been currently developed. In particular, the presence of a concise and informative table assists analytical chemists in utilizing the right chemometrics-based tool depending on the architectures and transduction.

Sara Tortorella and Stefano Cinti*

Cite This: *Anal. Chem.* 2021, 93, 2713−2722

Read Online

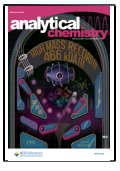
ACCESS | Metrics & More | Article Recommendations



DECENTRALIZED ANALYTICAL CHEMISTRY

The role of analytical chemistry to provide facile and affordable solutions for the major fields of action, namely, clinical, pharmaceutical, environmental, and agri-food, is continuously facing challenges.[1,2] The research and the economic efforts, along with the development of innovative and breakthrough point-of-need tools (including point-of-care and lab-on-chip devices), represents a hot topic in the analytical sciences. Although few of these efforts have been capable of generating marketable solutions, the glucose biosensor for diabetic patients and the lateral flow immunoassay strip for pregnancy tests still roughly encompasses the total market. Public authorities, nonprofit foundations, and private companies, e.g., EU Commission, NIH, Bill and Melinda Gates, Wellcome Trust, AIRC, Roche, Samsung, Google, etc., are committed to funding researchers who aim to enable nonspecialized customers and citizens, worldwide, to be actively involved in monitoring analytes. Devices to improve self-healthcare (diagnostics and personalized treatments), tools for evaluating environmental pollution and the effectiveness of remediation, and portable solutions to improve crop productivity while adapting to the effects of climate change are only a few of the contexts where analytical chemistry plays a leading role.[3−5] The COVID-19 pandemic is only the last example that highlights the necessity of user-friendly, rapid, and affordable devices for the use of nonspecialists, as already established by the WHO with the ASSURED criteria.[6,7] In addition to this, the obvious limitations existing in developing/remote countries, in the frame of the 2030 Agenda for Sustainable Development (UN), represent a clear objective to be extended within a multidisciplinary vision.[8] The research around the various architectures and principles of analytical methods, e.g., electrochemistry, colorimetry, fluorimetry, spectrophotometry, spectrometry, chromatography, has been merged to other disciplines such as biology, biochemistry, organic chemistry, material science, engineering, and microfabrication, with the

Published: January 26, 2021



https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

aim to reduce tasks for the end-user, to improve the reliability and (possibly) to reduce costs (e.g., the use of eco-friendly materials like paper-based substrates,[9−11] the synthesis of biomimetic nanomaterials,[12,13] the rational design of recognition probes (aptamers), the multiplexing approach, microfluidics for lowering sample treatment/chemicals use/waste management.)[14−16] Among all the aforementioned objectives, a common feature is recognized: making analytical processes more convenient in terms of (i) fabrication (use of synthetic materials instead of animal sources, e.g., oligonucleotide aptamers vs antibodies), (ii) application (in situ use, e.g., point-of-care vs laboratory-bounded), (iii) environment (reduction of waste production, e.g., paper vs plastic), (iv) social impact (improving citizen participation, e.g., non-specialists vs skilled personnel), and (v) economics (limiting the use of prime matters/maintenance, e.g., microfluidics vs bulky/expensive approaches). However, the path from conceptualization to market, through design and data analysis, still appears dependent on a univariate paradigm.[17] The choice of a target, optimization of a sensor, analysis of a signal, noise discrimination, and evaluation of experimental parameters, e.g., stability, reproducibility, shelf life, are mainly derived from linear correlation. For instance, the amount of an enzyme to develop a biosensor is optimized by assuming other experimental parameters are irrelevant and/or the presence of interfering species in complex matrixes might limit real application. Although this approach works in many cases, the adoption of statistical methods to understand the additive effect of multiple species, to evaluate the correlation of experimental variables on signal output, and to discriminate/classify multitargets simultaneously, represent a useful opportunity for moving toward a multivariate perspective.[18] Plenty of information can be extracted from data if more than one variable is considered at the same time: the understanding of how more inputs correlate with each other and affect the output can potentially improve both the analytical performances and the cost of portable devices.[19] The aim of this Feature is to highlight the use of chemometrics for the development of portable analytical devices, with a holistic description. Although other content on chemometrics applied to portable devices has been reported in the literature,[19−21] some novel aspects are included in this Feature: (1) the perspective is extended to aspects that are often not addressed, such as target identification and digitalization, and (2) the reported examples are focused on diverse disciplines around the world of portable devices, microfluidics, selective biosensors, nonspecific array, optical readouts, etc. All the steps from bench to market, namely, target identification, device optimization, signal treatment and data digitalization, can benefit from the adoption of chemometrics. Readers working in the field of point-of-needs would be able to consider the use of novel routes for enhancing their research. As reported in the preface of the book "Chemometrics in Electroanalysis",[21] Prof. Scholtz used the following words: "*Still, only a few electrochemists and electroanalysts make use of it, probably because their attention is completely absorbed by the purely electrochemical problems, leaving not much time to study chemometrics.*" This Feature is intended for those operating within the field of point-of-need devices that *still* do not consider the use of chemometrics to improve their outputs. Light theoretical descriptions are combined with practical evidence to support the realization of novel analytical tools.

CHEMOMETRICS AT A GLANCE

Chemometrics was defined for the first time by a young assistant professor writing a grant application as "the art of extracting chemically relevant information from data produced in chemical experiments".[22] It was 1972, and the professor was Svante Wold (professor of organic chemistry at Umeå University, Sweden), today frequently remembered as the "father of chemometrics". A few years later, together with his colleague Bruce Kowalski (professor of analytical chemistry, University of Washington, Seattle, WA), he founded the International Chemometrics Society. As clearly explained by Wold himself in different occasions, the main goal of this new discipline was to get chemically relevant information out of measured chemical data (e.g., design of experiment, DoE, multivariate analysis),[23−25] and to represent and display this information. These complementary tasks clearly demanded knowledge of statistics and applied mathematics, but the approach has always been the fit-for-use: "*We must remain chemists and adapt statistics to chemistry instead of vice versa. And chemometrics must continue to be motivated by chemical problem solving, not by method development.*"[22] From the 70s up to 90s, the development of computerized instruments (especially in analytical and physical-organic chemistry) made data acquisition easier and cheaper. In those years, chemometrics became established and a big effort was put on designing novel algorithms for data information extraction and optimization, analogous to what biologists, psychologists, and economists have done with biometrics, psychometrics, and econometrics, respectively.[26−28] In the late 90s, the application of computer and informatics technology in chemistry led to the coining of the counterpart of bioinformatics in chemistry: chemoinformatics.[29,30] Chemoinformatics is the use of informatics methods to solve chemical problems,[30,31] thus including chemical database systems and structures, computer assisted structural elucidation, computer-assisted drug and chemical synthesis design, and molecular modeling.[29,32,33] Mainly born as a tool to support analytical and physical-chemical data analysis, today chemometrics is applied in both academia and industry to broader areas of analytical chemistry, process optimization, drug design, biomarker discovery, material design, food science, digital and signal processing, image analysis, and omics sciences with the potential to revolutionize the very intellectual roots of problem solving.[33−35] The first big challenge for chemometrics is the reduction of the amount of experiments. Reasons for reducing the number of experiments are trivial: experiments are expensive, time-consuming, and sometimes pose ethical issues (e.g., animal experimentation). Minimizing the number of experiments, without compromising the information content therein, is arguably the main aim of a scientist. To this aim, we can use experimental design, also known as Design of Experiment (DoE): a mathematical framework for planning experiments by changing all involved variables simultaneously, thus extracting the maximum amount of information in the fewest number of experiments. Different mathematical strategies exist and have been extensively described elsewhere.[36−40] However, rather than the difficulty stemming from mathematical aspects, the main stumbling block is the mental attitude required to switch from changing from the one-variable-at-a-time strategy (OVAT) to DoE, which is still underrepresented in the scientific community.[24] When optimizing a biosensor, many variables must be taken into account: the pH, the concentration of the target analyte,

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

temperature, time of reaction, and other parameters depending on the working principle. DoE, by means of, e.g., factorial designs[40] or D-optimal design methodologies,[41] allows one to identify the minimum essential experiments needed to span all sources of the variation and suggests that resulting experimental data will identify the optimal conditions, the variables that most influence the results, and those that do not. In addition to DoE, chemometrics tools and, more generally, statistical tools of data-mining, are commonly used to extract hidden information and enlightening relationships among data. In the simplest case, data come or can be summarized in a block of data X, a $m \times n$ matrix in which for each of the m samples, n experimental variables, or molecular descriptors are reported. Starting from the assumption that the n variables frequently are more than the m samples, they are likely to be correlated, and they can be missing in some samples, multivariate analysis[25] (MVA) is the elected way to extract information through data analysis.[42] We can distinguish two approaches for the application of MVA: unsupervised and supervised. In the unsupervised setting, the goal is to explore the variance in a single block of data X. For that, a matrix factorization can be performed without any *a priori* knowledge (e.g., no information about the class label of data, the number of classes, etc.), so that natural patterns can be elucidated. This approach is ideal to explore data in an unbiased fashion, especially in an early phase of the investigation, when no or little information on variable or analytes involved in the process is available. Among unsupervised multivariate analysis tools, principal component analysis (PCA) is the workhorse in chemometrics. PCA is used to reduce the complexity and the dimensionality of a set of data contained in a matrix by rationalizing the variance and providing an overview of all observations or samples in the data table.[43] The idea around PCA is to reduce the dimensionality of a data set consisting of a large number of variables, by obtaining novel variables (principal components) that are obtained by a combination of former variables: it allows one to retain as much as possible the variability present in the data set and to reduce noise and redundancy. By inspecting a PCA model, groupings, trends, and outliers can also be found. However, in some cases we do have an *a priori* additional knowledge of the samples, for example, concentration, dose, age, gender. In this scenario, we can use supervised models to explore the variance in a block of data X that allows the prediction of a response block Y, that is, our additional knowledge. The latter may contain quantitative data, which puts one in the regression domain, or categorical data (i.e., healthy versus disease samples), which puts one in the classification domain. This method helps shift the question from "What is in there (X)?" to either "What is its relation to Y (quantitative)?" or "What is the difference between the classes in Y (categorical)?".[43] As a result, providing that no overfitting[44] is occurring (i.e., The model we are building to rationalize relationships among observations is too closely or purely fitting the training data with poor predictive ability on novel data), supervised methods can point to the variables that lead to the desired quantization or classification. Popular supervised multivariate analysis tools are partial least squares regression[45] (PLS, e.g., age, dose concentration) and its extension to classification problems, known as PLS discriminant analysis[46] (PLS-DA, e.g., control vs treated, healthy vs disease). Sparse and other variants of those techniques also exist.[47−49] When there is an high imbalance between the number of training samples in each class and/or where only one category is of interest (e.g., traceability problems, food authentication, and in the context of personalized medicine) class modeling tools are used instead.[50] Instead of looking for differences among samples belonging to different categories, class modeling techniques focus on the dis/similarities between the samples belonging to a particular category.[50,51] In the last decades, novel data mining techniques have been developed to identify relationships and trends in large, multidimensional big data sets as might be obtained from modern high-throughput screening (HTS) or untargeted omics analyses, namely, substructural analysis,[52] discriminant analysis,[53] neural networks,[54−56] decision trees,[57,58] support vector machine,[59] and kernel algorithms.[60] However, it is almost impossible to define the best modeling technique *a priori*: an initial benchmark study is necessary to determine the most appropriate one. Despite the mathematic complexity added to new algorithms, the final goal is still the same: to provide interpretable, thus useful, models with adequate predictive ability.

TARGET IDENTIFICATION

Point-of-need devices are built to sensitively, accurately, and selectively detect an analyte (or group of analytes) of interest. A question arises: how was that particular analyte (or group of analytes) identified as most related to a specific pathophysiological condition we aim to monitor? The process of target analyte identification or, more popularly nowadays, of biomarker discovery, deserves a dedicated research effort which can be speeded up by chemometrics. We can distinguish two possible approaches for target identification: hypothesis-based and discovery-based, Figure 1. The hypothesis-based approach is grounded on the mechanistic understanding of biochemical processes behind the pathophysiological condition of interest: understanding that diabetes mellitus increases blood glucose levels led to the identification of glycosylated hemoglobin as an ideal biomarker for diagnosis of diabetes.[61,62] The same happens for pregnancy tests.[63] Discovery-based

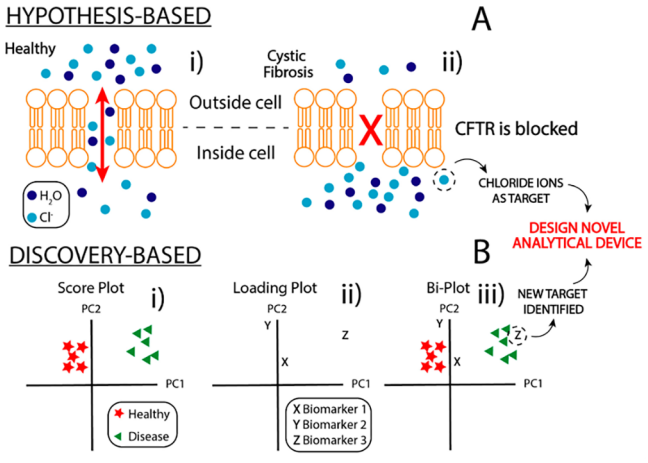


Figure 1. (A) Hypothesis-based approach focused on cystic fibrosis (CF). People affected by CF have CTFR (membrane protein) malfunctioning (ii), leading to chloride ions accumulation within cells. Sweat chloride detection is used to diagnose CF. (B) Discovery-based approach with the use of PCA: (i) score plot that displays variability among samples, (ii) loading plot that displays contribution of original variables (*X*, *Y*, *Z*, three possible biomarkers), and (iii) biplot that visualizes the correlation among samples and original variables.

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

approaches, on the other hand, aim to identify statistically significant changes in molecular species associated with the pathophysiological state of interest. For instance, the breast and ovarian cancer-associated gene BRCA1 was identified by positional cloning of a region on chromosome 17 that is frequently deleted in breast cancer.[64]

Today, in the high-throughput omics era, where generating data is arguably easier and faster than interpreting results, the discovery-based approach is predominant. Especially in the early stages of an investigation, appropriately sized data sets undergo statistical analysis to data reduction and classification in a purely data-driven fashion. In such a multivariable domain, multivariate analysis is the natural choice to identify a panel of complementary target analytes that can effectively discriminate the samples under investigation better than a single one (i.e., univariate approach).[65] Taking into account the correlation structure of the data and the synergies and antagonisms plausibly existing among the potential analytes, the multivariate approach outperforms the univariate one in sensitivity, specificity, and reliability and were successfully used for diagnostics and prognostic biomarkers discovery.[66] As an example, one of the few Food and Drug Administration (FDA) approved biomarkers[62] is the one for ovarian cancer (Ova1), discovered by artificial neural network (ANN) modeling of the plasma proteome of women with ovarian cancer compared to women with benign gynecological diseases.[67] As a result, a panel of five biomarkers was found to outperform the previously known ovarian cancer biomarker, CA125,[68] in the ability to discriminate between invasive ovarian cancer and benign lesions.[62,69] When the dimension of the cohorts is limited or the interest is focused on the phenomenological characterization of the disease,[51] multivariate biomarker discovery is achieved through the building of exploratory models. In this case, PCA and hierarchical cluster analysis (HCA) are predominantly used to reduce the dimensionality of the data and elucidating a natural pattern. Recently, a Sparse Mean approach was proposed as most sensitive and best able to identify the specifically perturbed variables in PCA-based methods.[70] When multiple sources of variability are present, PCA may suffer from an interpretational problem and other strategies can be used, e.g., ANOVA simultaneous component analysis, (ASCA),[71] and ANOVA principal component analysis (ANOVA-PCA).[72] However, such unsupervised, exploratory methods may not be the most straightforward choice for target identification, as they are not designed to specifically find differences among groups, while target analytes and biomarkers are supposed to be unique molecular signature of a certain group.[73] Among supervised classification methods, PLS-DA, support-vector machine, random forest, and artificial neural networks are used to force the method to provide the desired classification as well as to predict the classification of new samples. When a statistically significant discrimination among classes (e.g., disease and healthy) is found, it means that a mathematical relationship between the data and the categorical variable y (e.g., the class) was established and therefore it can be used to predict the class of novel samples. As anticipated, supervised methods suffer from the risk of overfitting, which arises when the fitting of training data is so well that both the predictive features of the data and noise are incorporated into the model, which will imply poor model performance in the prediction stage. In order to verify that the model holds a true biochemical meaning and avoid overfitting, we can focus on variable selection[74] and validation to verify that the predictive features identified are not noise and to test the predictive ability of the model beyond the training data, respectively.[44,75] All in all, even if supervised methods are preferred in the target identification step, it is good practice to start the analysis with unsupervised methods: if the desired classification is already visible in a PCA scores plot, for instance, the supervised algorithms can be applied with less probability of overfitting. A discovery-based approach can then evolve into a hypothesis-based approach, since it is not sufficient to prove that an analyte can discriminate two or more groups of samples: a biochemical mechanistic explanation is needed to support the discovery. Indeed, one of the bottlenecks of a discovery-based, data-driven target identification is to validate the robustness and to prove clinical applicability of the proposed markers, thus to prove interpretability of the model. In this context, it is crucial to combine the data-driven approach with expert knowledge throughout the entire process of target identification: from sample sizing to collection, from data modeling to result interpretation and validation.

DESIGN AND OPTIMIZATION OF THE DEVICE

As written above, a common way to develop point of need devices is represented by a single variable optimization. This approach appears inconvenient by two perspectives: number of experiments and reliability of optimization.[76] To optimize a device composed of N variables, L levels, and with a number of R repetitions, $N \times L \times R$ experiments are required: for instance, 75 experiments are needed to optimize (variable-by-variable) 5 variables with 5 levels, repeated 3 times. Even if the OVAT approach might work in some cases, the number of experiments increases quickly, along with time and cost.[77] In addition, the presence of interaction among variables is not taken into account at all. The lack of information related to variable correlation might lead to a "falsely" optimized final device, thus negatively weighing on the performance. Generally, if the interactions among variables are high, then a great difference is observed between the optimizations obtained by univariate and multivariate approaches. To overcome this limitation, DoE allows to observe all the variables simultaneously, by adopting statistical multivariate methods that have the goal of lowering resources and improving outcomes.[78,79] However, even if the multivariate vision contains the above-discussed advantages, a clarification should be given: if N variables are considered, and each of them is investigated at L levels, all the possible combinations are L^N, e.g., 2 variables with 10 values lead to $10^2 = 100$ experiments, namely, full factorial. In this case, the multivariate approach to define the correlation of just two variables produces a high number of experiments. The adoption of DoE might help analytical chemists understanding the effect of variables on response. Designs are obtained by combining the variables through well-defined rules. If the aim is to evaluate the effect of variables on the response, especially when a process is unknown, a Plackett−Burman (P−B) design can be used for screening experiments:[80] it is known as a screening design, and it is intended as linear combinations of two levels of each variable, i.e., the upper level is signed as "+" and the lower level is signed as "−". This is a very economic approach, for screening the contribution of a high number of variables (N) using a number of experiments equal to $N + 1$ that is a multiple of 4, e.g., 11 variables can be screened using 12 experiments (Figure 2A). However, it is very important to highlight that this approach (1) is useful to individuate the

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

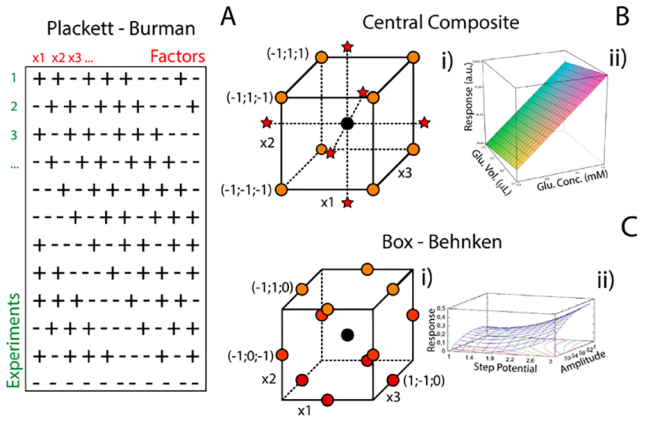


Figure 2. (A) P−B screening design algorithm for 11 factors, (B) (i) representation of a CCD on three variables and (ii) a response surface for microfluidic platform optimization. Reproduced with permission from ref 81. Copyright 2017 Springer-Verlag Berlin Heidelberg. (C) (i) BBD scheme and (ii) a response surface for pencil-based electrochemical device for metal ions. Reproduced with permission from ref 82. Copyright 2015 Elsevier B.V.

most relevant variables on response to be further investigated, (2) considers all the interactions negligible, and (3) is effective for linear data behavior.

If the contribution of a factor is relevant, this should not be inserted in this DoE, while also existing interactions should be considered while performing a P−B design. When data are not linear, e.g., quadratic behavior is present, central composite, and Box−Behnken experimental designs can be successfully used for obtaining response surfaces (the relation between different experimental variables and the responses) to define an optimum. The central composite design (CCD) is adopted on a lower number of factors (generally 2−5) with respect to the P−B design. CCD allows one to estimate the constants, the linear terms, the interactions between variables, and the quadratic terms, according to the selected model (usually, the interactions among more than two terms are not taken into account).[83] It should be thought as a cube of objects (as for factorial design) plus a second set of objects distributed in the form of a star which goes beyond the limits of the cube to provide that estimation of curvature (Figure 2B). In this case the number of experiments (N) is defined by the following equation: $N = 2^F + 2F + C$, where F is the number of factors and C is the number of central points (the points where all factors are set up at their center value). Additionally, CCD permits one to reuse previous factorial experiments (blocking):[84] if the experiments are too long, one can decide to carry out two blocks of experiments, i.e., cube points, star points. Alternatively, the Box−Behnken design (BBD) represents a valuable choice to the CCD:[85] BBD tries to minimize the effects of extreme values like these provided by the star points in CCD. For this purpose, a cube and some central points are still used but, unlike CCD, samples are not positioned in the vertices but in the middle of the edges, and star samples are not used (Figure 2C). BBD needs less experiments than analogous CCD: BBD is very useful when extreme experiments are undesired, and "blocking" is generally not available. The number of experiments required for BBD of F factors and C central points is given by the following formula $N = 2F \times (F - 1) + C$. This design avoids combination for which all factors are simultaneously at their highest or lowest levels, thus avoiding experiments performed under extreme conditions. For instance, it is commonly utilized to optimize parameters while developing portable colorimetric device that are built on paper substrate, i.e., microfluidic paper-based devices:[86] the optimization of the flow geometry and the amount of deposited reagents on the paper-based strips can be finely optimized, resulting in performance boost and time/cost savings (only 73 experiments were used) while analyzing uric acid in human urine. These multivariate methods, whereby all factors are varied simultaneously, represent a good starting point for beginners. The obtained data can be conveniently analyzed in combination with the analysis of variance (ANOVA) to understand the variation induced by the different variables and to obtain the optimal compromise between the number of experiments and chemical meaning.[87]

DATA ANALYSIS

Let us think about a single measurement like detecting blood glucose with a portable strip: the value of produced current is correlated to the glucose concentration within the blood droplet. This is an example when using univariate methods can be enough: the meaning of the variable (i.e., current) is the same among different samples.[88] Instead, a more complex example is highlighted when data are composed by information on the pH, target concentration, and color of the sample, as in the case of a microfluidic device.[89] In this case, data are heterogeneous with diverse magnitude, scale, and meaning: to obtain information from multinature data, a multivariate approach is essential. It should be clear that specific conditions like those relative to environmental pollution, health disease, and food authentication are often the results of multivariables, thus the availability of large sets of data needs to be extracted and interpreted to get information.[90] PCA is very useful for this. It is largely used for pattern recognition of data set acquired with multiarray systems such as the electronic nose and tongue.[91−93] As reported in Figure 3A, PCA combined to a lab-on-chip device composed of sensitized beads for recognition of interleukin-1β (IL-1β), C-reactive protein (CRP), and metalloproteinase-8 (MMP-8) allowed one to discriminate between periodontally healthy and unhealthy patients and increase the diagnostic value of IL-1β and MMP-8 biomarkers for periodontitis.[94] In addition, a pioneering paper demonstrated that it is possible to classifying wine samples from different regions, through the use of electrochemical capillary electrophoresis on a chip.[97] However, the PCA's exploratory nature is not enough to address specific issues: perhaps, it is useful within food science when a qualitative answer is required, but it fails to provide models for quantitative prediction. To overcome this, supervised chemometric approaches can be used. PLS is largely adopted for multivariate regression, for defining mathematical relationships among variables and providing quantitative predictions.[98,99] As shown in Figure 3B, PLS has been coupled to a mobile device for interpreting the thermal stability of raw milk by means of the alizarol test. The smartphone app, namely, "PhotoMetrix Pro", is freely available in the Google Play Store. PLS allowed a satisfactory agreement when compared with the reference method (potentiometric).[95]

PLS has been successfully applied to a microfluidic system with a ^{1}H NMR-based metabolomic footprint, named as "metabolomics-on-a-chip", to identify metabolomic markers[100] and to develop a small-molecule toxicity-oriented database. PLS demonstrated to reduce the coefficient of variation for the

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

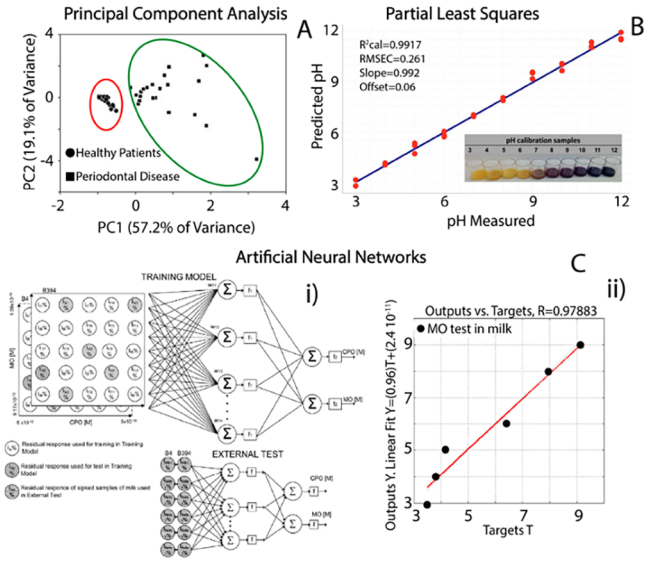


Figure 3. (A) Score plot of PCA demonstrates the different behavior among (circle) healthy patients and (square) periodontal-affected ones. Reproduced with permission from ref 94. Copyright 2007 John Wiley and Sons. (B) PLS correlations between the predicted and measured saffron samples with a portable electronic nose made with 10 metal oxide sensors. Reproduced with permission from ref 95. Copyright 2018 Springer Science Business Media, LLC, part of Springer Nature. (C) (i) ANN network constructed based on a set of data which is divided in training and external tests with different concentrations of AChE mixtures, i.e., chlorpyriphos-oxon (CPO) and malaoxon (MO), (ii) correlation graph between expected and target concentrations of MO in external test with milk. Reproduced with permission from ref 96. Copyright 2012 Elsevier B.V.

determination of propionaldehyde in wine from 33 to 15%, using an electrochemical biosensor.[101] The combination of PLS-DA with a microfluidic paper-based device, allowed to overcome the limitation of univariate approaches, that were not able to simultaneously detect acetate, cyanide, fluoride, and phosphate ions in aqueous solution.[102] Instead, a common tool adopted for nonlinear (but also linear) response is represented by the artificial neural networks (ANNs): they are algorithms simulating the biological neuronal system. Differently from previous chemometrics tools, ANNs do not require *a priori* knowledge in the model.[103] They are very useful both for data exploration and qualitative/quantitative prediction for chemical sensors whose answer is the result of a series of complex interacting phenomena. As shown in Figure 3C, ANN has been used in combination with an automated flow screen-printed AChE-inhibition biosensor for the detection of chlorpyriphos-oxon and malaoxon. The final configuration with two neurons as the input layer (use of two enzymes), 10 neurons as the hidden layer (mixtures of pesticides), and 2 neurons as the output layer (amount of pesticides) allowed one to obtain good quantification of mixtures in milk.[96] The use of ANN also allowed one to consider the metallic interactions among four metals, e.g., Zn, Cd, Pb, and Cu, and simultaneously detect them in raw propolis samples through the use of a pencil-based electrochemical sensor.[82] However, users should be aware of the previously discussed overfitting issue that affects those techniques: that is the reason why validation with a relevant number of objects is highly required. In Table 1, simple cases for users interested in approaching chemometrics tools are reported.

Table 1. Chemometrics Tools Applied to Portable Devices Design and Optimization

tool	type of device	chemometrics remarks	ref
CCD	PPy/HRP-SPE for Ochratoxin A	Optimal setup with lowest experiments	107
CCD	Colorimetric μPAD for glucose	Interactions evaluation, and 1.2% absolute error	81
CCD	Inkjet-printed PAD for isoniazid	Only 46 experiments to optimize the platform	108
BBD	Colorimetric μPAD for uric acid.	Setup optimization: reagents and geometry	76
PCA	Sensitized beads for Periodontitis	New biomarkers for periodontitis	94
PCAHCA	Smartphone for amines	Discrimination of three amines through color maps	118
PCAANN	AuNP-SPE for Tetracycline, cefixime	Discrimination of antibiotics mixture	109
PLS	μH NMR for NH_3, DMSO, phenol	Toxicity-oriented "metabolomic-on-a chip" database	88
PLS	D-SPE for propionaldehyde	Coefficient of variation from 33 to 15%	89
PLS	MIP-optosense for 1−2-naphthylamine	Lowering interferents, no pretreatment needed	111
PLS	Printed tongue for Cd, Pb, Tl, Bi	Simultaneous detection and data reduction	116
PLSANN	Pencil Electrode for Zn, Cd, Pb, Cu	Increase the linearity of the data	93
PLSANN	Color array for volatile N_2-based	Inaccurate predictions minimized	112
ANN	Multiarray for Cd, Pb, Hg	Decrease of complexity of the input data	113
ANN	μEllman assay for pesticides	Differentiate five different pesticides in mixture	119
LDA	Colorimetric array for glyphosate	100% discrimination of herbicide anions	114
PLS-DA	FTIR for morphine and thebaine	"Signature" peaks in the poppy IR spectra	115
PLS-DA	NIR for explosives	No handling variability due to human handling	110
LDA	Multifluo array for chloropropanol	Discriminate four species in mixture	117

Highly multivariate data coming from designed experiments, where both the variables (*x* block) and responses (*y* block) are multivariate, frequently in a metabolomics study, might be modeled by multilevel methods, including the above-mentioned ASCA[71] and ANOVA-PCA.[72] In addition, ANOVA-target projection (ANOVA-TP)[104] is well suited for testing the statistical significance of the studied effects and straightforward visualization and accurate estimation of between- and within-class variance, and repeated measures ANOVA (rMANOVA)[105] is applicable to clinical and personalized medicine investigations. In the context of personalized medicine, the class-modeling approach as Statistical Health Monitoring (SHM) can be used as for metabolomics studies: the metabolic profile of an individual is compared with respect to that of healthy people in a multivariate manner to detect abnormal metabolite/pattern concentrations.[106]

DIGITALIZATION

When thinking about the "next 20 years" scenario, we can foresee with confidence that advances in technologies, computerization, and miniaturization are likely to increase. The advent of the internet of things (IoT) and of innovative strategies based on information and communication tech-

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

nologies (ICTs), the development of technologies like wearables, digital biosensors, smart houses, and smart cities will make it possible to monitor everyone in real time.[120] Specifically referring to science, all areas of research will become data-intensive, emphasis will shift from data generation to data analysis, and knowledge of data-mining techniques will be essential to carry out research, thus bringing new challenges to researchers.[121,122] The combination of biosensors with IoT and ICTs strategies is also essential to generate population health data that can be used, for instance, to predict the outbreak of infectious diseases.[123] Gartner summarizes these concepts in its definition of big data as "high-volume, high-velocity, and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation".[124] In this scenario, to quote Gasteiger, "*the application of chemoinformatics is only limited by your own imagination!*"[30] Big data techniques, machine learning, signal theory, hierarchical architecture for the detection of security incidents in a security information system are the basis for the entire workflow. In the past 2 decades, the combination of multivariate techniques with the LASSO (least absolute shrinkage and selection operator) operator gave rise to the so-called sparse models, now popular in the chemometric community.[47,125] These methods adapt multivariate techniques to the huge data dimensionality, generating simple and easier to interpret models. Indeed, while the common tendency of data mining tools is to make the analysis as simple as possible for the end-user, leading in many cases to "black boxes" in which advanced data interpretation is very limited, chemometrics approaches like the ones mentioned above offer the unique opportunity of clearly interpreting and visualizing statistical analysis outcomes, and evaluating its robustness. This is what makes chemometrics "sexy" for the years to come: the features of being an already up-to-date tool for solving real complex problems in an effective and still interpretable thus user-friendly way. In addition, chemometrics offer the possibility of integrating and interpreting the complex multidimensional information provided by different sensors/devices through data fusion techniques.[126−128] Data can be fused by simple concatenation of different sensor data (namely, low-level data fusion); by fusing the features extracted from the original data, e.g., via MVA or features selection strategies (midlevel); by merging the different model responses only, after each data set has been modeled independently (high-level).[129,130] When using data fusion approaches, the trade-off to be found is in enhancing the quantity and the quality of the information content which can be extracted without including higher amounts of noise, not predictive information. For instance, from information provided by wearable sensors, level data fusion techniques and inference methods are used for activity recognition for Parkinson's disease monitoring,[131] fall detection and prediction, to physiological monitoring for early risk detection and intervention.[132,133] Clearly, also new issues arise: ethical and regulatory issues concerning the requirements and specifications of data analysis components, the user, and in e-Health applications, patient consent, data, and privacy protection.[113] Digital sensors and biometric monitoring should clearly empower citizens and hold the promise of huge potential benefits, but in order to fully exploit these devices, a cultural effort has to be done to inform and gain users' compliance and co-operation. In the digital and IoT era, chemometrics will be even closer to the end users in their everyday life, not only providing a mean to personalized, user-centered data analysis but also preserving privacy through edge analytics and understanding why and how their data are analyzed.

CONCLUSIONS

The use of chemometrics tools represent a great resource for developing novel devices for decentralized analysis: depending on the analytical necessities and experimental settings, a variety of statistical-based approaches might extract plenty of useful information from data (sometimes not taken into account). The adoption of the mathematical tools behind chemometrics is reported to be essential for all the steps around portable devices, starting from conceptualization to final application: in fact, the research around analytical devices implementation could benefit from chemometrics through identifying relevant targets, optimizing architectures, analyzing data, and collecting information. Chemometrics can make important advances in developing analytical portable devices. It might facilitate (i) the discovery of novel targets, e.g., hypothesis-based and discovery-based approaches, (ii) the optimization of experiments, e.g., design of experiments, and (iii) the classification of data and their processing. In addition, in the digital era, point of needs will be increasingly connected and personalized: chemometrics is and will be essential to mine the collected big data, derive interpretable models for early risk detection and intervention, and grant privacy and data protection. The aim of the paper is to offer nonchemometricians starting approaches for developing portable analytical devices. Although it is mainly focused on basic chemometric tools, data fusion approaches, and approaches able to deal with highly multivariate data coming from designed experiments such as ASCA, APCA, ANOVA-TP, and rMANOVA should be taken into account for outcomes from different devices. Chemometrics represents a multidisciplinary pursuit, incorporating chemical, mathematical, and computational sciences. The adoption of models to describe, process, and differentiate data should never be independent from the chemical perspective, and focus must be put on chemical interpretability and predictive ability. The years to come bring a new challenge for chemists: to bridge the gaps among different disciplines and to find solutions by embracing different cultures to the same scientific question. This Feature is a first step in this path.

AUTHOR INFORMATION

Corresponding Author

Stefano Cinti − *Department of Pharmacy, University of Naples "Federico II", 80131 Naples, Italy; BAT Center−Interuniversity Center for Studies on Bioinspired Agro-Environmental Technology, University of Napoli "Federico II", 80055 Naples, Italy;* orcid.org/0000-0002-8274-7452; Email: Stefano.cinti@unina.it

Author

Sara Tortorella − *Molecular Horizon srl, 06084 Bettona, Perugia, Italy;* orcid.org/0000-0001-9691-8323

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c04151

Author Contributions

The manuscript was written through contributions of all authors.

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

Notes

The authors declare no competing financial interest.

Biographies

Sara Tortorella graduated in Chemistry from the University of Perugia (Italy), and in 2017 she received her Ph.D. in Biotechnology in the group of Cheminformatics and Molecular Modeling led by Prof. Gabriele Cruciani. During her studies, she joined excellent and top-world research groups at Northwestern University (Chicago, US), University of York (York, UK), Barcelona Biomedical Research Park (Barcelona, ES), and GlaxoSmithKline (Philadelphia, US). Her expertise is the use of chemometric and cheminformatic for material and drug design, molecular modeling, MS and MS imaging lipidomics data analysis, software design. She is the coordinator of the "Diffusione della Cultura Chimica" of the Società Chimica Italiana (Dissemination of Chemical Culture group of the Italian Chemical Society), young member of the Italian Metabolomics Network board, and Science Communication Manager of EpiLipidNet COST action CA19105.

Stefano Cinti is an Assistant Professor at the Department of Pharmacy, University of Naples "Federico II". He obtained a Ph.D. in Chemical Sciences in 2016 in the group headed by Prof. Giuseppe Palleschi at University of Rome "Tor Vergata". He leads the Uninanobiosensors Lab (uninanobiosensors.com) at the University of Naples "Federico II", and his research interests include the development of electrochemical sensors, paper-based devices, point-of-care, nanomotors and nanomaterials. During his research activity, he had the opportunity to spend a period abroad in Finland, the U.K., the U.S., Germany, and Spain. He published more than 40 papers in peer-reviewed journals, with a H-index of 25 and >1800 citations. Among all the recognitions, in 2019 he was named Best Young Researcher in Analytical Chemistry (by the Italian Chemical Society), and in 2020 he was included in the World's Top 2% Scientists. He is a member of the board of the Chemical Cultural Diffusion Group and of the Young Group of the Italian Chemical Society.

■ ACKNOWLEDGMENTS

S.C. acknowledges the MIUR Grant "Dipartimento di Eccellenza 2018-2022" to the Department of Pharmacy of University of Naples "Federico II". Authors acknowledge Julian Ramirez for proofreading the manuscript.

■ REFERENCES

(1) Wang, S.; Lifson, M. A.; Inci, F.; Liang, L.-G.; Sheng, Y.-F.; Demirci, U. *Expert Rev. Mol. Diagn.* **2016**, *16*, 449−459.
(2) Chin, C. D.; Linder, V.; Sia, S. K. *Lab Chip* **2012**, *12*, 2118−2134.
(3) Nayak, S.; Blumenfeld, N. R.; Laksanasopin, T.; Sia, S. K. *Anal. Chem.* **2017**, *89*, 102−123.
(4) Gubala, V.; Harris, L. F.; Ricco, A. J.; Tan, M. X.; Williams, D. E. *Anal. Chem.* **2012**, *84*, 487−515.
(5) Dai, Y.; Liu, C. C. *Angew. Chem.* **2019**, *131*, 12483−12496.
(6) Urdea, M.; Penny, L. A.; Olmsted, S. S.; Giovanni, M. Y.; Kaspar, P.; Shepherd, A.; Wilson, P.; Dahl, C. A.; Buchsbaum, S.; Moeller, G.; Hay Burgess, D. C. *Nature* **2006**, *444*, 73−79.
(7) Weiss, C.; Carriere, M.; Fusco, L.; Capua, I.; Regla-Nava, J. A.; Pasquali, M.; Scott, J. A.; Vitale, F.; Unal, M. A.; Mattevi, C.; et al. *ACS Nano* **2020**, *14*, 6383−6406.
(8) https://www.un.org/sustainabledevelopment/development-agenda/
(9) Cinti, S.; Moscone, D.; Arduini, F. *Nat. Protoc.* **2019**, *14*, 2437−2451.
(10) Parolo, C.; Sena-Torralba, A.; Bergua, J. F.; Calucho, E.; Fuentes-Chust, C.; Hu, L.; Rivas, L.; Álvarez-Diduk, R.; Nguyen, E. P.; Cinti, S.; et al. *Nat. Protoc.* **2020**, *15*, 3788−3816.
(11) Cinti, S. *Anal. Bioanal. Chem.* **2019**, *411*, 4303−4311.
(12) Labib, M.; Sargent, E. H.; Kelley, S. O. *Chem. Rev.* **2016**, *116*, 9001−9090.
(13) Wu, Y.; Belmonte, I.; Sykes, K. S.; Xiao, Y.; White, R. J. *Anal. Chem.* **2019**, *91*, 15335−15344.
(14) Yetisen, A. K.; Akram, M. S.; Lowe, C. R. *Lab Chip* **2013**, *13*, 2210−2251.
(15) Xu, D.; Huang, X.; Guo, J.; Ma, X. *Biosens. Bioelectron.* **2018**, *110*, 78−88.
(16) Song, Y.; Lin, B.; Tian, T.; Xu, X.; Wang, W.; Ruan, Q.; Guo, J.; Zhu, Z.; Yang, C. *Anal. Chem.* **2019**, *91*, 388−404.
(17) Turner, A. P. *Chem. Soc. Rev.* **2013**, *42*, 3184−3196.
(18) Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R. *Anal. Bioanal. Chem.* **2017**, *409*, 5891−5899.
(19) Jalali-Heravi, M.; Arrastia, M.; Gomez, F. A. *Anal. Chem.* **2015**, *87*, 3544−3555.
(20) Martynko, E.; Kirsanov, D. *Biosensors* **2020**, *10*, 100.
(21) Díaz-Cruz, J. M.; Esteban, M.; Ariño, C. *Chemometrics in Electroanalysis*; Springer, 2019.
(22) Wold, S. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 109−115.
(23) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; Wiley, 2003.
(24) Leardi, R. *Anal. Chim. Acta* **2009**, *652*, 161−172.
(25) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. In *Chemometrics*; Springer Netherlands: Dordrecht, The Netherlands, 1984; pp 17−95.
(26) Geladi, P.; Esbensen, K. *J. Chemom.* **1990**, *4*, 337−354.
(27) Esbensen, K.; Geladi, P. *J. Chemom.* **1990**, *4*, 389−412.
(28) Kiralj, R.; Ferreira, M. M. C. *J. Chemom.* **2006**, *20*, 247−272.
(29) Chen, W. L. *J. Chem. Inf. Model.* **2006**, *46*, 2230−2255.
(30) Gasteiger, J. In *Handbook of Chemoinformatics*; Wiley-VCH Verlag GmbH, 2008; pp 3−5.
(31) Gasteiger, J. *Mol. Inf.* **2014**, *33*, 454−457.
(32) Oprea, T. I. *Chemoinformatics in Drug Discovery*; Oprea, T. I., Ed.; Methods and Principles in Medicinal Chemistry; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005.
(33) *Applied Chemoinformatics*; Engel, T., Gasteiger, J., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2018.
(34) Lavine, B. K.; Workman, J. Chemometrics: Past, Present, and Future. In *Chemometrics and Chemoinformatics*; ACS Symposium Series, Vol. *894*; American Chemical Society, 2005; pp 1−13.
(35) Lavine, B. K.; Workman, J. *Anal. Chem.* **2013**, *85*, 705−714.
(36) Box, J. F. *Am. Stat.* **1980**, *34*, 1−7.
(37) Trygg, J.; Wold, S. *Introduction to Statistical Experimental Design - What is it? Why and Where is it Useful?*, 1996; https://www.win.tue.nl/~adibucch/6BV04/tutorialTryggWold.pdf.
(38) L. Eriksson, E.; Johansson; Wold, N. K.; Wikstrom, C.; Wold, S. *Design of Experiments,Principles and Applications*; Carlson, R., Ed.; Umetrics AB, Umea Learnways AB: Stockholm, Sweden, 2001; Vol. *15*.
(39) Lanati, A.; Poli, C.; Imberti, M.; Menegon, A.; Grohovaz, F. A design of experiment approach to optimize an image analysis protocol for drug screening. In *Mathematical Models in Biology*; Springer: Cham, Switzerland, 2015; pp 65−84.
(40) Box, G. E. P.; Hunter, J. S.; Hunter, W. G. *Statistics for Experimenters: Design, Innovation, and Discovery*, 1st ed.; Wiley, 1978.
(41) de Aguiar, P. F.; Bourguignon, B.; Khots, M. S.; Massart, D. L.; Phan-Than-Luu, R. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 199−210.
(42) Smilde, A.; Bro, R.; Geladi, P. *Multi-Way Analysis with Applications in the Chemical Sciences*; Wiley: Chichester, U.K., 2004.
(43) Trygg, J.; Holmes, E.; Lundstedt, T. *J. Proteome Res.* **2007**, *6*, 469−479.
(44) Faber, N. M.; Rajkó, R. *Anal. Chim. Acta* **2007**, *595*, 98−106.
(45) Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.
(46) Barker, M.; Rayens, W. *J. Chemom.* **2003**, *17*, 166−173.
(47) Camacho, J.; Saccenti, E. *J. Chemom.* **2018**, *32*, e2964.

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

(48) Lê Cao, K.-A.; Rossouw, D.; Robert-Granié, C.; Besse, P. *Stat. Appl. Genet. Mol. Biol.* **2008**, *7*, 35.
(49) Trygg, J.; Wold, S. *J. Chemom.* **2002**, *16*, 119−128.
(50) De Luca, S.; Bucci, R.; Magrì, A. D.; Marini, F. In *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*; Wiley, 2006; pp 1−24.
(51) Calvani, R.; Marini, F.; Cesari, M.; Tosato, M.; Anker, S. D.; von Haehling, S.; Miller, R. R.; Bernabei, R.; Landi, F.; Marzetti, E. *J. Cachexia Sarcopenia Muscle* **2015**, *6*, 278−286.
(52) Cramer, R. D.; Redl, G.; Berkoff, C. E. *J. Med. Chem.* **1974**, *17*, 533−535.
(53) McFarland, J. W.; Gains, D. J. In *Comprehensive Medicinal Chemistry*; Ramsden, C. A., Ed.; Pergamon Press: New York, 1990; pp 667−689.
(54) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Nature* **1986**, *323*, 533−536.
(55) Schneider, G.; Wrede, P. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175−222.
(56) Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*, 2nd ed.; John Wiley & Sons, Inc.: New York, 1999.
(57) Salzberg, S. L. *Mach. Learn.* **1994**, *16*, 235−240.
(58) Berk, R. A. In *Statistical Learning from a Regression Perspective*; Springer New York: New York, 2008; pp 1−65.
(59) Saeh, J. C.; Lyne, P. D.; Takasaki, B. K.; Cosgrove, D. A. *J. Chem. Inf. Model.* **2005**, *45*, 1122−1133.
(60) Hann, M. M.; Leach, A. R.; Harper, G. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856−864.
(61) *Diabetes Care* **2010**, *33* Suppl 1, S4−S10,.
(62) McDermott, J. E.; Wang, J.; Mitchell, H.; Webb-Robertson, B.-J.; Hafen, R.; Ramey, J.; Rodland, K. D. *Expert Opin. Med. Diagn.* **2013**, *7*, 37−51.
(63) Canfield, R. E.; O'Connor, J. F.; Birken, S.; Krichevsky, A.; Wilcox, A. J. *Environ. Health Perspect.* **1987**, *74*, 57−66.
(64) Friedman, L. S.; Ostermeyer, E. A.; Lynch, E. D.; Szabo, C. I.; Anderson, L. A.; Dowd, P.; Lee, M. K.; Rowell, S. E.; Boyd, J.; King, M. C. *Cancer Res.* **1994**, *54*, 6374−6382.
(65) Zhang, Z.; Yu, Y.; Xu, F.; Berchuck, A.; van Haaften-Day, C.; Havrilesky, L. J.; de Bruijn, H. W. A.; van der Zee, A. G. J.; Woolas, R. P.; Jacobs, I. J.; et al. *Gynecol. Oncol.* **2007**, *107*, 526−531.
(66) Robotti, E.; Manfredi, M.; Marengo, E. *J. Proteomics Bioinform.* **2014**, *S3*, 003.
(67) Zhang, Z.; Barnhill, S. D.; Zhang, H.; Xu, F.; Yu, Y.; Jacobs, I.; Woolas, R. P.; Berchuck, A.; Madyastha, K. R.; Bast, R. C., Jr *Gynecol. Oncol.* **1999**, *73*, 56−61.
(68) Buas, M. F.; Gu, H.; Djukovic, D.; Zhu, J.; Drescher, C. W.; Urban, N.; Raftery, D.; Li, C. I. *Gynecol. Oncol.* **2016**, *140*, 138−144.
(69) Zhang, Z.; Chan, D. W. *Cancer Epidemiol., Biomarkers Prev.* **2010**, *19*, 2995−2999.
(70) Koeman, M.; Engel, J.; Jansen, J.; Buydens, L. *Sci. Rep.* **2019**, *9*, 1123.
(71) Smilde, A. K.; Jansen, J. J.; Hoefsloot, H. C.; Lamers, R. J. A.; Van Der Greef, J.; Timmerman, M. E. *Bioinformatics* **2005**, *21*, 3043−3048.
(72) Harrington, P. D. B.; Vieira, N. E.; Espinoza, J.; Nien, J. K.; Romero, R.; Yergey, A. L. *Anal. Chim. Acta* **2005**, *544*, 118−127.
(73) Hendriks, M. M. W. B.; Eeuwijk, F. A. va.; Jellema, R. H.; Westerhuis, J. A.; Reijmers, T. H.; Hoefsloot, H. C. J.; Smilde, A. K. *TrAC, Trends Anal. Chem.* **2011**, *30*, 1685−1698.
(74) Kvalheim, O. M.; Arneberg, R.; Bleie, O.; Rajalahti, T.; Smilde, A. K.; Westerhuis, J. A. *J. Chemom.* **2014**, *28*, 615−622.
(75) Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A. *Metabolomics* **2012**, *8*, 3−16.
(76) Leardi, R. *Anal. Chim. Acta* **2009**, *652*, 161−172.
(77) Ferreira, S. L.; Lemos, V. A.; de Carvalho, V. S.; da Silva, E. G.; Queiroz, A. F.; Felix, C. S.; da Silva, D. L. F.; Dourado, G. B.; Oliveira, R. V. *Microchem. J.* **2018**, *140*, 176−182.
(78) Bezerra, M. A.; Santelli, R. E.; Oliveira, E. P.; Villar, L. S.; Escaleira, L. A. *Talanta* **2008**, *76*, 965−977.
(79) Vera Candioti, L.; De Zan, M. M.; Camara, M. S.; Goicoechea, H. C. *Talanta* **2014**, *124*, 123−138.
(80) Vander Heyden, Y.; Nijhuis, A.; Smeyers-Verbeke, J.; Vandeginste, B. G. M.; Massart, D. L. *J. Pharm. Biomed. Anal.* **2001**, *24*, 723−753.
(81) Avoundjian, A.; Jalali-Heravi, M.; Gomez, F. A. *Anal. Bioanal. Chem.* **2017**, *409*, 2697−2703.
(82) Pierini, G. D.; Pistonesi, M. F.; Di Nezio, M. S.; Centurión, M. E. *Microchem. J.* **2016**, *125*, 266−272.
(83) Hamedpour, V.; Leardi, R.; Suzuki, K.; Citterio, D. *Analyst* **2018**, *143*, 2102−2108.
(84) NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/ (Accessed Dec 30, 2020).
(85) Ferreira, S.L.C.; Bruns, R.E.; Ferreira, H.S.; Matos, G.D.; David, J.M.; Brandao, G.C.; da Silva, E.G.P.; Portugal, L.A.; dos Reis, P.S.; Souza, A.S.; dos Santos, W.N.L. *Anal. Chim. Acta* **2007**, *597*, 179−186.
(86) Hamedpour, V.; Postma, G. J.; van Den Heuvel, E.; Jansen, J. J.; Suzuki, K.; Citterio, D. *Anal. Bioanal. Chem.* **2018**, *410*, 2305−2313.
(87) Lundstedt, T.; Seifert, E.; Abramo, L.; Thelin, B.; Nyström, Å.; Pettersen, J.; Bergman, R. *Chemom. Intell. Lab. Syst.* **1998**, *42*, 3−40.
(88) Ni, Y.; Kokot, S. *Anal. Chim. Acta* **2008**, *626*, 130−146.
(89) Jayawardane, B. M.; Wei, S.; McKelvie, I. D.; Kolev, S. D. *Anal. Chem.* **2014**, *86*, 7274−7279.
(90) Forina, M.; Oliveri, P.; Casale, M.; Lanteri, S. *Anal. Chim. Acta* **2008**, *622*, 85−93.
(91) Biancolillo, A.; Marini, F. *Front. Chem.* **2018**, *6*, 576.
(92) Di Natale, C.; Paolesse, R.; Macagnano, A.; Mantini, A.; D'Amico, A.; Legin, A.; Lvova, L.; Rudnitskaya, A.; Vlasov, Y. *Sens. Actuators, B* **2000**, *64*, 15−21.
(93) Wasilewski, T.; Migoń, D.; Gebicki, J.; Kamysz, W. *Anal. Chim. Acta* **2019**, *1077*, 14−29.
(94) Christodoulides, N.; Floriano, P. N.; Miller, C. S.; Ebersole, J. L.; Mohanty, S.; Dharshan, P.; Griffin, M.; Lennart, A.; Ballard, K. L. M.; King, C. P., Jr; et al. *Ann. N. Y. Acad. Sci.* **2007**, *1098*, 411−428.
(95) Helfer, G. A.; Tischer, B.; Filoda, P. F.; Parckert, A. B.; dos Santos, R. B.; Vinciguerra, L. L.; Ferrão, M. F.; Barin, J. S.; da Costa, A. B. *Food Anal. Methods* **2018**, *11*, 2022−2028.
(96) Mishra, R. K.; Alonso, G. A.; Istamboulie, G.; Bhand, S.; Marty, J. L. *Sens. Actuators, B* **2015**, *208*, 228−237.
(97) Scampicchio, M.; Mannino, S.; Zima, J.; Wang, J. *Electroanalysis* **2005**, *17*, 1215−1221.
(98) Bevilacqua, M.; Bro, R.; Marini, F.; Rinnan, Å.; Rasmussen, M. A.; Skov, T. *TrAC, Trends Anal. Chem.* **2017**, *96*, 42−51.
(99) Camacho, J.; Picó, J.; Ferrer, A. *J. Chemom.* **2008**, *22*, 299−308.
(100) Shintu, L.; Baudoin, R.; Navratil, V.; Prot, J. M.; Pontoizeau, C.; Defernez, M.; Blaise, B. J.; Domange, C.; Péry, A. R.; Toulhoat, P.; et al. *Anal. Chem.* **2012**, *84*, 1840−1848.
(101) Rojas, J.; Fontana Tachon, A.; Chevalier, D.; Noguer, T.; Marty, J.L.; Ghommidh, Ch. *Sens. Actuators, B* **2004**, *102*, 284−290.
(102) Jiménez-Carvelo, A. M.; Salloum-Llergo, K. D.; Cuadros-Rodríguez, L.; Capitán-Vallvey, L. F.; Fernández-Ramos, M. D. *Microchem. J.* **2020**, *157*, 104930.
(103) Abbasitabar, F.; Zare-Shahabadi, V.; Shamsipur, M.; Akhond, M. *Sens. Actuators, B* **2011**, *156*, 181−186.
(104) Marini, F.; de Beer, D.; Joubert, E.; Walczak, B. *J. Chromatogr. A* **2015**, *1405*, 94−102.
(105) van Der Leeden, R. *Qual. Quant.* **1998**, *32*, 15−29.
(106) Engel, J.; Blanchet, L.; Engelke, U. F.; Wevers, R. A.; Buydens, L. M. *PLoS One* **2014**, *9*, e92452.
(107) Alonso-Lomillo, M. A.; Dominguez-Renedo, O.; Ferreira-Goncalves, L.; Arcos-Martinez, M. J. *Biosens. Bioelectron.* **2010**, *25*, 1333−1337.
(108) Hamedpour, V.; Leardi, R.; Suzuki, K.; Citterio, D. *Analyst* **2018**, *143*, 2102−2108.
(109) Asadollahi-Baboli, M.; Mani-Varnosfaderani, A. *Measurement* **2014**, *47*, 145−149.
(110) Risoluti, R.; Gregori, A.; Schiavone, S.; Materazzi, S. *Anal. Chem.* **2018**, *90*, 4288−4292.

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722

(111) Valero-Navarro, A.; Damiani, P. C.; Fernández-Sánchez, J. F.; Segura-Carretero, A.; Fernández-Gutiérrez, A. *Talanta* **2009**, *78*, 57−65.
(112) Khulal, U.; Zhao, J.; Hu, W.; Chen, Q. *RSC Adv.* **2016**, *6*, 4663−4672.
(113) González-Calabuig, A.; Guerrero, D.; Serrano, N.; del Valle, M. *Electroanalysis* **2016**, *28*, 663−670.
(114) Hamedpour, V.; Sasaki, Y.; Zhang, Z.; Kubota, R.; Minami, T. *Anal. Chem.* **2019**, *91*, 13627−13632.
(115) Turner, N. W.; Cauchi, M.; Piletska, E. V.; Preston, C.; Piletsky, S. A. *Biosens. Bioelectron.* **2009**, *24*, 3322−3328.
(116) Pérez-Ràfols, C.; Serrano, N.; Díaz-Cruz, J. M.; Ariño, C.; Esteban, M. *Sens. Actuators, B* **2017**, *250*, 393−401.
(117) Wong, S. F.; Low, K. H.; Khor, S. M. *Talanta* **2020**, *218*, 121169.
(118) Bueno, L.; Meloni, G. N.; Reddy, S. M.; Paixao, T. R. *RSC Adv.* **2015**, *5*, 20148−20154.
(119) Mwila, K.; Burton, M. H.; Van Dyk, J. S.; Pletschke, B. I. *Environ. Monit. Assess.* **2013**, *185*, 2315−2327.
(120) Qi, J.; Yang, P.; Min, G.; Amft, O.; Dong, F.; Xu, L. *Pervasive Mob. Comput.* **2017**, *41*, 132−149.
(121) Eisenstein, M. *Nature* **2015**, *527*, S2−S4.
(122) Seife, C. *Nature* **2015**, *518*, 480−481.
(123) Radin, J. M.; Wineinger, N. E.; Topol, E. J.; Steinhubl, S. R. *Lancet Digit. Heal.* **2020**, *2*, e85−e93.
(124) *Gartner Glossary*, https://www.gartner.com/en/information-technology/glossary/big-data (accessed 2020-10-02).
(125) Camacho, J.; Rodríguez-Gómez, R. A.; Saccenti, E. *J. Comput. Graph. Stat.* **2017**, *26*, 501−512.
(126) Van Mechelen, I.; Smilde, A. K. *Chemom. Intell. Lab. Syst.* **2010**, *104*, 83−94.
(127) Forshed, J.; Idborg, H.; Jacobsson, S. P. *Chemom. Intell. Lab. Syst.* **2007**, *85*, 102−109.
(128) Biancolillo, A.; Boqué, R.; Cocchi, M.; Marini, F. In *Data Fusion Methodology and Applications*, Vol. *31*; Cocchi, M., Ed.; Elsevier, 2019; pp 271−310.
(129) Silvestri, M.; Elia, A.; Bertelli, D.; Salvatore, E.; Durante, C.; Li Vigni, M.; Marchetti, A.; Cocchi, M. *Chemom. Intell. Lab. Syst.* **2014**, *137*, 181−189.
(130) Vera, L.; Aceña, L.; Guasch, J.; Boqué, R.; Mestres, M.; Busto, O. *Talanta* **2011**, *87*, 136−142.
(131) Rodriguez-Martin, D.; Sama, A.; Perez-Lopez, C.; Catala, A.; Cabestany, J.; Rodriguez-Molinero, A. *Expert Syst. Appl.* **2013**, *40*, 7203−7211.
(132) Smolinska, A.; Blanchet, L.; Coulier, L.; Ampt, K. A. M.; Luider, T.; Hintzen, R. Q.; Wijmenga, S. S.; Buydens, L. M. C. *PLoS One* **2012**, *7*, e38163.
(133) King, R. C.; Villeneuve, E.; White, R. J.; Sherratt, R. S.; Holderbaum, W.; Harwin, W. S. *Med. Eng. Phys.* **2017**, *42*, 1−12.

https://dx.doi.org/10.1021/acs.analchem.0c04151
Anal. Chem. 2021, 93, 2713−2722