Avoiding nonsense in electronic taste sensing

Andrey Legin[1], Dmitry Kirsanov[1], Manel del Valle[2]

[1] Institute of Chemistry, Saint-Petersburg State University, Russian Federation

[2] Sensors and Biosensors Group, Department of Chemistry, Universitat Autònoma de Barcelona, Edifici Cn, 08193, Bellaterra, Barcelona, Spain

Abstract

Last years have watched the birth and profusion of analytical systems known as "taste sensors" or "electronic tongues". Regardless of the fancy names, these "taste sensors" and "electronic tongues" are primarily analytical instruments and require responsible handling and application in order to provide for meaningful results. In spite of intensively reported scientific activities and even commercial availability of such devices, multisensor systems are still not widely used in routine laboratory and industrial practice but are rather research instruments. The purpose of this report is to critically analyze different aspects of multisensor systems' studies and to highlight problematic issues requiring special attention from the research community to keep staying objective with reported results and declared conclusions.

Keywords

chemical sensors; sensor arrays; multisensor systems; electronic tongue; taste sensor;

1. Introduction

From the very beginning one of the main drivers, first of "electronic noses" [1] and later "taste sensors" (TS) and "electronic tongues" (ET) studies [2-4], was a challenging idea that a relatively simple artificially made multisensor system (MS) may mimic and even substitute the taste sense of mammals, even of human beings. Well over two decades have passed by, and some devices based on this idea perform so and some other do not.

The basic idea behind MS approach to "human-like" sensing is to employ a set of cross-sensitive chemical sensors that will provide a complex information-rich response from a sample. This response may be related to qualitative and quantitative sample parameters by using multivariate data processing tools (chemometrics).

The research in TS, ETs and functionally similar MS rapidly evolved through the last decades [5-7]. Numerous applications of such systems to food quality control, pharmaceutical development, environmental monitoring and other fields were reported. The other side of the popularity coin is the pop-up of research papers with questionable conclusions due to superficial research organization. Careless attention to important issues related to good analytical chemistry practices casts a shadow over the whole field. This is what the authors have certain concerns about, which is reflected in the title of this paper. In no way we aim to accuse the colleagues in unsuitable practices. There are plenty of well-organized multisensor studies with convincing content, but there are also others to tip-off.

We would like to suggest our viewpoint to the problems of MS for liquid phase analysis. The issues described below are not something imaginary but they are real-world problems that one may find in many published papers. This paper, not being a conventional review, is an attempt to generalize some common approaches of MS's research and to highlight the most important issues related to responsible development of such analytical devices.

MS for analysis of liquids, most widely known now as ET or TS emerged as a novel analytical technology in early 90's [2, 3]. The principle domain of these studies related to "unmanned" assessment of taste and flavor by artificial sensory devices was declared already in the first publications and the terms TS and ET obviously correspond to this.

Since then, a continuous growth of research papers in this field was observed. The annual number of papers containing "electronic tongue" or "taste sensor" in the title, abstract or keywords (Scopus™), increased quite significantly from just 25 in 2000 to about 150 in 2018. Over 50 papers on "bioelectronic tongues" published mostly in the last decade adjust closely to this field also. The area looks now mature with different teams contributing to this research in many locations. Researchers from at least 22 countries all over the world published ten or more papers on these topics.

This growth is not surprising considering the main advantage of multisensor systems: an opportunity to perform rapid machine assessment of complex integral sample parameters, including those related to taste and flavor, without immediate involvement of living beings. Several other important analytical advantages can also be pointed out: fast and technically simple measuring procedure, often it can be direct measurement; simple (or none) sample preparation; potential applicability for on-line measurements; inexpensive sensor technologies. Curiously, there is a widespread problem pertaining to the area, which, likely, originates in the field of human psychology. Since MS were traditionally called TS and ET many people, even in research community, would tend taking artificial devices for natural sensory systems. We believe it is a very serious mistake, though quite common. If the system is called a "taste sensor", it does not mean that it necessarily measures tastes! This delusion is often accompanied by a number of methodology failures described below.

It should be clearly understood that irrespectively of human-like names such as TS or ET and multiple other attempts to "animate" such devices one is dealing with lifeless analytical instruments that would perform correctly if only normal experimental rules and reasonable approaches are going to be followed. In fact, the term "multisensor systems" is the most correct one from the point of view of analytical chemistry to describe all kinds and types of "electronic tongues", "taste sensors" and the like devices.

TS or ET even properly configured will perform not exactly as the human sense but as a functional analogue of such. Moreover, if the sensors of your MS were not sensitive to the

components of your sample, the system would not work correctly, even if calling it TS or ET!

The purpose of this paper is describing the important issues related to MS studies and providing some views on good practices in this research field and commenting other to be avoided. It has been organized according a traditional analytical study, including samples, sensors, the measurements themselves and data processing.

< Fig.1 >

2. Sample's related issues

Though many considerations related to sampling are typical not only for MS but also for other methods of analytical chemistry dealing with real-world analytes, some of these aspects should be mentioned here for better understanding of the features of MS methodology and the results obtained using such devices.

The number of analyzed samples is ultimately important for any scientific paper on analytical chemistry. However, some papers are claiming to solve certain tasks, such as taste assessment and quantification dealing with a very limited (and obviously insufficient!) number of samples. In such cases it is hard to yield a solid proof of system's performance. A typical problem of a number of papers in this area is an overoptimistic approach to performance of TS and ET. The "taste sensor" used to be a nice illustration of this issue. The sensitivity of the sensors of TS e.g. to quinine and a few other bitter substances (e.g. [8-10]) does not guarantee the sensitivity of TS to bitter taste at large, stipulated by any kind of bitter substances. This has to be proven and confirmed. The capability of generalization, crucially important for human taste sensation is significantly understudied for artificial sensory systems, though widely and unreasonably declared. Unfortunately, this seems to be one of the most serious problems of TS and ET so far, whatever controversial declarations one can find in some published papers.

A crucially important feature of real-world samples such as foodstuffs is their natural-born, intrinsic variability. The samples of the same product, e.g. juice, even of the same producer, might be rather similar but would never be identical. The results of the analysis of such samples by a MS would inevitably contain additional uncertainty related to inherent reproducibility issues with the samples.

Another problem to be mentioned are the studies employing MS for discrimination of a limited set of apparently different samples, such as e.g. water, milk and orange juice. When the considered area had just emerged, few of such trivial experiments conducted with a few samples could have been considered as a "proof of concept" confirming that MS approach works. However, such primitive studies after many years of successful development of the field can hardly be justified.

The problem of a low number of samples may sometimes be related to their uniqueness, but also to improper planning of the experiments. Insufficient number of samples implies an obvious problem for responsible multivariate data processing and dramatically reduces the value and representativeness of derived models. Any study dealing with identification/classification of sample varieties has to be properly planned and developed showing sound statistical proofs related to consistency and reproducibility of the task done, and this will be highly depending on the replicas done and/or the number of independent samples per class.

Many samples analyzed by MS are unique, especially those that were obtained e.g. from producing companies or from various natural sources. The consequences of this situation are twofold. Firstly, it is impossible to correctly reproduce many of the experiments reported, even though they were described in proper details. This is an unusual situation in science papers and assumes high degree of reliability and reproducibility of MS used in such experiments. Unfortunately, this issue is more a matter of trust than the result of dedicated experiments. Many reported results cannot be verified or even reproduced in independent studies with the same set of samples, thus seriously decreasing the possibility of result's confirmation. Secondly, it is highly complicated to compare performance of the instruments

and quality of the results obtained by different research teams, on different samples, with different MS and under different experimental conditions. The whole area falls apart into isolated smaller fields because it is hard to compare and generalize many results and to consider them jointly.

In such a situation, we believe that collaborative efforts (e.g. [11, 12]) are one of the possible ways out of this cumbersome situation. Serious inter-laboratory efforts comparing, verifying and standardizing the output of different MS will be definitely of the highest value for future responsible analytical practice. Another suggestion is that the authors would make the raw data from their experiments available (e.g. as supplementary material), so that anyone interested could perform the processing in order to arrive to the same (or even better) results.

3. Sensors applied in multisensor systems

Cross-sensitivity of the sensors employed in a MS is a crucial feature. It is also known in the literature as non-selectivity [13, 14], global selectivity ([15] and multiple references therein), cross-selectivity [16, 17] and other similar names. All these diverse names are describing the same phenomenon stipulating a very important feature of such sensors – they can exhibit a measurable, reproducible and stable response not only to a single analyte (substance), but also to a group, a spectrum of analytes. Reproducible sensitivity along with stability of performance is an obvious prerequisite for reliable functioning of a MS of any type.

In order to provide for meaningful results the sensors of the array should possess pronounced and reproducible sensitivity towards substances responsible for the target features under analysis. This means that an attempt to quantify the astringent taste in red wine with a set of sensors with no sensitivity towards polyphenolics will never be successful. It is very important that one should possess a set of sensors, but not just a single astringent (or bitter, or salty, or sweet, etc.) sensitive sensor, since NONE of the sensing

materials known so far would be sensitive to ALL possible classes of astringent substances that might be found in natural-born samples.

The sensor array of MS should be customized for each particular application. This is not always true in real experimental practice since one has to have the widest choice of the sensors applicable for different samples, tasks and analytes. Standardized and small sensor arrays, especially commercial ones, are not really optimized for numerous practical tasks in terms of sensor array sensitivity and cross-sensitivity, though the commercial claims [18, 19]. Even chemical composition of e.g. beers and mineral waters are quite different, an attempt to analyze them with a single and not optimal sensor array formed, for both types of samples, by exactly the same few sensors may hardly be successful.

< Fig.2.>

We are deeply convinced that so far, there are no universal sensor arrays that may be applied for quantification of taste and flavor attributes in any kind of samples. Let us consider sweet taste. Fig. 2 shows chemical structures of various sweet substances. They are all sweet but very different in structure and properties; moreover, from chemical point of view, it is not realistic to synthesize a receptor that could provide for effective binding of all these diverse compounds. A universal sweet (or any other) taste sensor is hardly possible. Thus, application of small and standardized sensor arrays without giving proper respect to chemical nature of samples and sensor's sensitivity and cross-sensitivity is a doubtful way to ensure robust performance of MS. Much more realistic alternative to "universal" sensor systems is tailored development of sensor arrays for particular application, or at least particular type of analytes/samples, taking into account specific chemistry behind. This is a time consuming and highly complex research, performed rarely in the field.

One should have a wide enough choice of cross-sensitive sensors to realize this approach. In relation to this, in the multiway approaches, the point that must be verified is that the system is well overdetermined (enough information is carried by the sensors to

resolve the analytes sought) and no autocorrelation between sensors occur. If we use several sensors, and their responses are just proportional one to the other, the system will be ill conditioned and its performance will be reduced.

It is important to notice that there is no widely accepted and workable theory of cross-sensitive sensors. The existing theories e.g. that of potentiometric ion-selective electrodes (e.g. [20]) are of limited value for such sensors because these theories are not dealing with cross-sensitive low selective sensors and would fail in complex multicomponent media with unknown interferences. Moreover, most of the theories were developed so far just for discrete sensors and they are not useful for multisensor arrays at all.

Curiously, some resonating ideas concerning ion-selective electrodes suitable for array application were suggested as long ago as in 1988 [21] but were disregarded by the researchers, though could have suggested a starting point of dedicated sensor research.

Most potentiometric sensors suggested for MS so far were actually inspired more by ion-selective electrodes rather than by any biological sensors or any other analogues (e.g. [21]). A wide use of the term "lipid sensors" changes nothing in this aspect and seems a word play rather than an original idea [2, 15, 22]. Many organic molecules with a non-polar backbone and functional groups capable to dissociate may be called lipids but this fact does not bring us any closer to biological sensory systems. At last, those "lipid sensors" are very close to ionophore-free potentiometric sensor, which have been in use for years [23].

Another important problem of the sensors employed for MS is a possibility of unconventional contamination or unknown adverse effects induced by analytes. Most of the analytes studied by MS are organic substances, such as drugs, or complex mixtures, including foodstuffs. Contamination of sensing materials by these substances, surely, may occur but is rarely reported. Consequences of such contamination and even serious chemical changes that may happen in the sensing materials must be considered. Otherwise, the effects might be taken by noise of whatever nature or specific interaction of an analyte with the sensing material having little in common with meaningful multicomponent response.

Response stability of the sensors is not discussed properly in the vast majority of published cases either. Long-term stability of the sensors is neither duly reported. It must be noted that even if sensors used in MS were once properly checked in some laboratory solutions and their long-term stability was studied there, the performance of the same sensors may significantly differ in complex solutions, particularly those containing organics. Thus, reliable and stable long-term calibration and re-calibration sequences adjusted to certain features of particular analytes have to be developed and applied.

If MS design takes into account the issues mentioned above then MS may produce meaningful results about the content of certain components and it may generate reliable chemical information about integral complex parameters, such as taste and flavor attributes of the sample. If MS includes not enough sensitive, unstable and unreliable sensors, rapidly drifting or deteriorating ones then there will be no chance to end up with anything, besides garbage, even if you call such device a "taste sensor" or an "electronic tongue".

4. Measurement process

A widespread delusion concerning measurement protocol is that MS must be applied necessarily under precise conditions of calibration method. Trying to follow the same sample pretreatment protocol as for sensory panel (e.g. dealing with the samples without appropriate dilution), or the same as for some reference method, without taking care about chemistry behind these operations, may sometimes damage sensors or make them work under too hostile conditions. There are many examples where simple sample pretreatment procedures was quite favorable for establishing nice correlations with human sensory panel marks, much better than with intact samples. Dealing with beers [24, 25] kieselguhr filtering eliminates CO_2 bubbles and counteracts their sticking to sensor surfaces. This made sensor's performance dramatically more stable without any significant change of parameters of the analyte. In experiments with wines [26-28] slight dilution of the wines by distilled water was very favorable for sensor's stability and reproducibility and this simple procedure improved significantly the quality of results. Normally, in multisensor studies, the

achievement of correlation between a sensor array and a reference method of whatever kind is sought. If such a correlation is much better when employing a minor sample preparation, this procedure must definitely be used.

It is important mentioning that a random order of measurements in replicated samples is quite necessary for ET studies as well as in any other analytical field. It may help avoiding possible cumulative effects of the samples on the sensors and such randomization is necessary for good analytical practice. Unfortunately, this side of the story is rarely reported and it is hard to say at all if this protocol is widely applicable to MS.

Advanced data treatment procedures are efficient tools, especially in identifying patterns in the samples. If samples were prepared/measured with a certain structure, this is a first observable that may be noticed. Randomization of samples in preparation and measurement is the only way to get reliable results.

Signal acquisition time is an important issue, especially for potentiometric sensors. In complex analytes a true thermodynamic equilibrium between sensor membrane and the sample may take much longer time than in a simple solution. The degree of stationary state achieved surely depends on the chemical nature and concentration of the samples and can be significantly different both in different samples and even in different standard (calibration) solutions. The measurements might be taken each time at different parts of dynamic equilibrium curve, at different distances from the final equilibrium point and this is not improving precision and quality of the results. This implies the necessity of strict measurement time control in order to provide for reproducible MS response.

A misconception sometimes found is related to the number of voltammetric sensors in the array. A unique voltammetric sensor is able to extract a large amount of information, given the vector nature of the voltammogram, but the ET approach assumes a number of electrodes with cross-response features. The described situation is nothing more than a multivariate treatment of a voltammogram [29], an approach also frequent in the sensor field. This diagnostic is repeated if there are a number of different electrochemical

techniques used with a single sensor; again this setup does not follow the consensus definitions.

Reproducibility of the measurements with MS and detailed methodology of experiments were scarcely reported. Sometimes only a few measurements were performed in a sample, while the samples were also few and the results appeared being quite uncertain [30-33]. Since this might be accompanied by a lack of replicas, if any at all, some reported results would become highly doubtful and unreliable. The failure to follow responsible analytical (and even common sense) foundations in some experiments is pushing the whole area in the direction of exotic curiosities rather than valid scientific methods.

5. Data analysis and results interpretation

Responsible application of appropriate data analysis techniques is indispensable part of MS methodology. Since the response of a sensor array is a set of numbers, a vector, not a scalar, the use of chemometric tools is justified. A very nice overview of the popular data processing methods can be found in [34]. In fact, the careless use of data processing has been claimed as one of the main drawbacks for ENs and ETs [35], as this can be a source of significant uncertainties and misinterpretations of multisensor results. Below we will discuss several most typical cases.

After the measurements with ET one tries to relate MS data with reference values of target quality parameters to see whether it is possible to predict sample features based on MS's response in those samples. The observed correlation should be validated for statistical significance using independent test set of samples, i.e. by prediction of their parameters and comparison with reference technique.

In spite of the fact that the number of various chemometric techniques is huge, only a limited number of such methods have gained wide acceptance in multisensor studies. The most popular tool is PCA (principal component analysis) [36]. PCA allows for reduction of data dimensionality while keeping all useful information, for exploration of hidden data structure or for highlighting significant sensors in a sensor array. Typically, PCA is applied

to the results in order to evaluate the potential of sensor array in recognition of the studied samples. This can be based on geographical [37] or biological origin [38], authenticity [39] and numerous other characteristics. If PCA score plot shows the meaningful clustering of samples then one can claim that MS is capable of sample recognition. An important issue regarding these PCA score plots is that numerous papers employ them as a proof of sample classification ability of MS. Strictly speaking this is not so, since PCA is not a classification technique, unless some metrics for sample attribution to a particular class are introduced. Moreover, classification performance of MS requires thorough validation and calculation of certain figures of merit, like sensitivity, selectivity, or total percentage of misclassified samples [40]. This validation should be done with sufficiently large and representative independent test set. PCA as such does not allow performing these operations and real classification methods should be employed instead. SIMCA (soft independent modeling of class analogy), LDA and QDA (linear and quadratic discriminant analysis), kNN (k nearest neighbors) techniques can be used for classification [41].

PLS-DA (partial least squares discriminant analysis) is another popular classification tool; however, it also requires certain precautions. First, PLS-DA score plots should not be used as a demonstration of system performance [42]. PLS algorithm looks for the variance associated with the difference in classes even if this variance is related to e.g. noise. In such situation special attention has to be paid to the number of latent variables in PLS-DA model, when it is quite large (>5) it can be a potential sign of model overfitting. Another important issue regarding PLS-DA is that the method is poorly applicable for one-class classification problems [43, 44] since it requires representative sampling of both classes in order to capture variance structure. When dealing with one-class problems it is often hardly possible. A representative sample set for authentic samples is realistic while that for non-authentic ones is very unrealistic because any possible type of non-authentic sample has to be considered for representativeness. As an alternative approach to the tasks where one-class classification is needed, the use of SIMCA can be recommended instead. This method is aimed at the

modeling of variance inside the target class, and thus it is better adopted for the purposes of authenticity analysis.

PLS is probably the most common tool for regression analysis in multisensor studies and it is employed in the majority of the reports to derive mathematical models relating sensor array response with numerical values of certain sample quality. A very important issue here is the validation of regression models. Validation will allow judging on future predictive performance of a particular model and thus, proper investigation of this performance is of paramount importance for any practical application. There are several ways of model validation depending on the split of available samples into calibration and validation sets. Because availability of reference data on analyzed samples is quite often limited because of "time and money" considerations, the researchers widely apply cross-validation to verify the validity of their models. This approach implies the interchangeable use of the same samples in both calibration and validation sets [45]. In its ultimate version – full cross-validation, or so called "leave-one-out" (LOO) cross validation – each sample is becoming a validation set once and the results over all samples are averaged in RMSECV value (root-mean-squared error of cross-validation). Such type of validation is usually criticized in chemometrics for the proneness to yield biased overoptimistic results with respect to predictive performance assessment [45]. Validation with independent test set of samples, which were not employed for calibration, is recommended instead. When number of samples is limited, LOO can be an option to judge on model performance, however, in this case it is not recommended to make some far-reaching conclusions on the real predictive performance of the system. Another option to confirm the validity of the observed correlations between MS response and some quality parameter of interest when the number of available samples is small can be a permutation test [46]. In this test, the Y-vector with reference values is randomly permuted several dozens of times and the observed RMSECV values are compared to original Y-vector that is supposed to have the lowest RMSECV. Another result will normally imply the absence of meaningful correlation between system response and sample quality. One more issue regarding the validation of the models relates

to the data split. In some papers, several replicated measurements with MS in a single sample are considered as physically different and independent measurements, these replicated measurements in the same sample are then being split between calibration, and validation sets. Responsible laboratory practice suggests, however, very different approach, which assumes different, totally independent samples, randomly ordered multiple replicas and appropriate experimental planning. Current lack of these obvious elements of good laboratory practice in MS deteriorates the value of some nice research ideas.

Another wide spread data processing technique that may be used with MS are Artificial Neural Networks (ANNs), which have demonstrated to be highly versatile tools, as they can be used for qualitative identification/classification or for numerical quantification of a concentration/property.

Quite often, the results of regression modeling in MS studies are provided in terms of RMSEP (root-mean-squared error of prediction). It must be pointed out that this metrics has the dimensionality of the target predicted value and thus it has to be always reported together with the range of variation of target value. Otherwise the claims like "RMSEP was 0.5 which is a good result" tell nothing useful to the readers as they can describe the calibration range of 0.6 – 0.8 and then such results are hardly good.

When the dynamic range of an estimated parameter is large, RMSEP can also become inappropriate measure of model performance. Measuring some value in a range 1 – 200 a.u., RMSEP of 5 may seem to be a reasonable precision. However, at the lower border of the range, when studied parameter is 1-2 this RMSEP assumes several hundred percent relative error. Vice versa, when studied parameter is around 200, this RMSEP is below five percent relative error. This happens due to averaging step in RMSEP calculation. In order to provide a more practical estimation of modeling error in such cases MRE (mean relative error) can be calculated. In this case averaging of all relative errors observed for each sample in the test set is done and thus the MRE value may give more relevant information.

A special comment is needed when the sensors in ET are voltammetric; in this case the measured data is of high complexity and high dimensionality since a voltammetric sensor

will provide a vector as a measure, i.e. a number of currents per each applied voltage. In contrast, a measure from a single potentiometric sensor is simpler; it is just a scalar, some steady state potential. This high dimensionality is not easily accepted by data processing alternatives, establishing a correspondence between a vector of measured data and a (simpler) vector of predicted properties. There are particular algorithms such as n-PLS suitable for this situation [47], but they are not widely used and demand a higher level of expertise. A further way to proceed is to compress the voltammetric signal into a reduced number of scalars (or coefficients), and use then the vector built appending all scalars from all sensors as the information for modeling (see Fig 3). This compression/reduction of the data can be accomplished through different approaches. Some of the simplest ones are the use of PCA or PLS, or even the windowed slicing integral, that uses a number of segmented areas of the voltammogram [48]. The use of Fourier transform or Wavelet transform (also used in video compression) can be another way of reducing in a factor of ca. 90% the size of original data.

< Fig.3.>

A questionable procedure is described in the literature hat can lead to invalid results if not properly used. This is the feature extraction approach, which corresponds to measuring significant characteristics of the raw signal, a procedure that is routinely used in the work with gas sensors and ENs. The point is that, in voltammetry, there are specific features, that are related to the peaks recorded for the reduction and oxidation processes (such as potentials for the maximum current, V_1, V_2, V_3, maximum current values i_1, i_2, i_3, even widths of the observed peaks, w_1, w_2, w_3). The use of these features play efficiently the role of the data compression stage, therefore they can be perfectly licit and will arrive to correct results. But one still detects attempts to decompose the voltammetric signal in features, normally inspired in the time transient displayed by gas sensors, only that their equivalent are non-sense in voltammetric ET or TS systems. Among these, one has seen maximum and

minimum recorder currents (i_{min}, i_{max}, although these can be related to solvent redox processes), width between oxidation and reduction wave (Δi), slope of the latter (β), initial slope increase of signal (α) etc. This feature extraction only demonstrated the authors are familiar with electronic nose data treatment, but they lack familiarity of the working principles of voltammetric sensors [39]. Figure 4 illustrates the improper (A) and proper (B) features of voltammetric signals for the reduction of information in ET and TS systems. And even worst, it happens that chemometric tools used can be so powerful that they can extract correspondence between sought properties/concentrations and certain noise patterns in the extracted features, then arriving to response models with reasonable statistic indicators, only highly questionable because of validity of initial data [49]. Such inconsistencies obviously pass on only because the proper validation strategies are not performed with all needed rigor.

< Fig.4. >

A final issue worth mentioning in this section is related to Occam's razor principle. Some papers in literature tend using very sophisticated data processing algorithms to handle MS data. While there is nothing bad about these algorithms themselves, a clear motivation for the employment of this particular approach is often not provided. Moreover, comparison with conventional tools is not given and it is unclear whether some new trendy algorithm outperforms "cheap and dirty" PLS indeed. A famous chemometric concept GIGO (garbage in – garbage out) must never be ignored; if the data do not contain the information the one is looking for – it really does not matter what type of preprocessing, processing and post-processing will be applied, the result will be all junk.

6. Future research – calibration transfer, drift correction, theory

In spite of the fast development of MS technology and massive body of literature on the topic, TS and ET are still far from being a routinely accepted method for laboratory analysis. Several problems hinder MS introduction into daily analytical routine. First group of problems relates to the long-term response stability of sensor arrays. As it was discussed above, upon the interaction with complex multicomponent samples, sensor surfaces may get contaminated/affected with certain sample components due to adsorption/sorption/electrochemical reactions and this may lead to sensor's performance deterioration and drift. The changes in sensor readings in turn may invalidate the multivariate classification/regression models. In order to avoid frequent recalibration (which can be a tricky task in case of ET as the availability of the appropriate calibration samples can be limited and the required number of such samples is often large) various mathematical methods can be applied [50]. This question is not well studied in MS literature yet, and so far, only several reports addressing these problems are available [51, 52]. Further dedicated studies are definitely needed in this respect.

Another problem associated with possible mass production of MS relates to the necessity of multivariate calibration transfer protocols. This would allow using the calibration model established by the instrument producer for a single master instrument together with all other sensor arrays of the same type without the need for a separate calibration for each one. While the calibration transfer protocols are already well established for spectroscopic instruments [53], their application for multisensor systems was just started recently [12, 52, 54].

Another important problem relates to the lack of formal description of traditional analytical figures of merit, like selectivity, sensitivity, detection limits for MS. This issue hinders reliable comparison of various MS systems between each other. Due to the multivariate nature of MS response, these parameters are not easy to assess and the theoretical basis for such assessment is still under development [55]. It is also important to highlight that in case of global quality parameters of samples (like taste descriptors) some

performance metrics besides RMSEP values are even hard to introduce. The multisensor research area definitely lacks strict theoretical description.

We believe that future MS research will focus (but of course not limit) on these issues.

7. Conclusions

The electronic taste sensing is a rapidly developing field of research in analytical chemistry. Various fancy terms like “electronic tongue” or “taste sensor” were widely employed to describe this approach based on combination of multisensor arrays and chemometric data processing. These somewhat easy and superficial terms may provoke corresponding superficial attitude from research community especially from newcomers to the field. In this paper, we tried to give our vision of appropriate analytical practices leading to convincing results in “electronic tongue” studies. Being analytical instruments, “electronic tongues” of whatever type require responsible planning and application of all good analytical practices in order to arrive to meaningful conclusions. This relates to all conventional steps in multisensor studies: goal setting and experimental design, sample preparation steps, choice of particular sensors for arrays, measurement protocols and data processing.

Acknowledgements

M.del Valle thanks the support from program ICREA Academia and support from Spanish Ministry of Economy and Innovation, MINECO (Madrid) through project CTQ2016-80170-P. D. Kirsanov and A. Legin acknowledge the support from RFBR-BRICS project ##18-53-80010. The authors are grateful to Prof. Miriam Pein-Hackelbusch for inspiring discussions of the topic of this paper.

References

[1] K. Persaud, G. Dodd, Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose, Nature 299 (1982) 352-355. doi: 10.1038/299352a0.

[2] K. Hayashi, M. Yamanaka, K. Toko, K. Yamafuji, Multichannel taste sensor using lipid membranes, Sensors and Actuators B: Chemical 2 (1990) 205-213. doi: 10.1016/0925-4005(90)85006-k.

[3] A.V. Legin, E.A. Bychkov, B.L. Seleznev, Y.G. Vlasov, Development and analytical evaluation of a multisensor system for water quality monitoring, Sensors and Actuators B: Chemical 27 (1995) 377-379. doi: 10.1016/0925-4005(94)01621-n.

[4] A. Legin, A. Rudnitskaya, Y. Vlasov, C. Di Natale, F. Davide, A. D'Amico, Tasting of beverages using an electronic tongue, Sensors and Actuators B: Chemical 44 (1997) 291-296. doi: 10.1016/s0925-4005(97)00167-6.

[5] M. del Valle, Electronic tongues employing electrochemical sensors, Electroanalysis 22 (2010) 1539-1555. doi: 10.1002/elan.201000013.

[6] D. Ha, Q. Sun, K. Su, H. Wan, H. Li, N. Xu, F. Sun, L. Zhuang, N. Hu, P. Wang, Recent achievements in electronic tongue and bioelectronic tongue as taste sensors, Sensors and Actuators B: Chemical 207 (2015) 1136-1146. doi: 10.1016/j.snb.2014.09.077.

[7] M.L.R. Méndez, Electronic noses and tongues in food science, Academic Press, Amsterdam, 2016. doi: 10.1016/c2013-0-14449-2.

[8] K. Toko, Taste sensor with global selectivity, Materials Science and Engineering: C 4 (1996) 69-82. doi: 10.1016/0928-4931(96)00134-8.

[9] T. Uchida, Y. Miyanaga, H. Tanaka, K. Wada, S. Kurosaki, T. Ohki, M. Yoshida, K. Matsuyama, Quantitative evaluation of the bitterness of commercial medicines using a taste sensor, Chemical & Pharmaceutical Bulletin 48 (2000) 1843-1845. doi: 10.1248/cpb.48.1843.

[10] T. Uchida, A. Tanigake, Y. Miyanaga, K. Matsuyama, M. Kunitomo, Y. Kobayashi, H. Ikezaki, A. Taniguchi, Evaluation of the bitterness of antibiotics using a taste sensor, Journal of Pharmacy and Pharmacology 55 (2003) 1479-1485. doi: 10.1211/0022357022106.

[11] M. Pein, D. Kirsanov, P. Ciosek, M. del Valle, I. Yaroshenko, M. Wesoły, M. Zabadaj, A. Gonzalez-Calabuig, W. Wróblewski, A. Legin, Independent comparison study of six different electronic tongues applied for pharmaceutical analysis, Journal of Pharmaceutical and Biomedical Analysis 114 (2015) 321-329. doi: 10.1016/j.jpba.2015.05.026.

[12] M. Khaydukova, C. Medina-Plaza, M.L. Rodriguez-Mendez, V. Panchuk, D. Kirsanov, A. Legin, Multivariate calibration transfer between two different types of multisensor systems, Sensors and Actuators B: Chemical 246 (2017) 994-1000. doi: 10.1016/j.snb.2017.02.099.

[13] F. Winquist, J. Olsson, M. Eriksson, Multicomponent analysis of drinking water by a voltammetric electronic tongue, Analytica Chimica Acta 683 (2011) 192-197. doi: 10.1016/j.aca.2010.10.027.

[14] M. Zabadaj, I. Ufnalska, K. Chreptowicz, J. Mierzejewska, W. Wróblewski, P. Ciosek-Skibińska, Performance of hybrid electronic tongue and HPLC coupled with chemometric analysis for the monitoring of yeast biotransformation, Chemometrics and Intelligent Laboratory Systems 167 (2017) 69-77. doi: 10.1016/j.chemolab.2017.05.003.

[15] K. Toko, Y. Tahara, M. Habara, Y. Kobayashi, H. Ikezaki, Taste sensor, Wiley. Singapore, 2016, pp. 87-174.

[16] M.L. Rodríguez-Méndez, V. Parra, C. Apetrei, S. Villanueva, M. Gay, N. Prieto, J. Martinez, J.A. de Saja, Electronic tongue based on voltammetric electrodes modified with materials showing

complementary electroactive properties. Applications, Microchimica Acta 163 (2008) 23-31. doi: 10.1007/s00604-007-0907-8.

[17] F. Dini, G. Magna, E. Martinelli, G. Pomarico, C. Di Natale, R. Paolesse, I. Lundström, Combining porphyrins and pH indicators for analyte detection, Analytical and Bioanalytical Chemistry 407 (2015) 3975-3984. doi: 10.1007/s00216-014-8445-9.

[18] Intelligent Sensor Technology, Inc (INSENT), technology company (Japan). http://www.insent.co.jp/en/, (accessed: 10 September 2019).

[19] Alpha MOS, technology company (France). https://www.alpha-mos.com, (accessed: 10 September 2019).

[20] E. Bakker, P. Bühlmann, E. Pretsch, Carrier-based ion-selective electrodes and bulk optodes. 1. General characteristics, Chemical Reviews 97 (1997) 3083-3132. doi: 10.1021/cr940394a.

[21] K. Beebe, D. Uerz, J. Sandifer, B. Kowalski, Sparingly selective ion-selective electrode arrays for multicomponent analysis, Analytical Chemistry 60 (1988) 66-71. doi: 10.1021/ac00152a017.

[22] K. Toko, Taste sensor, Chemical Senses 30 (2005) i256-i257. doi: 10.1093/chemse/bjh212.

[23] N. Ishibashi, T. Masadome, T. Imato, Surfactant-selective electrode based on poly(vinyl chloride) membrane plasticized with o-nitrophenyl octyl ether, Analytical Sciences 2 (1986) 487-488. doi: 10.2116/analsci.2.487.

[24] E. Polshin, A. Rudnitskaya, D. Kirsanov, A. Legin, D. Saison, F. Delvaux, F.R. Delvaux, B.M. Nicolaï, J. Lammertyn, Electronic tongue as a screening tool for rapid analysis of beer, Talanta 81 (2010) 88-94. doi: 10.1016/j.talanta.2009.11.041.

[25] A. Rudnitskaya, E. Polshin, D. Kirsanov, J. Lammertyn, B. Nicolai, D. Saison, F.R. Delvaux, F. Delvaux, A. Legin, Instrumental measurement of beer taste attributes using an electronic tongue, Analytica Chimica Acta 646 (2009) 111-118. doi: 10.1016/j.aca.2009.05.008.

[26] A. Legin, A. Rudnitskaya, L. Lvova, Y. Vlasov, C. Di Natale, A. D'Amico, Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis and correlation with human sensory perception, Analytica Chimica Acta 484 (2003) 33-44. doi: 10.1016/s0003-2670(03)00301-5.

[27] A. Rudnitskaya, I. Delgadillo, A. Legin, S.M. Rocha, A.-M. Costa, T. Simões, Prediction of the Port wine age using an electronic tongue, Chemometrics and Intelligent Laboratory Systems 88 (2007) 125-131. doi: 10.1016/j.chemolab.2006.07.005.

[28] A. Rudnitskaya, L.M. Schmidtke, A. Reis, M.R.M. Domingues, I. Delgadillo, B. Debus, D. Kirsanov, A. Legin, Measurements of the effects of wine maceration with oak chips using an electronic tongue, Food Chemistry 229 (2017) 20-27. doi: 10.1016/j.foodchem.2017.02.013.

[29] V. Martina, K. Ionescu, L. Pigani, F. Terzi, A. Ulrici, C. Zanardi, R. Seeber, Development of an electronic tongue based on a PEDOT-modified voltammetric sensor, Analytical and Bioanalytical Chemistry 387 (2007) 2101-2110. doi: 10.1007/s00216-006-1102-1.

[30] A. Taniguchi, Y. Naito, N. Maeda, Y. Sato, H. Ikezaki, K. Toko, Detection of cyanide using the taste sensor, IEEJ Transactions on Sensors and Micromachines 119 (1999) 587-592. doi: 10.1541/ieejsmas.119.587.

[31] M. Kataoka, K. Yoshida, Y. Miyanaga, E. Tsuji, E. Tokuyama, T. Uchida, Evaluation of the taste and smell of bottled nutritive drinks, International Journal of Pharmaceutics 305 (2005) 13-21. doi: 10.1016/j.ijpharm.2005.08.019.

[32] S. Iiyama, M. Yahiro, K. Toko, Measurements of soy sauce using taste sensor, Sensors and Actuators B: Chemical 66 (2000) 205-206. doi: 10.1016/s0925-4005(00)00345-2.

[33] W. Xiao, Z. Chen, X. Jiang, H. Zhao, F. Chu, H. Hou, A taste sensor based on surface imprinted TiO2 membrane, Progress in Biomedical Optics and Imaging - Proceedings of SPIE 8329 (2012) 83290W. doi: 10.1117/12.922998.

[34] P. Oliveri, M. Chiara Casolino, M. Forina, Chemometric brains for artificial tongues, Academic Press. Amsterdam, 2010, pp. 57-117.

[35] M. Badertscher, E. Pretsch, Bad results from good data, TrAC Trends in Analytical Chemistry 25 (2006) 1131-1138. doi: 10.1016/j.trac.2006.09.003.

[36] R. Bro, A.K. Smilde, Principal component analysis, Analytical Methods 6 (2014) 2812-2831. doi: 10.1039/c3ay41907j.

[37] Z. Kovács, I. Dalmadi, L. Lukács, L. Sipos, K. Szántai-Kőhegyi, Z. Kókai, A. Fekete, Geographical origin identification of pure Sri Lanka tea infusions with electronic nose, electronic tongue and sensory profile analysis, Journal of Chemometrics 24 (2010) 121-130. doi: 10.1002/cem.1280.

[38] F. Souayah, N. Rodrigues, A.C.A. Veloso, L.G. Dias, J.A. Pereira, S. Oueslati, A.M. Peres, Discrimination of olive oil by cultivar, geographical origin and quality using potentiometric electronic tongue fingerprints, Journal of the American Oil Chemists' Society 94 (2017) 1417-1429. doi: 10.1007/s11746-017-3051-6.

[39] M. Bougrini, K. Tahri, T. Saidi, N. El Alami El Hassani, B. Bouchikhi, N. El Bari, Classification of honey according to geographical and botanical origins and detection of its adulteration using voltammetric electronic tongue, Food Analytical Methods 9 (2016) 2161-2173. doi: 10.1007/s12161-015-0393-2.

[40] D.G. Altman, J.M. Bland, Statistics notes: Diagnostic tests 1: sensitivity and specificity, BMJ 308 (1994) 1552-1552. doi: 10.1136/bmj.308.6943.1552.

[41] R.G. Brereton, Chemometrics. Data analysis for the laboratory and chemical plant, John Wiley & Sons, Ltd, Chichester, 2003. doi: 10.1002/0470863242.

[42] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, Journal of Chemometrics 28 (2014) 213-225. doi: 10.1002/cem.2609.

[43] O.Y. Rodionova, A.V. Titova, A.L. Pomerantsev, Discriminant analysis is an inappropriate method of authentication, TrAC Trends in Analytical Chemistry 78 (2016) 17-22. doi: 10.1016/j.trac.2016.01.010.

[44] P. Oliveri, Class-modelling in food analytical chemistry: Development, sampling, optimisation and validation issues – A tutorial, Analytica Chimica Acta 982 (2017) 9-19. doi: 10.1016/j.aca.2017.05.013.

[45] K.H. Esbensen, Multivariate data analysis-in practice. An Introduction to multivariate data analysis and experimental design, 5th edn ed., CAMO AS, Oslo, 2001. doi: 10.1002/cem.692.

[46] H. van der Voet, Comparing the predictive accuracy of models using a simple randomization test, Chemometrics and Intelligent Laboratory Systems 25 (1994) 313-323. doi: 10.1016/0169-7439(94)85050-x.

[47] L.A. Li, Y. Yu, J. Yang, R. Yang, G. Dong, T. Jin, Voltammetric electronic tongue for the qualitative analysis of milk adulterated with urea combined with multi-way data analysis, International Journal of Electrochemical Science 10 (2015) 5970-5980. doi:

[48] X. Cetó, F. Céspedes, M. del Valle, Comparison of methods for the processing of voltammetric electronic tongues data, Microchimica Acta 180 (2013) 319-330. doi: 10.1007/s00604-012-0938-7.

[49] K.L. Goodner, J.G. Dreher, R.L. Rouseff, The dangers of creating false classifications due to noise in electronic nose and similar multivariate analyses, Sensors and Actuators B: Chemical 80 (2001) 261-266. doi: 10.1016/s0925-4005(01)00917-0.

[50] A. Rudnitskaya, Calibration update and drift correction for electronic noses and tongues, Frontiers in Chemistry 6 (2018). doi: 10.3389/fchem.2018.00433.

[51] V. Panchuk, L. Lvova, D. Kirsanov, C.G. Gonçalves, C. Di Natale, R. Paolesse, A. Legin, Extending electronic tongue calibration lifetime through mathematical drift correction: Case study of microcystin toxicity analysis in waters, Sensors and Actuators B: Chemical 237 (2016) 962-968. doi: 10.1016/j.snb.2016.07.045.

[52] A. Rudnitskaya, A.M.S. Costa, I. Delgadillo, Calibration update strategies for an array of potentiometric chemical sensors, Sensors and Actuators B: Chemical 238 (2017) 1181-1189. doi: 10.1016/j.snb.2016.06.075.

[53] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Transfer of multivariate calibration models: a review, Chemometrics and Intelligent Laboratory Systems 64 (2002) 181-192. doi: 10.1016/s0169-7439(02)00085-0.

[54] M. Khaydukova, V. Panchuk, D. Kirsanov, A. Legin, Multivariate calibration transfer between two potentiometric multisensor systems, Electroanalysis 29 (2017) 2161-2166. doi: 10.1002/elan.201700190.

[55] F. Allegrini, A.C. Olivieri, Sensitivity, prediction uncertainty, and detection limit for artificial neural network calibrations, Analytical Chemistry 88 (2016) 7807-7812. doi: 10.1021/acs.analchem.6b01857.

CAPTIONS FOR FIGURES

Figure 1. Issues to consider in multisensor system's studies

Figure 2. Chemical structure of some widespread sweet substances. They are all sweet but very different in chemical structure and properties.

Figure 3. General strategy to cope with the high dimensionality situation occurring when working with voltammetric ET or TS systems.

Figure 4. Bad (A) and proper (B) features extracted from voltammetric measures.

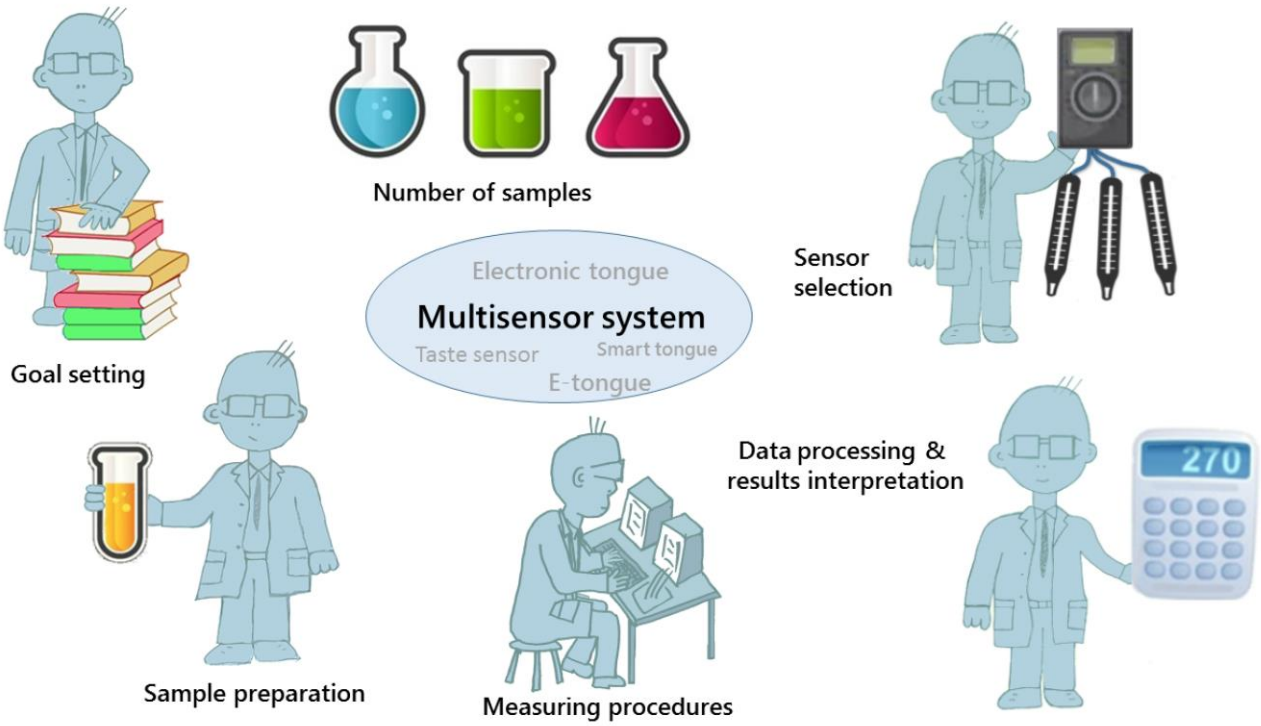
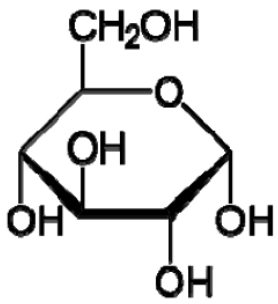
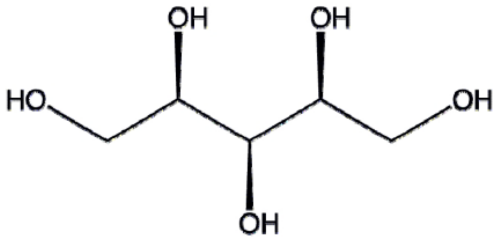

Goal setting
Number of samples
Electronic tongue
Multisensor system
Taste sensor
Smart tongue
E-tongue
Sensor selection
Sample preparation
Measuring procedures
Data processing &
results interpretation
270

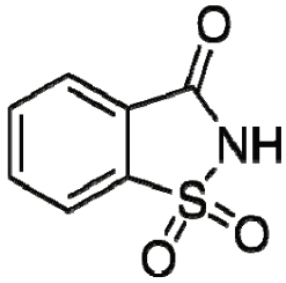
Figure 1



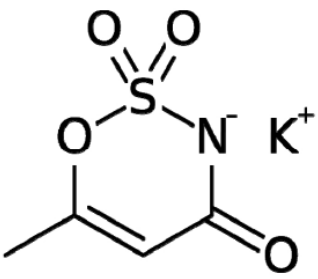
Glucose



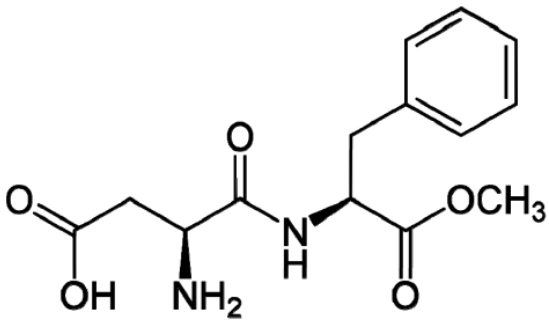
Xylitol



Saccharin



Acesulfame K



Aspartame

Figure 2

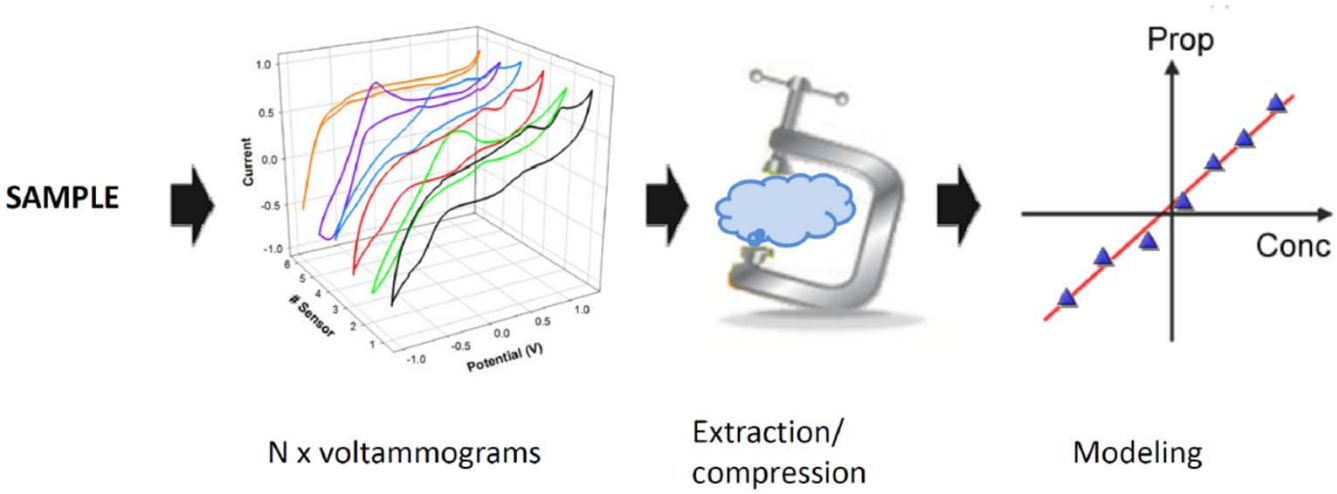


Figure 3

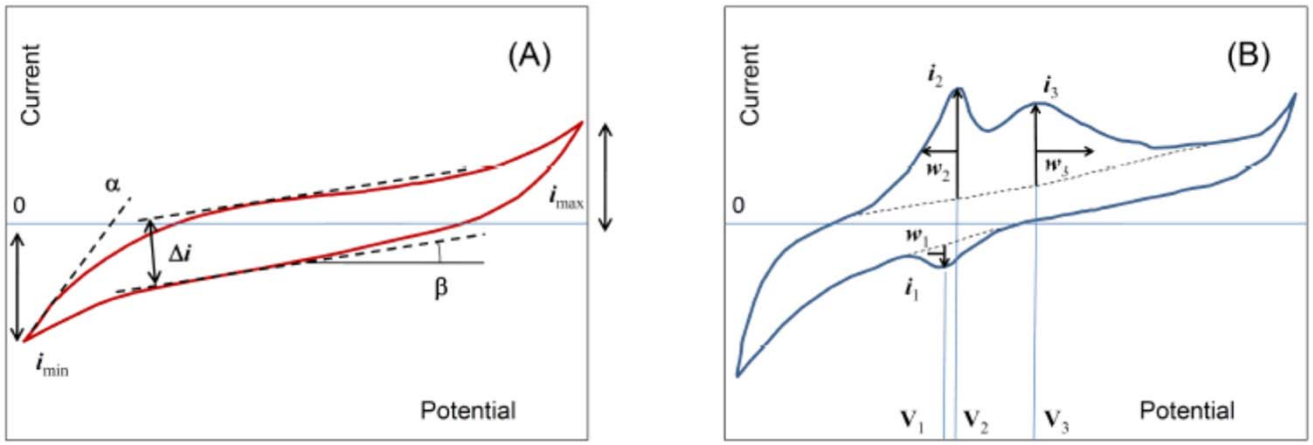


Figure 4