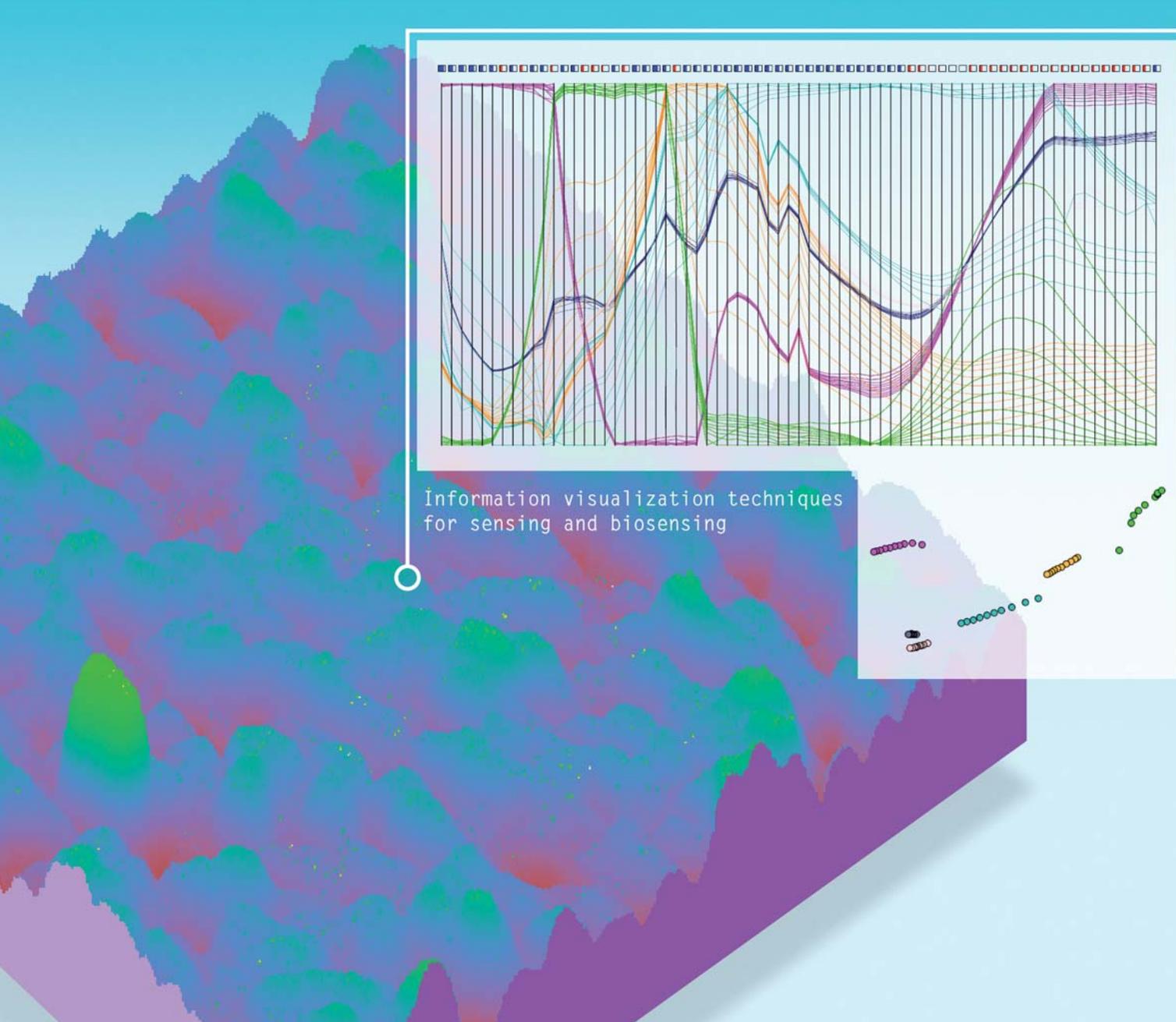


Analyst

Interdisciplinary detection science

www.rsc.org/analyst

Volume 136 | Number 7 | 7 April 2011 | Pages 1265–1524



ISSN 0003-2654

RSC Publishing

PAPER

Osvaldo N. Oliveira Jr. *et al.*
Information visualization techniques
for sensing and biosensing

Cite this: *Analyst*, 2011, **136**, 1344

www.rsc.org/analyst

PAPER

Information visualization techniques for sensing and biosensing

Fernando V. Paulovich,^a Marli L. Moraes,^b Rafael Mitsuo Maki,^a Marystela Ferreira,^b Osvaldo N. Oliveira Jr.^c and Maria Cristina F. de Oliveira^a

Received 25th October 2010, Accepted 3rd January 2011

DOI: 10.1039/c0an00822b

The development of new methods and concepts to visualize massive amounts of data holds the promise to revolutionize the way scientific results are analyzed, especially when tasks such as classification and clustering are involved, as in the case of sensing and biosensing. In this paper we employ a suite of software tools, referred to as *PEx-Sensors*, through which projection techniques are used to analyze electrical impedance spectroscopy data in electronic tongues and related sensors. The possibility of treating high dimension datasets with *PEx-Sensors* is advantageous because the whole impedance vs. frequency curves obtained with various sensing units and for a variety of samples can be analyzed at once. It will be shown that non-linear projection techniques such as *Sammon's Mapping* or *IDMAP* provide higher distinction ability than linear methods for sensor arrays containing units capable of molecular recognition, apparently because these techniques are able to capture the cooperative response owing to specific interactions between the sensing unit material and the analyte. In addition to allowing for a higher sensitivity and selectivity, the use of *PEx-Sensors* permits the identification of the major contributors for the distinguishing ability of sensing units and of the optimized frequency range. The latter will be illustrated with sensing units made with layer-by-layer (LbL) films to detect phytic acid, whose capacitance data were visualized with Parallel Coordinates. Significantly, the implementation of *PEx-Sensors* was conceived so as to handle any type of sensor based on any type of principle of detection, representing therefore a generic platform for treating large amounts of data for sensors and biosensors.

1. Introduction

Sensing and biosensing involve primarily tasks of classification and clustering, through which similar samples should be clustered together, far apart from very dissimilar samples. This is particularly important for applications – such as in clinical diagnosis – where the samples are made of multi-component, complex liquids, as in blood or urine exams. Various methods have been used to improve the distinguishing ability of sensors, especially in biosensors where molecular recognition capabilities are exploited between the materials forming the sensing units and analytes in the sample under analysis.^{1–3} The principles of detection in such sensors may vary widely, encompassing electrical, electrochemical and optical measurements, in addition to spectroscopies.^{4–6} In any of these methods, a large amount of data is generated, but to facilitate analysis on the part of the operator, normally only a minute part of the data is treated. In clinical diagnosis, for example, the measurements involved in the exams yield as a final outcome a mere table with some

characteristics and numbers – e.g. the concentration of cholesterol or glucose in the blood sample.

This scenario may change in the next few years with the realization that robust computational and statistical methods may be employed to handle massive amounts of high-dimensional data.⁷ We have recently used information visualization methods to enhance the performance of biosensors, in which crosstalk effects could be eliminated from the results obtained with biosensors based on field-effect devices⁸ and phytic acid could be detected with distinct sensing units.⁹ Furthermore, highly sensitive biosensors employing impedance measurements allowed for the detection of tropical diseases.¹⁰ In this paper, we show that information visualization techniques may also be used to optimize sensing performance. The system chosen for proof-of-principle experiments was a sensor array made with nanostructured films to detect phytic acid. Before discussing the results in Section 3, we provide background knowledge on Information Visualization and on a suite of software tools, referred to as *PEx-Sensors*, to manipulate data from sensors and biosensors.

2. Projection techniques and the *PEx-Sensors* platform

Information Visualization techniques provide visual representations of multidimensional data capable of conveying information

^aInstituto de Ciências Matemáticas e de Computação, USP CP 668, 13560-970 São Carlos, SP, Brazil

^bUniversidade Federal de São Carlos, campus de Sorocaba, 18052-780 Sorocaba, SP, Brazil

^cInstituto de Física de São Carlos, USP CP 369, 13560-970 São Carlos, SP, Brazil

in a straightforward, intuitive manner. They exploit the high-bandwidth human visual channel by mapping data objects into glyphs or icons placed into a (usually) two- or three-dimensional visual space, and data attributes into visually perceptible properties of such glyphs, such as color, size, or shape. Human visual perception works naturally helping users to identify patterns and relationships. A variety of mapping strategies have been proposed, targeted at different data types, user tasks and application domains.^{7,11,12} A major goal shared by all mapping solutions in Information Visualization is interactivity, in order to support user-driven exploration of the visual representations in trying to find answers to specific questions or to raise new hypotheses based on the data. The visual mapping process typically involves a data pre-processing stage for cleaning, selecting, formatting and normalizing the raw data, followed by a mapping of the data objects into visual objects and choice of visual properties to represent data attributes, followed by user interaction. The process is iterative in the sense that any of these stages may happen multiple times before a particular user task is successfully executed.

In this work we employ techniques to create visual representations of the high-dimensional data generated by the sensing units, generally known as multidimensional projection techniques. These techniques resemble the so-called ‘dimension reduction’ strategies widely employed to reduce data dimensionality for analysis, of which *Principal Component Analysis*¹³ is perhaps one of the best-known examples. A major distinction between projection techniques in general and dimension reduction strategies is that the latter are not necessarily related to the problem of creating visual representations, though they might be applied in this context. We refer to multidimensional projections as a range of techniques that support effective data reduction to two or three spatial dimensions only and, as such, allow creating low-dimensional views. Moreover, an additional requirement arises in a visualization scenario: techniques must be computationally efficient on the number of data objects and data attributes, in order to handle large data sets while preserving user interactivity.

Different rationales may be adopted to project the data down to two or three dimensions. *Principal Component Analysis*, for example, is a linear technique based on second-order statistics (co-variances). It seeks for mutually orthogonal linear combinations of the data attributes that account for most of the data variance. *Fastmap*¹⁴ is another well-known technique for dimension reduction, which works by recursively projecting the high-dimensional points into a reduced number of mutually orthogonal directions. It is characterized by being fast, *i.e.*, it has computational cost linear on the number of data objects.

Multidimensional Scaling (MDS),¹⁵ on the other hand, refers to a family of projection techniques that adopt an optimization approach, seeking to minimize an error function that expresses a property to be preserved in the dimension reduction process. A common property to preserve is some function of the pairwise data distances, taken as a measure of data dissimilarity in the high dimensional space. We adopt MDS strategies, with the goal of deriving a visual representation of the data samples that conveys their global (dis)similarity. Users may infer groups of similar objects from this representation – called a data

map – which provides a powerful aid to analyze the behavior of different sensor configurations.

Formally, let $X = \{x_1, x_2, \dots, x_n\}$ be a given set of n data samples, with $\delta(x_i, x_j)$ being a dissimilarity function between two given samples. Let $Y = \{y_1, y_2, \dots, y_n\}$ be the corresponding mappings of X into a low-dimensional visual layout, and $d(y_i, y_j)$ a distance function defined between any two elements of Y . An MDS projection can be described as an injective mapping function $f: X \rightarrow Y$ that seeks to minimize some error measure based on the differences $|\delta(x_i, x_j) - d(y_i, y_j)| \forall x_i, x_j \in X$. There are many forms to implement the mapping function f ^{6,17} and to formulate the error measure, originating distinct classes of projection techniques. The choice may depend on characteristics of dataset X and its distance distribution, as well as on computational requirements. In any case, the desired ideal of preserving the original distances is achieved up to a certain extent, as it is normally impossible to faithfully retain all the data relationships as in the original data space.

In this work, we consider both $\delta(x_i, x_j)$ and $d(y_i = f(x_i), y_j = f(x_j))$ to be, respectively, the Euclidean distances computed between the electrical responses of samples x_i and x_j , obtained with impedance spectroscopy, and between the mapped samples. For a spectrum of size m , each data sample may be interpreted as point embedded into an m -dimensional space. The data maps illustrated were obtained with two projection techniques: a classical projection known as *Sammon’s Mapping* (SM),¹⁸ and a recent technique called *Interactive Document Map* (IDMAP), originally conceived for mapping document collections.¹⁹ Both are briefly introduced in the following. Other techniques are not discussed further, and the reader is referred to ref. 17 for additional references and a comparison amongst distinct projection approaches.

Sammon’s Mapping (SM) employs the error function S shown in eqn (1), which may be interpreted as a measure of the amount of information loss incurred in the projection process, hence being named a loss function. This function is minimized using an iterative non-linear steepest descent optimization approach based on the function gradient to find a (local) minimum.

$$S_{\text{SM}} = \frac{1}{\sum_{i < j} \delta(x_i, x_j)} \sum \frac{(d(y_i, y_j) - \delta(x_i, x_j))^2}{\delta(x_i, x_j)} \quad (1)$$

The IDMAP technique operates by mapping the samples to the low-dimensional visual space with a computationally inexpensive, albeit not necessarily precise in this context, dimension reduction strategy, *e.g.*, *Fastmap*, and then improves the initial placement with a precise strategy known as *Force Scheme*.²⁰ This strategy mimics a force-directed placement approach based on a mass-spring model with attraction and repulsion forces, similarly to graph-layout approaches.²¹ The rationale behind this strategy is as follows: for each projected sample $y_i \in Y$, a vector $\vec{v}_{ij} = (y_j - y_i)$, $\forall y_i \neq y_j$ is computed, then y_i is moved on the direction of \vec{v}_{ij} . The amount of movement is given by

$$S_{\text{IDMAP}} = \frac{\delta(x_i, x_j) - \delta_{\min}}{\delta_{\max} - \delta_{\min}} - d(y_i, y_j) \quad (2)$$

where δ_{\min} and δ_{\max} are the minimum and maximum distances between the samples. By successively applying this process to all

samples, the difference $|\delta(x_i, x_j) - d(y_i, y_j)| \forall x_i, x_j \in X$ is reduced, resulting in a more precise placement of the points, as far as distance preservation is concerned, as compared to the original space.

Notice that the different cost functions of SM and IDMAP result in different behaviors. IDMAP seeks to uniformly preserve the difference $|\delta(x_i, x_j) - d(y_i, y_j)|$ for each pair of samples (δ_{\min} and δ_{\max} are normalization factors that do not affect the distance relationships), whereas SM assigns greater importance to smaller sample distances by dividing the (quadratic) difference $(d(y_i, y_j) - \delta(x_i, x_j))^2$ by the factor $\delta(x_i, x_j)$. Thus, SM may be preferable, e.g., if the distance values among samples are distributed over a wide range, in order to minimize the effect of large variations on the resulting projection.

The data maps provide a global view of the behavior of a set of data samples, each sample depicted as a point in a two-dimensional visual space. However, one still needs to find out how the electrical responses affect sample behavior. For this, a classical visualization technique may be employed, such as the *Parallel Coordinates*,²² suitable for depicting multidimensional data objects into a visual representation that preserves the individual data dimensions. *Parallel Coordinates* maps the m -dimensional space onto a two-dimensional space by drawing m equally spaced parallel axes, aligned with one of the principal axes. The axes map the data attributes (dimensions); each one is linearly scaled within its data range values. Each data object is presented as a polygonal line that intersects each of the axes at the point corresponding to the item's data value at the corresponding dimension. This visual representation is very effective to reveal data distributions and correlation amongst attributes.

The visualizations shown were created on a Java-based tool, named *PEx-Sensors*, that implements a variety of classic and novel multidimensional projection techniques – including the aforementioned ones – as well as an implementation of *Parallel Coordinates*. It also incorporates several data pre-processing methods required for handling sensor data, e.g., standardization and other procedures for data normalization. In Section 4 we illustrate how data maps obtained in this platform, combining projections views coupled with *Parallel Coordinates* views, offer a powerful aid to investigate sensor behavior under different conditions. *PEx-Sensors* is freely available at <http://www.icmc.usp.br/~paulovic/pexsensors/>.

3. Experimental details of the sensing measurements

3.1. Preparation of sensing units

The sensing units were fabricated with nanostructured films deposited onto 50 pairs of gold interdigitated electrodes having 10 µm width and being 10 µm apart from each other. The films were obtained using the layer-by-layer (LbL) method,²³ as described by Moraes *et al.*⁹ Three sensing units were used: a bare electrode, one coated with a 5-bilayer LbL film of phytase from *Aspergillus ficuum* (EC 3.1.3.8) alternated with poly(allylamine) hydrochloride (PAH), and another coated with a 5-bilayer LbL film made with PAH alternated with poly(vinyl sulfonate) (PVS). Prior to the adsorption of these 5-bilayer LbL films, a 2-bilayer PAH/PVS cushion was deposited onto the electrode to reduce substrate effects.²⁴

3.2. Experimental procedures to obtain the data

The electrical response was obtained by impedance spectroscopy using a Solartron 1260A impedance/gain phase analyzer in a frequency range from 1 to 10⁷ Hz. Each electrode was immersed in a 100 mM sodium acetate pH 5.5 buffer solution containing different molar concentrations of phytic acid (0, 10⁻⁶, 10⁻⁵, 10⁻⁴, 10⁻³ and 10⁻² mol l⁻¹). The measurements were obtained after 10 min of the electrode immersion, which is the period of time for the system to stabilize. Ten impedance spectra were collected for each sample. The real and imaginary parts of these impedance data were used to calculate the capacitance (C) and loss (G/ω) values. Here, we used only the capacitance values.

4. Using *PEx-Sensors* for analyzing sensing performance

We have recently found that optimization of sensors and biosensors based on impedance spectroscopy and nanostructured films requires a systematic search not only of suitable materials to interact with the analytes but also dielectric relaxation properties that are favorable for distinguishing similar samples.⁹ A surprising result in the latter work was the fact that a sensing unit containing a layer-by-layer (LbL) film of polyelectrolytes performed equally well in detecting phytic acid as a unit comprising the enzyme phytase, which is known to interact specifically with the acid. This was explained by analyzing the loss curve that showed shifts in the relaxation processes in the presence of the biomolecules. It was concluded that while the molecular recognition interaction between phytase and phytic acid should have enhanced the sensing ability, the final electrical properties of the phytase-containing film were not as favorable for detection. Therefore, optimizing the performance may require analysis of various parameters, in addition to the film-forming materials for the sensing units.

The massive amounts of data generated with the sensors and the need for performance optimization mentioned above require robust approaches for data treatment. Here we illustrate how visualization techniques implemented in *PEx-Sensors* may be employed to visualize the data for distinct sensors and identify the frequency range where distinction is best for each type of sensor. Fig. 1 shows projection views of capacitance measurements obtained at different frequencies, from samples of phytic acid at different concentrations. Each sample is thus characterized by a set of capacitance values taken at a spectrum of frequencies, using three different sensor configurations. The IDMAP projection was employed in all cases to obtain the 2D layout of the samples (71 in total), based on similarity. Fig. 1(a) shows the layout for the samples measured with the bare electrode, while Fig. 1(b) and 1(c) give the plots for the measurements with the sensing unit containing the PAH/Phytase and the PAH/PVS films, respectively. Note that even the bare electrode is sufficient to provide some distinction of the samples with various concentrations of phytic acid, if the appropriate projection method is used.

While poorer distinctions were obtained for PCA and *Sammon's Mapping* (results not shown), upon using IDMAP for visualizing the capacitance one could distinguish some samples, which is promising from the technological point of view, as low

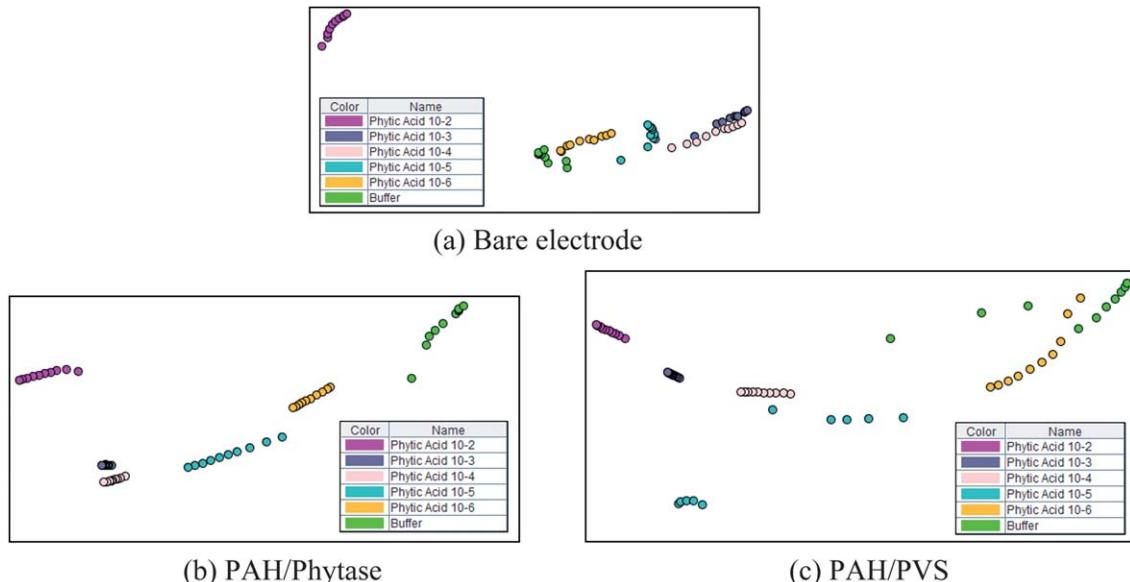


Fig. 1 Projections of measurements from three sensors with samples of phytic acid at different concentrations. Note that the axes are not labeled as in multidimensional projections, what matters is the distance between data points. The closer the points are the more similar the samples they represent are. The axes, on the other hand, have no specific meaning.

cost sensors may be fabricated. It should be stressed that this sensitivity can only be reached using projection techniques that handle the whole spectrum of measurements, rather than a specific frequency interval.

Distinction is improved when the same projection technique is applied to the data from the PAH/Phytase sensor, as one should expect owing to the molecular-recognition capability. For PAH/PVS, on other hand, the distinguishing ability is better than the bare electrode, but not as good as the one containing phytase. This result differs from the one reported in our earlier work.⁹ In the latter, rather than using only the capacitance data, we considered both the imaginary and real parts of the

measurements, and then full distinction could be achieved with either the PAH/PVS or the PAH/Phytase sensors. Because here we are interested in analyzing performance and identifying ways to improve it, the lower distinction ability of the PAH/PVS serves the purpose. Hence, concentrating on the data for this sensor, we shall try and identify the role of measurements taken at different frequency intervals for achieving distinction. A simple way to investigate this issue is to generate *Parallel Coordinates* (PC) visualizations of the space of measurements, as shown in Fig. 2 for the PAH/PVS data. It shows the capacitance values varying the frequency relative to all the samples. By simply inspecting the distribution of the lines in the axes one observes the different

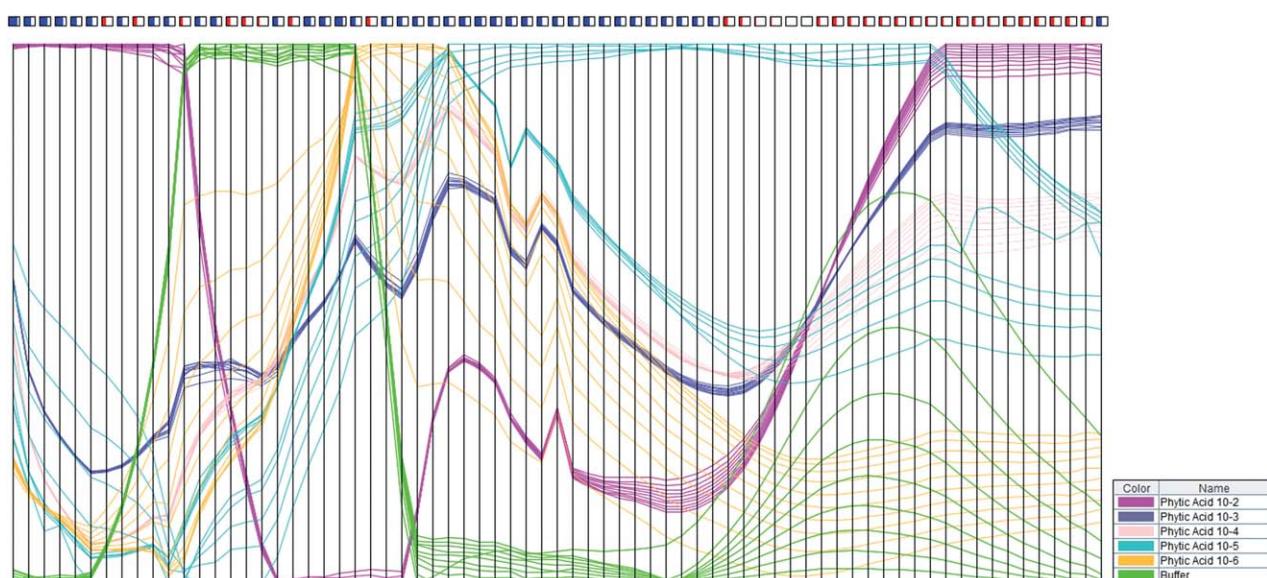


Fig. 2 *Parallel Coordinates* visualization of PAH/PVS data. The same effect of separation and grouping the samples obtained on the projection (Fig. 1(c)) can be viewed on the frequency spectrum.

information about the samples carried by each measure. The PC view somehow reflects the point sample distribution in the projection – *e.g.*, notice how in the central region the blue samples are separated into two distinct visual groups, and the orange ones are spread throughout, just as in the projection view of the samples in Fig. 1(c).

The little boxes on top of the axes show a representation of the silhouette coefficient²⁵ computed for each data attribute (*i.e.*, a measured value at a particular frequency). The silhouette is a metric that assesses the quality of a data cluster, taking values in the interval $[-1, 1]$, with higher values indicating better quality. Given a cluster of data points, in this case the groups of samples into different concentrations, the silhouette value for each point is a measure of how similar that point is to points in its own cluster, as compared to points in other clusters, according to some defined similarity metrics (Euclidean distance, in our case). An average of the values obtained for the cluster points may be assigned as a measure for the cluster as a whole. The silhouette is computed as:

$$S = \frac{1}{n} \sum_{i=1}^n \frac{(b_i - a_i)}{\max(a_i, b_i)}$$

where a_i is calculated as the average of the distances between the i^{th} data point and all other points belonging to the same cluster, and b_i is the minimum distance between the i^{th} data point and all other points belonging to the other clusters.

For the sensor data, we computed the silhouette coefficient considering each individual frequency value for each sample and depict resulting values on top of each axis in the PC visualization. Boxes in blue indicate values above zero, and red depicts values below zero. A white box indicates no information can be inferred on how that particular frequency affects the outcome, whereas the more filled-in blue boxes identify frequencies with high distinction capability.

This procedure represents an objective assessment of how much each frequency contributes to cluster distinction, that is, how much each frequency contributes to separate the different samples based on their concentration: the higher the values obtained for the silhouette coefficient, the better the distinction is. This information from the silhouette value for each frequency was input into a genetic optimization algorithm with the goal of automatically detecting the best frequencies for distinction. (The

approach resembles the one adopted by Lletí *et al.*²⁶ who also use a genetic algorithm that optimizes cluster silhouettes to select features for k -means cluster analysis.) Shortly, this algorithm employs a heuristic to generate combinations of k frequencies, with k defined as an input parameter, choosing the combination which reaches the best value of a cost function. In our case the silhouette coefficient – the distance required to calculate the silhouette – is computed considering the k frequencies, not the individual frequencies. The number of ‘best’ frequencies is the only parameter required. Fig. 3 shows both the PC visualization of 10 selected frequencies and the IDMAP projection considering these frequencies for the PAH/PVS data. The average silhouette value computed from the individual silhouette values for these 10 frequencies is 0.69663864. Note the improvement in the layout’s distinguishability of the different concentrations of phytic acid, as compared to Fig. 1(c). It is also worth noticing in Fig. 3(b) that one of the selected frequencies has a low silhouette value (as indicated by the red box on top of corresponding axis). It is interesting that, although individually it has poor discrimination capability, taken in association with others it is actually one of the best to improve distinction. Nonetheless, the points representing substances at lower concentrations are still spread in the layout, and therefore further improvement should be sought.

We then considered the set of all measures (213 in total) obtained from the three sensor configurations shown in Fig. 1, and run the genetic optimization algorithm to identify the 10 best frequencies for distinction, based on their silhouette metric. Fig. 4 shows the IDMAP layout produced from employing the selected frequencies suggested as the best ones, and the resulting IDMAP projection, which has now both separated the points at different concentrations and agglomerated the points at the same concentration, as desired (indicating that in these frequencies the sensors exhibit a very consistent response). The average silhouette value has sensibly improved, reaching 0.9116712. Again, out of the 10 selected frequencies, one has a poor silhouette metric if it is considered alone. It should be stressed that the cluster corresponding to 10^{-2} M of phytic acid is placed very far apart from the other data points. Therefore, at high concentrations, the electrical response appears to be governed entirely by the ionic concentration. Fig. 4(c) shows the resulting projection removing these high concentration samples (the same 10 frequencies previously selected are employed). The good distinguishability of the groups of samples with different concentrations is kept,

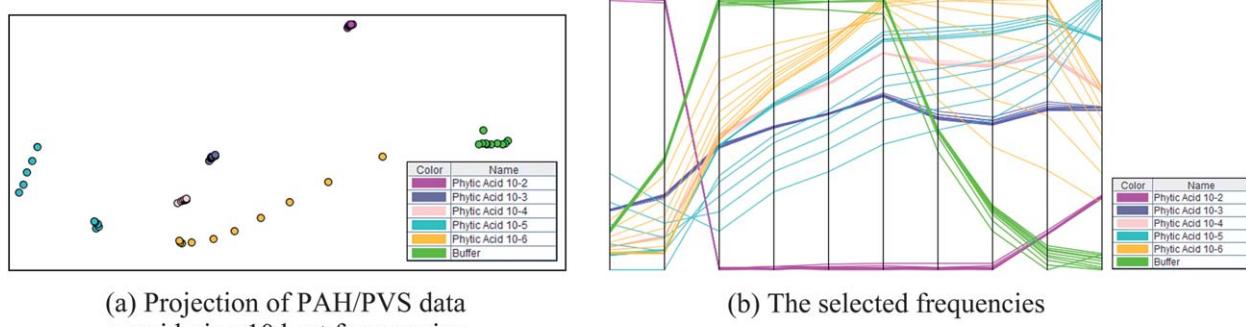


Fig. 3 Result of selecting frequencies which best differentiate the phytic acid concentrations using the PAH/PVS sensor.

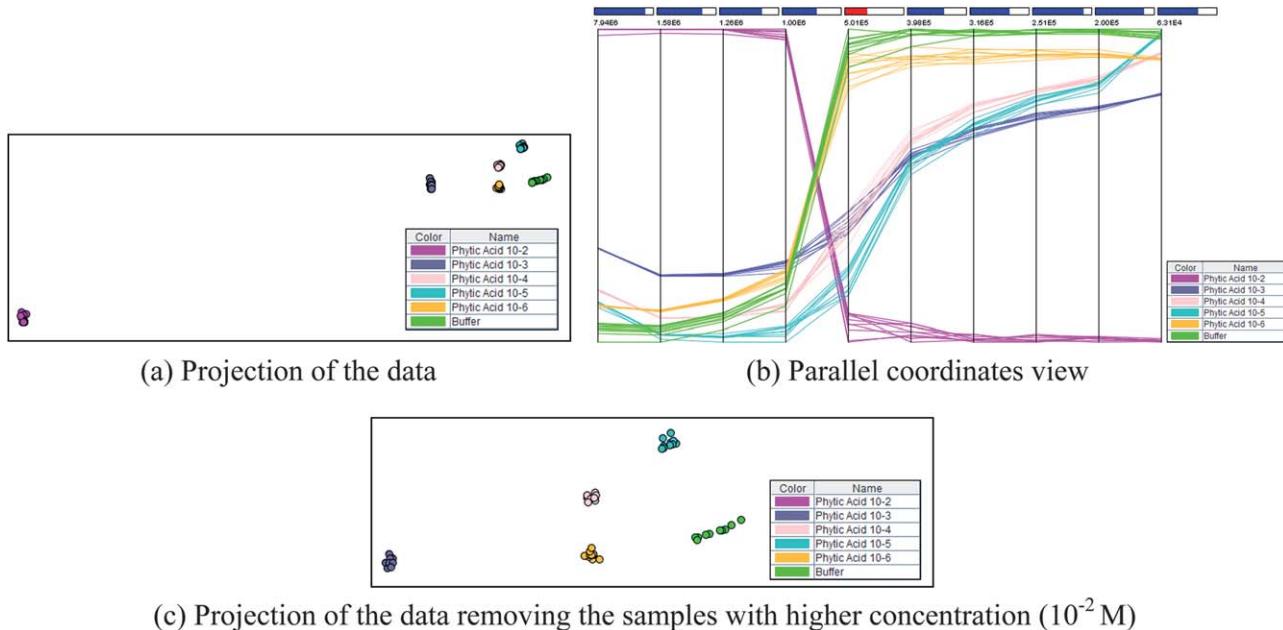


Fig. 4 Selecting amongst all sensors the best frequencies to distinguish various phytic acid concentrations.

indicating that even for lower concentrations, the selected frequencies define good results in terms of grouping and separating the samples.

The results discussed here demonstrate that simple visualization methods can help uncovering important information regarding data measurements, including the identification of frequencies and frequency intervals that play a major role in selectivity for different sensor configurations. Moreover, they illustrate how computational methods can be used to define in which frequencies a sensor, or a set of sensors, presents (almost) the same response, helping the manufacturing of reproducible sensors. This latter point is extremely important because one of the main drawbacks of biosensors and sensor arrays – such as electronic tongues²⁷ – is the lack of reproducibility of sensing units. While extremely sensitive to any changes in the liquid, the sensors made with nanostructured films or bare metallic electrodes²⁸ can never provide the same electrical response, even when they are nominally identical. Therefore, whenever one or more sensing units have to be replaced, all the knowledge acquired with the previous set of sensors is lost. The only possible way of retrieving such knowledge is to map the response of the new units into the response of the replaced units, and this can be possibly done with information visualization methods, as proposed here.

5. Conclusions and perspectives

In this paper we have shown that sensitivity and selectivity may be enhanced through usage of multiple high-dimensional visualization techniques. With suitable data projections in some cases even bare electrodes may be sufficient for a reasonable discrimination of similar samples. Moreover, *Parallel Coordinates* views of the data measurements allow observing and comparing the behavior of different data frequencies and their role in sensitivity. These and possibly other visualization techniques may be

employed in a coordinated manner to investigate sensor behavior and optimize sensor performance, for instance by helping identify the frequency ranges where distinction is best. Significantly, since sensitivity and selectivity depend on the type of unit and sample, it is useful to employ projection techniques to explore the data space and determine which method yields the higher performance. We also demonstrated that complementary computational tools (e.g., a genetic optimization algorithm) can be employed to automatically determine the relevant measurements in this context.

Finally, we emphasize that the information visualization methods implemented in *PEx-Sensors* are completely generic, and may be applied to any type of sensor and biosensor. In sensing based on spectroscopic measurements, for instance, the whole spectrum for each sample may be processed rather than relying on information of specific bands. This will be especially important for single molecule detection, e.g. obtained with surface-enhanced Raman scattering,²⁹ and for the analysis of biological samples such as bacteria that were identified using their molecular fingerprints.³⁰ Therefore, one may envisage sophisticated systems for clinical diagnosis – or similar applications in analytical chemistry – in which not only massive amounts of data may be treated but also their interpretation integrated with other sources of information. For instance, image processing tools for exams based on different types of tomography may be combined with the information visualization and artificial intelligence methods – such as the genetic algorithms used in this paper – thus leading to expert systems capable of providing diagnostic performance that would be impossible to get only with separated, individual sensing systems.

Acknowledgements

This work was supported by FAPESP, CNPq, Capes, INEO and nBioNet (Brazil).

References

- 1 K. Ariga, J. P. Hill, M. V. Lee, A. Vinu, R. Charvet and S. Acharya, *Sci. Technol. Adv. Mater.*, 2008, **9**, 014109.
- 2 S. Vaddiraju, I. Tomazos, D. J. Burgess, F. C. Jain and F. Papadimitrakopoulos, *Biosens. Bioelectron.*, 2010, **25**, 1553.
- 3 J. R. Siqueira Jr., L. Caseli, F. N. Crespiholo, V. Zucolotto and O. N. Oliveira Jr., *Biosens. Bioelectron.*, 2010, **25**, 1254.
- 4 M. L. Moraes, O. N. Oliveira Jr., U. P. Rodrigues-Filho and M. Ferreira, *Sens. Actuators, B*, 2008, **131**, 210.
- 5 V. Zucolotto, A. P. A. Pinto, T. Tumolo, M. L. Moraes, M. S. Baptista, A. Riul, A. P. U. Araujo and O. N. Oliveira Jr., *Biosens. Bioelectron.*, 2006, **21**, 1320.
- 6 D. Grieshaber, R. MacKenzie, J. Vörös and E. Reimhult, *Sensors*, 2008, **8**, 1400.
- 7 M. C. F. de Oliveira and H. Levkowitz, *IEEE Trans. Visualization Comput. Graphics*, 2003, **9**, 378.
- 8 J. R. Siqueira Jr., R. M. Maki, F. V. Paulovich, C. F. Werner, A. Poghossian, M. C. F. de Oliveira, V. Zucolotto, O. N. Oliveira Jr. and M. J. Schöning, *Anal. Chem.*, 2010, **82**, 61.
- 9 M. L. Moraes, R. M. Maki, F. V. Paulovich, U. P. Rodrigues Filho, M. C. F. de Oliveira, A. Riul Jr., N. C. de Souza, M. Ferreira, H. L. Gomes and O. N. Oliveira Jr., *Anal. Chem.*, 2010, **82**, 3239.
- 10 A. C. Perinotto, R. M. Maki, M. C. Colhone, F. R. Santos, V. Migliaccio, K. R. Daghistanli, R. G. Stabeli, P. Ciancaglini, F. V. Paulovich, M. C. F. de Oliveira, O. N. Oliveira Jr. and V. Zucolotto, Novel Biosensors for Rapid and Efficient Diagnosis of Leishmaniasis: Innovations in Bioanalytics for a Neglected Disease, *Anal. Chem.*, 2010, **82**, 9763.
- 11 S. K. Card, J. D. Mackinlay and B. Shneiderman, *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc., San Francisco, 1999.
- 12 C. Chen, *Information Visualization: Beyond the Horizon*, Springer, 2004.
- 13 I. T. Jolliffe, *Principal component analysis*, Springer-Verlag, 1986.
- 14 C. Faloutsos and K.-I. Lin, FastMap: a fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets, in *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data (SIGMOD'95)*, New York, NY, USA, CM, 1995, pp. 163–174.
- 15 T. F. Cox and M. A. A. Cox, *Multidimensional scaling*, Chapman and Hall/CRC, second edn, 2000.
- 16 F. V. Paulovich, L. G. Nonato, R. Minghim and H. Levkowitz, *IEEE Trans. Visualization Comput. Graphics*, 2008, **14**, 564.
- 17 F. V. Paulovich, C. T. Silva and L. G. Nonato, *IEEE Trans. Visualization Comput. Graphics*, 2010, **16**, 1281.
- 18 J. W. Sammon, *IEEE Trans. Comput.*, 1969, **18**, 401.
- 19 R. Minghim, F. V. Paulovich and A. A. Lopes, in *Proceedings of Visualization and Data Analysis; IS&T/SPIE Symposium on Electronic Imaging*, San Jose/California, 2006, 6060, pp. S1–S12.
- 20 E. Tejada, R. Minghim and L. G. Nonato, *Inf. Vis.*, 2003, **2**, 218.
- 21 T. M. J. Fruchterman and E. M. Reingold, *Software: Pract. Exper.*, 1991, **21**, 1129.
- 22 A. Inselberg and B. Dimsdale, Parallel coordinates: a tool for visualizing multidimensional geometry, in *Proceedings of the IEEE Visualization 1990 (Vis'90)*, 1990, p. 361–375.
- 23 G. Decher, *Science*, 1997, **277**, 1232.
- 24 R. F. M. Lobo, M. A. Pereira-da-Silva, M. Raposo, R. M. Faria and O. N. Oliveira Jr., *Nanotechnology*, 2003, **14**, 101.
- 25 P.-N. Tan, M. Steinbach and V. Kumar, *Introduction to Data Mining*, Addison-Wesley Longman Publishing Co., Inc., Boston, 2005.
- 26 R. Lletí, M. C. Ortiz, L. A. Sarabia and M. S. Sánchez, *Anal. Chim. Acta*, 2004, **515**, 87.
- 27 A. Riul, A. M. Gallardo Soto, S. V. Mello, S. Bone, D. M. Taylor and L. H. C. Mattoso, *Synth. Met.*, 2003, **132**, 109.
- 28 N. K. L. Wiziack, L. G. Paterno, F. J. Fonseca and L. H. C. Mattoso, *Sens. Actuators, B*, 2007, **122**, 484.
- 29 C. J. L. Constantino, T. Lemma, P. A. Antunes and R. Aroca, *Anal. Chem.*, 2001, **73**, 3674.
- 30 I. O. Osorio-Roman, R. F. Aroca, J. Astudillo, B. Matsuhiro, C. Vazquez and J. M. Perez, *Analyst*, 2010, **135**, 1997.