Multi-Ion-Sensing Emulator and Multivariate Calibration Optimization by Machine Learning Models

IVAN NY HANITRA[ID]**, (Graduate Student Member, IEEE), FRANCESCA CRISCUOLO**[ID]**,
SANDRO CARRARA**[ID]**, (Fellow, IEEE), AND GIOVANNI DE MICHELI**[ID]**, (Life Fellow, IEEE)**
Integrated Systems Laboratory (IC-IINFCOM-LSI1), Swiss Federal Institute of Technology Lausanne (EPFL), Station 14, 1015 Lausanne, Switzerland
Corresponding author: Ivan Ny Hanitra (ivan.nyhanitra@epfl.ch)

ABSTRACT One paramount challenge in multi-ion-sensing arises from ion interference that degrades the accuracy of sensor calibration. Machine learning models are here proposed to optimize such multivariate calibration. However, the acquisition of big experimental data is time and resource consuming in practice, necessitating new paradigms and efficient models for these data-limited frameworks. Therefore, a novel approach is presented in this work, where a multi-ion-sensing emulator is designed to explain the response of an ion-sensing array in a mixed-ion environment. A case study is performed emulating the concurrent monitoring of sodium, potassium, lithium, and lead ions, in a medium representative of sweat samples. These analytes are relevant examples of sweat ion-sensing applications for physiology, therapeutic drug monitoring, and heavy metal contamination. It is demonstrated that calibration datasets output by the emulator explain accurately the experimental response of polymeric solid-contact ion-selective electrodes, where root-mean-squared error of $1.37, 1.44, 1.78, 2\,mV$ are obtained, respectively, for Na^+, K^+, Li^+, Pb^{2+} sensor calibration in artificial sweat. Besides, synthetic datasets of custom size are generated to train, validate, and evaluate different types of multivariate regressors. A Multi-Output Support Vector Regressor (M-SVR) is proposed as a compact, accurate, robust, and efficient multivariate calibration model. It features 13.22% normalized root mean squares, and 20.29% mean root squares improvement compared to a simple linear regression model. It is an unbiased estimator for medium to large datasets, and its average generalization error is of 3.22%. Besides, M-SVR models have a lower computational complexity than single-output SVR or neural network models, making them a suitable solution for memory and energy-constrained edge devices used for continuous and real-time multi-ion monitoring.

INDEX TERMS Ion interference modeling, machine learning, multi-ion-sensing emulator, multivariate calibration, multi-output support vector regressor.

I. INTRODUCTION

Potentiometric ion-sensors are increasingly applied in many fields, leveraging progress in all-solid-state sensing and technology [1]. Namely, they are used for environmental monitoring of pollutants [2], for agricultural soil analysis [2], for water quality control [3], or for biomedical applications such as physiology and healthcare monitoring [4], [5]. The sensor consists in an Ion-Selective Electrode (ISE) coated with a polymeric membrane that selectively entraps the target analyte. Research efforts are focused on improving

The associate editor coordinating the review of this manuscript and approving it for publication was Ze Ji[ID].

sensing technology and transduction mechanisms [6], but ion interference is strongly limiting sensing performances because of ISE intrinsic selectivity bounds [7]. Interference arises from the electrolytes inherently constituting the sample, whose composition could vary in a non-predictable way with the environment. Besides, accurate ion-monitoring becomes more intricate when the target electrolyte is extremely diluted in the sample, the analyte concentration being of the same order as sensor Limit Of Detection (LOD).

As a result, multivariate calibration, that consists in binding the multivariate response of the ion-sensing array to the concentration of the target ions, is becoming less accurate. To overcome ion interference and increase sensing accuracy,

the concept of electronic tongue has been introduced. It is a recent trend for which sensor arrays are coupled with chemometric tools. [8] More precisely, it is defined by IUPAC as "a multisensor system, which consists of a number of low-selective sensors, and uses advanced mathematical procedures for signal processing based on Pattern Recognition and/or Multivariate data analysis—Artificial Neural Networks (ANNs), Principal Component Analysis (PCA), and so forth–." [9] Several works are proposed in literature where ANN architectures are coupled to sensor arrays in order to determine simultaneously multi-ions in samples subject to ion interference [2], [10]–[17]. The quality of results obtained with these black-box chemometric models depends on the size and the representativeness of the training dataset. Unfortunately, the acquisition of big data is time and resource consuming in practice, entailing high labor costs of chemical assays. Indeed, the datasets in the aforementioned works are restricted to 15 up to 100 samples. Some approaches have been proposed to cope with this data scarcity issue. Data fusion techniques are used in [18] to extract relevant features from data-limited ion-sensors, and generate synthetic samples from Cumulative Distribution Functions (CDFs) of the concentration of the target ions, before feeding the training set to different machine learning algorithms implemented for ammonium ion measurement.

In this work, a novel approach is proposed, where a multi-ion-sensing emulator is designed to explain the response of an ion-sensor array subject to ion interference from a mixed-ion sample. A compact version of the phase-boundary potential model [19], that is valid for polymeric ISEs, is built at the core of the emulator. Datasets of different size are generated and fed to linear and non-linear regressors that are implemented to improve sensing accuracy. A case study is carried out, emulating the concurrent monitoring of sodium, potassium, lithium, and lead ions, in an environment representative of sweat samples. These ions are relevant examples of sweat ion-sensing applications for physiology, therapeutic drug monitoring, and trace heavy metal contamination. Namely, the main minerals present in perspiration are sodium, potassium, and calcium [20]. During physical exercise, a depletion of sodium and potassium ions is observed [21]. An excessive loss of these ions could lead to muscle cramps, dehydration, or to hypokalemia and hyponatremia [22]. Besides, sweat ion-sensing could be applied to therapeutic drug monitoring. For instance, lithium salts are administrated to people suffering from bipolar disorder [23]. The therapeutic window of the drug is narrow ($0.8 - 1.5\,mM$ [24]), necessitating an accurate and continuous lithium monitoring. Moreover, some trace heavy metals such as lead are present in sweat, but in diluted amount (below $283\,\mu g/L$). Nevertheless, massive lead contamination from food, water, or the environment could result in memory trouble, pain, numbness, or behavioral problems [25], thus requiring an accurate lead monitoring system.

The remainder of the manuscript is organized as follows. First, some related works on multivariate calibration optimization in multi-ion-sensing are discussed in Section II to highlight the challenges in data-constrained systems for accurate and simultaneous monitoring of multi-ions. Then, the multi-ion-sensing emulator is presented in Section III, with the description of the compact phase-boundary potential model that is at its core. The methodology carried out to generate synthetic datasets out of the emulator is described in Section IV. Next, the proposed Multi-output Support Vector Regressor (M-SVR) is detailed in Section V. The results of the case study are discussed in Section VI, and the conclusions are reported in Section VII.

II. CHALLENGES AND RELATED WORKS

One of the main challenge in accurate ion-sensing arises from ion interference, due to the background electrolytes in the sample, that creates non-linearity in sensor response. Moreover, in multi-ion-sensing, cross-interference is not negligible, and monitoring analytes that are diluted in the sample is intricate. As a result, multivariate calibration is commonly optimized by machine learning models. Some related works on multi-ion-sensing are listed in Table 1, aiming mainly at environmental, agriculture, and water quality monitoring. Polymeric ISEs are fabricated and characterized, enabling the monitoring of ions of interest, and in some works, the tracking of interfering ions as well. Then, synthetic training datasets are acquired from the sensors. They are of relative small size, considering the huge time and chemical resource needed to acquire big data. Nevertheless, the calibration dataset should be representative enough of real samples. Factorial design, random design in the concentration range of the target ions, or usage of CDFs of the concentration of the ions are strategies implemented to obtain statistically relevant synthetic training sets with a limited amount of samples. The test sets are acquired either from real samples, or with synthetic design procedures. Multivariate calibration is performed mostly with feed-forward Back-Propagation Neural Network (BPNN) of one to three hidden layers, trained with a Bayesian regularization algorithm. More complicated architectures are investigated to incorporate a priori knowledge from the sensors (charge balance, electrical conductivity constraints) [12], or genetic independent component analysis [10], [11].

III. EMULATOR DESIGN FOR MULTI-ION-SENSING

The approach proposed in this work to overcome the data-limited constraints in the training of chemometric tools deployed for multivariate calibration is to design an emulator of synthetic datasets, explaining the responses of polymeric ISEs exposed to mixed-ions samples. First, the compact modeling of ISE response is described, then the design of the emulator is detailed.

A. MODELING OF INTERFERENCE IN ION-SENSING

In potentiometric sensing, ISEs are functionalized with an ion-selective membrane that entraps the target ion. The electrode transduces the thermodynamic activity of the target

TABLE 1. Related works on multivariate calibration in multi-ion-sensing.

Multi-ion-sensing application	Dataset design	Multivariate calibration models	Ref
NH_4^+, Na^+, K^+, for pollution monitoring in river Cu^{2+}, Pb^{2+}, Zn^{2+}, Cd^{2+} in open air waste streams	*Training*: 27 synthetic samples (factoral design) *Validation*: 10 synthetic samples (random sampling) *Test*: real samples acquired from a river	BPNN[1] (1 hidden layer)	[2]
K^+, Na^+, Ca^{2+}, Mg^{2+} for agriculture water quality monitoring	*Training*: 16 synthetic samples (factoral design) *Validation*: 15 synthetic samples (factoral design) *Test*: 20 real irrigation water samples	PCA + genetic ICA[2] + BPNN[1] (1 hidden layer)	[10]
Na^+, K^+, Ca^{2+}, Mg^{2+}, NH_4^+, Cl^-, NO_3^-, SO_4^{2-}, carbonate, and pH, for environmental monitoring	*Dataset*: 67/98 synthetic samples (random sampling) *Training/Validation/Test*: 0.7/0.15/0.15	BPNN[2] constrained with a priori knowledge (charge balance, electrical conductivity)	[12]
NO_3^-, NO_2^-, NH_4^+ monitoring in water plus interfering ions (Na^+, K^+, Cl^-)	*Training*: 37 synthetic samples (factoral design) *Test*: 15 synthetic samples (random sampling)	BPNN[1]	[15]
S^{2-} and ClO_4^- monitoring for water pollution	*Training*: 59 samples, *Test*: 20 samples, obtained from microvolume additions of 2 concentration levels	BPNN[1] (1 hidden layer)	[17]
Ca^{2+}, Mg^{2+}, Na^+, K^+ for water monitoring	*Dataset*: 50 synthetic samples generated from individual and joint CDFs[3] of concentration of the ions	MLP, SVR, RF[4] regressors	[18]
Na^+ and K^+ for physiology Li^+ for therapeutic drug monitoring Pb^{2+} for environment contamination	*Training*: 68 up to 680 samples (factoral design) *Validation*: 5 splits, 20% of training set (random sampling) *Test*: 20% of training set (random sampling)	Linear regressors, SVR, M-SVR, MLP models	**This work**

[1] Back-Propagation Neural Network [2] Independent Component Analysis [3] Cumulative Distribution Function [4] Random Forest

ion into an electric potential. The Open Circuit Potential (OCP) between the sensing electrode and an inert reference electrode, which has a stable potential, is measured under quasi-zero current conditions [6]. Sensor calibration consists in mapping the measured OCP with the target ion activity. The latter refers to the ability of the analyte to react with other ions, and is related to the effective analyte concentration in the sample by the activity coefficient. In the following, ion activity is the chemical property of interest in the multivariate calibration.

ISE response is usually described by the semi-empirical Nicolsky-Eisenman model [26]. However, it is not very accurate when ions of different charges are present in the sample. Namely, large deviations from acquired sensor data are observed in ion mixtures [27]. In this work, the phase-boundary potential model is used to describe ion-sensor response of a polymeric ISE in mixed-ion solutions. A comprehensive description of the model is presented in [19]. The hereunder analytical derivation is carried out to get a compact ion-sensing model, highlighting the key parameters explaining sensor output distortion due to ion interference. An exhaustive derivation of the model is reported in Appendix A. The phase-boundary potential model is based on ion-exchange considerations at the sample/membrane interface. Transmembrane ion fluxes are ignored in this model. When the target ion is entrapped in the polymeric membrane, an electrical potential is built up at the interface to counter-balance the ion fluxes from the sample phase to the membrane phase, and to preserve electroneutrality. The phase-boundary potential at equilibrium for any ion j is shown in (1). It is obtained by equating the electrochemical potential in the membrane and sample phase, where Φ is the electrical potential, ϵ_j^0 is the standard potential of ion j of valence z_j, of activity $a_j(m)$ and $a_j(aq)$, in the membrane and sample phase, respectively. $s = \frac{RT}{F} ln\,10$ is the

Nernst slope.

$$E_{PB} \overset{def}{=} \Phi(m) - \Phi(aq)$$
$$= \epsilon_j^0 + \frac{1}{z_j} s \log\left(\frac{a_j(aq)}{a_j(m)}\right). \quad (1)$$

Next, it is assumed that the membrane ionic strength is unaltered during ion exchange. By applying membrane electroneutrality, with multiple ions interacting with the membrane phase (interfering ions),

$$10^{-z_j E_{PB}/s} \sum_j a_j(aq)\, 10^{z_j E_j^0/s} = 1, \quad (2)$$

where $E_j^0 = \epsilon_j^0 + \frac{1}{z_j} s \log\left(\frac{1}{a_j(m)}\right)$ is the apparent standard potential of ion j. All interactions of ion j with the ionophore of the membrane are embedded in $a_j(m)$. Then, the membrane selectivity is quantified by the selectivity coefficient

$$\log K_{I,J}^{pot} = \frac{z_I}{s}(E_J^0 - E_I^0). \quad (3)$$

It represents how much the membrane is selective towards a target ion I than an interferent ion J. Substituting (3) in (2), the compact equation (4), as shown at the bottom of the next page, is obtained, considering monovalent and divalent ions in the sample. The contributions of the target ion I, monovalent, and divalent interfering ions are put in evidence. The selectivity coefficients could be seen as weighting factors of the activity of interfering ions. Equation (4) is solved for E_{PB}, and yields (5) and (6), as shown at the bottom of the next page, for a monovalent and divalent target ion, respectively.

An ideal Nernstian response is recovered if all interfering electrolytes have a null activity (no interferent), or if $\forall j$, $\log K_{I,j}^{pot} \longrightarrow -\infty$ (perfectly selective membrane). The ion-selective membrane properties are embedded in the

constant E_I^0 that could be seen as an offset potential independent of $a_I(aq)$. The parameters to model ion interference are reduced to the selectivity coefficients log $K_{I,J}^{pot}$, and the activity of the interfering ions $a_J(aq)$. Eventually, the OCP between an ISE for target ion I, and the reference electrode is

$$E_I = K_{cell} + E_{PB},\qquad(7)$$

where K_{cell} takes into account the constant potentials in the galvanic cell, including reference electrode potential and electrode contact potential drops.

B. MULTI-ION-SENSING EMULATOR

A Graphical Unit Interface (GUI) implementation of an ion-sensor calibration curves emulator has been presented in [28]. Its main features are summarized in the following. The tool allows for selecting generic monovalent or divalent primary ions and interfering ions. The selectivity coefficients of the ISEs are tunable, as well as the offset parameter that is fitted from real ISE calibration curves. The phase-boundary potential model and Nicolsky-Eisenman model could be used to output the calibration data, where a Gaussian noise equivalent to sensor signal-to-noise ratio is added. The investigation tool features parametric analysis of the impact of interferent ion activity, and selectivity coefficient of an ISE towards an interferent, in the distortion of the calibration curves due to ion interference. Database of calibration curves from in-vitro measurements could be loaded to compare the emulator outputs to real calibration data.

The investigation tool described previously is modified into a synthetic dataset generator in order to produce the training dataset for the multivariate calibration models. The building blocks of the emulator are pictured in Fig. 1. The compact phase-boundary potential model takes as input parameters the selectivity coefficients of each target ion against each interferent, and the offset parameter encompassing membrane composition. Both parameters are experimentally extracted from real ISE calibration curves. The synthetic dataset is obtained following a factorial design where the activity of each constituent in the sample is discretized in L levels in the detection range of the analyte, for the given application. The constituents include the primary ions and the interfering ones. The output of the dataset generator consists of calibration curves of each sensor at discrete level points.
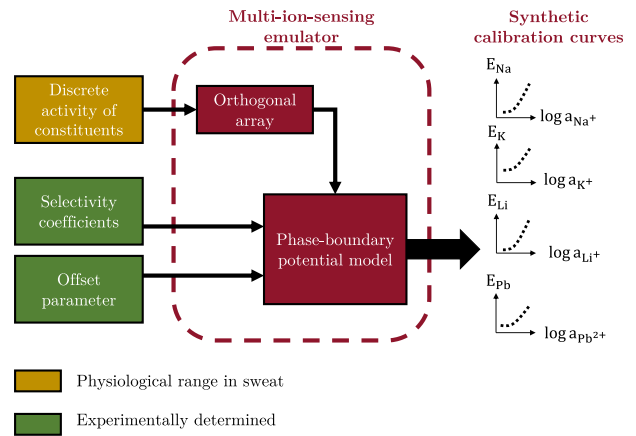

FIGURE 1. Building blocks of multi-ion-sensing emulator of Na^+, K^+, Li^+, and Pb^{2+} monitoring in synthetic sweat samples.

TABLE 2. OA_{18} ($6^1 \times 3^6$).

	C0	C1	C2	C3	C4	C5	C6
R0	0	0	0	0	0	0	0
R1	0	1	1	1	1	1	1
R2	0	2	2	2	2	2	2
R3	1	0	0	1	1	2	2
R4	1	1	1	2	2	0	0
R5	1	2	2	0	0	1	1
R6	2	0	1	0	2	1	2
R7	2	1	2	1	0	2	0
R8	2	2	0	2	1	0	1
R9	3	0	2	2	1	1	0
R10	3	1	0	0	2	2	1
R11	3	2	1	1	0	0	2
R12	4	0	1	2	0	2	1
R13	4	1	2	0	1	0	2
R14	4	2	0	1	2	1	0
R15	5	0	2	1	2	0	1
R16	5	1	0	2	0	1	2
R17	5	2	1	0	1	2	0

IV. DESIGN OF SYNTHETIC DATASETS

A case study is performed to emulate simultaneous monitoring of Na^+, K^+, Li^+, and Pb^{2+} in ion mixtures representative of sweat samples. The methodology applied to design synthetic training, validation, and test datasets are described.

The multi-ion-sensing emulator described in Section III is used to generate synthetic OCP signals in controlled

$$10^{z_I(E_I^0-E_{PB})/s}a_I(aq) + 10^{(E_I^0-E_{PB})/s}\sum_{j\ s.t.\ z_j=1}(K_{I,j}^{pot})^{\frac{1}{z_I}}a_j(aq) + 10^{2(E_I^0-E_{PB})/s}\sum_{j\ s.t.\ z_j=2}(K_{I,j}^{pot})^{\frac{2}{z_I}}a_j(aq) = 1.\qquad(4)$$

$$E_{PB} = E_I^0 + s\log\left(\frac{1}{2}\left(a_I(aq) + \sum_{j\ s.t.\ z_j=1}K_{I,j}^{pot}a_j(aq)\right) + \sqrt{\frac{1}{4}\left(a_I(aq) + \sum_{j\ s.t.\ z_j=1}K_{I,j}^{pot}a_j(aq)\right)^2 + \sum_{j\ s.t.\ z_j=2}(K_{I,j}^{pot})^2a_j(aq)}\right).\qquad(5)$$

$$E_{PB} = E_I^0 + s\log\left(\frac{1}{2}\sum_{j\ s.t.\ z_j=1}(K_{I,j}^{pot})^{\frac{1}{2}}a_j(aq) + \sqrt{\frac{1}{4}\left(\sum_{j\ s.t.\ z_j=1}(K_{I,j}^{pot})^{\frac{1}{2}}a_j(aq)\right)^2 + a_I(aq) + \sum_{j\ s.t.\ z_j=2}K_{I,j}^{pot}a_j(aq)}\right).\qquad(6)$$

TABLE 3. Discrete levels of the activity of the seven constituents (*activity* $\times 1e-3$).

Ions	Na^+ [4]	K^+ [4]	Li^+ [24]	Pb^{2+} [25]	NH_4^+ [29]	Ca^{2+} [30]	Mg^{2+} [20]
Detection range	[7; 95] mM	[3; 8] mM	[2.4; 4.5] mM	[1.4; 85.4] µM	–	–	–
Nominal concentration	39.1 mM	5.12 mM	0.1 mM	0.24 µM	3.4 mM	0.37 mM	53.5 µM
L0	1	0.1	0.1	1e-3	0.1	1e-2	1e-3
L1	5	0.5	0.25	5e-3	0.5	5e-2	5e-3
L2	10	1	0.5	1e-2	1	0.1	1e-2
L3	25	2.5	1.25	2.5e-2	2.5	0.25	2.5e-3
L4	50	5	2.5	5e-2	5	0.5	5e-2
L5	100	10	5	0.1	10	1	0.1

TABLE 4. Selectivity coefficients of platinum-nanostructured solid-contact ISEs obtained from fixed-interference method.

				$\log K_{I,J}^{pot}$			
I \ J	Na^+	K^+	Li^+	Pb^{2+}	NH_4^+	Ca^{2+}	Mg^{2+}
Na^+	–	-2.2	-2.58	n.a.	-1.89	-2.57	-3.56
K^+	-1.8	–	-1.91	n.a.	-1.68	-4.6	-3.32
Li^+	-2.55	-2.56	–	n.a.	-3.08	-4.01	n.a.
Pb^{2+}	-2.89	-1.64	-2.9	–	-2.76	-4.28	-3.76

mixed-ion samples, in an automated way. Four sensors for $Na^+, K^+, Li^+,$ and Pb^{2+} monitoring are emulated in presence of $NH_4^+, Ca^{2+},$ and Mg^{2+}, that are the major ions constituting sweat samples [31]. The determination of ion mixture combinations requires a multi-factorial design of synthetic samples. In this work, the physiological range of activity of each constituent is divided into six levels. It is a good trade-off between capturing enough variability in the activity of the ions, and avoiding to generate exponential number of samples. Indeed, a full-factorial design involves $6^7 \sim 280'000$ ion combinations. This amount is not reasonable. Moreover, a full-factorial design will produce a redundant dataset. Therefore, Taguchi method is implemented to generate a representative subset of all constituent combinations, leveraging orthogonal arrays [32]. The $L_{18}(6^1 \times 3^6)$ orthogonal array used is reported in Table 2, where the factor C_0 comprises 6 levels, and the 6 other factors have 3 levels. The columns C_i are orthogonal, and orthogonality is preserved through column permutation. Thus, C_0 could be permuted over the seven constituents. Row R_0 is redundant while permuting columns, so it is removed. The activity of the seven constituents are quantized over the physiological or therapeutic range of the considered ions in sweat. The discrete levels of activities are reported in Table 3. As for selectivity coefficients, they are obtained from electrochemical characterization of polymeric solid-contact ISEs. Fixed interference method is carried out to compute the coefficients reported in Table 4, and the full characterization of $Na^+, K^+, Li^+,$ and Pb^{2+} sensors is published at [33]. The entire factorial-designed set is needed to train the multivariate calibration model since it includes all the statistical variability in the data that the regressor should learn. Therefore, cross-validation procedure or splitting of
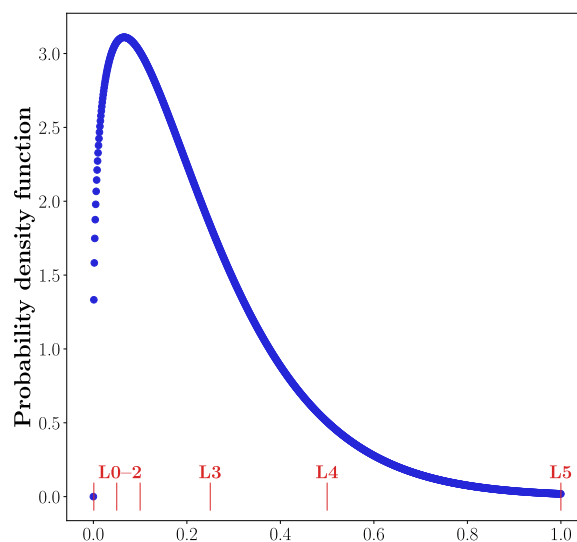
FIGURE 2. Weibull distribution ($k = 1.25, b = 6$) (blue), and discrete levels of each constituent scaled to its respective activity range(red).

the training set is avoided. Consequently, validation and test sets are designed independently. Random sampling in the activity window of the ions of interest is the usual approach adopted [2], [15]. In this work, the Weibull distribution is used. Its Probability Distribution Function (PDF) is plotted in Fig. 2, and it is defined as

$$f(x; k, b) = \begin{cases} b\, k\, x^{k-1} exp(-b\, x^k), & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (8)$$

where the scale and shape of the distribution are $\lambda = b^{-1/k}$ and k, respectively. This PDF is chosen so as to generate independent and identically distributed ion mixtures representative of sweat samples for which the activity of each constituent is around its nominal value, or rather in excess in the sample (e.g. due to physical exercise for physiology). In order to increase the robustness of the comparison between multivariate regressors, five validation splits, of one-fifth of the size of the training set each, are randomly generated to emulate a cross-validation. Thus, chemometric models are evaluated on each validation split, and the average and standard deviation of the metrics are computed. The final

validation of the model is performed on an additional test set of similar size as a validation split. Furthermore, datasets of different size are generated to assess the effect of dataset scarcity on the chemometric models. One notices that if the upper bound of the ion activity $L5$ in Table 3 is discarded, there are 10 combinations of 3-level ion activity to configure the orthogonal Table 2, allowing to tune the training set size. As a result, training sets of 68 up to 680 samples are generated following the aforementioned procedure.

V. MULTIVARIATE CALIBRATION MODELS

In this section, the multivariate calibration models used in this work are presented. In a multivariate calibration procedure, the signals output by an array of sensors are processed by a multivariate model that estimates the activity of selected analytes constituting the sample. The chemometric model is trained beforehand with OCP/activity observations, and is then inferred to predict the activity of the analytes in unknown samples [34]. For N observations, let $\mathbf{X} = \{\mathbf{x_n}\}_{n=1,\cdots,N}$, with $\mathbf{x_n} \in \mathbb{R}^P$, denote the tensor of OCP signals coming from P ISEs, and $\mathbf{Y} = \{\mathbf{y_n}\}_{n=1,\cdots,N}$, with $\mathbf{y_n} \in \mathbb{R}^M$, denote the tensor of activity of the M electrolytes of interest. The multivariate calibration problem can be formulated as

$$\mathbf{Y} = \mathbf{XW} + \mathbf{E}, \tag{9}$$

where $\mathbf{W}$ is the tensor of regression coefficients, and $\mathbf{E}$ is the error made in the prediction of the activity of the analytes. It is an inverse calibration problem since the controlled variables (activity of the ions) are estimated taking the dependent variables (OCP signals) as regressors.

A. TRADITIONAL LINEAR REGRESSION MODELS

Inverse least-squares regressor is the simplest first-order model used in multivariate calibration. Contrary to classical least-squares regression, the activity of all electrolytes constituting the sample do not need to be known during the calibration phase. As a result, the tensor of activity of the ions $\mathbf{Y}$ could be of any dimension M. The minimization objective is

$$\forall\, m \in [\![1; M]\!], \quad \min_{\mathbf{w}_{*,m}} \quad \|\mathbf{X}\mathbf{w}_{*,m} - \mathbf{y}_{*,m}\|_2^2, \tag{10}$$

where $\mathbf{w}_{*,m}$ and $\mathbf{y}_{*,m}$ are columns of $\mathbf{W}$ and $\mathbf{Y}$, respectively. Namely, the activity of each target ions is estimated independently, ignoring cross-correlations. Hence, this multivariate calibration model consists in multiple independent Ordinary Least-Squares (OLS) regressions, or Multiple Linear Regression (MLR). $\widehat{\mathbf{W}}$ could be computed by least-squares minimization of (10), or in a closed-form solution with the normal equation

$$\widehat{\mathbf{W}} = (\mathbf{X}^\mathrm{T}\mathbf{X})^{-1}\mathbf{X}^\mathrm{T}\mathbf{Y}, \tag{11}$$

where $\mathbf{X}^\mathrm{T}\mathbf{X}$ must be invertible. This holds if $\mathbf{X}$ is a full-rank matrix, or the columns of $\mathbf{X}$ are uncorrelated. The application of PCA to $\mathbf{X}$ will generate a matrix $\mathbf{T}$ that complies with the latter constraint. Namely, OCP signals are decomposed in

a set of successive orthogonal components that explain a maximum amount of their variance. The resulting Principal Component (PC) scores, denoted as $\mathbf{T}$, are the new input tensor variables. Combination of PCA and MLR refers to Principal Components Regression (PCR). Furthermore, Partial Least-Squares (PLS) regression is another popular chemometric tool [35]. It constructs latent variables for $\mathbf{X}$ and $\mathbf{Y}$ by taking into account the variance of $\mathbf{X}$ and the correlation between factors excerpted from both tensors.

Linear regressors are traditionally used for univariate calibration systems. However, in multivariate systems, and above all, in presence of non-linearity due to ion interference, non-linear regressors are privileged to increase multivariate calibration accuracy. They are described in the following.

B. SINGLE-OUTPUT SUPPORT VECTOR REGRESSION

Support Vector Machines (SVMs) are prominent tools for linear and non-linear input/output modeling, as it is the case of multivariate calibration subject to ion interference [36]. Standard SVR models perform unidimensional regression of the input tensor $\mathbf{X}$ to each output channel m, $\mathbf{y}_{*,m}$. For non-linear SVRs, $\mathbf{X}$ is implicitly mapped to a higher dimensional feature space by a kernel function [37]. Thus, the model learns a linear function in the space induced by the kernel. Gaussian Radial Basis Function (RBF) is a non-linear kernel commonly used for which $\mathbf{X}$ is mapped to an infinite dimensional Hilbert space [38]. The primal formulation of SVR minimization objective is

$$\forall\, m \in [\![1; M]\!], \quad \min_{\mathbf{w}_{*,m}, b_m} \frac{1}{2}\left\|\mathbf{w}_{*,m}\right\|^2 + C\sum_{n=1}^{N} L_\epsilon(u_{n,m}),$$
$$\text{with} \quad u_{n,m} = \mathbf{y}_{n,m} - (\Phi(\mathbf{x}_n)^\mathrm{T}\mathbf{w}_{*,m} + b_m). \tag{12}$$

L_ϵ is the Vapnik ϵ-insensitive loss function defined as $L_\epsilon(u) = \max\{0, |u| - \epsilon\}$. Φ is the non-linear function defining the kernel function $\kappa(\mathbf{x}_i, \mathbf{x}_j) \equiv \Phi(\mathbf{x}_i)^T \cdot \Phi(\mathbf{x}_j)$. C is a hyperparameter trading-off margin violations and minimization of the distance ϵ from the support vectors to the SVR model. In practice, a dual formulation of the SVR minimization is used, where quadratic programming leveraging kernel formulation yields a problem faster to solve.

C. PROPOSED MULTI-OUTPUT SUPPORT VECTOR REGRESSION

In single-output SVR problems, the loss function L is an L_1-based norm, hence, it needs to take into account each of the M dimensions independently which makes the solution complexity grow linearly with M. Moreover, such unidimensional model does not consider correlations between the output columns. Hence, it is less robust to noise and non-linearity in the dataset. In this work, a multivariate multi-output SVR is implemented to consider correlations and non-linear cross-relations between $\mathbf{X}$ and $\mathbf{Y}$, enabling a concurrent optimization of the regressor of each target ion activity.

The objective function to minimize is

$$\mathcal{L}(\mathbf{W}, \mathbf{b}) = \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{w}_{*,m}\|^2 + C \sum_{n=1}^{N} L(u_n),$$

$$\text{with } u_n = \|\mathbf{e}_n\|_2 \ \ s.t. \ \mathbf{e}_n = \mathbf{y}_n - (\mathbf{W}^{\mathrm{T}} \Phi(\mathbf{x}_n) + \mathbf{b}). \quad (13)$$

Iterative Reweighted Least-Squares (IRWLS) procedures permit the use of an arbitrary cost function L. L_2-based norm cost functions are appealing since the constraints for all M output dimensions could be considered into a single error vector, contrary to L_1-based Vapnik loss function described previously that need to take into account all M dimensions separately. As a result, IRWLS procedures yield one-dimension support vectors, instead of M-dimension for single-output SVRs. In this work, an IRWLS procedure is implemented with the quadratic and differentiable cost function

$$L(u) = \begin{cases} 0, & u < \epsilon. \\ u^2 - 2u\epsilon + \epsilon^2, & u \geq \epsilon. \end{cases} \quad (14)$$

The proof of convergence of such SVM is demonstrated in [39]. The IRWLS method is constructed by performing a first-order Taylor approximation of the error function (14) over the previous solution $(\mathbf{W}, \mathbf{b})$. A quadratic approximation of the former expansion gives the weighted least-squares minimization problem

$$\mathcal{L}'(\mathbf{W}, \mathbf{b}) = \frac{1}{2} \sum_{m=1}^{M} \|\mathbf{w}_{*,m}\|^2 + \frac{1}{2} \sum_{n=1}^{N} a_n u_n^2 + cste, \quad (15)$$

where $a_n = \frac{C}{u_n^k} \frac{d\,L(u)}{du}\Big|_{u_n^k} = \begin{cases} 0, & u_n^k < \epsilon. \\ \frac{2C(u_n^k - \epsilon)}{u_n^k}, & u_n^k \geq \epsilon. \end{cases}$

The latter function is minimized with respect to $(\mathbf{W}, \mathbf{b})$. It yields a linear system represented in matrix form as

$$\forall m \in [\![1; M]\!], \underbrace{\begin{bmatrix} \mathbf{K} + \mathbf{D}_a^{-1} & \mathbf{1} \\ \mathbf{a}^{\mathrm{T}}\mathbf{K} & \mathbf{1}^{\mathrm{T}}\mathbf{a} \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \mathbf{w}_{*,m}^{sol} \\ b_m^{sol} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{*,m} \\ \mathbf{a}^{\mathrm{T}}\mathbf{y}_{*,m} \end{bmatrix},$$

$$(16)$$

where $(\mathbf{K})_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{a} = [a_1, \cdots, a_n]^{\mathrm{T}}$, $(\mathbf{D}_a)_{ij} = a_i \delta(i - j)$. It is important to notice that the matrix $\mathbf{H}$ is independent of m, so it does not depend on the output channel. The kernel function used is the Gaussian RBF. The pseudo-code of the IRWLS procedure is reported hereunder.

IRWLS procedure is implemented in a quasi-Newton approach, for which the errors u_n^k and a_n are recomputed at each iteration k, so as the solution $(\mathbf{W}^{sol}, \mathbf{b}^{sol})$ of the reweighted least-squares problem. Each iteration has the complexity of an OLS problem.

D. MULTI-LAYER PERCEPTRONS MODEL

Multi-Layer Perceptrons (MLP) models are commonly used in complex multivariate calibration systems [2], [10]–[17]. MLP networks are built for the multivariate calibration of

---

Algorithm 1 Iterative Reweighted Least-Squares Procedure

Input: OCP signal $\mathbf{X}$, ions log-activity $\mathbf{Y}$, kernel function Φ, kernel parameters *param*, convergence tolerance *tol*, maximum number of iterations *max_iter*

Output: Trained weights $\mathbf{W}$, trained bias $\mathbf{b}$, number of iterations k

/* Initialization */

1 $(\mathbf{W}, \mathbf{b}) \leftarrow (\mathbf{0}, \mathbf{0})$; k $\leftarrow 0$;

2 $\mathbf{K} \leftarrow \mathbf{X}$ mapped to kernel Φ ;

3 **foreach** n **do**

4    $\mathbf{e}_n \leftarrow \mathbf{y}_n - (\mathbf{W}^{\mathrm{T}}\mathbf{K}_*, n + \mathbf{b})$;   // residual error

5    $u_n^k \leftarrow \|\mathbf{e}_n\|$;

6 $\{i\}^k \leftarrow \{n \text{ s.t. } u_n^k \geq \epsilon$;   // support vectors

7 **foreach** i **do**

8    $L(u_i^k) \leftarrow u_i^{k2} - 2u_i^k \varepsilon + \varepsilon^2$;   // to compute $\mathcal{L}$

9    $a_i \leftarrow \frac{2C(u_i^k) - e}{u_i^k}$;   // to compute $\mathcal{L}'$

/* Iterative procedure */

10 **while** *convergence tolerance is not reached* **and** k < max_iter **do**

11    $(\mathbf{W}^k, \mathbf{b}^k) \leftarrow (\mathbf{W}, \mathbf{b})$; $\mathbf{E}^k \leftarrow \mathbf{E}$; $\mathbf{u}^k \leftarrow \mathbf{u}$; $\{i\}^k \leftarrow \{i\}$;

12   build $\mathbf{H}$ tensor and solve $\mathcal{L}' \rightarrow (\mathbf{W}^{sol}, \mathbf{b}^{sol})$;

13   update $\mathbf{u}$, $\{i\}$, L, and $\mathcal{L}$;

14   minimize $\mathcal{L}$ with descending direction step η^k;

15   compute next step solution

$$\begin{bmatrix} \mathbf{W}k+1 \\ (\mathbf{b}^{k+1})\mathrm{T} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^k \\ (\mathbf{b}^k)\mathrm{T} \end{bmatrix} + \eta^k \begin{bmatrix} \mathbf{W}^{sol} - \mathbf{W}^k \\ (\mathbf{b}^{sol} - \mathbf{b}^k)\mathrm{T} \end{bmatrix};$$

16   recompute ai and update least-squares problem $\mathcal{L}'$;

17   k $\leftarrow$ k + 1;

---

potentiometric sensors subject to ion interference in order to benchmark the proposed M-SVR algorithm. The feed-forward NN architecture implemented is illustrated in Fig. 3. The input layer consists of four pass-through neurons that relay the input features $\mathbf{x}_n$ to the subsequent layers. The features are either the OCP signals of the $P = 4$ ion-sensors, or their PC scores after applying PCA. Two hidden layers of $N1$ and $N2$ units produce non-linear transformations of the dataset, and the output layer predicts the activity of the $M = 4$ target ions. The model parameters (weights and bias) are trained with a back-propagation training algorithm minimizing the mean-squared error loss function [40]. The hyper-parameters of the MLP models are optimized with a grid-search procedure. They are summarized in Table 5.

E. SOFTWARE AND HARDWARE

Data pre-processing pipeline and multivariate models are implemented within a Python 3.7 environment. PCA algorithm is implemented through a Singular Value Decomposition (SVD) where LAPACK routine [42] is used. MLR and OLS minimization problems are solved using SVD on
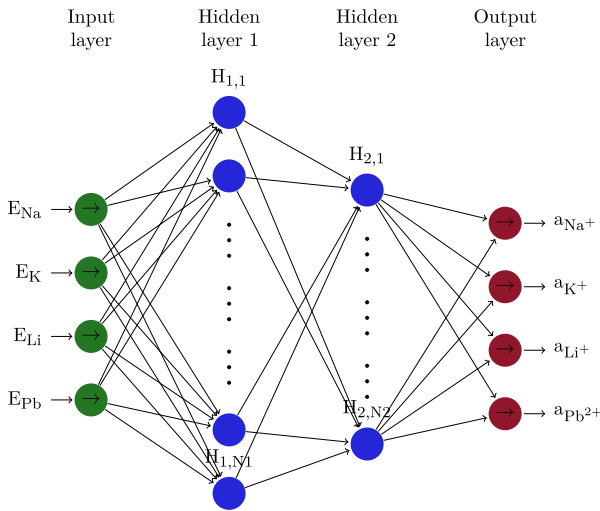
FIGURE 3. Feed-forward neural network architecture for simultaneous monitoring of sodium, potassium, lithium, and lead ions: the input features (OCP signals or their PC scores) are fed to the passthrough neurons of the input layer, then conveyed to two fully-connected hidden layers of $N1$ and $N2$ number of units, respectively. The output layer consists of four units with no activation function, providing the activity of the target ions.

TABLE 5. Configuration and hyper-parameters of MLP model, optimizer, and training procedure.

MLP model	Input layer: 4 pass-through neurons Hidden layer 1: **N1** neurons (σ = ELU) Hidden layer 2: **N2** neurons (σ = ELU) Output layer: 4 neurons Weights initializer: Glorot (normal distribution) [41] Regularizer: L_2 regularizer ($\alpha = 10^{-3}$)
Loss function	Mean-squared error
Optimizer	Type: stochastic gradient descent with momentum optimization Learning rate scheduling: $\eta(t) = \frac{\eta_0}{1+10^{-5}t}$
Training	Epochs: 500 (early-stopping) Batch size: 32

zero-mean and unit-variance $\mathbf{X}$ with LAPACK routine [42], while PLS is implemented with Non-Linear Iterative Partial Least-Squares (NIPALS) algorithm [43]. Single-output SVR is constructed leveraging LIBSVM library that supports quadratic programming [44]. As for MLP models, they are implemented through Keras high-level API [45], with TensorFlow 2.0 deep learning library as computational back-end. The experiments are run on a 2x Intel(R) Xeon(R) CPU E5-2690 v4 @ 2.60 GHz with 256 GB RAM.

VI. RESULTS AND DISCUSSION

In this section, the results of the case study are presented. The performance of the multi-ion-sensing emulator is assessed and compared with calibration curves obtained from polymeric solid-contact ISEs. Then, the synthetic datasets generated by the tool are shown. Next, the multivariate calibration
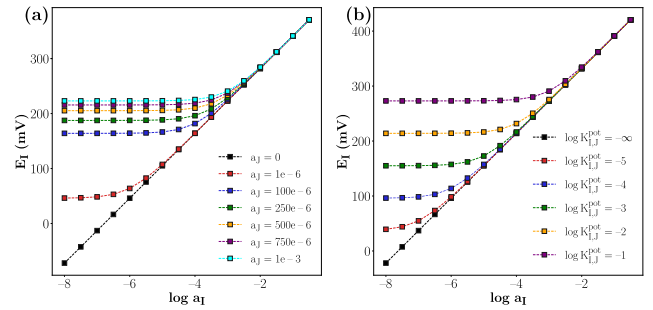

FIGURE 4. Modeling OCP distortion due to: (a) ion interference a_J swept from 0 to $1e-3$, with $z_I = 1$, $offset = 400\,mV$, $z_J = 1$, and $\log K_{I,J}^{pot} = -3$; (b) ISE selectivity $\log K_{I,J}^{pot}$ swept from $-\infty$ to -1, with $z_I = 1$, $offset = 450\,mV$, $z_J = 1$, and $a_J = 10e-6$.

optimization with linear and non-linear regressors is discussed.

A. MULTI-ION-SENSING EMULATOR AND ION INTERFERENCE MODELING

First, the ion-sensing emulator is used as investigation tool to evaluate the effect of ion interference on ISE response. Fig. 4 illustrates the impact of the two model parameters on calibration curve distortion. Namely, the activity of the interfering ion a_J, and the selectivity coefficient of the primary ion I towards the interfering ion J, $\log K_{I,J}^{pot}$. The background potential offset in diluted analyte ($\log a_I \leftarrow -\infty$) increases with the activity of the interfering electrolyte, and this effect is enhanced for a poorly-selective sensor. Moreover, the elbow of the calibration curve highlighting the change of sensor regime from flat OCP response (not detecting) to a Nernstian behavior is extrapolated as the sensor lower LOD [46]. It is shown that sensor lower LOD increases, the more the severity of ion interference is high. This is critical if the lower LOD reaches the physiological/therapeutic range of the target analyte. In practice a margin of 1 : 10 is needed between the sensor lower LOD and the lowest activity of the primary ion to be measured.

Next, the calibration curves generated by the emulator are compared with OCP responses acquired from batch of polymeric solid-contact ISEs in artificial sweat [33]. The results are displayed in Fig. 5, where the measured calibration curves exhibit inter-sensor variability, mainly for Li^+ and Pb^{2+} sensors. The simulated OCP response is computed by taking the same electrolyte background composition as the artificial sweat, and with the selectivity coefficients of the corresponding polymeric solid contact ISEs against the interfering ions. The offset parameter for each primary ion is fitted to the in-vitro calibration data. Then, the RMSE between the modeled and the experimental data is computed, and yields, 1.37, 1.44, 1.78, $2\,mV$, for Na^+, K^+, Li^+, Pb^{2+} sensors, respectively. As a result, the ion-sensing model explains accurately polymeric ISE responses subject to ion interference from artificial sweat background electrolyte.
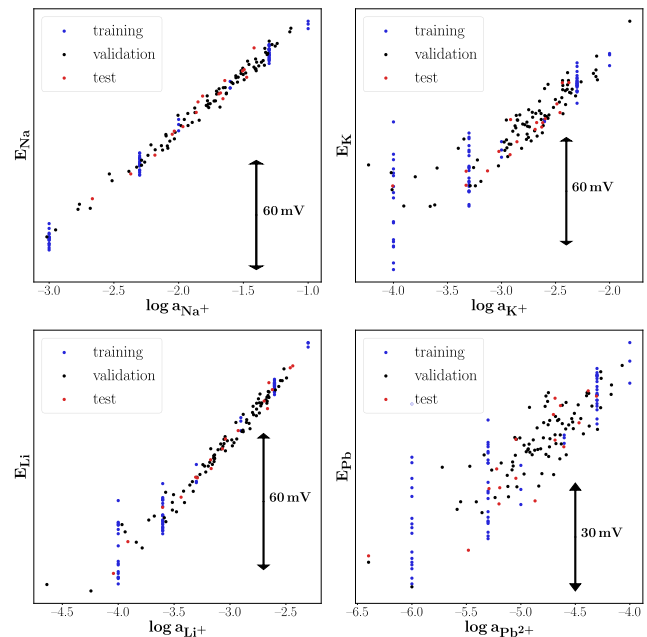
TABLE 6. Prediction accuracy obtained by applying linear regressors to datasets of different size.

Dataset	$\log a_{Na^+}$		$\log a_{K^+}$		$\log a_{Li^+}$		$\log a_{Pb^{2+}}$		Total	
	NRMSE	MRE	NRMSE	MRE	NRMSE	MRE	NRMSE	MRE	**NRMSE**	**MRE**
Dataset_1	3.64	3.04	9.72	8.19	3.11	2.76	6.68	5.82	**23.15**	**19.81**
Dataset_2	4.32	3.61	8.22	7.23	2.84	2.61	6.33	4.96	**21.71**	**18.41**
Dataset_3	3.76	3.40	8.74	7.15	3.00	2.52	5.44	4.36	**20.94**	**17.43**
Dataset_4	3.72	3.07	7.33	6.56	3.18	2.47	5.66	4.71	**19.88**	**16.81**
Dataset_5	3.17	2.59	8.92	7.85	2.60	2.21	5.17	3.85	**19.86**	**16.49**
Dataset_6	3.02	2.47	8.24	6.81	2.72	2.28	5.31	4.11	**19.29**	**15.68**
Dataset_7	3.55	2.99	7.21	5.95	3.68	2.20	4.98	4.10	**19.41**	**15.23**
Dataset_8	3.51	2.94	8.21	5.31	2.21	1.80	5.08	4.11	**19.02**	**14.16**
Dataset_9	3.14	2.63	6.98	5.15	3.36	1.98	5.04	4.11	**18.52**	**13.87**
Dataset_10	3.35	2.87	7.16	5.30	3.42	1.77	4.13	3.46	**18.06**	**13.40**
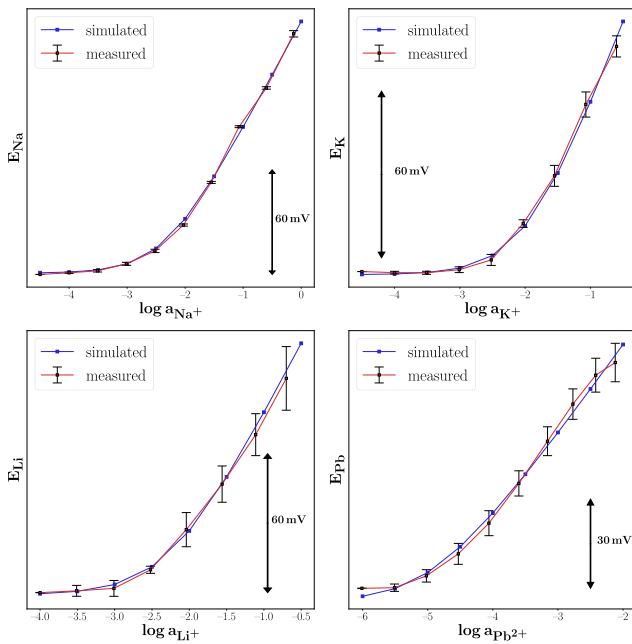
FIGURE 5. Comparison between the calibration curves output by the ion-sensing emulator and the calibration curves acquired from batch of polymeric solid-contact ISEs in artificial sweat. RMSE of 1.37, 1.44, 1.78, 2 *mV* are obtained for Na^+, K^+, Li^+, Pb^{2+} sensors, respectively.

FIGURE 6. Dataset_1: $N_{train} = 68$, $N_{val, split1} = 17$, $N_{test} = 17$.

B. GENERATION OF SYNTHETIC DATASETS

The multi-ion-sensing emulator has been demonstrated to model well ISE responses in nominal sweat background electrolyte. Thus, in this section, the tool is used to generate synthetic datasets following experimental design procedures described in Section IV. Ten datasets Dataset_n of training set size $N_{train} = 68 \cdot n$ ($n \in [\![1; 10]\!]$) are generated. The scarcest dataset is displayed in Fig. 6. The six levels of the training set are clearly visible since it has been generated through a factorial design. The composition of the validation and test sets is randomly designed in the same activity range as the training set, resulting in samples intermingled in between the factor levels. Emulated Na^+-ISE calibration data exhibits a Nernstian response since the physiological range in sweat is large (above tens of mM). Conversely, Pb^{2+} is detected as metal traces (tens of μM), so it is diluted in the sample. The emulated potential response shows more

potential dispersion. As for Li^+-ISE, the sensor has a good selectivity towards the constituents (see Table 4), so it is less affected by ion interference. It is not the case of K^+-ISE that is poorly selective, mainly towards Na^+ that is a predominant ion in the samples. This results in major OCP dispersion, accentuated at lower a_{K^+}. The same results are observed with larger datasets. The apparent linear correlation between E_X and $\log a_X$ that is observed in Fig. 6 suggests to compute the correlation coefficients between the two variables, for each ion, for the training set of the ten datasets. The results are displayed in Fig. 7. The lower correlation in $K+$ and Pb^{2+} channels is due to ion interference that is explained by the larger potential dispersion at lower ion activity. The correlation between sensor OCP and ion activity tends to increase with larger datasets.

C. LINEAR REGRESSION MODELS

Prior to multivariate calibration, PCA is applied to the input tensor **X**. Fig. 8 shows the scatter plots of the PC scores.
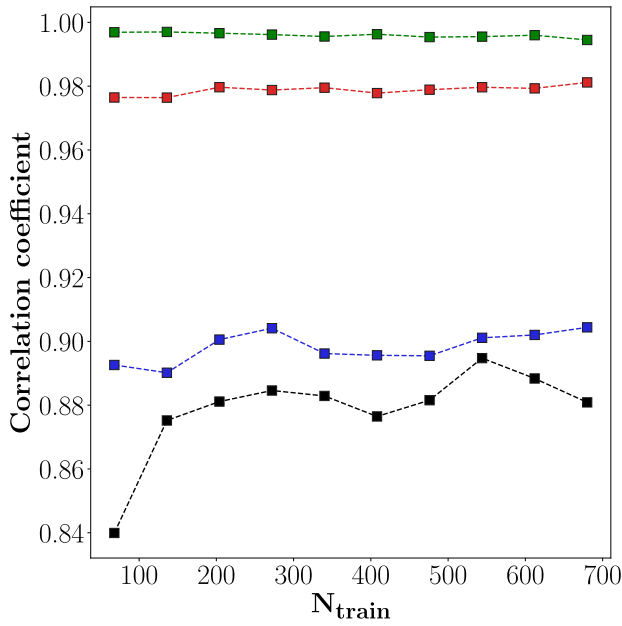
FIGURE 7. Pearson correlation coefficient between E_X and $\log a_X$ for Na^+ (green), K^+ (blue), Li^+ (red), and Pb^{2+} (black) channels.
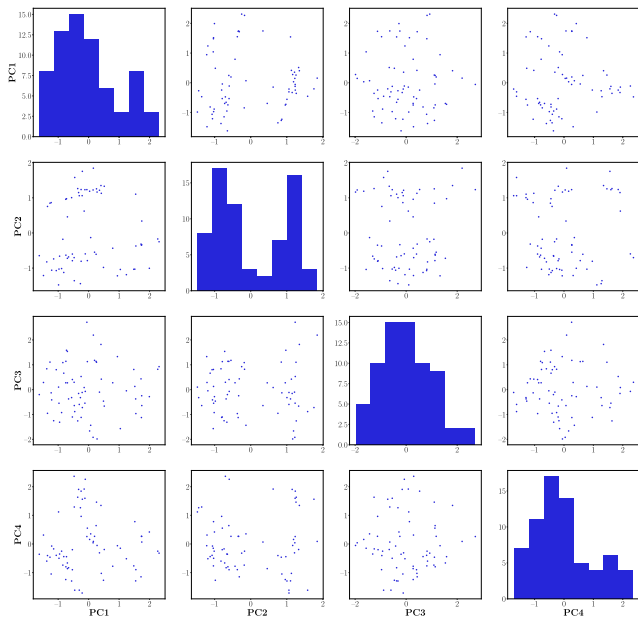

FIGURE 8. Scatter matrices of PCA scores for Dataset_1.

The whitened PC scores are uncorrelated and do not present higher order dependence. Moreover, the explained variance per PC yields 34.23%, 24.51%, 21.39%, and 19.87%, for $PC1$ to $PC4$, respectively. This indicates that all the PCs are needed to explain the variance present in the input tensor. It is important to underline that PCA is employed as a pre-processing step, and not as a dimensionality reduction technique ($P = 4$). Moreover, two metrics are introduced to assess the accuracy of the regressors. The Normalized Root Mean-Squared Error (NRMSE) and the Mean Relative

Error (MRE) are defined as

$$NRMSE = \frac{100}{\bar{\bar{y}}} \sqrt{\frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \left(\tilde{y}_n - \hat{\tilde{y}}_n \right)^2}, \quad (17)$$

$$MRE = \frac{100}{N_{test}} \sum_{n=1}^{N_{test}} \frac{|\tilde{y}_n - \hat{\tilde{y}}_n|}{\tilde{y}_n}, \quad (18)$$

where $\bar{\bar{y}}$ is the mean of the test set labels, $\tilde{y}_n$ and $\hat{\tilde{y}}_n$ are the true and predicted log-activity of the primary ions, respectively. Normalized metrics are required to balance error contributions from analytes highly concentrated (sodium ions), and diluted analytes in the sample (lead ions). In addition, the two metrics do not over-penalize outliers. NRMSE and MRE are computed for each primary ion, but total NRMSE and MRE of all four primary ions yield two compact model performance measure that are used during training.

An analysis is carried out by applying MLR to $\tilde{\mathbf{X}} = \{\{PC1\}, \{PC1, PC2\}, \{PC1, PC2, PC3\}, \{PC1, PC2, PC3, PC4\}\}$. The regressors are fitted to the training set and evaluated on the validation splits. The total NRMSE decreases with the dimension of the latent space. Likewise, PLS algorithm is applied to the input tensor $\mathbf{X}$, and the same factorial analysis is performed, suggesting to use the four-dimension latent space. Next, the fitted models are evaluated on the external test set. It results that factorial decomposition with PCR and PLS yield identical multivariate calibration accuracy when the full latent space is used. When a reduced latent space is used, PLS performs slightly better, as expected, since it is a more compact model than PCR. The metrics achieved by applying linear regressors on the ten datasets are reported in Table 6. It could be seen that K^+ and Pb^{2+} calibration are the less accurate ones due to the non-linear distortion added by severe ion interference. Besides, the metrics are improving steadily with larger datasets, for the four target ions. The metrics obtained with linear regression models are used as benchmark for the subsequent non-linear multivariate regressors implemented.

D. MULTIVARIATE CALIBRATION OPTIMIZATION

In this section, the non-linear multivariate calibration models are compared to simple MLR model. Namely, SVR, the proposed M-SVR, and MLP models are evaluated on the ten synthetic datasets. A cross-validation grid-search on the model hyper-parameters is carried out to identify the best models, while limiting over-fitting. More explicitly, the models are trained on the entire orthogonal training set, and evaluated on five validation splits. The best model is retained and evaluated on the external test set. For MLP models, the hyper-parameters space being wide, a coarse search is done before refining the grid-search on relevant hyper-parameters space. The results of multivariate calibration are displayed in Fig. 9.

We observe a clear improvement between linear regressors and non-linear ones. For total NRMSE, there is an average improvement of 16.27%, 13.22%, and 18.12%, and an improvement of total MRE of 22.23%, 20.29%, and
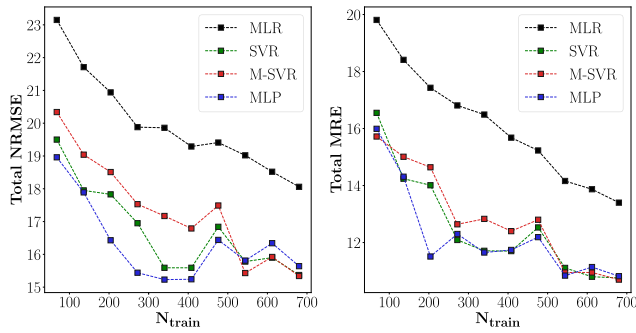
FIGURE 9. Total NRMSE and total MRE obtained by evaluating the best multivariate calibration models on the test set of the ten synthetic datasets.
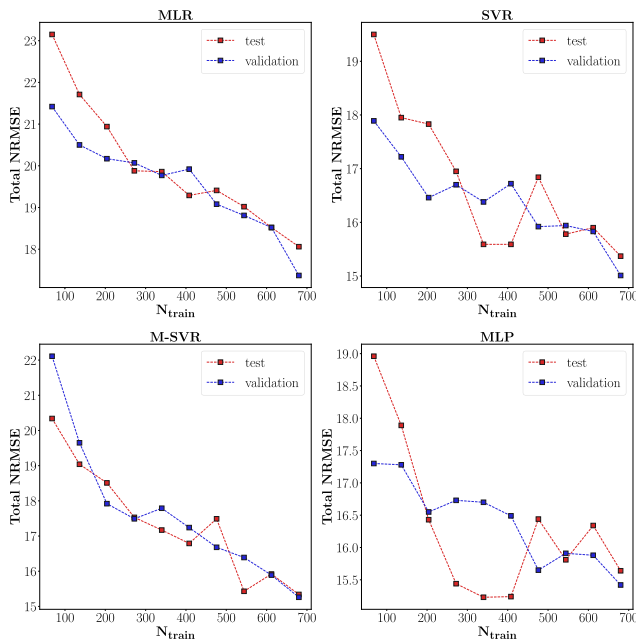

FIGURE 10. Total NRMSE obtained while assessing multivariate calibration models on five random validation splits and on a test set.

FIGURE 11. Total NRMSE obtained while applying PCA pre-processing or not before multivariate calibration.

23.92%, for SVR, M-SVR, and MLP, respectively. The accuracy improvement obtained with the proposed M-SVR is satisfactory given the lower complexity of the proposed model. Namely, M-SVR optimizes concurrently the calibration of the four channels with four-dimension support vectors. Conversely, single-output SVR optimizes independently one ion channel at a time with one-dimension support vectors, therefore four times more support vectors than with M-SVR. Besides, M-SVR is more robust to noise and correlation in the dataset since it considers the four ion channels when constructing the model regression hyperplane. It is also observed that with large datasets (more than 540 training samples), M-SVR performs better than SVR and MLP regressors.

Next, the quality of the models is assessed by comparing the scores obtained during the validation stage (metrics calculated with five validation splits of random samples) and the scores obtained with the test set. The results are displayed in Fig. 10. We observe that the generalization error is large
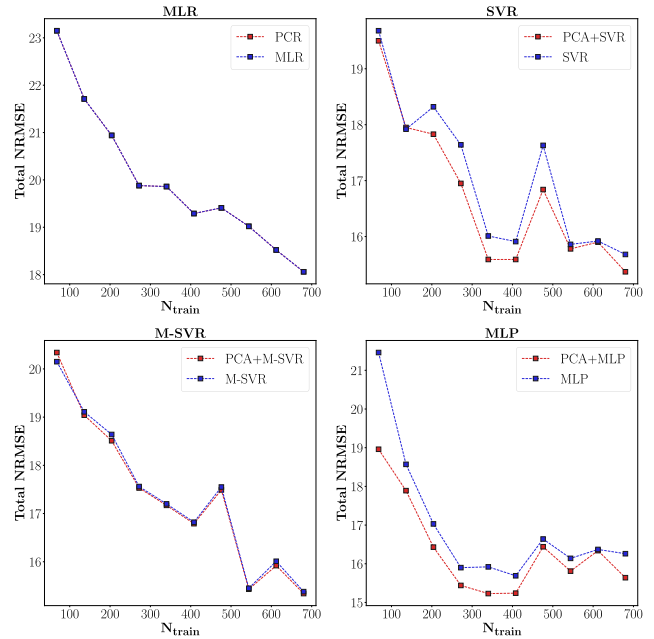
for the scarcest dataset, for which the models are more likely subject to over-fitting. Overall, the generalization error is the smallest for M-SVR with an average of 3.22%, while it is of 4.43% and 4.79%, for SVR and MLP models, respectively.

A further analysis is performed on the importance of PCA pre-processing. The scores obtained while applying PCA before the multivariate calibration, and without applying PCA are reported in Fig. 11. It is added that the input tensor **X** is scaled to zero-mean and unit-variance in both cases.

We observe that PCA does not influence the calibration accuracy for MLR and M-SVR, while improvement is observed with SVR and MLP models. MLP regressors are feature-based models, so the pre-processing of the input tensor is essential to feed uncorrelated data to the network model. SVR and M-SVR models are distance-based models, so scaling the input tensor is the essential pre-processing step.

The predicted ion activity vs target values when applying the optimal M-SVR model on Dataset_1 and Dataset_10 are plotted in Fig. 12 and 13, respectively. The scatter points fall along the 1:1 lines, with more dispersion for K^+ and Pb^{2+} ions that are subject to severe ion interference. For Dataset_1, the 95% confidence interval on the slopes of the model fit contains 1 for Li^+ and Pb^{2+} channels, and the intercepts contain 0 except for Pb^{2+}. These results are consistent with the previous conclusion that M-SVR models are less accurate with a scarcer dataset since the model is slightly biased. As for Dataset_10, the 95% confidence interval on the slopes of the model fit contains 1, and the intercepts contain 0 for the four channels. Therefore, M-SVR models applied to larger datasets are accurate and unbiased estimators. The computed coefficients of determination provide reliable information

TABLE 7. Training parameters of the multivariate calibration models.

| N_{train}	MLR # trainable parameters	SVR # support vectors	M-SVR # support vectors	# IRWLS iterations	MLP # trainable parameters	# training epochs
68	16	251	68	5	1329	96
136	16	543	125	5	3004	365
204	16	814	204	12	1604	187
272	16	1024	272	9	2179	235
340	16	1360	340	4	2579	177
408	16	1632	408	10	1904	221
476	16	1904	476	17	1229	192
544	16	2176	544	5	1104	203
612	16	2446	612	17	3179	223
680	16	2720	680	4	3179	209
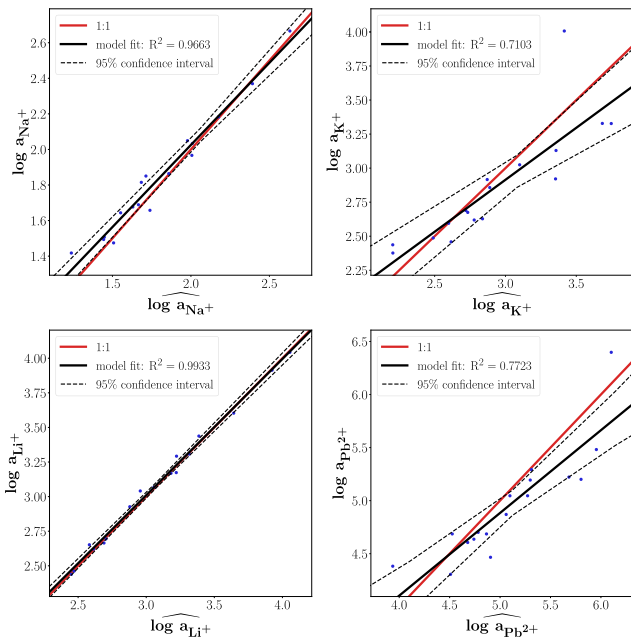
FIGURE 12. Scatter plot of predicted $\widehat{\log a_X}$ vs target $\log a_X$ for Na^+, K^+, Li^+, and Pb^{2+} channel, by applying the optimal M-SVR model on Dataset_1. The model fit and its 95% confidence interval are plotted, where R^2 is the coefficient of determination. The red plot is the 1:1 line.
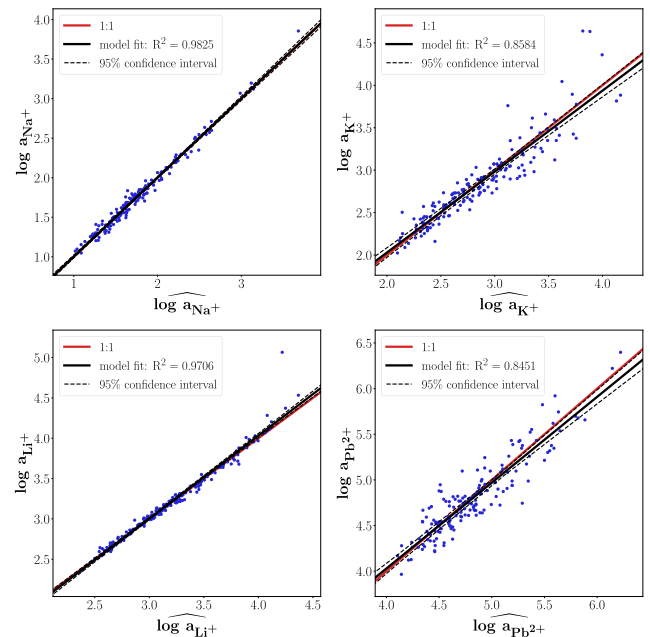
FIGURE 13. Scatter plot of predicted $\widehat{\log a_X}$ vs target $\log a_X$ for Na^+, K^+, Li^+, and Pb^{2+} channel, by applying the optimal M-SVR model on Dataset_10. The model fit and its 95% confidence interval are plotted, where R^2 is the coefficient of determination. The red plot is the 1:1 line.

about the variance in activity of the target ions explained by the multivariate calibration models.

E. ANALYSIS OF MODEL COMPLEXITY

Model complexity is a paramount feature to consider prior to deploying the multivariate calibration models onto memory and energy-constrained edge nodes, for real-time ion prediction. The trainable parameters of the multivariate calibration models are reported in Table 7. MLR model has $M \cdot P = 16$ trainable weights, independently of the size of the training set, and OLS minimization has a complexity of $\mathcal{O}(NM^2)$. The complexity of SVR models is indicated by the number of support vectors needed to build the regression hyperplane. More accurate models are obtained with small ϵ, so most of the training instances are used as support vectors. However, M-SVR has four-dimension support vectors for each training

instance since the four ion channels are considered simultaneously. On the contrary, four independent models are constructed with single-output SVR, explaining the four times larger amount of one-dimension support vectors. Besides, the complexity of IRWLS procedure is equivalent to an OLS minimization per iteration. The number of iterations required to reach convergence is reasonable, being less than 20. As for MLP models, the complexity of the model is set by the number of neurons in the model. The amount of wights and bias to estimate during the training phase is quite large, and tends to grow with larger training sets. The number of epochs before early-stopping is also reported. It does not exceed 400 epochs for a training batch size of 32 samples.

The training runtime of the different models is measured to compare the model complexity quantitatively. The models are trained 100 times with the best hyper-parameters found
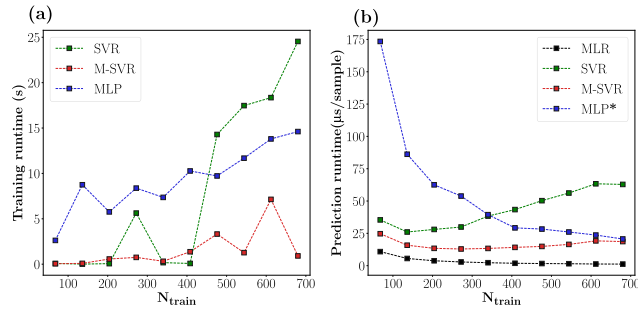
FIGURE 14. **(a) Training runtime and (b) prediction latency of multivariate calibration models. * values divided by 1.**

during grid-search. The results are displayed in Fig. 14.a. The training runtime of MLR models is not reported since it is constantly between 300 and 450 μs. We observe that the training runtime of SVR model is relatively fast for scarce datasets, but it increases geometrically with the number of training samples beyond 400 samples. Conversely, the training runtime of M-SVR models does not grow excessively and remains below 8 s. As for MLP models, they are the longer to train for scarce datasets, and the runtime grows steadily when increasing the training set. The multivariate calibration models could be trained offline and we could deploy the trained models directly onto edge devices for real-time ion monitoring. Consequently, the prediction latency of each model is measured for 100 runs, and normalized by the number of samples constituting the test set. The results are reported in Fig. 14.b. The prediction latency is obviously the shortest for MLR models. M-SVR models are compact models, so the prediction latency is inferior to 20 $\mu s/sample$. Conversely, single-output SVR embed M dense models, explaining the larger latency. Neural network models have the larger prediction latency (scale divided by ten to fit onto the graph), making these models the less suitable for real-time monitoring applications.

VII. CONCLUSION

This work presents a novel approach for multi-ion-sensing frameworks, where the accuracy of the multivariate calibration depends on the training dataset representativeness and size. Namely, the response of ion-sensing arrays subject to ion interference are explained by an ion-sensing emulator. The parameters of the tool are experimentally determined from a couple of in-vitro calibration curves acquired from polymeric solid-contact ISEs, and from the selectivity coefficients of the developed sensors. It is demonstrated that the calibration curves output by the emulator fit with the calibration data acquired in artificial sweat from Na^+, K^+, Li^+, Pb^{2+} sensors, with RMSE of 1.37, 1.44, 1.78, 2 mV, respectively. This legitimates the use of the tool for generating synthetic datasets of custom size, emulating multi-ion-sensing in sweat. Therefore, the proposed approach circumvents the acquisition of big calibration data that is highly expensive in terms of chemical resource and time.

Then, the emulated datasets are fed to a multivariate and multi-output SVR model. Multivariate calibration accuracy is increased, with NRMSE improvements of 13.22%, and MRE improvements of 20.29% with respect to a simple MLR model. The generalization error is quite small, being of 3.22%, and the regressor is statistically unbiased for medium to large datasets. The proposed model is more compact and more robust than a single-output SVR that embeds an independent regressor for each ion channel to predict. The latter does not consider the correlations between output channels, and above all, its computational complexity increases geometrically with the training set size, for large datasets. The proposed M-SVR model represents a robust, accurate, and low-complexity solution for memory and energy-constrained embedded devices, paving the way for continuous and real-time multi-ion monitoring.

Eventually, the multi-ion-sensing emulator and multivariate calibration models are designed in this work for biomedical and healthcare monitoring in sweat. But the flexibility of the tool enables its use for broader applications such as environmental or water quality monitoring.

APPENDIX A
## PHASE-BOUNDARY POTENTIAL MODEL

At equilibrium, the electrochemical potentials for any ion j, of valence z_j, in the sample and membrane phase are

$$\begin{cases} \mu_j(aq) = \mu_j^0(aq) + RTln(a_j(aq)) + z_jF\,\Phi(aq) \\ \mu_j(m) = \mu_j^0(m) + RTln(a_j(m)) + z_jF\,\Phi(m), \end{cases} \quad (A.1)$$

respectively, where μ_j^0 is the standard chemical potential, a_j is the ion activity, and Φ is the electrical potential. Assuming partition equilibrium of ion j between both membrane and sample phases, $\mu_j(aq) = \mu_j(m)$, one obtains the phase-boundary potential defined as $E_{PB} \overset{def}{=} \Phi(m) - \Phi(aq)$,

$$E_{PB} = \epsilon_j^0 + \frac{1}{z_j}s\log\left(\frac{a_j(aq)}{a_j(m)}\right), \quad (A.2)$$

where $\epsilon_j^0 = \frac{\mu_j^0(aq)-\mu_j^0(m)}{z_jF}$ is the standard potential, constant for a given ion. $s = \frac{RT}{F}ln\,10$ is the Nernst slope. Then, let us consider the ion-selective membrane composed of an ionophore L of charge z_L and concentration L_T, and an ion-exchanger R of charge z_R and molar concentration R_T. The ionophore includes ion carriers that selectively transport ions across the membrane. The ionophore forms complexes with the target ion I as

$$I^{z_j}(m) + nL \xrightarrow{\beta_n} jL_n, \quad (A.3)$$
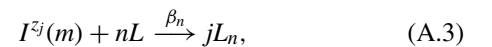
where β_n is the complex formation constant. The interactions between the ion-exchanger and the co-ions are neglected. The charge balance in the ion-selective membrane yields

$$z_RR_T + z_LL_T + \sum_n (nz_L + z_j)c_{jL_n} + z_jc_j(m) = 0, \quad (A.4)$$

where each contributions are, from the ion-exchanger, the ionophore, the complex formed with the target ion, and

the free form of the target ion in the membrane, respectively. c stands for compound concentration. The label (m) is added to emphasize that it refers to the concentration of ion j in the membrane and not in the sample phase. Inserting complex formation constants β_n and activity coefficients γ,

$$z_R \, R_T + z_L \frac{a_L}{\gamma_L} + z_j \frac{a_{j(m)}}{\gamma_{j(m)}} \left(1 + \sum_n \frac{nz_L + z_j}{z_j} \frac{\gamma_{j(m)}}{\gamma_{jL_n}} \beta_n a_L^n \right) = 0. \tag{A.5}$$

Next, let us consider that multiple ions from the sample phase (interfering ions) could interact with the membrane. It is assumed that the activity of the uncomplexed ionophore and the activity coefficient of the extracted ions remain unaltered. This is valid for a polymeric membrane with an ionophore in excess with respect to the ion-exchanger, and considering that the membrane ionic strength is not changing during ion exchange. (A.5) is summed over the ions extracted from the sample phase into the membrane, and yields

$$\sum_j z_j \frac{a_{j(m)}}{\gamma_{j(m)}} \left(1 + \sum_n \frac{nz_L + z_j}{z_j} \frac{\gamma_{j(m)}}{\gamma_{jL_n}} \beta_n a_L^n \right) = -z_R R_T - z_L \frac{a_L}{\gamma_L}. \tag{A.6}$$

Then, an apparent standard potential for each ion j, E_j^0, is introduced such that $E_{PB} = E_j^0 + \frac{s}{z_j} \log(a_{j(aq)})$. The apparent standard potential is expressed as

$$E_j^0 = \epsilon_j^0 + \frac{s}{z_j} \log \left(-\frac{z_j}{\gamma_{j(m)}} \frac{1 + \sum_n \frac{nz_L + z_j}{z_j} \frac{\gamma_{j(m)}}{\gamma_{jL_n}} \beta_n a_L^n}{z_R R_T + z_L \frac{a_L}{\gamma_L}} \right), \tag{A.7}$$

where the expression in the log is equal to $\frac{1}{a_{j(m)}}$. (A.7) is rewritten as

$$z_j \frac{k_j}{\gamma_{j(m)}} \left(1 + \sum_n \frac{nz_L + z_j}{z_j} \frac{\gamma_{j(m)}}{\gamma_{jL_n}} \beta_n a_L^n \right) 10^{-z_j E_j^0 / s}$$
$$= -z_R R_T - z_L \frac{a_L}{\gamma_L}, \tag{A.8}$$

where the free energy of ion transfer k_j is defined as $\epsilon_j^0 = \frac{s}{z_j} \log(k_j)$. Moreover, $a_{j(m)}$ is extracted from (A.2) and substituted in (A.6), yielding

$$\sum_j \frac{z_j}{\gamma_{j(m)}} k_j \left(1 + \sum_n \frac{nz_L + z_j}{z_j} \frac{\gamma_{j(m)}}{\gamma_{jL_n}} \beta_n a_L^n \right) 10^{-z_j E_{PB}/s} a_{j(aq)}$$
$$= -z_R R_T - z_L \frac{a_L}{\gamma_L}. \tag{A.9}$$

Inserting (A.8) in (A.9), one finds

$$10^{-z_j E_{PB}/s} \sum_j a_{j(aq)} 10^{z_j E_j^0 / s} = 1. \tag{A.10}$$

The selectivity coefficient of a membrane with target ion I, with respect to an interfering ion J, is defined as

$$\log K_{I,J}^{pot} = \frac{z_I}{s}(E_J^0 - E_I^0). \tag{A.11}$$

Equation (A.11) is substituted in (A.10), and gives the compact equation

$$10^{z_I(E_I^0 - E_{PB})/s} a_I(aq) + 10^{(E_I^0 - E_{PB})/s} \sum_{j \, s.t. \, z_j = 1} (K_{I,j}^{pot})^{\frac{1}{z_I}} a_j(aq)$$
$$+ 10^{2(E_I^0 - E_{PB})/s} \sum_{j \, s.t. \, z_j = 2} (K_{I,j}^{pot})^{\frac{2}{z_I}} a_j(aq) = 1, \tag{A.12}$$

limiting the sum expansion to monovalent and divalent ions. Equation (A.12) is solved for E_{PB}, with a second-order polynomial equation formulation. For $z_I = 1$,

$$E_{PB}$$
$$= E_I^0 + s \log \left(\frac{1}{2} \left(a_I(aq) + \sum_{j \, s.t. \, z_j = 1} K_{I,j}^{pot} a_j(aq) \right) \right.$$
$$\left. + \sqrt{\frac{1}{4} \left(a_I(aq) + \sum_{j \, s.t. \, z_j = 1} K_{I,j}^{pot} a_j(aq) \right)^2 + \sum_{j \, s.t. \, z_j = 2} (K_{I,j}^{pot})^2 a_j(aq)} \right), \tag{A.13}$$

while for $z_I = 2$,

$$E_{PB}$$
$$= E_I^0 + s \log \left(\frac{1}{2} \sum_{j \, s.t. \, z_j = 1} (K_{I,j}^{pot})^{\frac{1}{2}} a_j(aq) \right.$$
$$\left. + \sqrt{\frac{1}{4} \left(\sum_{j \, s.t. \, z_j = 1} (K_{I,j}^{pot})^{\frac{1}{2}} a_j(aq) \right)^2 + a_I(aq) + \sum_{j \, s.t. \, z_j = 2} K_{I,j}^{pot} a_j(aq)} \right). \tag{A.14}$$

APPENDIX B
SYNTHETIC DATASETS
The synthetic datasets generated are available at http://ieee-dataport.org/documents/multi-ion-sensing-emulator.

REFERENCES
[1] Ö. Isildak and O. Özbek, "Application of potentiometric sensors in real samples," *Crit. Rev. Anal. Chem.*, pp. 1–14, Jan. 2020, doi: 10.1080/10408347.2019.1711013.

[2] A. Mimendia, J. M. Gutiérrez, L. Leija, P. R. Hernández, L. Favari, R. Muñoz, and M. del Valle, "A review of the use of the potentiometric electronic tongue in the monitoring of environmental systems," *Environ. Model. Softw.*, vol. 25, no. 9, pp. 1023–1030, Sep. 2010.

[3] A. V. Mueller and H. F. Hemond, "Statistical generation of training sets for measuring NO$_3$-, NH$_4$+ and major ions in natural waters using an ion selective electrode array," *Environ. Sci., Processes Impacts*, vol. 18, no. 5, pp. 590–599, 2016.

[4] W. Gao, S. Emaminejad, H. Y. Y. Nyein, S. Challa, K. Chen, A. Peck, H. M. Fahad, H. Ota, H. Shiraki, D. Kiriya, D.-H. Lien, G. A. Brooks, R. W. Davis, and A. Javey, "Fully integrated wearable sensor arrays for multiplexed *in situ* perspiration analysis," *Nature*, vol. 529, no. 7587, pp. 509–514, Jan. 2016.

[5] M. Parrilla, M. Cuartero, and G. A. Crespo, "Wearable potentiometric ion sensors," *TrAC Trends Anal. Chem.*, vol. 110, pp. 303–320, Jan. 2019.

[6] J. Hu, A. Stein, and P. Bühlmann, "Rational design of all-solid-state ion-selective electrodes and reference electrodes," *TrAC Trends Anal. Chem.*, vol. 76, pp. 102–114, Feb. 2016.

[7] G. Dimeski, T. Badrick, and A. S. John, "Ion selective electrodes (ISEs) and interferences—A review," *Clinica Chim. Acta*, vol. 411, nos. 5–6, pp. 309–317, Mar. 2010.

[8] M. del Valle, "Sensor arrays and electronic tongue systems," *Int. J. Electrochem.*, vol. 2012, pp. 1–11, Oct. 2012.

[9] Y. Vlasov, A. Legin, A. Rudnitskaya, C. Di Natale, and A. D'Amico, "Nonspecific sensor arrays ('electronic tongue') for chemical analysis of liquids (IUPAC technical report)," *Pure Appl. Chem.*, vol. 77, no. 11, pp. 1965–1983, 2005.

[10] L. Wang, D. Yang, C. Fang, Z. Chen, P. J. Lesniewski, M. Mallavarapu, and R. Naidu, "Application of neural networks with novel independent component analysis methodologies to a prussian blue modified glassy carbon electrode array," *Talanta*, vol. 131, pp. 395–403, Jan. 2015.

[11] L. Wang, Y. Cheng, D. Lamb, M. Megharaj, and R. Naidu, "Application of ion selective electrode array to simultaneously determinate multi-free ions in solution," *Environ. Technol. Innov.*, vol. 15, Aug. 2019, Art. no. 100424.

[12] A. V. Mueller and H. F. Hemond, "Extended artificial neural networks: Incorporation of a priori chemical knowledge enables use of ion selective electrodes for *in-situ* measurement of ions at environmentally relevant levels," *Talanta*, vol. 117, pp. 112–118, Dec. 2013.

[13] L. Wang, Y. Cheng, D. Lamb, Z. Chen, P. J. Lesniewski, M. Megharaj, and R. Naidu, "Simultaneously determining multi-metal ions using an ion selective electrode array system," *Environ. Technol. Innov.*, vol. 6, pp. 165–176, Nov. 2016.

[14] A. Mimendia, J. M. Gutiérrez, J. M. Alcañiz, and M. del Valle, "Discrimination of soils and assessment of soil fertility using information from an ion selective electrodes array and artificial neural networks," *CLEAN Soil, Air, Water*, vol. 42, no. 12, pp. 1808–1815, Dec. 2014.

[15] L. Nuñez, X. Cetó, M. I. Pividori, M. V. B. Zanoni, and M. del Valle, "Development and application of an electronic tongue for detection and monitoring of nitrate, nitrite and ammonium levels in waters," *Microchem. J.*, vol. 110, pp. 273–279, Sep. 2013.

[16] J. Gallardo, S. Alegret, M. A. de Román, R. Muñoz, P. R. Hernández, L. Leija, and M. del Valle, "Determination of ammonium ion employing an electronic tongue based on potentiometric sensors," *Anal. Lett.*, vol. 36, no. 14, pp. 2893–2908, Jan. 2003.

[17] D. Wilson, M. N. Abbas, A. L. A. Radwan, and M. D. Valle, "Potentiometric electronic tongue to resolve mixtures of sulfide and perchlorate anions," *Sensors*, vol. 11, no. 3, pp. 3214–3226, Mar. 2011.

[18] A. M. Snauffer, U. Chauhan, K. Cogert, M. K. H. Winkler, and A. V. Mueller, "Data fusion for environmental process control: Maximizing useful information recovery under data limited constraints," *IEEE Sensors Lett.*, vol. 3, no. 1, pp. 1–4, Jan. 2019.

[19] E. Bakker, P. Bühlmann, and E. Pretsch, "The phase-boundary potential model," *Talanta*, vol. 62, no. 4, pp. 843–860, 2004.

[20] M. Chung, G. Fortunato, and N. Radasci, "Wearable flexible sweat sensors for healthcare monitoring: A review," *J. Roy. Soc. Interface*, vol. 16, pp. 1–15, Oct. 2018.

[21] B. Lara, C. Gallo-Salazar, C. Puente, F. Areces, J. J. Salinero, and J. Del Coso, "Interindividual variability in sweat electrolyte concentration in marathoners," *J. Int. Soc. Sports Nutrition*, vol. 13, no. 1, pp. 1–8, Dec. 2016.

[22] D. B. Speedy, T. D. Noakes, and C. Schneider, "Exercise-associated hyponatremia: A review," *Emergency Med.*, vol. 13, no. 1, pp. 17–27, Mar. 2001.

[23] A. Muneer, "Staging models in bipolar disorder: A systematic review of the literature," *Clin. Psychopharmacol. Neurosci.*, vol. 14, no. 2, pp. 117–130, May 2016.

[24] D. Gruson, A. Lallali, V. Furlan, A. M. Taburet, A. Legrand, and M. Conti, "Evaluation of a new lithium colorimetric assay performed on the dade behring dimension X-pand system," *Clin. Chem. Lab. Med.*, vol. 42, no. 9, pp. 1066–1068, Jan. 2004.

[25] M. E. Sears, K. J. Kerr, and R. I. Bray, "Arsenic, cadmium, lead, and mercury in sweat: A systematic review," *J. Environ. Public Health*, vol. 2012, pp. 1–10, Feb. 2012.

[26] L. Z. Benet, "Glass electrodes for hydrogen and other cations. Edited by george Eisenman. Marcel Dekker, Inc., 95 Madison Ave., New York, NY 10016, 1967. XII + 582pp. 15 × 22.5 cm. Price $24.75," *J. Pharmaceutical Sci.*, vol. 57, no. 7, p. 1271, Jul. 1968.

[27] M. Nägele, E. Bakker, and E. Pretsch, "General description of the simultaneous response of potentiometric ionophore-based sensors to ions of different charge," *Anal. Chem.*, vol. 71, no. 5, pp. 1041–1048, Mar. 1999.

[28] I. N. Hanitra, D. Demarchi, S. Carrara, and G. De Micheli, "Emulator design and generation of synthetic dataset in multi-ion sensing," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Oct. 2020, pp. 1–5.

[29] W. Ament, J. Huizenga, G. Mook, C. Gips, and C. J. Verkerke, "Lactate and ammonia concentration in blood and sweat during incremental cycle ergometer exercise," *Int. J. Sports Med.*, vol. 18, no. 1, pp. 35–39, Jan. 1997.

[30] R. Klesges, K. Ward, M. Shelton, and, "Changes in bone mineral content in male athletes: Mechanisms of action and intervention effects," *J. Amer. Med. Assoc.*, vol. 276, no. 3, pp. 226–230, 1996.

[31] M. Bariya, H. Y. Y. Nyein, and A. Javey, "Wearable sweat sensors," *Nature Electron.*, vol. 1, no. 3, pp. 160–171, Mar. 2018.

[32] R. N. Kacker, E. S. Lagergren, and J. J. Filliben, "Taguchi's orthogonal arrays are classical designs of experiments," *J. Res. Nat. Inst. Standards Technol.*, vol. 96, no. 5, p. 577, Sep. 1991.

[33] F. Criscuolo, I. Ny Hanitra, S. Aiassa, I. Taurino, N. Oliva, S. Carrara, and G. De Micheli, "Wearable multifunctional sweat-sensing system for efficient healthcare monitoring," *Sens. Actuators B, Chem.*, vol. 328, Feb. 2021, Art. no. 129017.

[34] A. C. Olivieri, *Introduction to Multivariate Calibration*. Cham, Switzerland: Springer, 2018.

[35] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.

[36] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation and signal processing," in *Proc. 9th Int. Conf. Neural Inf. Process. Syst.*, Cambridge, MA, USA: MIT Press, 1996, pp. 281–287.

[37] V. N. Vapnik, *Statistical Learning Theory*. Hoboken, NJ, USA: Wiley, 1998.

[38] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.

[39] M. Sanchez-Fernandez, M. de-Prado-Cumplido, J. Arenas-Garcia, and F. Perez-Cruz, "SVM multiregression for nonlinear channel estimation in multiple-input multiple-output systems," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2298–2307, Aug. 2004.

[40] D. E. Rumelhart and J. L. McClelland, Eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, vol. 1. Cambridge, MA, USA: MIT Press, 1986.

[41] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2010, pp. 249–256.

[42] J. J. Dongarra, J. W. Demmel, and S. Ostrouchov, "Lapack: A linear algebra library for high-performance computers," in *Computational Statistics*, Y. Dodge and J. Whittaker, Eds. Berlin, Germany: Physica-Verlag HD, 1992, pp. 23–28.

[43] P. Geladi and B. R. Kowalski, "Partial least-squares regression: A tutorial," *Analytica Chim. Acta*, vol. 185, pp. 1–17, Jan. 1986.

[44] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, Apr. 2011.

[45] F. Chollet. (2015). *Keras*. [Online]. Available: https://keras.io

[46] R. P. Buck and E. Lindner, "Recomendations for nomenclalture of ion-selective electrodes," *Pure Appl. Chem.*, vol. 66, no. 12, pp. 2527–2536, 1994.

IVAN NY HANITRA (Graduate Student Member, IEEE) received the B.S. degree in physics and chemistry from the Grenoble Institute of Technology, Grenoble, France, in 2014, and the M.S. degree in micro and nanotechnologies for integrated systems, a joint degree between the Politecnico di Torino, Turin, Italy, the Grenoble Institute of Technology, and EPFL, Lausanne, Switzerland, in 2016. He is currently pursuing the Ph.D. degree in microelectronics and microsystems. He joined the Integrated Systems Laboratory, EPFL, under the supervision of Prof. De Micheli, in October 2016, where he designed and realized multi-target electrochemical sensing platforms. His work is aiming at physical status monitoring and advanced healthcare diagnosis for wearable applications. His research interests includes the co-design and development of hardware and software interfaces for electrochemical sensors. A strong emphasis of his research is laid on developing efficient data processing tools and optimization algorithms within a multi-sensing paradigm.

FRANCESCA CRISCUOLO received the B.Sc. and M.Sc. degrees *(cum laude)* in materials engineering and nanotechnology from the Politecnico di Milano, Italy, and the Ph.D. degree from the Integrated System Laboratory, EPFL, Lausanne, Switzerland, in April 2020, with her thesis "Wearablemulti-electrode platform for ion-sensing". During her studies, she participated in several exchange programs in France, Czech Republic, The Netherlands, and Belgium. She spent one year with TU Delft, where she specialized in bio-nanomaterials and in nanoelectronics. In 2015, she worked with the Energy Storage Group, IMEC, Leuven, Belgium. She is currently a Postdoctoral Fellow with EPFL, where she focuses on novel electrochemical biosensors for healthcare and diagnostics.

GIOVANNI DE MICHELI (Life Fellow, IEEE) is currently a Professor and the Director of the Institute of Electrical Engineering, EPFL, Lausanne, Switzerland. His research interest includes several aspects of design technologies for integrated circuits and systems, such as synthesis for emerging technologies, networks on chips, and bio-sensing/systems. He was a recipient of the 2016 IEEE/CS Harry Goode Award, the 2016 EDAA Lifetime Achievement Award, and the 2012 IEEE/CAS Mac Van Valkenburg Award among other recognitions. He is also a Fellow of ACM and a member of the Academia Europaea and an International Honorary member of the American Academy of Arts and Sciences.

●●●

SANDRO CARRARA (Fellow, IEEE) is currently a Faculty with the EPFL, Lausanne, Switzerland, and a former Professor with the University of Genoa, Italy, and the University of Bologna, Italy. Along his carrier, he published seven books, with prestigious publishers like Springer/Nature and Cambridge University Press. He published more than 270 papers and 14 patents. He was a member of the BoG of the IEEE CASS and also a member of the IEEE Sensors Council. He was a recipient of the IEEE Sensors Council Technical Achievement Award. He is also the Editor-in-Chief of the IEEE Sᴇɴsᴏʀs Jᴏᴜʀɴᴀʟ and an Associate Editor of IEEE Tʀᴀɴsᴀᴄᴛɪᴏɴs ᴏɴ Bɪᴏᴍᴇᴅɪᴄᴀʟ Cɪʀᴄᴜɪᴛs ᴀɴᴅ Sʏsᴛᴇᴍs.