Classification of Proteomic MS Data as Bayesian Solution of an Inverse Problem

Pascal Szacherski, Jean-François Giovannelli, Laurent Gerfault, Pierre Mahé, Jean-Philippe Charrier, Audrey Giremus, Bruno Lacroix, Pierre Grangeat

▶ **To cite this version:**

Pascal Szacherski, Jean-François Giovannelli, Laurent Gerfault, Pierre Mahé, Jean-Philippe Charrier, et al.. Classification of Proteomic MS Data as Bayesian Solution of an Inverse Problem. IEEE Access, 2014, 2, pp.1248 - 1262. 10.1109/ACCESS.2014.2359979 . cea-01615512

HAL Id: cea-01615512
https://cea.hal.science/cea-01615512v1

Submitted on 12 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IEEE Access
The journal for rapid open access publishing

Received May 6, 2014, accepted June 23, 2014, date of publication September 25, 2014, date of current version October 30, 2014.

Digital Object Identifier 10.1109/ACCESS.2014.2359979

Classification of Proteomic MS Data as Bayesian Solution of an Inverse Problem

PASCAL SZACHERSKI[1,2], JEAN-FRANÇOIS GIOVANNELLI[2], LAURENT GERFAULT[1], PIERRE MAHÉ[3], JEAN-PHILIPPE CHARRIER[3], AUDREY GIREMUS[2], BRUNO LACROIX[3], AND PIERRE GRANGEAT[1], (Senior Member, IEEE)

Université Grenoble Alpes, Grenoble F-38000, France
IMS Laboratory, Université de Bordeaux, Talence 33400, France
Technology Research Department, bioMérieux, Marcy l'Etoile F-69280, France

Corresponding author: P. Szacherski (pszacherski@gmail.com)

This work was supported by the French National Research Agency through the Bayesian Hierarchical Inversion in Proteomics Project under Contract ANR-2010-BLAN-0313.

ABSTRACT The cells in an organism emit different amounts of proteins according to their clinical state (healthy/pathological, for instance). The resulting proteomic profile can be used for early detection, diagnosis, and therapy planning. In this paper, we study the classification of a proteomic sample from the point of view of an inverse problem with a joint Bayesian solution, called inversion-classification. We propose a hierarchical physical forward model and present encouraging results from both simulation and clinical data.

INDEX TERMS Statistical signal processing, inverse problems, mathematical modelling, classification algorithms, probability, proteins, proteomics, selective reaction monitoring, mass spectrometry, liquid chromatography.

I. INTRODUCTION

Cells of an organism emit different amounts of proteins according to their clinical state – e.g. sane/insane, effectiveness/ineffectiveness of a drug – caused by a genetic modification or dysfunction. In proteomics, secreted proteins or a cellular proteome in biological fluids such as blood or urine, in samples from biopsies are analysed. The resulting proteomic profile, i.e., the gathered information about protein concentration, can then be used for diagnosis, early detection, therapy planning and follow-up, drug development, etc. [1]. Nevertheless, the reconstruction of proteomic profiles remains a challenge due to small and variable concentration of the discriminant proteins, called biomarkers. Furthermore, biomarkers are present within a large protein content with abundance ratios of up to 10^8, hence the need for efficient recognition of biomarkers on molecular profiles [2], [3].

The use of a cascade of steps for the sample preparation and instruments such as liquid chromatography (LC) and mass spectrometry (MS) [4] introduces several sources of variability: serum is manually extracted from tubes by pipettes; tryptic digestion is a kinetic process [5] with random behaviour; and data acquisition is perturbed by electric noise, to name only three sources. As a consequence, the peak-like shapes that constitute the acquired data differ, for example, in form and position at every experiment. In spite of the variability of the LC-MS tandem, it is still the most commonly used technology in proteomics. A reliable diagnosis has to integrate efficiently variability and be robust with respect to it, be it of technical or biological nature.

The reconstruction task has been tackled by several methods: non-parametric methods such as area under peak, PLS, N-PLS, PARAFAC [6] and references therein, [7], MRCQuant [8], MZmine [9], VIPER [10], parametric methods based either on deterministic least square fitting or on other statistical estimation using for example Bayesian inference [11]–[13]. Reconstructed profiles can be used for differential analysis where different states are compared. They can also be used for diagnosis which has been shown specific and sensitive [11], in particular early disease detection before morphological symptoms break out.

In signal and data processing, diagnosis making can be tackled by classification or decision theory: given a certain knowledge on each class – either by the use of former experiments, of literature, of training cohorts –, a new sample is associated with a possible outcome (e.g., pathological or healthy).

In this paper, we will discuss an example of classification of serum samples using data from a tandem MS method

2169-3536 © 2014 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

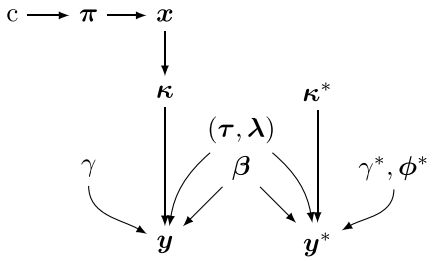

FIGURE 1. Graphical oriented hierarchical diagram for the forward model.

coupled with LC called *Selective Reaction Monitoring* (SRM). An in-depth introduction to SRM is beyond the scope of this paper. We refer the interested reader to [14]–[16] giving recent introductions to SRM. The main idea is that SRM proceeds by decomposing proteins into peptides which are then fragmented and measured. A single protein therefore gives rise to several measurements from the various peptides. This induces a hierarchy in the obtained data that needs to be dealt with to quantify the protein itself.

We will formulate the classification as the solution of an inverse problem [51]. The task is therefore to invert efficiently and robustly the analytical work flow in order to retrieve information about the protein as correctly as possible. An unknown clinical state (class) at the input of the work flow is observed indirectly through the output data. Inverting the underlying physical model will deliver an estimation for the class. Bayesian statistics are well adapted to this difficulty as instrumental and biological sources of variability are modelled by probability distributions, translating the knowledge and belief in the processes. Furthermore, unknown variables are optimally self-calibrated in a Bayesian sense. Finally, models can be updated easily so that the method is also flexible. The main contribution of this paper is hence the modelling of the SRM work flow and the *global* inversion of all variables of the model in one step, which is different from the classical, sequential approach, as we will explain later.

Section II describes the physical and statistical hierarchical forward model of the SRM acquisition process, including the justification of the priors. This data model is embedded into the classification process in section III, inducing the extension of the hierarchical model to training and estimating branches. We show how to solve the problem, and prepare the algorithmic implementation discussed in Section IV. We demonstrate the strength of our method in section V on both simulated and clinical cohorts. We conclude and give perspectives for our work in section VI.

II. THE PHYSICAL AND STATISTICAL MODELS IN SRM BASED PROTEOMICS

This section builds upon our previous modelling work [17] introduced to improve protein quantification and describes the hierarchical physical data generation model of SRM. We deduce the associated probabilistic model for SRM, resulting in several conditional independences. This will be at the core of the classification developments, discussed later in section III.

Remark 1 (Notations): Within this document, the following notations have been adopted.

- ***Generals:*** *We use MATLAB-like notation* $1:N$ *for the vectorised fusion of indexed parameters:* $\theta_{1:N} = [\theta_1, \theta_2, \ldots, \theta_N]$.
- ***Subscripts:*** *We use the subscript index* $p = 1, \ldots, P$ *to denote a protein,* $i = 1, \ldots, I$ *for a peptide,* $l = 1, \ldots, L$ *for a fragment. The subscript* $n = 1, \ldots, N$ *stands for an individual sample from a cohort of* N *samples.*
- ***Superscripts:*** *We use the superscript index* c *to denote the parameters issued from a class. When comparing* M *classes, the parameters associated with class* c^{m} *will be superscripted by* m *(which should not be confused with a mathematical power). A superscript asterisk* $*$ *stands for parameters associated with a standard molecule, a superscript star* $\star$ *for a true value.*
- ***Distributions:*** *We denote* $\mathcal{N}(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{\Gamma})$ *the multivariate normal distribution for* $\boldsymbol{x}$ *given its distribution mean* $\boldsymbol{m}$ *and the precision matrix* $\boldsymbol{\Gamma}$. *The gamma distribution*[1] *with shape* α *and scale* β *is denoted* $\mathcal{G}(x; \alpha, \beta)$ *for a scalar* x. *The uniform distribution with bounding multidimensional interval* $[\boldsymbol{m}, \boldsymbol{M}] = [m_1, M_1] \times \cdots \times [m_{\dim(\boldsymbol{x})}, M_{\dim(\boldsymbol{x})}]$ *is denoted* $\mathcal{U}(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{M})$. *We will denote the normal-Wishart distribution for a couple of parameters* $(\boldsymbol{x}, \boldsymbol{G}) \in \mathrm{R}^P \times \mathrm{R}^{P \times P}$ *by* $\mathcal{NW}(\boldsymbol{x}, \boldsymbol{G}; \boldsymbol{\mu}, \boldsymbol{\Lambda}, \eta, \nu)$. *The parameters* ν *and* $\boldsymbol{\Lambda}$ *describe the degrees of freedom and the scale matrix for the Wishart distribution on* $\boldsymbol{G}$; *the prior sample mean is* $\boldsymbol{\mu}$ *and the prior sample size (the number of prior measurements) on the* $\boldsymbol{G}$ *scale is* η. *Finally, for a discrete variable, the categorical distribution of the event* $X = x$ *is denoted* $\Pr(X = x) \equiv \Pr(x)$.

A. INSTRUMENTAL MODEL

We want to analyse the serum sample from a subject of clinical state $\mathsf{c} \in \mathcal{C}$, with protein concentration $\boldsymbol{x} \in \mathrm{R}^P$. Let its distribution be a multivariate normal distribution, parametrised by $\boldsymbol{\pi}^{\mathsf{c}} = [\boldsymbol{m}^{\mathsf{c}}, \boldsymbol{\Gamma}^{\mathsf{c}}]$ gathering the class mean vector and precision matrix respectively. We assume that the P proteins are all biomarkers for at least one pathology. The Human Proteome Organisation (HUPO) identified about 3000 proteins in human plasma [19], [20]. Typically, a much smaller number of proteins is targeted in SRM based proteomics. Within our clinical conditions P may be up to 60 biomarkers.

Proteins undergo a gain, due to the preparation, fractionating, freezing and other processes. The values of this gain for each protein are collected in the diagonal of the matrix $\boldsymbol{\psi} \in \mathrm{R}^{P \times P}$.

For specificity and instrumental reasons, the P proteins are *digested* into I peptides. We model the peptide quantity after digestion by a linear function $\mathbf{D}\boldsymbol{x}$, where $\mathbf{D} \in \mathrm{N}_0^{I \times P}$ is the digestion matrix, with the convention $\mathrm{N}_0 = \mathrm{N} \cup \{0\}$. One protein carries in general several peptides, and one given

[1] In our developments, we use the definition given by [18, Appendix 1].

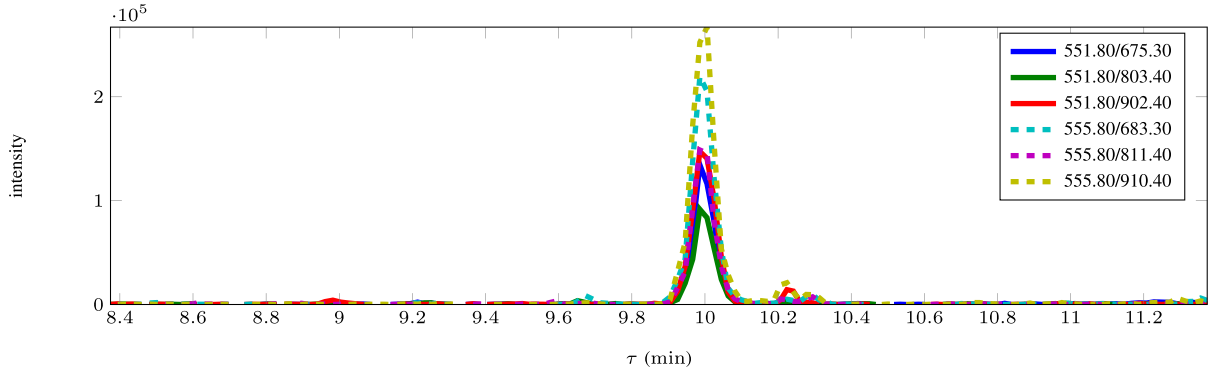

FIGURE 2. Superposed plots of SRM data for protein L-FABP, peptide TVVQLEGDNK. Plain lines correspond to data from native molecules, dashed lines to data from standard molecules. The legend entries are the transitions (i, l) in molecule mass.

peptide is carried by several proteins. Shared peptides are quite common in nature but difficult to handle. Therefore, researchers in proteomics look out for *proteotypic* peptides, i.e. peptides that are carried by one and only one of the targeted proteins in the sample in order to avoid ambiguities and to map them correctly. Hence, the matrix $\mathbf{D}$ is block-diagonal.

Output peptides are subjected to a digestion gain inflected by the kinetic process and possible interactions during the digestion. They are collected on the diagonal of $\boldsymbol{\chi} \in \mathrm{R}^{I\times I}$ and in an additive digestion noise $\boldsymbol{\varepsilon}_{\boldsymbol{\kappa}} \in \mathrm{R}^{I}$. By diagonalising the vectors, the peptide model writes, with $\boldsymbol{\kappa} \in \mathrm{R}^{I}$ the final peptide quantity, $\boldsymbol{\kappa} = (\boldsymbol{\chi}\mathbf{D}\boldsymbol{\psi})\boldsymbol{x} + \boldsymbol{\varepsilon}_{\boldsymbol{\kappa}}$.

By the use of a chromatography column, the peptide mixture is separated according to physico-chemical properties and affinity with the column surface. Injected at time $t_0 := 0$, peptide i elutes after time τ_i. The chromatography output signal $\mathscr{C}_i$ for a peptide i, taking into account a convolution due to instrument imperfectness, is modelled by a Gaussian function with mean τ_i and width λ_i, $\mathscr{C}_i(t;\tau_i,\lambda_i) = \exp\left(-\lambda_i(t-\tau_i)^2/2\right)$. Note that this is only an approximation of the real process whose shape function is a result of solving differential equations with a high number of arguments. Other models, such as bi-Gaussian [21], asymmetric peak shapes [22], splines [23], have been studied. Selecting the adequate chromatography model is a very challenging task as one can see by reading [24]. Nevertheless, approximation by a Gaussian shape is reasonable with respect to the data we have processed so far (see Fig. 2), to the number of involved parameters and to the computation efficiency. However, our approach can easily be extended to any other parametrised shapes.

After ionisation, a peptide precursor ion is isolated depending on its mass in the first quadrupole of the mass spectrometer operating in SRM mode. The ion enters the collision chamber where the precursor ions are fragmented by a collision gas, yielding L ions that we call fragments. The fragment indexed by l is associated with the peptide i. In our context, we call the couple (i, l) *transition*. (In a biological context, a transition is defined as pair of precursor and fragment molecule masses.) Fragments are subjected to a instrumental gain $\beta_l \in \mathrm{R}$ for all $l = 1, \ldots, L$, gathering several phenomena (mainly ionisation, fragmentation, collision gains).

The fragments reach the detector where the signal $\boldsymbol{y}_l(t)$ is recorded at time t, including an instrument, measurement noise $\boldsymbol{\varepsilon}_l(t)$. Let the noise be zero mean, white Gaussian of power γ_l^{-1}. Note that this approximation of the noise process is done for computational convenience, as it decreases highly the inversion complexity and yields nevertheless good results. Following [16], we model all fragments of one given peptide to have the same chromatographic profile since between the elution of the peptides from the chromatography column and the detector, the produced fragments are supposed not to change the profile.

Gains are very fluctuating and hardly known in advance. In order to monitor them, we need to calibrate the system. Generally, two types are used: internal and external calibration.

In our context, labelled AQUA peptides (Absolute QUAntification) are used as internal calibration. Gains and shared parameters up to the peptide level, i.e. instrumental gain $\boldsymbol{\beta}$, elution times $\boldsymbol{\tau}$ and elution peak widths $\boldsymbol{\lambda}$ can be monitored easily since native and labelled molecules share several parameters up to the peptide level. (More information on labelling methods in proteomics can be found in [25].) Although labelled and native molecules ought to have exactly the same behaviour in the instruments, modelled by the identical parameters $\boldsymbol{\tau}$, $\boldsymbol{\lambda}$ [16] and $\boldsymbol{\beta}$, we noticed in several experiments that the ratio between amplitudes of corresponding transition signals may not always be equal to one. In other words, between the fragmentation process and the arrival at the detector, the fragmentation gains of a given fragment can be different between the native and the labelled version. Hence, we introduce an adjustment factor ϕ_l^* per labelled fragment.

Gains beyond the peptide level have to be calibrated externally using a calibration sample. The preparation and digestion gains $\boldsymbol{\psi}$ and $\boldsymbol{\chi}$ are supposed not to vary within one given day of injection so that they are calibrated externally by the use of synthetic reference samples with the same,

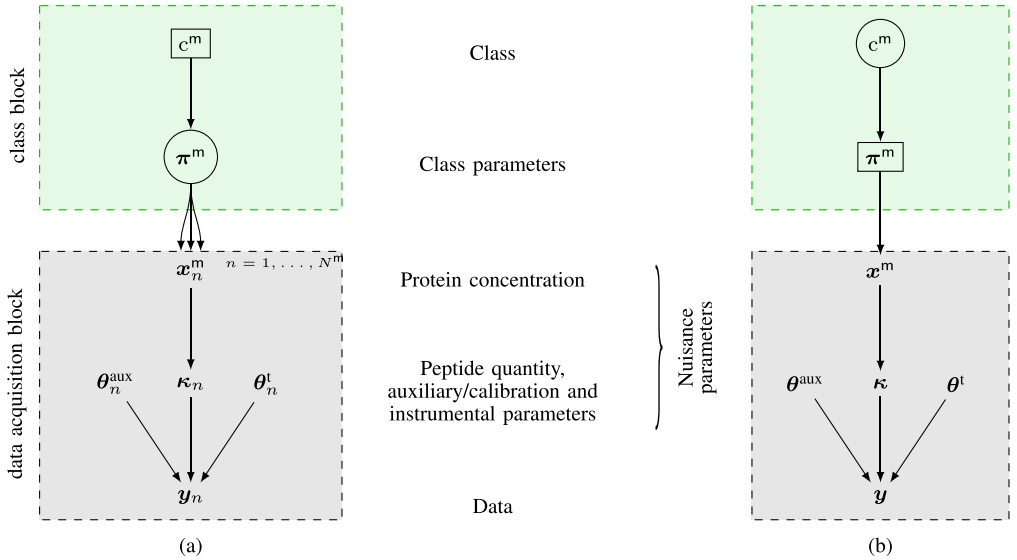

FIGURE 3. **Hierarchical diagram of the forward model. On the left for the training for any given class $c^m \in \mathcal{C}$, on the right for classifying a new sample. Boxed variables are known, circled variables are the parameters of interest. (a) Training. (b) Estimating.**

known native protein concentration and AQUA peptides for every day of injection. We will note the externally pre-calibrated parameters by a subscript index zero.

The final set of instrumental equations write

$$\boldsymbol{\kappa} = (\boldsymbol{\psi}_0 \mathbf{D} \boldsymbol{\chi}_0)\boldsymbol{x} + \boldsymbol{\varepsilon}_{\boldsymbol{\kappa}} \tag{1a}$$

$$\boldsymbol{y}_l(t) = \beta_l \kappa_i \mathscr{C}(t; \tau_i, \lambda_i) + \boldsymbol{\varepsilon}_l(t) \tag{1b}$$

$$\boldsymbol{y}_l^*(t) = \phi_l^* \beta_l \kappa_i^* \mathscr{C}(t; \tau_i, \lambda_i) + \boldsymbol{\varepsilon}_l^*(t). \tag{1c}$$

Parameters introduced above are characterised by their function and can be gathered as follows. The protein concentration $\boldsymbol{x}$ is a biological parameter, the peptide quantity $\boldsymbol{\kappa}$ a techno-biological one, denoted $\boldsymbol{\theta}^{\mathrm{tb}}$.[2] Furthermore, chromatographic parameters ($\boldsymbol{\tau}$, $\boldsymbol{\lambda}$) and gains ($\boldsymbol{\beta}$, $\boldsymbol{\phi}^*$) are combined in the technological parameter vector $\boldsymbol{\theta}^{\mathrm{t}}$. By writing a set of equations, we underline the implicit hierarchy of the model expressed schematically in Fig. 3.

Transitions in SRM experiments are independent. The parameter $\boldsymbol{\theta}^{\mathrm{t}}_{(i,l)}$ will regroup all technical parameters of the transition (i, l).

Fig. 2 shows SRM data of the L-FABP protein by superposing the signals for the three targeted transitions of peptide TVVQLEGDNK (plain) and its labelled equivalent (dashed). One can clearly see that the assumption of the same chromatographic profile for all fragments of the same precursor ion holds as long as the data sampling frequency is high enough.

In summary, we derived the physical, hierarchical forward model of an SRM data acquisition. The equations in (1c) represent the final model that will be used in the following statistical forward model.

[2]Even if these notations are redundant, they are introduced in order to outline the hierarchy.

B. PROBABILISTIC FORWARD MODEL

The forward data generation model including class variables follows a multilevel hierarchical structure, as shown on the right of Fig. 3: the first two levels are due to the class model, the other two levels to the data acquisition model, the last level corresponds to the data.

In order to quantify uncertainty, we set up a probabilistic forward model, including the (conditional) prior distributions for each parameter. The product of these distributions is the joint distribution for all parameters and the data involved in the data generation and it will be at the core of the inversion (see Sect. III).

Remark 2: Consider a hierarchical structure $\theta_1 \to \cdots \to \theta_h \to \theta_{h+1} \to \cdots \to \boldsymbol{y} \equiv \theta_{I+1}$.

For $h = 1, \ldots, H$, *a hierarchical prior* $\mathrm{p}(\theta_h \,|\, \theta_{h-1})$ *for a parameter* θ_h *of hierarchical level h is conjugated by its associated hierarchical likelihood* $\mathrm{p}(\theta_{h+1} \,|\, \theta_h)$ *if the conditional posterior* $\mathrm{p}(\theta_h \,|\, \theta_{h+1}, \theta_{h-1})$ *is of the same functional family as the prior.*

1) CLASS PARAMETERS

Class parameters $\boldsymbol{\pi}^{\mathsf{m}}$ are associated with a given class $c^m \in \mathcal{C}$. The class is modelled as a discrete random variable, C, that takes values in $\mathcal{C}$ with cardinality $\mathrm{card}(\mathcal{C}) = \mathsf{M}$. For all events $c^m \in \mathcal{C}$, define $\Pr(C = c^m) = p_m$, the sum of all p_m's equalling 1. In other words, $\Pr(C)$ is a categorical distribution with event probability vector $[p_1, \ldots, p_M]$.

The class parameter $\boldsymbol{\pi}$ represents the couple of mean and precision $(\boldsymbol{m}, \boldsymbol{\Gamma})$ of the protein concentration distribution. Therefore, and for the sake of conjugacy with other distributions, we propose a normal-Wishart distribution as prior for the class parameters

$$\mathrm{p}(\boldsymbol{m}, \boldsymbol{\Gamma} \,|\, c) = \mathcal{NW}(\boldsymbol{m}, \boldsymbol{\Gamma}; \boldsymbol{\mu}^c, \boldsymbol{\Lambda}^c, \eta^c, \nu^c). \tag{2}$$

2) BIOLOGICAL PARAMETER PRIOR

The native protein concentration $\boldsymbol{x}$ depends on the clinical state c where the class is represented by class parameters $\boldsymbol{\pi}^{\mathrm{c}} = \left[\boldsymbol{m}^{\mathrm{c}}, \boldsymbol{\Gamma}^{\mathrm{c}}\right]$: $\boldsymbol{x} \sim \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{\pi}^{\mathrm{c}}\right) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{m}^{\mathrm{c}}, \boldsymbol{\Gamma}^{\mathrm{c}})$.

3) TECHNO-BIOLOGICAL PARAMETER PRIOR

The native peptide quantity is directly related to the protein concentration by the digestion process (see (1a)), so that in a statistical model, there is dependence on $\boldsymbol{x}$. Using a multivariate normal distribution, this writes $\boldsymbol{\kappa} \sim \mathrm{p}(\boldsymbol{\kappa} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{\kappa}; (\boldsymbol{\psi}_0 \mathbf{D} \boldsymbol{\chi}_0)\boldsymbol{x}, \boldsymbol{\Gamma}_{\boldsymbol{\kappa}})$ where $\boldsymbol{\Gamma}_{\boldsymbol{\kappa}}$ is the precision matrix, informing about the uncertainty of the digestion process.

4) TECHNICAL PARAMETER PRIORS

The noise inverse variance γ_l (resp. the labelled noise inverse variance γ_l^*) is distributed under a gamma density $\gamma_l \sim \mathrm{p}(\gamma_l) = \mathcal{G}(\gamma_l; \alpha, \beta)$ (resp. $\gamma_l^* \sim \mathrm{p}(\gamma_l^*) = \mathcal{G}(\gamma_l^*; \alpha_l^*, \beta_l^*)$).

The global gain parameter vector $\boldsymbol{\beta}$ has a multivariate normal distribution with mean $\boldsymbol{m}_{\boldsymbol{\beta}}$ and inverse variance $\boldsymbol{\Gamma}_{\boldsymbol{\beta}}$: $\boldsymbol{\beta} \sim \mathrm{p}(\boldsymbol{\beta}) = \mathcal{N}(\boldsymbol{\beta}; \boldsymbol{m}_{\boldsymbol{\beta}}, \boldsymbol{\Gamma}_{\boldsymbol{\beta}})$.

Elution times, in minutes, are in general known within a given time interval. Moreover, special experimental protocols (Scheduled SRM) do detect fragments only in a prefixed time range. The elution time vector $\boldsymbol{\tau}$ associated with the peptides is modelled as distributed under a multidimensional uniform distribution with interval bounds τ_i^{m} for the lower and τ_i^{M} for the upper one for each peptide, yielding the lower bound vector $\boldsymbol{\tau}^{\mathrm{m}}$ and the upper bound vector $\boldsymbol{\tau}^{\mathrm{M}}$: $\boldsymbol{\tau} \sim \mathrm{p}(\boldsymbol{\tau}) = \mathcal{U}(\boldsymbol{\tau}; [\boldsymbol{\tau}^{\mathrm{m}}, \boldsymbol{\tau}^{\mathrm{M}}])$.

In our expertise, a range of possible values for the chromatographic peak widths is known. Therefore, peak widths are distributed under a multidimensional uniform distribution with interval bounds λ^{m} and λ^{M} for each dimension: $\boldsymbol{\lambda} \sim \mathrm{p}(\boldsymbol{\lambda}) = \mathcal{U}(\boldsymbol{\lambda}; [\lambda^{\mathrm{m}}, \lambda^{\mathrm{M}}]^I)$. Typical values in order to exclude too flat or too sharp chromatographic peaks are $\left[\lambda^{\mathrm{m}}, \lambda^{\mathrm{M}}\right] = [20, 100]\mathrm{min}^{-2}$, i.e. the deviation believed to be between 0.1 and 0.22 min.

The adjustment gain is supposed to be close to 1 with slight variations. The ratio ϕ_l^* is distributed under a normal density with mean $m_{\phi^*} = 1$ and low precision γ_{ϕ^*}.

The other hyperparameters, however, are chosen as to yield weakly informative priors in order to translate little knowledge and to avoid non informative priors.

5) LIKELIHOOD

For each $l = 1, \ldots, L$, the data $\boldsymbol{y}_l$ associated with transition (i, l) is corrupted by a white zero-mean Gaussian noise. The likelihood is defined by a change of variable based on (1b) from the noise distribution which is Gaussian:

$$\boldsymbol{y}_l \sim \mathrm{p}\left(\boldsymbol{y}_l \mid \boldsymbol{\theta}^{\mathrm{t}}_{(i,l)}\right) = \mathcal{N}\left(\boldsymbol{y}_l; \beta_l, \kappa_i, \mathscr{C}\left(\tau_i, \lambda_i\right), \gamma_l\right). \quad (3)$$

In an analogue manner, by the use of (1c), the distribution of the data y_l^* for the labelled fragment is $\boldsymbol{y}_l^* \sim \mathrm{p}\left(\boldsymbol{y}_l^* \mid (\boldsymbol{\theta}^{\mathrm{t}}_{(i,l)})^*\right) = \mathcal{N}\left(\boldsymbol{y}_l^*; \beta_l, \phi_l^*, \kappa_i^* \mathscr{C}\left(\tau_i, \lambda_i\right), \gamma_l\right)$.

6) JOINT DISTRIBUTION FOR THE DATA AND THE PARAMETERS

Gathering the information of the previous paragraphs, we express the joint distribution of the data acquisition process, including the data variable and the parameters, as product of all hierarchical priors. By taking into account conditional independences, we have

$$\begin{aligned}&\mathrm{p}(\boldsymbol{y}, \boldsymbol{\theta}^{\mathrm{t}}, \boldsymbol{\theta}^{\mathrm{tb}}, \boldsymbol{x}, \boldsymbol{\pi}, \mathrm{c})\\ &\quad = \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{\theta}^{\mathrm{t}}, \boldsymbol{\theta}^{\mathrm{tb}})\, \mathrm{p}(\boldsymbol{\theta}^{\mathrm{tb}} \mid \boldsymbol{x})\, \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{\pi})\, \mathrm{p}(\boldsymbol{\pi} \mid \mathrm{c})\, \Pr(C = \mathrm{c}). \quad (4)\end{aligned}$$

III. METHODOLOGY: CLASSIFICATION AS PARAMETER ESTIMATION PROBLEM

In this section, we introduce the core of the paper: classification seen as a resolution of an inverse problem by the use of Bayesian statistics. For supervised classification, two tasks have to be fulfilled.

1) **Training**, i.e. by the use of a set of labelled data (data with known associated class, called cohort), estimate class parameters for each class in $\mathcal{C}$.
2) **Classifying**, i.e. by the knowledge of all class parameters $\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^{\mathrm{M}}$ and the probability vector $[p_1, \ldots, p_{\mathrm{M}}]$, estimate which class a sample belongs to.

There are several reasons why we split these tasks and do not favour a joint train-and-classify method whilst other methods keep them joined. Firstly, we may have cohorts for several clinical states without having a sample to classify. In this case, we simply want to describe the distribution. This may be the case if we have several replicates of one sample and wish to estimate the common protein concentration distribution. Secondly, if we were to classify several samples in different times, we can easily reuse the results of the training task. Thirdly, thanks to the choice of the normal-Wishart distribution, we can update the trained distributions by setting the prior to the formerly trained values and adding the new training samples. In this way, we do not have to do a training step on the whole set of old and new training samples.

A. DIFFERENT CLASSIFICATION STRATEGIES

Classification is a major stream in signal processing, machine learning and their applications. However, among all classification methods, although they have the same intention (i.e. to classify a sample), methodologies are not always the same.

When parameters are not accessed directly, we access only to transformed, attenuated or amplified, distorted versions of the hidden parameter of interest, i.e. to an indirect measurement of what we initially intended to measure. This is the case for example in astronomy by the use of an optical instrument (SPIRE in [26]), in tomography where the X-ray transform of an object is observed [27, Ch. 12], [28], in posture detection where accelerato-, magneto- or gyrometric data is observed [29], in super-resolution where the observation is a decimated, low-resolution image or video [30], and of course in proteomics by the use of chromatography and spectrometry [17], [31]–[33], as described previously.

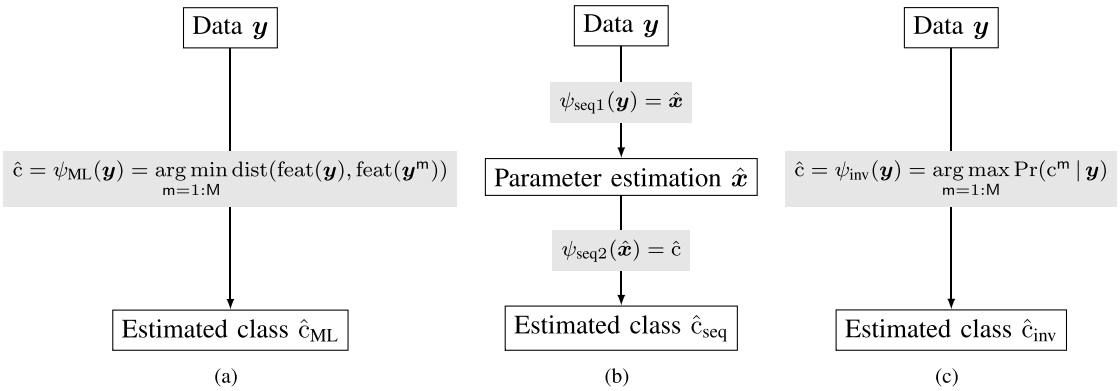

FIGURE 4. Schemes for different classification types: (a) depicts the machine learning approach (compare against representative data using a distance function on features), (b) the sequential approach (use intermediate estimations) and (c) the inversion approach (invert the physical model, here using Bayesian statistics).

In machine learning, the classification process is mainly carried out in the *data* or *feature space* and incorporates at most ad-hoc models of the physical acquisition process (Fig. 4(a)). The classification datum is compared to representative data for each class with respect to the considered features, and the decision is made by minimising a distance function in the feature space. By doing so, the final consequence, i.e. a feature of the data, is used as characteristic parameter for each class. In Brain-Computer-Interface, [34] for example use, within a Riemannian manifold, the centroid of a training set of covariance matrices as feature of each class. A new acquisition is then compared to them using the Riemannian distance. We will refer to this group of processing methods as ''machine learning approach''.

In biostatistics, the logistic regression is often used for classification (Fig. 4(b)). Protein concentrations associated to the classification data are estimated and are afterwards regressed with respect to the estimations within the training sets. Note that the estimation can be carried out by any method or estimation function that returns an estimate of the concentration. This may include a quantification step by inversion which leads to working in the *parameter space* [17], or – as in the proteomic state of the art – by several sequential estimations (noise filtering, then peak identification, then peptide identification, after this peptide quantification, etc). However, the latter leads to a loss of information which may in certain cases be of high importance: classification is done on the sequence of intermediate estimations, leaving out possible sources of variability and correlation between hierarchy levels. Note that the ''sequential approach'' is not limited to the logistic regression since it is as well used for classification for example in proteomics [35], in chemical out-of-specification tests [36], in classification of corrupted signals [37]; partition methods such as fuzzy *c*-means [38] can easily be adapted to integrate a sequential classification.

We propose to go beyond this by working in a *joint space* without sequential estimations for each data, referred to as ''Inversion-Classification'' (Fig. 4(c)). Estimation is done by the inversion of the forward data model, explaining data generation from the causes (input, parameters, here: the clinical state) over the instruments (biological, technico-biological, technological parameters) to the consequences (output, data, here: SRM spectra). Among several other possible choices, we propose a Bayesian framework for the estimation process which allows for integration of several sources of variability by means of probability densities and allows for optimal solutions given the probabilistic model.

What is the advantage in coupling classification and inverse problem? The inverse problem methodology allows for *joint* inference on *all* involved parameters, i.e. the continuous (instrument, concentration) and the discrete (class) variables, as will be demonstrated in this paper. As several parameters constitute the model, the Bayesian framework will be useful to separate and quantify uncertainty on each of them. Furthermore, the hierarchical structure between class, parameters and data can be exploited successfully since the Bayesian theory is particularly adapted for this: while each of the individual components adds complexity to the data model, not considering them turns out to prohibit exploitation of the information they confer. As a consequence, a *global* optimal solution to the problem is proposed, as opposed to successive sequential estimates.

B. CLASSIFICATION SEEN AS AN INVERSE PROBLEM

The forward model predicts the consequence given the causes. The inverse problem is about evaluating the causes from the consequences, i.e. reconstructing the input given the knowledge of the output. To find the best solution, the proposed method takes various sources of information into account: the physical forward model and the data, but also prior knowledge on the parameters and their uncertainty. Fusing these sources relies on Bayesian inference where each quantity in play is given a probability and results in the joint probability density for all variables introduced in section II-B. From the joint density we infer all the other densities and relations among variables. Their development is detailed in the following paragraphs.

1) TRAINING

Within the training task where only the class parameters $\boldsymbol{\pi}$ are estimated for each class, the situation is the following:

- the data comes from a cohort of size N, i.e. a sample of the population;
- within the cohort, M classes coexist;
- each data sample of the cohort is labelled by the corresponding class without error;
- each class c^{m} is represented by the class parameter $\boldsymbol{\pi}^{\mathsf{m}}$ which is the hyperparameter for the distribution of the protein concentrations.

We define data sets according to the classes $\mathfrak{Y}^{\mathsf{m}} = \{\boldsymbol{y}_n \,|\, \boldsymbol{y}_n \in \mathrm{c}^{\mathsf{m}}, n = 1, \ldots, N\} \subseteq \mathcal{Y}^{N^{\mathsf{m}}}$. The cardinality of $\mathfrak{Y}^{\mathsf{m}}$ is $\mathrm{card}(\mathfrak{Y}^{\mathsf{m}}) = N^{\mathsf{m}}$, members of $\mathfrak{Y}^{\mathsf{m}}$ will be denoted $\boldsymbol{y}_n^{\mathsf{m}}$ $(n = 1, \ldots, N^{\mathsf{m}})$ and $\sum_{\mathsf{m}=1}^{\mathsf{M}} N^{\mathsf{m}} = N$.

Each data sample is acquired independently from the others, and parameters are i.i.d. Hence, the joint distribution for the training step separates into a product of $\mathsf{M}(\sum_{\mathsf{m}} N^{\mathsf{m}} + 1)$ distributions:

$$\mathrm{p}\left(\mathfrak{Y}^{1:\mathsf{M}}, \boldsymbol{\pi}^{1:\mathsf{M}}, [(\boldsymbol{x})_{1:N^{\mathsf{m}}}^{\mathsf{m}}]^{1:\mathsf{M}}, [(\boldsymbol{\theta}^{\mathrm{t}})_{1:N^{\mathsf{m}}}^{\mathsf{m}}]^{1:\mathsf{M}}\right) = \prod_{\mathsf{m}=1}^{\mathsf{M}} \mathrm{p}\left(\boldsymbol{\pi}^{\mathsf{m}}\right) \cdot \left[\prod_{n=1}^{N^{\mathsf{m}}} \mathrm{p}\left(\boldsymbol{y}_n^{\mathsf{m}}, (\boldsymbol{x})_n^{\mathsf{m}}, (\boldsymbol{\theta}^{\mathrm{t}})_n^{\mathsf{m}}\right)\right]. \quad (5)$$

The joint distribution (5) is indeed the product of the prior for the class parameters and the product of N joint distributions for the data acquisition as seen in sect. II-B6.

From the last equation, we deduce that the global, multi-class training task breaks down into M separate, independent, mono-class training tasks. For this reason, we will focus on only one class c which means developing (5); we may hence omit the class index m in the remainder of this section.

The prior distribution for the class parameters is normal-Wishart which is conjugated by a normal hierarchical likelihood which will come in very handy in the algorithmic exploitation.

Remark 3 (Wishart and Riemann): The Wishart distribution is often chosen as sampling distribution for precision matrices.

Why is this a reasonable choice? Apart from practical reasons due to the conjugation of the prior by the likelihood, there is another rationale. Within the argument of the exponential function in the Wishart distribution, the trace of a matrix resulting from the multiplication by A with the inverse of B is computed. Covariance and precision matrices live in a Riemannian space. The Riemannian distance between two covariance matrices is calculated as the sum of the logarithms of the eigenvalues of the product between A and B^{-1}. The trace is invariant to a change of basis, hence $\mathrm{tr}(AB^{-1}) = \mathrm{tr}(D)$ *where D is a diagonal matrix porting the eigenvalues of AB^{-1}, and the logarithm is a monotonically increasing function. Hence, by maximising the Wishart distribution of a matrix with respect to a scale matrix, their Riemannian distance is minimised which makes the Wishart distribution a reasonable choice.*

Let $\psi : \mathcal{Y}^N \to \Omega_{\boldsymbol{\pi}}$ be an estimation function. The output of this function is denoted $\hat{\boldsymbol{\pi}} = \psi(\mathfrak{Y})$. Let $\|\boldsymbol{\pi}\|^2 = \|(\boldsymbol{m}, \boldsymbol{\Gamma})\|_2^2$ where $\boldsymbol{\Gamma}$ has been vectorised. It is trivial to see that by developing this expression, we have $\|\boldsymbol{\pi}\|^2 = \|\boldsymbol{m}\|_2^2 + \|\boldsymbol{\Gamma}\|_{\mathrm{fro}}^2$ where $\|\cdot\|_{\mathrm{fro}}$ is the Frobenius matrix norm. The Bayesian estimator is defined through its associated loss function. We choose the quadratic loss

$$\mathrm{L_q}(\boldsymbol{\pi}^\star, \psi(\mathfrak{Y})) = \left\|\boldsymbol{\pi}^\star - \psi(\mathfrak{Y})\right\|^2 \quad (6)$$

where $\boldsymbol{\pi}^\star$ is the true, hypothetical value of the class parameters.

The estimator that minimises the Bayesian risk, i.e. the mean loss over the joint distribution of class parameters and data, is the posterior mean (PM) [39, Ch. 2.5]:

$$\hat{\boldsymbol{\pi}}_{\mathrm{PM}} = \psi_{\mathrm{PM}}(\mathfrak{Y}) = \mathrm{E}_{\Pi \,|\, \mathfrak{Y}}\,(\boldsymbol{\pi} \,|\, \mathfrak{Y}) = \int_{\Omega(\boldsymbol{\pi})} \boldsymbol{\pi}\, \mathrm{p}(\boldsymbol{\pi} \,|\, \mathfrak{Y})\, d\boldsymbol{\pi}. \quad (7)$$

We access the posterior $\mathrm{p}(\boldsymbol{\pi} \,|\, \mathfrak{Y})$ by marginalisation of the nuisance parameters $\boldsymbol{x}_{1:N}$, $\boldsymbol{\theta}_{1:N}^{\mathrm{tb}}$ and $\boldsymbol{\theta}_{1:N}^{\mathrm{t}}$ out of the joint posterior $\mathrm{p}(\boldsymbol{\pi}, \boldsymbol{x}_{1:N}, \boldsymbol{\theta}_{1:N}^{\mathrm{tb}}, \boldsymbol{\theta}_{1:N}^{\mathrm{t}} \,|\, \mathfrak{Y})$ which is proportional to the joint distribution (5).

2) CLASSIFYING

In the classifying task, class parameters $\boldsymbol{\pi}^{1:\mathsf{M}}$ will be needed since they determine the conditional priors for the protein concentration $\mathrm{p}\left(\boldsymbol{x} \,|\, \boldsymbol{\pi}^{\mathsf{m}}\right)$. The class parameters are known with certitude so that from the joint distribution we can write simply $\mathrm{p}(\boldsymbol{x} \,|\, \mathrm{c}^{\mathsf{m}})$[3] depending on the class.

In this task, we consider the following situation:

- the class parameters $\boldsymbol{\pi}^{\mathsf{m}}$ and the categorical probability p_{m} are known (by exact knowledge or training) for each class $\mathsf{m} = 1, \ldots, \mathsf{M}$;
- we process a data sample $\boldsymbol{y}$ of unknown class.

Which class does the data sample belong to? To answer this question, we construct a classifier $\psi : \mathcal{Y} \to \mathcal{C}$ from the data space into the class space, mapping a data sample to one class $\hat{\mathrm{c}} = \psi(\boldsymbol{y})$.

Gathering the priors and the hypotheses from the previous sections, we can easily write the joint probability distribution

$$\mathrm{p}(\mathrm{c}, \boldsymbol{x}, \boldsymbol{\theta}^{\mathrm{t}}, \boldsymbol{y}) = \Pr(\mathrm{c})\, \mathrm{p}(\boldsymbol{x} \,|\, \mathrm{c})\, \mathrm{p}(\boldsymbol{\theta}^{\mathrm{t}} \,|\, \boldsymbol{x})\, \mathrm{p}(\boldsymbol{y} \,|\, \boldsymbol{\theta}^{\mathrm{t}}). \quad (8)$$

We can deduce all involved distributions from this joint density by conditioning rule and/or by marginalisation.

Our proposed classifier is constructed within a Bayesian framework and thus based upon the posterior probability for each class $\mathsf{m} = 1, \ldots, M$:

$$\Pr(C = \mathrm{c}^{\mathsf{m}} \,|\, \boldsymbol{y}) = \int_{\Omega_{\boldsymbol{x},\boldsymbol{\theta}^{\mathrm{t}}}} \mathrm{p}(C = \mathrm{c}^{\mathsf{m}}, \boldsymbol{x}, \boldsymbol{\theta}^{\mathrm{t}} \,|\, \boldsymbol{y})\, d(\boldsymbol{x}, \boldsymbol{\theta}^{\mathrm{t}}), \quad (9)$$

where $\Omega_{\boldsymbol{x},\boldsymbol{\theta}^{\mathrm{t}}}$ denotes the joint parameter space of biological and technical parameters. In words, these parameters are marginalised out of the joint posterior distribution. By this

[3] Given the relation between class and class parameters, it is equivalent to write $\mathrm{p}(\boldsymbol{x}, \boldsymbol{\pi}^{\mathsf{m}} \,|\, \mathrm{c}^{\mathsf{m}})$ and $\mathrm{p}(\boldsymbol{x} \,|\, \mathrm{c}^{\mathsf{m}})$ by marginalising the fixed parameter $\boldsymbol{\pi}$.

means, biological and technical sources of variability, i.e. sources of uncertainty, are integrated during the estimating task.

The optimal classifier ψ_{opt} minimises the Bayesian risk, that is the mean loss function:

$$\psi_{\text{opt}} = \arg\min_{\psi \in \Psi} \mathrm{R}(\psi(\boldsymbol{y}), \mathrm{c}) = \arg\min_{\psi \in \Psi} \mathrm{E}_{C,Y}\left[\mathrm{L}(\psi(\boldsymbol{y}), \mathrm{c})\right].$$

By the use of a 0–1 loss function

$$\mathrm{L}_{01}(\psi, \mathrm{c}^\star) = \begin{cases} 0 & \text{if } \psi(\boldsymbol{y}) = \mathrm{c}^\star, \\ 1 & \text{if } \psi(\boldsymbol{y}) \neq \mathrm{c}^\star, \end{cases} \tag{10}$$

the optimal classifier is given by

$$\hat{\mathrm{c}} = \psi_{\text{MAP}}(\boldsymbol{y}) = \arg\max_{\mathsf{m}=1,\ldots,\mathsf{M}} \Pr(C = \mathrm{c}^{\mathsf{m}} \,|\, \boldsymbol{y}). \tag{11}$$

corresponding to a Maximum A Posteriori (MAP) decision: the estimated class is the one with the highest posterior probability whilst

1) taking into account of biological and technical uncertainty sources by marginalisation, and
2) expressing the certainty (as degree of belief) of the estimating task thanks to the computed probability.

Remark 4 (Inverse approach, Naïve Bayes and logistic regression): As stated above, although aiming at the same result, the inverse and the sequential approaches differ in their processing. However, if we set $\boldsymbol{y} = \hat{\boldsymbol{x}}$, the construction coincides after several other restrictions with the Naïve Bayes (NB) and the logistic regression (LR). NB needs independence between entries, leading to a normal distribution with a diagonal precision matrix. By the use of normal priors and a normal likelihood, and assuming furthermore the same variances across all classes, the mathematical expression of the previous development coincides with a construction for the logistic regression. Hence, the latter is a special, very restricted case of the Bayesian classification approach.

C. SUMMARY

Two tasks are fulfilled sequentially for the classification process: training and classifying. Given a labelled cohort, the class parameters $\hat{\boldsymbol{\pi}}^{\mathsf{m}}$ are estimated according to (7). Then, given the class parameters and a new sample, the class of the sample $\hat{\mathrm{c}}$ is estimated using (11).

These tasks require difficult, analytically impracticable integration and marginalisation. This is mainly due to the product of several distributions leading to a non standard form, the non-conjugacy of the distributions of certain technical parameters, and the high dimension of the problem. In the following section, we use stochastic sampling in order to transform this rather tricky integration exercise into an easy-to-solve sampling exercise.

IV. ALGORITHMIC IMPLEMENTATION

In order to approximate the posterior distribution and to carry out marginalisation (posterior calculation) and integration (posterior mean), we resort to a stochastic sampling technique, namely a *Monte Carlo Markov Chain* (MCMC) [40], [41], which is very well adapted to our application. If we are able to sample the joint posterior distribution in either task, marginalisation is done by keeping only the samples of the variables of interest.

However, neither the joint posterior nor the marginalised posteriors in both training and estimating are easily accessible for sampling: the resulting distributions are highly multivariate and of non standard form due to the product of several different distributions. Hence, a Gibbs structure is adapted in order to transform the problem of sampling jointly all variables into the one of sampling sequentially one variable (or group of variables) after another. For this purpose, conditional posteriors on each variable have to be calculated, i.e. each variable given all other variables and data. The forward model is hierarchical, so that the conditional posterior simplifies from ''one given all others'' to ''one given the variables in the adjacent hierarchical layers'' [39, Ch. 10]. This is a considerable simplification since it brings in less dependences and easier expressions for the conditional posteriors due to conjugacy of prior distributions on the one hand, and the hierarchy on the other hand.

Hence, having chosen mainly normal ($\boldsymbol{x}$, $\boldsymbol{\beta}$, $\boldsymbol{\phi}^*$), gamma ($\boldsymbol{\gamma}$), normal-Wishart ($\boldsymbol{\pi}$) and categorical (c) distributions for the parameters, they are indeed conjugated by the associated hierarchical likelihoods. Therefore, the conditional posteriors for these variables are standard ones. For those where conjugacy cannot be reached ($\boldsymbol{\tau}$, $\boldsymbol{\lambda}$), one *Metropolis-Hastings* step per parameter and per iteration is included. In our developments, a sequential random walk with a Gaussian kernel is used due to its efficient compromise between convergence speed and computation time per iteration.

By sampling a discrete variable such as for the class parameter, we add some difficulty to our problem. The estimating task can be identified with a model choice problem: Given model $\mathcal{M}^{\mathsf{m}} \equiv \mathrm{c}^{\mathsf{m}}$, the protein concentration is distributed under $\mathrm{p}(\boldsymbol{x} \,|\, \mathcal{M}^{\mathsf{m}}) \equiv \mathrm{p}(\boldsymbol{x} \,|\, \mathrm{c}^{\mathsf{m}})$, for all competing models $\mathsf{m} = 1, \ldots, \mathsf{M}$. By the implicit estimation of the discrete variable c in the sampling process, the proposed algorithm is a simplified version of a *Reversible Jump MCMC* [39], [42] where all the models have the same dimension. Moreover, jumping from one class to another is always possible independently on the current class, and the transformation between classes is the identity function. Note that another way of computing the posterior would have been to compute the evidence $\mathrm{p}(\boldsymbol{y} \,|\, \mathrm{c}^{\mathsf{m}})$ for each class, demanding M MCMC chains to be evaluated with fixed class parameter. This can, however, lead to numerical inefficiency.

By the Monte Carlo integration theory, the posterior mean for the class parameters $\boldsymbol{\pi}$ is given by averaging the class parameter samples $\boldsymbol{\pi}^{(k)}$:

$$\hat{\boldsymbol{\pi}} = \frac{1}{\mathrm{card}(\mathcal{K}_t)} \sum_{k \in \mathcal{K}_t} \boldsymbol{\pi}^{(k)}$$

where $\mathcal{K}_t$ is the sample index set. Due to the Markov Chain property, the first samples are not distributed under the target distribution and are still influenced by the initialisation of the algorithm. This is the reason for leaving out the first K_0 samples, called *burn-in time*, that are needed for the MCMC to reach the stationary, target distribution. $\mathcal{K}_t$ can then be, for example, all iterations from K_0+1 to K_0+K.

In estimating, the posterior for the class parameter is a discrete, categorical distribution which can be visualised as histogram of the class samples; hence, the estimated class is the most frequently sampled one, i.e.

$$\hat{\mathrm{c}} = \underset{\mathsf{m}=1,\dots,\mathsf{M}}{\arg\max}\ \Pr(C=\mathrm{c}^{\mathsf{m}}\,|\,\boldsymbol{y}) = \underset{\mathsf{m}=1,\dots,\mathsf{M}}{\arg\max}\ \mathrm{card}(\mathcal{K}_c^{\mathsf{m}})$$

where $\mathcal{K}_c^{\mathsf{m}}$ gathers the sample indices $k > K_0$ with $\mathrm{c}^{(k)} = \mathrm{c}^{\mathsf{m}}$.

The algorithms are sketched in Tab. 1 for the Training part and Tab. 2 for the Classifying part.

TABLE 1. Class parameter estimation.

Require: $\boldsymbol{y}_{1:N}$, priors for $\boldsymbol{\pi}$, $\boldsymbol{x}$, $\boldsymbol{\kappa}$, $\boldsymbol{\theta}^{\mathrm{inst}}$
Maximal number of iterations: K^{M}.
Minimal number of burn-in: K_0.
Number of samples taken into consideration: K.

function $\hat{\boldsymbol{\pi}} = \textsc{Train}(\boldsymbol{y}_{1:N})$
  $k=0$, initialise parameters
  while $k \le K^{\mathrm{M}}$ **do**
    $k \leftarrow k+1$
    for $n = 1,\dots,N$ **do**
      sample $(\boldsymbol{\theta}_n^{\mathrm{inst}})^{(k)} \sim \mathrm{p}(\boldsymbol{\theta}^{\mathrm{inst}}\,|\,\boldsymbol{y}_n, \boldsymbol{\kappa}_n^{(k-1)})$ ▷ Hybrid Gibbs.
      sample $\boldsymbol{\kappa}_n^{(k)} \sim \mathrm{p}(\boldsymbol{\kappa}\,|\,\boldsymbol{y}_n, (\boldsymbol{\theta}_n^{\mathrm{inst}})^{(k)}, \boldsymbol{x}_n^{(k-1)})$ ▷ Explicit sampling.
      sample $\boldsymbol{x}_n^{(k)} \sim \mathrm{p}(\boldsymbol{x}\,|\,\boldsymbol{\kappa}_n^{(k)}, \boldsymbol{\pi}^{(k-1)})$ ▷ Explicit sampling.
    end for
    sample $\boldsymbol{\pi}^{(k)} \sim \mathrm{p}(\boldsymbol{\pi}\,|\,\boldsymbol{x}_{1:N}^{(k)})$ ▷ Explicit sampling.
    if Burn-in condition met **and** $k \ge K_0$ **then**
      Let $K^{\mathrm{M}} \leftarrow k+K$ and $\mathcal{K} := \{k+1,\dots,k+K\}$
    end if
  end while
  return $\hat{\boldsymbol{\pi}} = \dfrac{1}{K}\displaystyle\sum_{k\in\mathcal{K}} \boldsymbol{\pi}^{(k)}$
end function

Remark 5 (Nuisance parameter estimation): In a MCMC, we have to sample parameters to access to the joint posterior. For the given estimators, we marginalise the nuisance parameters by considering only the samples of the parameter of interest. Since samples for nuisance parameters are drawn anyway, we can compute e.g. *posterior mean estimations for the nuisance parameters by averaging the drawn samples. Note that, in the classification case, only samples from the estimated class should be used in order not to compute an estimation from a distribution mixture. They are just the samples indexed by* $\mathcal{K}_c^{\hat{\mathsf{m}}}$, $\hat{\mathsf{m}}$ *designing the estimated class index.*

V. RESULTS

We apply our devised classification to several experimental campaigns. Firstly, we will work on simulated data in order to compare the results to the ground truth, i.e. the estimated values can be compared directly to true, known values. We will simulate cohorts of different sizes in order to give an approximation of the optimal cohort size for our study.

Secondly, the method will then be tested on a clinical data set, provided by a campaign on colorectal cancer patients (for the case cohort) and from a blood donation centre (for the control cohort).

However, we need to introduce the error evaluation criterion for the classification.

TABLE 2. Class estimation.

Require: $\boldsymbol{y}$, categoric distribution for C, parameters $\boldsymbol{\pi}$ for $\boldsymbol{x}$, priors for $\boldsymbol{\kappa}$, $\boldsymbol{\theta}^{\mathrm{inst}}$
Maximal number of iterations: K^{M}.
Minimal number of burn-in: K_0.
Number of samples taken into consideration: K.

function $\hat{\mathrm{c}} = \textsc{Classify}(\boldsymbol{y})$
  $k=0$, initialise parameters
  while $k \le K^{\mathrm{M}}$ **do**
    $k \leftarrow k+1$
    sample $(\boldsymbol{\theta}^{\mathrm{inst}})^{(k)} \sim \mathrm{p}(\boldsymbol{\theta}^{\mathrm{inst}}\,|\,\boldsymbol{y}, \boldsymbol{\kappa}^{(k-1)})$ ▷ Hybrid Gibbs.
    sample $\boldsymbol{\kappa}^{(k)} \sim \mathrm{p}(\boldsymbol{\kappa}\,|\,\boldsymbol{y}, (\boldsymbol{\theta}^{\mathrm{inst}})^{(k)}, \boldsymbol{x}^{(k-1)})$ ▷ Explicit sampling.
    sample $\boldsymbol{x}^{(k)} \sim \mathrm{p}(\boldsymbol{x}\,|\,\boldsymbol{\kappa}^{(k)}, \boldsymbol{\pi}^{(k-1)})$ ▷ Explicit sampling.
    sample $\mathrm{c}^{(k)} \sim \Pr(\mathrm{c}\,|\,\boldsymbol{x}^{(k)})$ ▷ Explicit sampling.
    if Burn-in condition met **and** $k \ge K_0$ **then**
      Let $K^{\mathrm{M}} \leftarrow k+K$ and $\mathcal{K} := \{k+1,\dots,k+K\}$
    end if
  end while
  for $\mathsf{m} = 1,\dots,\mathsf{M}$ **do**
    $\mathcal{K}^{\mathsf{m}} = \{k \in \mathcal{K}\,|\,\mathrm{c}^{(k)} = \mathrm{c}^{\mathsf{m}}\}$
    $\Pr(C=\mathrm{c}^{\mathsf{m}}\,|\,\boldsymbol{y}) = \frac{\mathrm{card}\,\mathcal{K}^{\mathsf{m}}}{K}$
  end for
  return $\hat{\mathrm{c}} = \underset{\mathsf{m}=1,\dots,\mathsf{M}}{\arg\max}\ \Pr(C=\mathrm{c}^{\mathsf{m}}\,|\,\boldsymbol{y})$
end function

A. ERROR EVALUATION

We evaluate classification performances of several estimators by the use of the loss function introduced in (10). By sampling a class c from the categorical distribution and the associated data $\boldsymbol{y}$, we can approximate the Bayesian risk, by averaging the output of the losses:

$$\mathrm{R}(\psi, \mathrm{c}) \approx \frac{1}{R}\sum_{r=1}^{R} \mathrm{L}\left(\psi(\mathrm{y}^{[r]}), \mathrm{c}^{[r]}\right). \tag{12}$$

While we do have true labels in both clinical and simulated campaigns, we only have true values for the underlying concentration distributions in simulation. For this case and for a given, fixed class $\mathrm{c}^\star$, let $\boldsymbol{\pi}^\star = [\boldsymbol{m}^\star, \boldsymbol{\Gamma}^\star]$ be the class parameters, and $\hat{\boldsymbol{\pi}} = [\hat{\boldsymbol{m}}, \hat{\boldsymbol{\Gamma}}] = \psi(\mathfrak{Y})$ the estimated class parameters given the cohort $\mathfrak{Y} \sim \mathrm{p}(\mathfrak{Y}\,|\,\boldsymbol{\theta}^{\mathrm{t}}_{1:N}, \boldsymbol{x}_{1:N}, \boldsymbol{\pi}^\star)$ with size N. We evaluate the training by analysing the training loss function introduced in (6) and approximate the Bayesian risk by sampling class parameters and creating associated cohorts. As we have shown, the norm for the class parameter decomposes into the sum of Euclidean norm for the mean and Frobenius norm for the precision which we will consider separately.

Note that, since all estimations depend on a cohort with a specific size N, all estimations and hence all errors also depend on the size. In order to simplify the notation, we omit the mention of N; however, the influence of the cohort size will be analysed.

B. RESULTS ON TRAINING SIMULATED SRM DATA

For the sake of clarity, we will call "population truth" the true class parameters or distribution, "cohort truth" the class parameters or distribution obtained by the use of the simulated protein concentration draws, and "estimations" the class parameters or associated distribution obtained by our Bayesian Hierarchical Inversion.

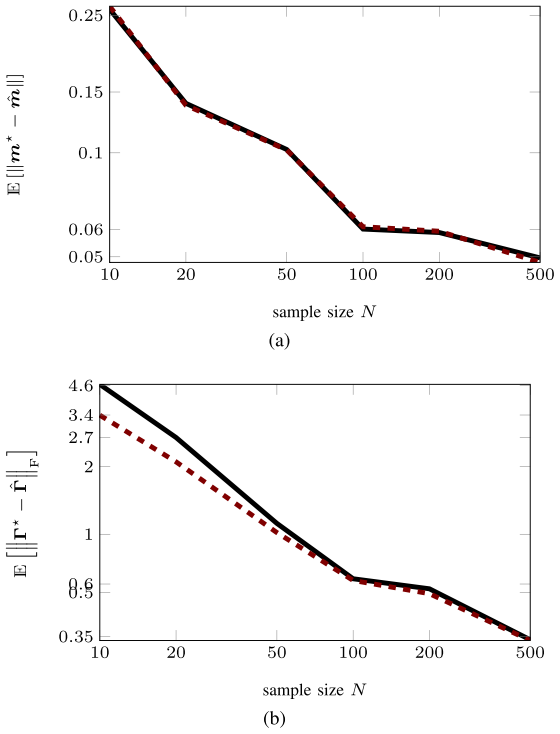

FIGURE 5. Evolution of the class parameters estimation errors for one class of Situation 2 (2 proteins $\xrightarrow{1\ \ 3}$ 6 peptides $\xrightarrow{1\ \ 3}$ 18 fragments) for the cohort truth (dashed) and the estimation (solid). – (a): mean error on the class mean (Euclidean norm). (b): mean error on the class precision (Frobenius norm).

In the simulation case, we are able to compare estimations and the cohort truth. We illustrate the results for the training stage where we consider two dependent proteins. Fig. 5 shows the evolution of estimation errors depending on the cohort size N for the class mean and the class precision separately. As one can see, although we do not have direct access to the protein concentration, the parameter estimations given by our algorithm and by the population truth are very close. In particular, both the mean vector and the non-diagonal precision matrix have been estimated correctly, showing a very similar evolution as in Fig. 5 by applying the Euclidean norm and the Frobenius distance respectively as performance indicator.

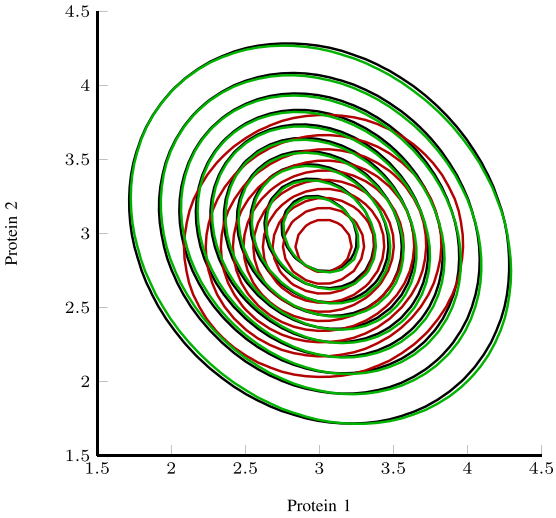

FIGURE 6. Reconstruction of the class distribution; black: distribution on the population level; red: learnt distribution with $N = 10$; green: with $N = 500$.

The reconstruction in Fig. 6 shows for two examples the influence of the cohort size for the good estimation performance. We reach nearly a perfect estimation for the class parameters with a huge cohort size whereas the moderate size of ten individuals is not sufficient for retrieving the mean and the precision, especially the non-diagonal entries of the precision matrix.

TABLE 3. Mean computation time for the training step in seconds.

N	10	20	50	100	200	500
time (minutes)	2	4	13	37	119	928

The mean process time for this two-protein example as a function of the cohort size is given in Table 3. One can see that computation time increases drastically. In the following, we will thus compare performances up to $N = 100$ training samples.

C. RESULTS ON CLASSIFYING SIMULATED SRM DATA

We compare the Inversion-Classification (IC) methods against three state-of-the-art methods: Naïve Bayes (NB) and logistic regression (LR, using the `logregBayes` routine from the free Matlab toolbox http://code.google.com/p/pmtk3/pmtk3) [43], computing probabilities for the classes, and the fuzzy c-means (FCM) [38, Ch. 2], returning a degree of belief which is not based on probability. The competing methods work on intermediate estimations for protein concentrations, established by our robust Bayesian Inversion-Quantification integrating technological variability [44]. The estimated class is the one with highest probability (IC, NB, LR) or degree of belief (FCM). Class parameters are trained using subcohorts of 100 samples per class, which is close to sizes within our clinical data bases. We use different data for training and estimating.

Methods will be compared on their mean loss, approximated by the use of $R = 3000$ data samples. To demonstrate robustness, simulations have four different mean noise levels and fluctuating gains and peak characteristics so as to imitate real data, yielding Signal-Noise-Ratios between 0 and 15 dB. We will draw simulations for $\mathsf{M} = 3$ classes, $P = 2$ proteins, 2 peptides per protein, 3 fragments per peptide, i.e. $I = 4$ peptides, $L = 12$ data traces. Class parameters have been chosen to result in overlapping class distributions, as shown in Fig. 7.

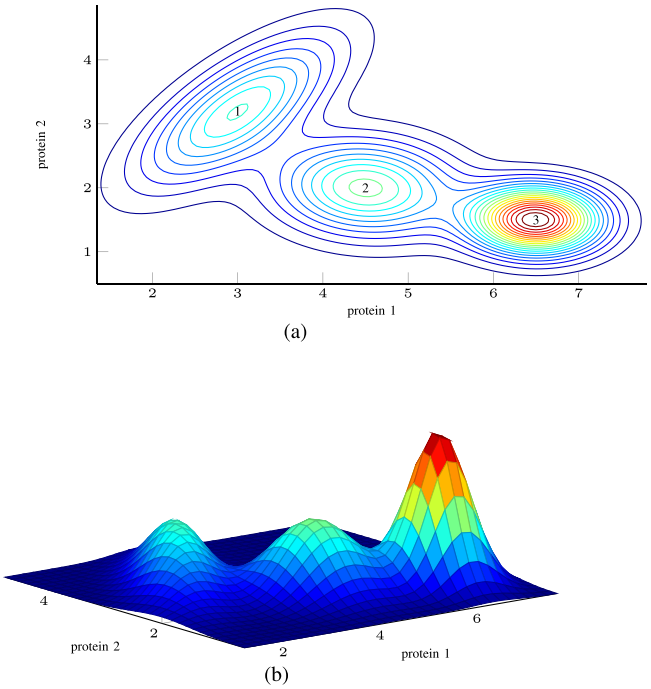

FIGURE 7. **Two-protein-distribution used in the evaluation of the classification method (contour plot on the top, three dimensional representation on the bottom).**

TABLE 4. **Evolution of the risk for each classifier depending on the signal to noise ratio.**

SNR [dB]	IC	NB	LR	FCM
0.84	0.33	0.60	0.49	0.56
1.47	0.13	0.42	0.41	0.39
3.04	0.08	0.19	0.12	0.15
6.11	0.04	0.06	0.05	0.04
14.3	0.04	0.05	0.05	0.04

The comparison of the methods is given in Table 4 where the mean loss is given as function of the mean SNR. Overall, the performance increases with decreasing noise level. Furthermore, one can see that the IC outperforms the sequential methods, whatever the SNR. In average, the mean loss is very low despite the overlapping class distributions. When strong additive noise is used (mean SNR of 0.84 dB), 33 % of the test samples are misclassified by IC against 60 % for NB, 49 % for LR and 56 % for FCM. Remember that these results are obtained for strongly corrupted data, the main information of the peptides being drowned in noise, and are still much better than choosing a class randomly. With a mean SNR of 1.47 dB, 13 % are misclassified by IC which is three times better than the competing performances. These two examples show that the IC does very well even near the limits of detection and quantification. We reach nearly perfect classification with equivalent performances for all methods in presence of very weak noise thanks to good, robust quantification, yielding less than 5 % error. This result was expected: firstly, the competing methods benefit from a robust quantification method, integrating variabilities; however, if a state-of-the-art method such as Peak Maximum or Area under Peak had been chosen, the classification methods – especially for strong noise – would have been less comparable. Secondly, class distributions overlap which creates naturally confusion. However, the analysis of the misclassified samples shows that the probability of the estimated class is rather small which should be enough not to retain this automatised diagnosis and demand a new examination in a clinical context.

D. RESULTS ON CLASSIFYING CLINICAL DATA

This subsection will consider real clinical data from a Colorectal Cancer cohort, collected by bioMérieux, France. We have 90 case sample from individuals that have been tested positive (represented in red in the following). Furthermore, we have 114 control samples that have either been tested negative or are issued from a blood donor centre (represented in blue). This study will use two possible biomarkers of the colorectal cancer, *L-FABP* (Liver-type Fatty-Acid-Binding Protein) and Protein X (renamed for possible patent reasons). The following results are obtained using a 20-fold random cross-validation on ten per cent of the database. The learnt distributions on all samples per class is presented in Fig. 8. One can see that their topology is very interesting. The black lines present the conic separating the two classes [45].

Table 5 summarises the performances by presenting an empirical frequency table for the considered classifiers. The true negatives is high for all four of them, although FCM shows the least performance, while the Bayesian Inversion-Classification does not show any error in this study. The true positives however have a different tendency. The FCM performs well here while the NB shows mediocre results.

The mean risk for the four methods are given by $\mathrm{R}^{\mathrm{IC}} = 0.095$, $\mathrm{R}^{\mathrm{NB}} = 0.1125$, $\mathrm{R}^{\mathrm{LR}} = 0.10$, $\mathrm{R}^{\mathrm{FCM}} = 0.15$; the minimal value is obtained by our method.

One notices that the results are partially similar (with the exception of the fuzzy c-means algorithm). This might be partially due to the fact that the competing methods benefit from a Bayesian Inversion-Quantification step, compensating the technical variability. This shows also that, compared to the simulated case, our data situate in a low noise area thanks to good sample preparation and data acquisition.

E. DISCUSSION

Comparing the results of our method to other ones, one has to admit that the inversion of a physical model does provide a

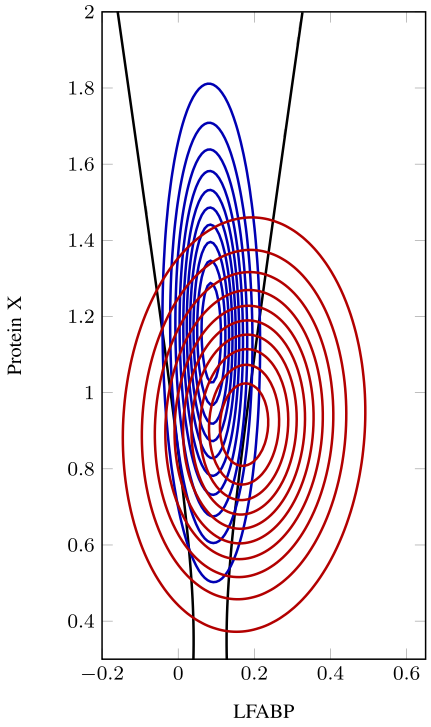

FIGURE 8. Bi-dimensional density for each class of the colorectal data set. The black hyperbole corresponds to the class separator determined by the Inversion-Classification method.

TABLE 5. Empirical frequency tables for the four classifiers. The variables $c^\star$ and $\hat{c}$ stand for the true and the estimated class respectively. (a) Inversion classification (b) Naïve bayes (c) Logistic regression (d) Fuzzy c-means.

(a)

$c^\star \downarrow$ \ $\hat{c} \rightarrow$	control	case
control	1	0
case	0.38	0.62

(b)

$c^\star \downarrow$ \ $\hat{c} \rightarrow$	control	case
control	0.985	0.015
case	0.45	0.55

(c)

$c^\star \downarrow$ \ $\hat{c} \rightarrow$	control	case
control	0.95	0.05
case	0.35	0.65

(d)

$c^\star \downarrow$ \ $\hat{c} \rightarrow$	control	case
control	0.6	0.4
case	0.2	0.8

powerful classification. By the use of a Bayesian framework, not only do we comprehend uncertainty, occurring within the instrument and often omitted; we can also interpret the results in a probabilistic way. We give the probability for the estimated class, based upon the marginalisation of nuisance parameters, whereas other estimation methods require explicit nuisance parameter estimations. Errors are evaluated in a probabilistic sense by the use of the samples drawn by MCMC. The user can provide prior information on the classes and parameters which robustifies the algorithm outcome further.

We see also that the hypotheses of independent protein concentrations as done by NB or of homoscedasticity as done by LR make the performances decrease. The IC keeps as much information as possible from each internal stage, obtaining thus the best, the most global results [46], [47].

Moreover, in a similar study [48], algorithms have been compared using ROC curves where one can deduce the sensitivity of a classifier at given specificity, and vice versa. By changing the loss function which is an important part of our method, one can optimise the outcome at one's needs and hence change the compromise between false positives and false negatives.

VI. CONCLUSION AND PERSPECTIVES

A. CONCLUSION

This paper introduces the joint use of an inverse problem strategy and Bayesian methods for the solution of a classification problem. We based our exposé on a serum classification use case in proteomics. After a short introduction of the used instruments and the physical model, the method was described in theory and practice. The final section showed encouraging results both on simulated data and clinical data. The devised method outperforms other competing methods by efficiently integrating the variability sources, propagating information from each hierarchical step, avoiding intermediate estimation steps, and finding a global best solution.

The hierarchical nature of the model is generic as it links several parameter groups. We have already used these notations in [49] and [32] for an LC-MS process, involving other parameters within the similar hierarchical structure. Hence, all developments for the "inversion-classification" of this paper can be extended – nearly as is – to problems that are structured in a similar way, not only in proteomics, but as well in genomics, metabolics, and other application domains where a physical model has to be inverted.

B. PERSPECTIVES

Several perspectives can be drawn. Of course, the eternal quest of the best model has to be mentioned. We do know that the data does not have exactly a Gaussian shape, we do know that the noise is not idealistic, white, zero-mean. But we do also know that the assumptions made in our work deliver good approximations and interesting results while permitting fast computations. It would be interesting to consider a model that is adapted to fit even more the real data, e.g. by using a Laplacian noise distribution or a spline model for the chromatographic signals, and compare the gains and losses induced by this choice.

In this document, the class labels were known without error. The Bayesian framework permits extending our training step to a more general one, including mislabelled and missing data as well as label correction.

Furthermore, we worked on only one clinical data set so far. The BHI-PRO project will produce more results on other data acquired with different mass spectrometres including SRM, and with other proteins in order to discover and validate potential biomarkers [44], [46], [47], [50].

ACKNOWLEDGMENT

The authors would like to thank all researchers from the bioMérieux labs for acquiring the precious data. The authors would finally highlight the very precious remarks and critics by the reviewers of this paper.

REFERENCES

[1] M. Palmblad, A. Tiss, and R. Cramer, "Mass spectrometry in clinical proteomics—From the present to the future," *Proteomics, Clin. Appl.*, vol. 3, no. 1, pp. 6–17, Jan. 2009. [Online]. Available: http://doi.wiley.com/10.1002/prca.200800090

[2] N. L. Anderson and N. G. Anderson, "The human plasma proteome history, character, and diagnostic prospects," *Molecular Cellular Proteomics*, vol. 1, no. 11, pp. 845–867, Nov. 2002. [Online]. Available: http://www.mcponline.org/content/1/11/845

[3] R. Schiess, "Proteomic strategy for biomarker discovery," Ph.D. dissertation, ETH Zürich, Zürich, Switzerland, 2008.

[4] R. Aebersold and M. Mann, "Mass spectrometry-based proteomics," *Nature*, vol. 422, no. 6928, pp. 198–207, Mar. 2003. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature01511

[5] E. J. Finehout, J. R. Cantor, and K. H. Lee, "Kinetic characterization of sequencing grade modified trypsin," *Proteomics*, vol. 5, no. 9, pp. 2319–2321, Jun. 2005. [Online]. Available: http://doi.wiley.com/10.1002/pmic.200401268

[6] G. Strubel, "Reconstruction de profils moléculaires: Modélisation et inversion d'une chaîne de mesure protéomique," Ph.D. dissertation, École Polytechnique Grenoble, Grenoble, France, 2008.

[7] W. S. Noble and M. J. MacCoss, "Computational and statistical analysis of protein mass spectrometry data," *PLoS Comput. Biol.*, vol. 8, no. 1, p. e1002296, Jan. 2012. [Online]. Available: http://dx.doi.org/10.1371/journal.pcbi.1002296

[8] W. E. Haskins, K. Petritis, and J. Zhang, "MRCQuant—An accurate LC–MS relative isotopic quantification algorithm on TOF instruments," *BMC Bioinform.*, vol. 12, no. 1, p. 74, 2011.

[9] M. Katajamaa, J. Miettinen, and M. Orešič, "MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data," *Bioinformatics*, vol. 22, no. 5, pp. 634–636, 2006.

[10] M. Monroe, N. Tolić, N. Jaitly, J. Shaw, J. Adkins, and R. Smith, "VIPER: An advanced software package to support high-throughput LC–MS peptide identification," *Bioinformatics*, vol. 23, no. 15, pp. 2021–2023, 2007.

[11] P. Grangeat *et al.*, "First demonstration on NSE biomarker of a computational environment dedicated to lab-on-chip based cancer diagnosis," in *Proc. 58th ASMS Int. Conf.*, Salt Lake City, UT, USA, 2010.

[12] G. Strubel, J.-F. Giovannelli, C. Paulus, L. Gerfault, and P. Grangeat, "Bayesian estimation for molecular profile reconstruction in proteomics based on liquid chromatography and mass spectrometry," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Lyon, France, Aug. 2007, pp. 5979–5982. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4353710

[13] T. Schwarz-Selinger, R. Preuss, V. Dose, and W. von der Linden, "Analysis of multicomponent mass spectra applying Bayesian probability theory," *J. Mass Spectrometry*, vol. 36, no. 8, pp. 866–874, 2001.

[14] M.-Y. K. Brusniak, C. S. Chu, U. Kusebauch, M. J. Sartain, J. D. Watts, and R. L. Moritz, "An assessment of current bioinformatic solutions for analyzing LC–MS data acquired by selected reaction monitoring technology," *Proteomics*, vol. 12, no. 8, pp. 1176–1184, Apr. 2012. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/pmic.201100571/abstract

[15] M. Rauh, "LC–MS/MS for protein and peptide quantification in clinical chemistry," *J. Chromatogr. B*, vols. 883–884, pp. 59–67, Feb. 2012.

[16] V. Lange, P. Picotti, B. Domon, and R. Aebersold, "Selected reaction monitoring for quantitative proteomics: A tutorial," *Molecular Syst. Biol.*, vol. 4, p. 222, Oct. 2008.

[17] L. Gerfault, P. Szacherski, J.-F. Giovannelli, J.-P. Charrier, P. Mahé, and P. Grangeat, "A hierarchical SRM acquisition chain model for improved protein quantification in serum samples," in *Proc. RECOMB-CP*, San Diego, CA, USA, Apr. 2012. [Online]. Available: http://proteomics.ucsd.edu/recombcp2012/

[18] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer-Verlag, 1985.

[19] G. S. Omenn *et al.*, "Overview of the HUPO plasma proteome project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database," *Proteomics*, vol. 5, no. 13, pp. 3226–3245, Aug. 2005. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16104056

[20] G. S. Omenn, R. Aebersold, and Y.-K. Paik, "7(th) HUPO world congress of proteomics: Launching the second phase of the HUPO plasma proteome project (PPP-2) 16-20 august 2008, Amsterdam, The Netherlands," *Proteomics*, vol. 9, no. 1, pp. 4–6, Jan. 2009. [Online]. Available: http://doi.wiley.com/10.1002/pmic.200800781

[21] T. Yu and H. Peng, "Quantification and deconvolution of asymmetric LC–MS peaks using the bi-Gaussian mixture model and statistical model selection," *BMC Bioinform.*, vol. 11, p. 559, Nov. 2010. [Online]. Available: http://www.biomedcentral.com/1471–2105/11/559

[22] Z. Pápai and T. L. Pap, "Analysis of peak asymmetry in chromatography," *J. Chromatogr. A*, vol. 953, nos. 1–2, pp. 31–38, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021967302001218

[23] J. Listgarten, R. M. Neal, S. T. Roweis, P. Wong, and A. Emili, "Difference detection in LC–MS data for protein biomarker discovery," *Bioinformatics*, vol. 23, no. 2, pp. e198–e204, Jan. 2007. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/23/2/e198

[24] V. B. Di Marco and G. G. Bombi, "Mathematical functions for the representation of chromatographic peaks," *J. Chromatogr. A*, vol. 931, nos. 1–2, pp. 1–30, Oct. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0021967301011360

[25] V. Brun, C. Masselon, J. Garin, and A. Dupuis, "Isotope dilution strategies for absolute quantitative proteomics," *J. Proteomics*, vol. 72, no. 5, pp. 740–749, Jul. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S1874391909001079

[26] F. Orieux, J.-F. Giovannelli, T. Rodet, A. Abergel, H. Ayasso, and M. Husson. (May 2011). "Super-resolution in map-making based on a physical instrument model and regularized inversion. Application to SPIRE/Herschel." [Online]. Available: http://arxiv.org/abs/1103.3698

[27] J. Idier, Ed., *Bayesian Approach to Inverse Problems*. New York, NY, USA: Wiley, 2008.

[28] P. Grangeat, Ed., *Tomography*, 1st ed. New York, NY, USA: Wiley, Oct. 2009.

[29] P. Jallon, "A graph based algorithm for postures estimation based on accelerometers data," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug./Sep. 2010, pp. 2778–2781.

[30] M. Protter, M. Elad, H. Takeda, and P. Milanfar, "Generalizing the nonlocal-means to super-resolution reconstruction," *IEEE Trans. Image Process.*, vol. 18, no. 1, pp. 36–51, Jan. 2009. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4694003

[31] K.-A. Do, P. Müller, and M. Vanucci, *Bayesian Inference for Gene Expression and Proteomics*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[32] P. Szacherski, J.-F. Giovannelli, and P. Grangeat, "Joint Bayesian hierarchical inversion-classification and application in proteomics," in *Proc. IEEE Workshop Statist. Signal Process.*, Jun. 2011, pp. 121–124. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5967636

[33] R. Pérenon, A. Mohammad-Djafari, L. Duraffourg, and P. Grangeat, "Quantification moléculaire par spectrométrie de masse à base de NEMS: Modélisation et inversion du problème," in *Proc. 23rd Colloq. GRETSI*, Bordeaux, France, 2011.

[34] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain—Computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6046114

[35] M. Guindani, K.-A. Do, P. Müeller, and J. S. Morris, "Bayesian mixture models for gene expression and protein profiles," in *Bayesian Inference for Gene Expression and Proteomics*. Cambridge, U.K.: Cambridge Univ. Press, 2006, pp. 238–253.

[36] I. Kuselman, F. Pennecchi, C. Burns, A. Fajgelj, and P. de Zorzi, "Investigating out-of-specification test results of chemical composition based on metrological concepts," *Accreditation Qual. Assurance*, vol. 15, no. 5, pp. 283–288, Nov. 2009. [Online]. Available: http://www.springerlink.com/index/10.1007/s00769–009-0618–4

[37] N. Parrish, M. R. Gupta, and H. S. Anderson, "Robust classification of signal estimates given a channel model," in *Proc. IEEE Statist. Signal Process. Workshop (SSP)*, Nice, France, Jun. 2011, pp. 273–276.

[38] P. P. Wang, D. Ruan, and E. E. Kerre, *Fuzzy Logic: A Spectrum of Theoretical & Practical Issues*. New York, NY, USA: Springer-Verlag, 2007.
[39] C. P. Robert, *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. New York, NY, USA: Springer-Verlag, 2007.
[40] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (Texts in Statistical Science), 2nd ed. London, U.K.: Chapman & Hall, Jul. 2003.
[41] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer-Verlag, 2004.
[42] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995. [Online]. Available: http://biomet.oxfordjournals.org/cgi/doi/10.1093/biomet/82.4.711
[43] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer-Verlag, 2009.
[44] P. Szacherski *et al.*, "MRM protein quantification and serum sample classification," in *Proc. 61st ASMS Conf.*, Minneapolis, MN, USA, Jun. 2013, pp. 1–2.
[45] F. Adjed, "Classification, apprentissage et sélection de modèles pour un mélange de populations appliqués en protéomique," IMS, Bordeaux, France, Tech. Rep., 2012.
[46] P. Grangeat *et al.*, "Convergence entre l'analyse biostatistique et les méthodes d'inversion hiérarchique bayésienne pour la recherche et la validation de biomarqueurs par spectrométrie de masse," in *Proc. 24th Colloq. GRETSI*, Brest, France, Sep. 2013.
[47] L. Gerfault *et al.*, "Statistical analysis of Bayesian hierarchical inversion for MRM protein quantification and QDA serum sample classification," in *Proc. 62nd ASMS Conf. Mass Spectrometry Allied Topics*, Baltimore, MD, USA, Jun. 2014.
[48] L. Gerfault *et al.* "Assessing MRM protein quantification and serum sample classification performances of a Bayesian hierarchical. Inversion method on a colorectal cancer cohort," in *Proc. EuPA Sci. Meeting*, Saint-Malo, France, Oct. 2013.
[49] P. Szacherski, J.-F. Giovannelli, L. Gerfault, and P. Grangeat, "Apprentissage supervisé robuste de caractéristiques de classes. Application en protéomique," in *Proc. 23rd Colloq. GRETSI*, Bordeaux, France, 2011, Sep. 2011.
[50] F. Adjed, J.-F. Giovannelli, A. Giremus, N. Dridi, and P. Szacherski, "Variable selection for a mixed population applied in proteomics," in *Proc. 38th Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 1153–1157.
[51] P. Szacherski, "Reconstruction de profils protéiques pour la recherche de biomarqueurs," Ph.D. dissertation, Dept. Comput. Sci., Univ. Bordeaux I, Talence, France, Dec. 2012.

PASCAL SZACHERSKI was born in Dortmund, Germany, in 1983. He received the B.Sc. degree in mathematics from the University of Dortmund, Dortmund, in 2007, the M.Sc. degree in harmonic signal processing and control from the University of Bordeaux, Bordeaux, France, in 2009, and the Ph.D. degree from the University of Bordeaux, in 2012, for his work on biomarker discovery by the use of Bayesian methods for inverse problems, carried out in collaboration with the CEA Leti and within the BHI-PRO project. He is currently with InvenSense, Inc., San Jose, CA, USA, developing algorithms for end-user applications using inertial sensor data and data fusion. His research interests include statistical signal/data processing, Bayesian data analysis for inverse problems, computational proteomics, and various classification problems.

JEAN-FRANÇOIS GIOVANNELLI was born in Béziers, France, in 1966. He received the degree from the École Nationale Supérieure de l'Électronique et de ses Applications, Cergy, France, in 1990, and the Ph.D. degree and the H.D.R. degree in physics with a minor in signal-image processing from the Université Paris-Sud, Orsay, France, in 1995 and 2005, respectively. From 1997 to 2008, he was an Assistant Professor with the Université Paris-Sud, where he was also a Researcher with the Laboratoire des Signaux et Systèmes, Groupe Problèmes Inverses. He is currently a Professor with the Université de Bordeaux, Bordeaux, France, where he is also a Researcher with the Laboratoire d'Intégration du Matériau au Système, Groupe Signal-Image. He is interested in regularization and Bayesian methods for inverse problems in signal and image processing, mainly unsupervised and myopic problems. His application fields essentially concern astronomical, medical, proteomics, and geophysical imaging.

LAURENT GERFAULT received the degree in electronic and signal processing from the University of Science of Lyon, Lyon, France, and the Ph.D. degree in acoustic from the National Institute of Applied Sciences of Lyon, Villeurbanne Cedex, France, in 2000, for the study of imaging of ultrasound contrast agents. In 2000, he joined the Laboratory of Electronics and Information Technology at the Atomic Energy and Alternative Energies Commission, Grenoble, France, as a Research Engineer in Image and Signal Processing. He was involved in X-ray and gamma imaging projects, biofluorescence microsystems, and proteomic studies. The common aspect through these projects is signal processing using physical modeling.

PIERRE MAHÉ was born in Compiègne, France, in 1979. He received the degree from the Institut National des Sciences Appliquées of Rouen, Saint-Étienne du Rouvray, France, in 2002, and the Ph.D. degree from the École des Mines de Paris, Paris, France, in 2006, for his work on kernel methods with applications in chemoinformatics. Since 2008, he has been a Research Scientist with the Department of Bioinformatics Research, bioMérieux, Craponne, France. His research interests involve statistical learning for the analysis of high-dimensional and structured data, with applications in clinical microbiology and biomarker discovery.

JEAN-PHILIPPE CHARRIER received the Biological Engineering degree and the D.E.A. degree in industrial process engineering from Compiègne Technology University, Compiègne, France. He joined bioMérieux, Craponne, France, in 1992, where he is currently a Principal Scientist in Mass Spectrometry with the Innovation Unit, Department of Technology Research. Since 1995, he has been involved in developing proteomics research with bioMérieux, mainly in prostate and colorectal cancer, and microbiology. During the last years, he contributed to the development of VITEK-MS platform for microbiological identification using matrix-assisted laser desorption/ionization time-of-flight and protein quantization using multiple reaction monitoring (MRM) and cubed MRM.

AUDREY GIREMUS received the Engineering degree and the Ph.D. degree in signal processing from the École Nationale Supérieure de l'Aéronautique et de l'Espace, Toulouse, France, in 2002 and 2005, respectively. She is currently an Associate Professor with the University of Bordeaux, Bordeaux, France. Since 2006, she has been with the Signal and Image Research Group, IMS Laboratory, Detroit, MI, USA. Her research interests include statistical signal processing and optimal filtering techniques applied to navigation, mobile communications, and biomedical engineering.

BRUNO LACROIX received the advanced master's degree from École Polytechnique, Palaiseau, France, in 1989, and the Ph.D. degree in bioinformatics from the University of Paris VI, Paris, France, in 1992. He joined bioMérieux, Craponne, France, in 1997, where he was appointed as the Senior Director with the Department of I&S/Technology Research in 2012.

PIERRE GRANGEAT (M'85–SM'07) received the Telecommunication Engineering and Ph.D. degrees from Télécom ParisTech, Paris, France, in 1981 and 1987, respectively, and the H.D.R. degree from the Institut National Polytechnique de Grenoble, Grenoble, France, in 1993. His main research interests are in information processing and inverse problems for biomedical technologies. He joined the Laboratory of Electronics and Information Technology (LETI), Atomic Energy and Alternative Energies Commission (CEA), Grenoble, France, in 1982, first as a Research Engineer on reconstruction algorithms for cone-beam 3-D image reconstruction, second, from 1987 to 2004, as a Project Manager on medical 3-D and 4-D tomography using X-ray, gamma ray, or positron, with several industrial partners. Since 2004, his researches have been directed toward information processing applied to microsystems for biology and healthcare. His one main research topic is molecular profiling and clinical proteomics associated with lab-on-chip measurement. From 2006 to 2010, he was the Scientific Director of the LOCCANDIA European project (IST-2005-43202) Lab-On-Chip protein profiling for CANcer DIAgnosis and the Project Manager of the CEA project CAPSI dedicated to the development of a high-sensitivity proteomic analytical chain based on integrated components. From 2011 to 2013, he was the Coordinator of the ANR BHI-PRO project on Bayesian hierarchical inversion for mass spectrometry applied to discovery and validation of new protein biomarkers. He was appointed as the CEA Research Director in 2002. Since 2004, he has been with the Scientific Direction, Microtechnology for Biology and Healthcare Division, CEA-LETI. Since 2009, he has been with the Internal Scientific Committee of LETI.