# BAYESIAN LINEAR UNMIXING OF TIME-EVOLVING GENE EXPRESSION DATA USING A HIDDEN MARKOV MODEL

*Cécile Bazot*$^{(1)}$*, Nicolas Dobigeon*$^{(1)}$*, Jean-Yves Tourneret*$^{(1)}$ *and Alfred O. Hero III*$^{(2)}$

$^{(1)}$ University of Toulouse, IRIT/INP-ENSEEIHT, Toulouse, France
$^{(2)}$ University of Michigan, EECS Dept., Ann Arbor, USA
{cecile.bazot, nicolas.dobigeon, jean-yves.tourneret}@enseeiht.fr, hero@umich.edu

## ABSTRACT

This paper describes a new hierarchical temporal Bayesian model and a Markov chain Monte Carlo (MCMC) algorithm for gene factor analysis. Each data sample is decomposed as a linear combination of characteristic gene signatures (also called *factors*) with appropriate proportions, or *factor scores*, following a linear mixing model (LMM). The particularity of the proposed algorithm is that the LMM model is combined with a hidden Markov model (HMM) to take into account temporal dependencies between the samples. The proposed HMM structure is motivated by the behavior of host molecular response following exposure to an infectious agent. The complexity of the posterior distribution resulting from the proposed HMM is alleviated by using a hybrid Gibbs sampler that generates samples distributed according to this distribution. These samples are then used to approximate the standard Bayesian estimators of the unknown parameters. The performance of the proposed method is illustrated by simulations conducted on synthetic data and on a real public dataset.

***Index Terms*—** Bayesian inference, factor analysis, hidden Markov model, time-evolving gene expression data

## 1. INTRODUCTION AND PROBLEM STATEMENT

During the last decades, factor analysis methods have been widely studied and applied to gene microarray samples for discovering the patterns of differential expression in time course experiments. The aim of these methods is to find an interpretable decomposition of an observation matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{R}^{G\times N}$ whose columns (resp. rows) correspond to samples (resp. gene expression levels). Each observed sample vector $\mathbf{y}_i$ ($i = 1, \ldots, N$), of $G$ gene expression levels, is assumed to satisfy a linear mixing model (LMM)

$$\mathbf{y}_i = \sum_{r=1}^{R} \mathbf{m}_r a_{r,i} + \mathbf{n}_i \tag{1}$$

where $\mathbf{m}_r = [m_{1,r}, \ldots, m_{G,r}]^T$ denotes the $r$th gene signature vector, also referred to as *factor*, $a_{r,i}$ is the contribution (or *factor score*) of the $r$th gene signature in the $i$th observed sample, $R$ is the number of gene signatures present in the chip and $\mathbf{n}_i$ denotes a residual error. Considering $N$ samples, the LMM model can be rewritten as $\mathbf{Y} = \mathbf{MA} + \mathbf{N}$ where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \in \mathbb{R}^{R\times N}$ represents the factor score matrix, $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R] \in \mathbb{R}^{G\times R}$ the factor loading matrix and $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_N] \in \mathbb{R}^{G\times N}$ the matrix of residual errors. The proposed model also incorporates non-negativity constraints on the factors ($m_{g,r} \geq 0$) and factor scores ($a_{r,i} \geq 0$), as well as a sum-to-one constraint on the factor scores ($\sum_{r=1}^{R} a_{r,i} = 1$), as motivated in [1, 2]. As in other Bayesian gene factor analysis methods, the residual error vectors $\mathbf{n}_i$ are assumed to be independent and identically distributed (i.i.d.) zero-mean Gaussian with covariance matrix $\mathbf{\Sigma} = \sigma^2\mathbf{I}_G$

$$\mathbf{n}_i|\sigma^2 \sim \mathcal{N}\left(\mathbf{0}_G, \sigma^2\mathbf{I}_G\right) \tag{2}$$

where $\mathbf{I}_G$ is the identity matrix of dimension $G \times G$ and $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$ denotes the multivariate Gaussian distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{\Sigma}$.

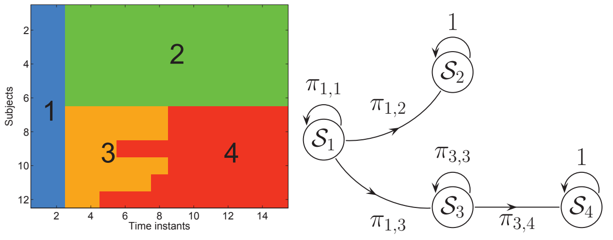
The objective of linear unmixing is to estimate the factor matrix $\mathbf{M}$ and the factor score matrix $\mathbf{A}$ jointly from the available data samples contained in $\mathbf{Y}$. Such approach has already been developed in [1, 2] for gene expression microarrays where the authors particularly focused on the estimation of the number of factors $R$ additionally to the unmixing. In this paper, we extend the method proposed in [1] to exploit temporal dependencies between samples, using a hidden Markov model (HMM). Indeed, HMMs are useful and popular tools for analyzing time-varying data, also called *time-evolving* data in the biostatistics literature. They have been recently adapted for gene expression data analysis in other contexts [3, 4].

This paper is organized as follows. Section 2 describes the proposed HMM designed to incorporate temporal dependencies. Section 3 presents the Bayesian model based on the HMM. Section 4 studies an hybrid Gibbs sampler generating samples distributed according to the posterior distribution associated with the proposed Bayesian model. The resulting algorithm is applied on both synthetic and real time-evolving gene expression data in Section 5. Conclusions are reported in Section 6.

## 2. DEFINING TEMPORAL DEPENDENCIES USING HMMS

The observation matrix $\mathbf{Y}$ is composed of $N$ columns corresponding to the $N$ samples collected on $S$ individuals at $T$ time instants, so that $N = ST$. Previous results on real time-evolving gene expression data [1, 5] have shown that the individual host molecular responses cluster into $K = 4$ states denoted as $\mathcal{S}_1, \ldots, \mathcal{S}_K$ and defined as follows: before inoculation ($\mathcal{S}_1$), post-inoculation asymptomatic ($\mathcal{S}_2$), pre-onset-symptomatic (before significant symptoms occur) ($\mathcal{S}_3$) and post-onset-symptomatic ($\mathcal{S}_4$). To identify the state of a given individual $s$ at a given instant $t$, we introduce a discrete latent variable $z_{s,t}$ that takes its values in the finite set $\{1, \ldots, K\}$. Hence, $z_{s,t} = k$ if and only if the $t$th sample of the $s$th subject is in the $k$th state $\mathcal{S}_k$. Let $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_S]^T \in \mathbb{R}^{S\times T}$ denote the label matrix giving information about the state of all samples, for each subject ($s = 1, \ldots, S$) and time instant ($t = 1, \ldots, T$). The vector $\mathbf{z}_s = [z_{s,1}, \ldots, z_{s,T}]$ is the label vector of the $s$th subject states. A schematic view of this classification process is depicted in Fig. 1(a), where the state of a given individual over time (resp. at a given time instant over individuals) appears in the rows (resp. columns) of this classification matrix.

To exploit the temporal evolution of the molecular responses, these $K$ states are modeled using an HMM assigned to the latent variables gathered in the label matrix $\mathbf{Z}$ (see [6] for more details on HMMs). Fig. 1(b) shows the proposed $K$-state HMM associated with the temporal structure depicted in Fig. 1(a).



(a) (b)

**Fig. 1**. Left: classification matrix. Right: the 4-state Markov model.

Based on this directed graph, assuming the transition probabilities are independent of the considered subject, the state transition probability matrix $\mathbf{\Pi}$ can be decomposed as

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{1,1} & \pi_{1,2} & \pi_{1,3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \pi_{3,3} & \pi_{3,4} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

with the initial state probability vector $\boldsymbol{\pi}^{(0)}$

$$\boldsymbol{\pi}^{(0)} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

and where $\pi_{k,k'} = \mathrm{P}[z_{s,t} = k' \,|\, z_{s,t-1} = k]$ for $1 \leq k, k' \leq K$ and $t = 2, \ldots, T$. Note also that $\pi_{1,3} = 1 - \pi_{1,1} - \pi_{1,2}$ and $\pi_{3,4} = 1 - \pi_{3,3}$. The HMM state transition matrix (3) and diagram in Fig. 1(b) are motivated by the well-known susceptible-infectious-recovered (SIR) model [7] for host molecular response. In particular, the estimated transition probabilities $\pi_{k,k'}$ might be used for clinical interpretation. However, the proposed transition matrix is easily generalized to other models with the caveat that the matrix should be sparse in small sample size situations.

## 3. HIERARCHICAL BAYESIAN MODEL

This section introduces the hierarchical Bayesian model used to estimate the unknown parameters $\mathbf{\Theta} = \{\mathbf{M}, \mathbf{A}, \mathbf{Z}, \sigma^2\}$. This model is based on the likelihood of the observations and on prior distributions for the unknown parameters and hyperparameters.

### 3.1. Likelihood

The model and the statistical properties of the error vectors $\mathbf{n}_i$ defined in Section 1 yield a conditionally Gaussian distribution for the $i$th observed sample, i.e., $\mathbf{y}_i|\mathbf{M}, \mathbf{a}_i, \sigma^2 \sim \mathcal{N}\left(\mathbf{M}\mathbf{a}_i, \sigma^2\mathbf{I}_G\right)$. Assuming the $N$ samples are independent, the likelihood function of $\mathbf{Y}$ can be written as

$$f(\mathbf{Y}|\mathbf{M}, \mathbf{A}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{GN/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\|\mathbf{y}_i - \mathbf{M}\mathbf{a}_i\|^2\right] \tag{5}$$

where $\|\cdot\|$ is the standard $l_2$-norm.

### 3.2. Parameter priors

In this section, we introduce prior distributions for the unknown parameters contained in $\mathbf{\Theta} = \{\mathbf{M}, \mathbf{A}, \mathbf{Z}, \sigma^2\}$.

#### *3.2.1. Factor loading prior*

Following [1] and [8], we propose to estimate the projections $\mathbf{t}_r$ of the factors $\mathbf{m}_r$ $(r = 1, \ldots, R)$ onto a lower-dimensional subspace $\mathcal{V}_{R-1}$ of $\mathbb{R}^{G-1}$ of dimension $R-1$. More precisely, $\mathcal{V}_{R-1}$ is identified by a standard dimension reduction method such as the principal component analysis (PCA). The factors $\mathbf{m}_r$ and their corresponding projections $\mathbf{t}_r$ are related by $\mathbf{t}_r = \mathbf{P}(\mathbf{m}_r - \bar{\mathbf{y}})$ where $\bar{\mathbf{y}}$ is the empirical mean of the observed samples and $\mathbf{P}$ is the $(R-1) \times G$ projection matrix onto $\mathcal{V}_{R-1}$. The projected factors $\mathbf{t}_r$ are then assigned a multivariate Gaussian distribution (MGD) $\mathcal{N}_{\mathcal{T}_r}\left(\mathbf{e}_r, s_r^2\mathbf{I}_{r-1}\right)$ truncated on the set $\mathcal{T}_r$. The truncation on the set $\mathcal{T}_r$ (defined in [8]) ensures that all the components of the factor signatures are positive. The mean vectors $\mathbf{e}_r$ are fixed using available prior knowledge or provided by an endmember extraction algorithm, e.g. the vertex component analysis (VCA) [9] for hyperspectral imaging. Considering *a priori* independence between the $R$ projected factors, the joint prior distribution for $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_R]$ is $f(\mathbf{T}) = \prod_{r=1}^{R} f(\mathbf{t}_r)$.

#### *3.2.2. Factor score prior*

As demonstrated in [5], molecular host responses of asymptomatic and symptomatic subjects mainly differ in expression levels of the factors. Consequently, the prior distributions of the factor score vectors $\{\mathbf{a}_i\}_{i=1,\ldots,N}$ are assumed to be distinct for scores associated with different states $\mathcal{S}_1, \ldots, \mathcal{S}_K$. Moreover, to promote interpretability of the results as in [4], the factor score vectors $\{\mathbf{a}_i\}_{i=1,\ldots,N}$ are assumed to satisfy the non-negativity and the sum-to-one constraints defined in Section 1. Therefore, a Dirichlet distribution is a natural prior distribution for the factor score vectors $\mathbf{a}_i$ $(i = 1, \ldots, N)$ conditionally to the label $k$ assigned to the $i$th sample[1] $\mathbf{y}_i$

$$\mathbf{a}_i|z_i = k, \boldsymbol{\delta}_k \sim \mathcal{D}_R\left(\boldsymbol{\delta}_k\right)$$

where $\mathcal{D}_R\left(\boldsymbol{\delta}_k\right)$ is a Dirichlet distribution with parameters $\boldsymbol{\delta}_k = [\delta_{1,k}, \ldots, \delta_{R,k}]^T$. Assuming all factor score vectors $\{\mathbf{a}_i\}_{i=1,\ldots,N}$ are *a priori* independent, the joint prior distribution for the factor score matrix $\mathbf{A}$ is

$$f\left(\mathbf{A}|\mathbf{Z}, \Delta\right) = \prod_{k=1}^{K}\prod_{i\in\mathcal{C}_k} f\left(\mathbf{a}_i|z_i = k, \boldsymbol{\delta}_k\right)$$

with $\Delta = [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K] \in \mathbb{R}^{R\times K}$ and $\mathcal{C}_k = \{i = 1, \ldots, N | z_i = k\}$ denotes the subset of sample indexes associated with the $k$th label.

#### *3.2.3. Label prior*

The prior probabilities of the latent variables $z_i$ $(i = 1, \ldots, N)$ are given by the initial state matrix $\boldsymbol{\pi}^{(0)}$ and the transition state matrix $\mathbf{\Pi}$ previously defined in (4) and (3). Some state transition probabilities ($\pi_{1,1}$, $\pi_{1,2}$ and $\pi_{3,3}$) are unknown and will be estimated using a hierarchical Bayesian algorithm.

#### *3.2.4. Noise variance prior*

As in common practice [1, 8], a conjugate inverse-Gamma distribution with parameters $\nu/2$ and $\gamma/2$ is chosen as prior distribution for the noise variance

$$\sigma^2|\nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right). \tag{6}$$

The shape parameter $\nu$ will be fixed to $\nu = 2$ whereas the scale parameter $\gamma$ will be an adjustable hyperparameter.

Assuming *a priori* independence between the individual parameters, the following prior is obtained

$$f(\mathbf{\Theta}|\Delta, \gamma) = \mathrm{P}\left[\mathbf{Z}\right] f(\mathbf{T}) f(\mathbf{A}|\Delta) f(\sigma^2|\nu, \gamma). \tag{7}$$

[1] Note that, for conciseness, the latent variable $z_i$ is here indexed by a single index for brevity. Of course, there is a direct relationship between this index $i$ and the couple of indices $(s, t)$ introduced in Section 2.

### 3.3. Hyperparameter priors

The accuracy of the proposed Bayesian estimation algorithm depends on the values of the hyperparameters $\Delta$, $\mathbf{\Pi}$ and $\gamma$. The approach investigated here consists of assigning appropriate priors to these hyperparameters (also referred to as *hyperpriors*). Due to the lack of prior information for these hyperparameters, we have chosen non-informative hyperpriors. More precisely, the hyperparameters $\Delta$ of the factor score vectors are assigned an improper uniform distribution on $\mathbb{R}^{+}$, i.e.,

$$f(\Delta) \propto \mathbf{1}_{\mathbb{R}_{+}^{RK}}(\Delta)$$

where $\propto$ stands for "proportional to".

Denote as $\boldsymbol{\pi}_1 = [\pi_{1,1}, \pi_{1,2}, \pi_{1,3}]$ and $\boldsymbol{\pi}_3 = [\pi_{3,3}, \pi_{3,4}]$ the unknown state transition probability subvectors of the matrix $\mathbf{\Pi}$. Following [10], a Dirichlet distribution with parameter vector $\boldsymbol{\alpha_i}$ $(i = 1, 3)$ is chosen as prior distribution for each of these unknown state transition probability subvector $\boldsymbol{\pi}_i$, i.e.,

$$\boldsymbol{\pi}_1 | \boldsymbol{\alpha_1} \sim \mathcal{D}_3(\boldsymbol{\alpha_1}), \quad \boldsymbol{\pi}_3 | \boldsymbol{\alpha_3} \sim \mathcal{D}_2(\boldsymbol{\alpha_3}).$$

Due to the lack of information regarding $\mathbf{\Pi}$, all values of the Dirichlet parameter vectors $\{\boldsymbol{\alpha_i}\}_{i=1,3}$ are assumed to be equal to 1.

A non-informative Jeffreys' prior is assigned to the noise hyperparameter $\gamma$

$$f(\gamma) \propto \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^{+}}(\gamma).$$

Assuming that all the individual hyperparameters of this hierarchical Bayesian model are a priori independent, the full posterior distribution of the hyperparameter vector $\mathbf{\Psi} = \{\Delta, \mathbf{\Pi}, \gamma\}$ is

$$f(\mathbf{\Psi}) = f(\Delta) f(\mathbf{\Pi}) f(\gamma). \tag{8}$$

### 3.4. Posterior distribution

The joint posterior distribution of the unknown parameter vector $\mathbf{\Theta} = \{\mathbf{T}, \mathbf{A}, \mathbf{Z}, \sigma^2\}$ and the hyperparameter vector $\mathbf{\Psi} = \{\Delta, \mathbf{\Pi}, \gamma\}$ can be computed from the hierarchical structure

$$f(\mathbf{\Theta}, \mathbf{\Psi} | \mathbf{Y}) \propto f(\mathbf{Y} | \mathbf{\Theta}) f(\mathbf{\Theta} | \mathbf{\Psi}) f(\mathbf{\Psi}) \tag{9}$$

where $f(\mathbf{Y}|\mathbf{\Theta})$, $f(\mathbf{\Theta}|\mathbf{\Psi})$ and $f(\mathbf{\Psi})$ have been respectively defined in (5), (7) and (8). Due to the constraints enforced on the data and the proposed temporal HMM model, the joint posterior distribution $f(\mathbf{\Theta}, \mathbf{\Psi}|\mathbf{Y})$ defined in (9) is far too complex to derive analytical expressions for the Bayesian estimators of the unknown parameters. To alleviate this problem, it is natural to use Markov chain Monte Carlo (MCMC) methods [11] to generate samples asymptotically distributed according to (9) and to compute Bayesian estimators using these generated samples.

## 4. HYBRID GIBBS SAMPLER

This section presents a Metropolis-within-Gibbs (MwG) sampling strategy that generates random samples asymptotically distributed according to the joint posterior distribution of interest $f(\mathbf{\Theta}, \mathbf{\Psi}|\mathbf{Y})$ defined in (9). The Gibbs sampler iteratively generates samples distributed according to the full conditional distributions of the target distribution. The principle of the MwG sampler is to use a Metropolis move for any conditional distribution that cannot be sampled directly. The different steps of the MwG proposed for the Bayesian unmixing of gene expression data are detailed below.

### 4.1. Sampling from $f(\mathbf{T}|\mathbf{A}, \sigma^2, \mathbf{Y})$

Gibbs moves are used to sample from $f(\mathbf{T}|\mathbf{A}, \sigma^2, \mathbf{Y})$ (see [8] for details).

### 4.2. Sampling from $f\left(\mathbf{a}_i | \mathbf{M}, z_i = k, \sigma^2, \boldsymbol{\delta}_k, \mathbf{y}_i\right)$

For each sample $i$ $(i = 1, \ldots, N)$, the conditional posterior distribution of the factor score vector $\mathbf{a}_i$ is

$$\begin{aligned} & f\left(\mathbf{a}_i | \mathbf{M}, z_i = k, \sigma^2, \boldsymbol{\delta}_k, \mathbf{y}_i\right) \\ & \quad \propto \prod_{r=1}^{R} a_{r,i}^{\delta_{r,k}-1} \times \exp\left(-\frac{\|\mathbf{y}_i - \mathbf{M}\mathbf{a}_i\|^2}{2\sigma^2}\right) \mathbf{1}_{\mathcal{A}}(\mathbf{a}_i). \end{aligned} \tag{10}$$

Since generating samples according to (10) is not straightforward, we propose to use a Metropolis-Hastings step. We choose an MGD truncated on the simplex $\mathcal{A}$ as the proposal distribution for the first $R-1$ scores, $\mathbf{a}_{1:R-1,i} = [a_{1,i}, \ldots, a_{R-1,i}]$. Then, the sum-to-one constraint enforced on the scores allows the remaining coefficient to be computed, i.e., $a_{R,i} = 1 - \sum_{r=1}^{R-1} a_{r,i}$.

### 4.3. Sampling from $\mathbf{P}\left[z_{s,t} = k | z_{s,t-1}, \mathbf{a}_i, \boldsymbol{\delta}_k, \boldsymbol{\pi}^{(\mathbf{0})}, \mathbf{\Pi}\right]$

For the $t$th sample of subject $\#s$ (associated with $\mathbf{y}_i$ and $\mathbf{a}_i$) straightforward computations yield to the following result

$$\begin{aligned} & \mathrm{P}\left[z_{s,t} = k | z_{s,t-1}, \mathbf{a}_i, \boldsymbol{\delta}_k, \boldsymbol{\pi}^{(\mathbf{0})}, \mathbf{\Pi}\right] \\ & \quad \propto \mathrm{P}[z_{s,t} = k | z_{s,t-1}] f\left(\mathbf{a}_i | z_{s,t-1} = k, \boldsymbol{\delta}_k\right) f\left(\mathbf{y}_i | \mathbf{a}_i, \sigma^2\right) \\ & \quad \propto \mathrm{P}[z_{s,t} = k | z_{s,t-1}] \frac{\Gamma\left(\sum_{r=1}^{R} \delta_{r,k}\right)}{\prod_{r=1}^{R} \delta_{r,k}} \prod_{r=1}^{R} a_i^{\delta_{r,k}-1} \mathbf{1}_{\mathcal{A}}(\mathbf{a}_i). \end{aligned} \tag{11}$$

The probabilities $\mathrm{P}[z_{s,t} = k | z_{s,t-1}]$ are defined in (3) and (4). Sampling from this discrete conditional distribution can be achieved by drawing a value in the finite set $\{1, \ldots, K\}$ with the normalized probabilities (11).

Unfortunately we have to cope with the *label switching* problem that can occur when assigning labels $\mathcal{S}_2$ and $\mathcal{S}_3$. This is a common problem due to the lack of identifiability in HMM models such as ours (see [12, p. 478] for more details on label switching). To solve the label switching problem, a common approach is to enforce constraints. Here, since the fluctuations of the factor scores associated with symptomatic subjects are expected to be greater than those associated with asymptomatic subjects, the variances of asymptomatic scores are enforced to be lower than those of symptomatic scores, avoiding the label switching problem.

### 4.4. Sampling from $f\left(\sigma^2 | \mathbf{M}, \mathbf{A}, \gamma, \mathbf{Y}\right)$

Using (5) and (6), the conditional distribution of the noise variance $\sigma^2 | \mathbf{M}, \mathbf{A}, \gamma, \mathbf{Y}$ is the following inverse-Gamma distribution

$$\sigma^2 | \mathbf{M}, \mathbf{A}, \gamma, \mathbf{Y} \sim \mathcal{IG}\left(\frac{GN}{2}, \frac{1}{2} \sum_{i=1}^{N} \|\mathbf{y}_i - \mathbf{M}\mathbf{a}_i\|^2\right),$$

that is easy to sample.

### 4.5. Sampling from $f(\delta_{r,k}|\mathbf{a}_r,\mathbf{Z})$

For each factor $r$ ($r = 1,\ldots,R$) and each state $\mathcal{S}_k$ ($k = 1,\ldots,K$), the Dirichlet parameters $\delta_{r,k}$ are generated according to

$$\begin{aligned} f(\delta_{r,k}|\mathbf{a}_r,\mathbf{Z}) &\propto f(\delta_{r,k}) \prod_{i\in\mathcal{C}_k} f(\mathbf{a}_i|z_i=k,\boldsymbol{\delta}_k) \\ &\propto \prod_{i\in\mathcal{C}_k}\left[\frac{\Gamma(\sum_{r=1}^{R}\delta_{r,k})}{\Gamma(\delta_{r,k})} a_{r,i}^{\delta_{r,k}-1}\right]\mathbf{1}_{\mathbb{R}^+}(\delta_{r,k}). \end{aligned} \tag{12}$$

A Metropolis-Hastings step is employed to generate samples distributed according to (12). More precisely, samples are generated using a random-walk with a truncated Gaussian instrumental distribution $\mathcal{N}(0,w^2)$. The variance $w^2$ is fixed in order to obtain an acceptance rate between 0.15 and 0.50, as recommended in [13, p. 55].

### 4.6. Sampling from $f(\boldsymbol{\Pi}|\mathbf{Z})$

Straightforward mathematical manipulations lead to a Dirichlet distribution as conditional distribution for the unknown state transition probability vector $\boldsymbol{\pi}_i$ ($i = 1,3$) with parameters $\boldsymbol{\alpha}_i + N_i$, where $N_1 = [n_{1,1}, n_{1,2}, n_{1,3}]$, $N_3 = [n_{3,3}, n_{3,4}]$, and $n_{i,j} = \#\{t \,|\, z_{s,t} = j \text{ and } z_{s,t-1} = i\}$. More precisely, $n_{1,2}$ corresponds to the number of asymptomatic subjects and $n_{1,3} = n_{3,4}$ is the number of symptomatic subjects.

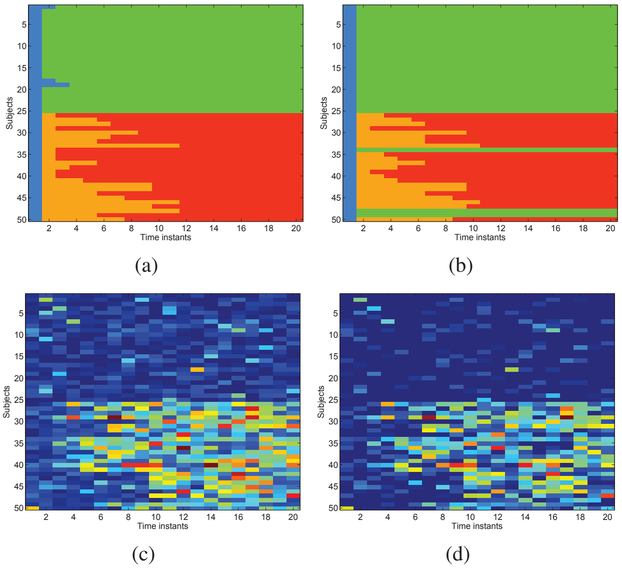
## 5. SIMULATION RESULTS

### 5.1. Synthetic data

The performance of the proposed algorithm is first illustrated on a synthetic dataset consisting of $N = 1000$ samples, more precisely $S = 50$ subjects and $T = 20$ time instants. Each sample vector is composed of exactly $R = 4$ factors, with $G = 12000$ gene expression levels. To generate realistic signatures, the factors have been extracted from previous results obtained on real time-evolving gene expression dataset and have been mixed using the LMM model (1). The synthetic state map, represented in Fig. 2(a), has been generated according to the 4-state Markov chain in Fig. 1(b) with $\boldsymbol{\pi}_1 = [0.1, 0.45, 0.45]$, and $\boldsymbol{\pi}_3 = [0.8, 0.2]$. The observed vectors have been corrupted by an i.i.d. Gaussian noise sequence with signal-to-noise ratio SNR $= 20$ dB. The hidden mean vectors $\mathbf{e}_r$ ($r = 1,\ldots,R$) have been chosen as the PCA projections of signatures previously identified by VCA [9].

The proposed algorithm has been run with $\mathrm{N}_{\mathrm{mc}} = 1000$ MCMC iterations (with $\mathrm{N}_{\mathrm{bi}} = 100$ burn-in iterations). The MAP estimators of the unknown parameters have then been computed from the generated samples. For instance, the marginal MAP estimates $\widehat{\mathbf{Z}}$ of the state matrix $\mathbf{Z}$ depicted in Fig. 2(b) are globally in good agreement with the actual states (Fig. 2(a)). The corresponding MAP estimates of the inflammatory factor scores are displayed in Fig. 2(d). These estimates also agree with the ground truth shown in Fig. 2(c). The MMSE estimates of the unknown transition probabilities are $\widehat{\boldsymbol{\pi}}_1 = [0.02, 0.55, 0.43]$ and $\widehat{\boldsymbol{\pi}}_3 = [0.83, 0.17]$. The confusion matrix displayed in Table 1 shows the performance of the classifier based on the estimated label matrix $\widehat{\mathbf{Z}}$. From this confusion matrix, one can compute the overall accuracy of the classification, i.e., the percentage of correctly classified samples. The overall accuracy is 92.5%.

The performance of the proposed model is compared with the non-temporal Bayesian linear unmixing model developed in [5], by using the following criteria:



(a)

(b)

(c)

(d)

**Fig. 2**. Top left: actual synthetic label matrix $\mathbf{Z}$. Top right: marginal MAP estimate of the label matrix $\widehat{\mathbf{Z}}$. Bottom left: actual synthetic inflammatory factor score $\mathbf{A}$. Bottom right: MAP estimate of the inflammatory factor score $\widehat{\mathbf{A}}$.

**Table 1**. Confusion matrix for the state classification ($\mathcal{S}_1,\ldots,\mathcal{S}_4$).

| | | Actual $\mathbf{Z}$ | | | | |
|---|---|---|---|---|---|---|
| | | $\mathcal{S}_1$ | $\mathcal{S}_2$ | $\mathcal{S}_3$ | $\mathcal{S}_4$ | Total |
| Estimated $\widehat{\mathbf{Z}}$ | $\mathcal{S}_1$ | **50** | 0 | 0 | 0 | 50 |
| | $\mathcal{S}_2$ | 4 | **471** | 15 | 42 | 532 |
| | $\mathcal{S}_3$ | 0 | 0 | **98** | 12 | 110 |
| | $\mathcal{S}_4$ | 0 | 0 | 2 | **306** | 308 |
| | Total | 54 | 471 | 115 | 360 | 1000 |

- the factor mean square errors (MSE) $\mathrm{MSE}_r^2 = \frac{1}{G}\|\hat{\mathbf{m}}_r - \mathbf{m}_r\|^2$, $r = 1,\ldots,R$ where $\hat{\mathbf{m}}_r$ is the estimated $r$th factor loading vector,
- the global MSE of factor scores $\mathrm{GMSE}_r^2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{a}_{r,i} - a_{r,i})^2$, $r = 1,\ldots,R$ where $\hat{a}_{r,i}$ is the estimated proportion of the $r$th factor in the $i$th sample,
- the reconstruction error (RE) $\mathrm{RE} = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|^2$ where $\hat{\mathbf{y}}_i = \sum_{r=1}^{R}\hat{\mathbf{m}}_r\hat{a}_{r,i}$ is the estimate of $\mathbf{y}_i$,
- the computational time.

The results reported in Table 2 show that the proposed algorithm performs similarly or better than its non-temporal version. The proposed model has also the great advantage of providing a classification of the samples according to the state of a subject at a given time instant (Fig. 2(b)) since, unlike the static model [5], the states are random variables with a posterior distribution.

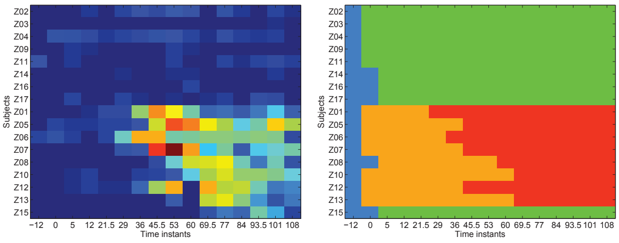
**Table 2**. Comparative measures between the proposed temporal algorithm and its non-temporal version.

| | Temporal | Non-temporal |
|---|---|---|
| $\mathrm{MSE}_r^2(\times 10^{-1})$ | 0.04 | 0.06 |
| | 10.52 | 12.13 |
| | 1.70 | 1.70 |
| | 9.31 | 9.31 |
| $\mathrm{GMSE}_r^2(\times 10^{-2})$ | 1.31 | 3.76 |
| | 0.91 | 2.49 |
| | 0.90 | 0.84 |
| | 0.44 | 0.48 |
| RE $(\times 10^2)$ | 7.03 | 6.30 |
| Time (in hours) | 5.87 | 5.53 |

### 5.2. Real data

This section illustrates the algorithm applied to the publicly available time-evolving gene expression dataset (GEO series accession number GSE30550). This dataset consists of the gene expression levels of $N = 267$ affymetrix chips collected at $T = 16$ time instants on $S = 17$ healthy human volunteers experimentally infected with influenza A/H3N2/Wisconsin (see [14] for more details). Each sample consisted of over $G = 12000$ gene expression values normalized according to the procedure in [5].

The proposed algorithm was run with $\mathrm{N}_{\mathrm{mc}} = 1000$ Monte Carlo iterations, including a burn-in period of $\mathrm{N}_{\mathrm{bi}} = 100$ iterations. The number of factors was determined using the algorithm described in [1] yielding $R = 4$. As the time of inoculation was known, the state probability $\pi_{1,1}$ was fixed to $\pi_{1,1} = 2/T$. As in previous analysis [5, 14], the proposed algorithm identifies a strong factor, also called the *inflammatory component*. The factor score vector associated with this inflammatory component is shown in Fig. 3(a) as an image whose columns (resp. rows) correspond to a specific time point (resp. subject) across the $S = 17$ subjects (resp. the $T = 16$ time instants). Note that the subjects have been reordered following the prior classification proposed in [14]. The corresponding estimated classification matrix $\widehat{\mathbf{Z}}$ is shown in Fig. 3(b). The proposed algorithm can be used to estimate the unknown state transition probabilities. For instance, the MMSE estimates are $\widehat{\boldsymbol{\pi}}_1 = [0.14, 0.45, 0.41]$ and $\widehat{\boldsymbol{\pi}}_3 = [0.84, 0.16]$.



(a)

(b)

**Fig. 3**. Left: estimated factor scores associated with the inflammatory component. Right: estimated classification map $\widehat{\mathbf{Z}}$.

Fig. 3 shows that the proposed algorithm clearly separates the subjects exhibiting symptoms (associated with the last 9 rows) from those who remain asymptomatic (associated with the first 8 rows). Subject #15 (associated with the last row) is classified as an asymptomatic subject, which seems to be coherent with the associated factor scores (see [5, 14]). The proposed temporal method has been compared with its non-temporal version [1]. Table 3 reports the Fisher linear discriminant measure [15, p. 119] computed between the post-onset-symptomatic samples ($\mathcal{C}_4$) and the other samples, REs and computational times for the two considered algorithms. The two algorithms provide similar results in terms of estimation performance. However, the temporal algorithm has the advantage that it generates a posterior distribution of the state. Results obtained by the proposed temporal algorithm on the H3N2 dataset also provide similar results in terms of discovering an inflammatory factor and separating sick from healthy individuals than previous analysis [5].

**Table 3**. Comparative measures on real H3N2 dataset.

| | Temporal | Non-temporal |
|---|---|---|
| Fisher linear discriminant ($\times 10^{-2}$) | 5.24 | 5.28 |
| RE ($\times 10^{-2}$) | 6.59 | 6.51 |
| Time (in $s$) | 3295 | 2808 |

## 6. CONCLUSIONS

This paper proposed a new hierarchical time-evolving Bayesian unmixing algorithm for longitudinal time series measurements. Time dependencies were considered by using a 4-state hidden Markov model. The resulting Markov model was combined with other prior information and statistical properties of the observed data to build an appropriate posterior distribution. A hybrid Gibbs sampler was implemented to generate random samples asymptotically distributed according to this joint posterior distribution. MAP estimators of the model parameters and hyperparameters were computed using these samples. Simulation results performed on synthetic and real gene time-series illustrated the accuracy of the proposed temporal algorithm in terms of unmixing and classification. Future works include i) the estimation of the number of factors jointly with the other parameters and the classification map, ii) the consideration of non-stationary states in the HMM exploiting the fact that the state transitions be non-Markovian.

## 7. REFERENCES

[1] C. Bazot, N. Dobigeon, J.-Y. Tourneret, and A. O. Hero, "Unsupervised Bayesian analysis of gene expression patterns," in *Rec. 44th IEEE Asilomar Conf. Signals, Systems and Computers (Asilomar)*, Pacific Grove, CA, Nov. 2010, pp. 364–368.

[2] C. Bazot, N. Dobigeon, J.-Y. Tourneret, and A. O. Hero, "Bernoulli-Gaussian model for gene expression analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 5996–5999.

[3] A. Schliep, A. Schönhuth, and C. Steinhoff, "Using hidden Markov models to analyze gene expression time course data," *Bioinformatics*, vol. 19, no. suppl 1, pp. i255–i263, 2003.

[4] Q. Huang, L.-Y. Wu, J.-B. Qu, and X.-S. Zhang, "Analyzing time-course gene expression data using profile-state hidden Markov model," in *Proc. IEEE Int. Conf. Systems Biology (ISB)*, Sept. 2011, pp. 351–355.

[5] Y. Huang, A. K. Zaas, A. Rao, N. Dobigeon, P. J. Woolf, T. Veldman, N. C. Oien, M. T. McClain, J. B. Varkey, B. Nicholson, L. Carin, S. Kingsmore, C. W. Woods, G. S. Ginsburg, and A. O. Hero, "Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza A infection," *PLoS Genetics*, vol. 8, no. 7, Aug. 2011.

[6] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[7] B. Coburn, B. Wagner, and S. Blower, "Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1)," *BMC Medicine*, vol. 7, no. 1, pp. 30, 2009.

[8] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[9] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. and Remote Sensing*, vol. 43, no. 4, pp. 898–910, April 2005.

[10] N. Dobigeon, J.-Y. Tourneret, and J. D. Scargle, "Joint segmentation of multivariate Poissonian time series applications to burst and transient source experiments," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sept. 2006.

[11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996.

[12] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[13] G. O. Roberts, "Markov chain concepts related to sampling algorithms," in *Markov chain Monte Carlo in practice*, pp. 45–57. Chapman & Hall, London, 1996.

[14] A. K. Zaas, M. Chen, J. Varkey, T. Veldman, A. O. Hero, J. Lucas, Y. Huang, R. Turner, A. Gilbert, R. Lambkin-Williams, N. C. Øien, B. Nicholson, S. Kingsmore, L. Carin, C. W. Woods, and G. S. Ginsburg, "Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans.," *Cell host & microbe*, vol. 6, no. 3, pp. 207–17, Sept. 2009.

[15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, New York, 2nd edition, 2000.