

# Bayesian Blind Separation and Deconvolution of Dynamic Image Sequences Using Sparsity Priors

Ondrej Tichý\* and Václav Šmídl, *Member, IEEE*

**Abstract**—A common problem of imaging 3-D objects into image plane is superposition of the projected structures. In dynamic imaging, projection overlaps of organs and tissues complicate extraction of signals specific to individual structures with different dynamics. The problem manifests itself also in dynamic tomography as tissue mixtures are present in voxels. Separation of signals specific to dynamic structures belongs to the category of blind source separation. It is an underdetermined problem with many possible solutions. Existing separation methods select the solution that best matches their additional assumptions on the source model. We propose a novel blind source separation method based on probabilistic model of dynamic image sequences assuming each source dynamics as convolution of an input function and a source specific kernel (modeling organ impulse response or retention function). These assumptions are formalized as a Bayesian model with hierarchical prior and solved by the Variational Bayes method. The proposed prior distribution assigns higher probability to sparse source images and sparse convolution kernels. We show that the results of separation are relevant to selected tasks of dynamic renal scintigraphy. Accuracy of tissue separation with simulated and clinical data provided by the proposed method outperformed accuracy of previously developed methods measured by the mean square and mean absolute errors of estimation of simulated sources and the sources separated by an expert physician. MATLAB implementation of the algorithm is available for download.

**Index Terms**—Blind source separation, computer-aided detection and diagnosis, functional imaging, probabilistic and statistical methods.

## I. INTRODUCTION

**D**YNAMIC scintigraphic and molecular imaging is a tool for examination of organ, tissue, cellular and intracellular functions at various levels of spatial and time resolution. Its aim is to describe physiological functions in the body by simple diagnostic parameters derived from sophisticated pharmacokinetic models. On input, pharmacokinetic models require the

curves reflecting local concentration of a tracer (radiopharmaceutical or molecular probe) as functions of time at specific sites in individual tissues, compartments, and distribution spaces, formally described as dynamic image components. However, tissue-specific time-activity curves (TAC) are not directly observable due to projection overlaps of the tissues in dynamic scintigraphy [1], or tissue mixtures in the image plane in dynamic positron emission tomography or dynamic and functional magnetic resonance tomography [2]–[4].

Thus, one more step is required between collecting observed data and using pharmacokinetic model, i.e., extraction of the TACs from image data. In dynamic scintigraphy, this step is usually performed by a human operator who manually draws a region of interest (ROI), an area including the tissue of interest over which the image signal is integrated to a sum representing one point on TAC [5]. It is easy to understand that this single step is a sort of a bottleneck of the whole process: it is subjective (dependent on the operator's knowledge, experience and skills) and prone to errors due to overlapping dynamic structures [6]. Including several dynamic components in one ROI (assumed to include just one component) invalidates the method producing false results. In attempt to avoid inclusion of more than one dynamic structure into a single ROI, the operator often needs to define very small ROIs that are only few pixels large and produce weak noisy signal. Sometimes there is no such area in the image from where the signal of a single dynamic structure can be safely extracted.

In the past, many attempts were made to define the ROIs and extract TACs automatically or semiautomatically avoiding user interaction (for more recent reports and reviews, see [7]–[10]). Recent attempts to extract image-derived input function belong to this category as well [11]–[13]. Despite all of these methods were reported to provide good results and some of them are used in their authors' departments, none has been generally accepted to be used in clinical practice.

An important option to facilitate separation of dynamic components in image sequence is application of mathematical and physiological models, for example factor analysis model assuming observed images and curves as linear combination of dynamic components [10], or a linear model assuming organ curve as convolution of input function (signal from the blood) and tissue kernel [14]–[16]. Such models are employed in the methods aiming at blind source separation as in the independent component analysis [17], blind source separation with positivity constraint [18], nonnegative matrix factorization [19], or blind source separation and deconvolution [20], [21].

Manuscript received July 03, 2014; revised August 22, 2014; accepted August 23, 2014. Date of publication August 28, 2014; date of current version December 24, 2014. This work was supported by the Czech Science Foundation, Grant GA13-29225S. Collection of data for the database of dynamic renal scintigraphy (<http://www.dynamicrenalstudy.org>) was supported by the Czech Science Foundation Grant 303/07/0950. *Asterisk indicates corresponding author.*

\*O. Tichý is with the Institute of Information Theory and Automation, 182 08 Prague, Czech Republic (e-mail: otichy@utia.cas.cz).

V. Šmídl is with the Institute of Information Theory and Automation, 182 08 Prague, Czech Republic.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2014.2352791

Specialized methods of blind source separation has been developed for analysis of medial image sequences [22]–[24]. In [23], the sparsity prior on the source images was used for blind estimation of the ROIs. In [22], the mixture model of the source images pixels was complemented by a convolution model of the TAC with exponential parametrization of the input function and the convolution kernels. Another parametrization of the convolution kernels using piecewise linear function [25] has been presented in [24]. The main disadvantage of explicit parametrization of the input function and the convolution kernels is that the true functions may differ from the assumed parametric model if the parametric model is not sufficiently flexible. A more flexible model was introduced in [26] using sparsity priors [27], where the only restriction on the convolution kernel is its sparsity. However, the proposed method of estimation is incapable to recover the correct number of underlying sources which results in artifacts (a single source is estimated as a sum of two identical sources). In this paper, we use the model of [26]; however, we propose a more complex estimation method respecting correlation of the convolution kernels.

The presented method is compared to the state-of-the-art methods on simulated phantom study as well as clinical data from renal scintigraphy. Specifically, we present the results of the blind source separation with positivity constraints (BSS+) [18], factor analysis with integrated ROIs estimation model (FAROI) [23], CAM-CM algorithm [22], and the sparse blind source separation and deconvolution (S-BSS-DC) model [26]. The set of 12 dynamic image sequences of computer phantoms simulated using Monte Carlo method [28] available at [29] is used. Advantage of simulated data is a knowledge of true TACs of respective dynamic structures that can be used to validate the algorithms. The performance of the methods with clinical data was tested with 19 dynamic renal studies in the patients in which reference TACs were obtained by an experienced physician following the recommended guidelines [30], [31].

## II. MATHEMATICAL MODEL

Let us first summarize the assumptions of the model of blind source separation and deconvolution [26]. In nuclear imaging, the source of image signal is a tracer (radiopharmaceutical). In the body, the tracer is distributed in several compartments of its distribution space. Each compartment is characterized by its specific dynamics. The compartment may or may not correspond to a specific anatomical structure and may contain one or several different tissues. Observed data (image sequences recorded over the period of time after intravenous injection of the tracer) represent noisy observations of the compartments in the body that, in the projection images, usually overlap each other.

In the model, the signal source to be separated from a mixture corresponds to a single compartment. Each compartment is represented by its image reflecting spatial distribution and by its TAC reflecting time distribution of its specific signal. TACs observed over individual compartments result from convolution of a common input function and the compartment-specific kernels. Both input function and source kernels are unknown.

The aim of analysis is to extract the images and TACs corresponding to individual signal sources (i.e., compartments) from observed data. In the model [26], the only assumption about the source images and convolution kernels is their sparsity. It is assumed that the compartments cover only part of the image and their kernels only part of the observation period. Sparsity is modeled using prior distribution known as automatic relevant determination (ARD) [32]. Therefore, the model is formulated as a hierarchical Bayesian model.

### A. Model of Observed Data

An observed image at time  $t$  is stored as a vector  $\mathbf{d}_t \in \mathbf{R}^{p \times 1}$  where the pixels are stored columnwise. The vector  $\mathbf{d}_t$  is assumed to be a noisy observation of a superposition of  $r$  source images  $\mathbf{a}_1, \dots, \mathbf{a}_r, \mathbf{a}_k \in \mathbf{R}^{p \times 1}$ , weighted by their activities  $x_{1,t}, \dots, x_{r,t}$  stored in a row vector  $\bar{\mathbf{x}}_t \in \mathbf{R}^{1 \times r}$ . Formally

$$\mathbf{d}_t = \mathbf{a}_1 x_{1,t} + \mathbf{a}_2 x_{2,t} + \dots + \mathbf{a}_r x_{r,t} + \mathbf{e}_t = A \bar{\mathbf{x}}_t^T + \mathbf{e}_t \quad (1)$$

where  $A = [\mathbf{a}_1, \dots, \mathbf{a}_r] \in \mathbf{R}^{p \times r}$ ,  $()^T$  denotes transposition of vector or matrix in this paper, and  $\mathbf{e}_t$  is the observation noise. A sequence of observed images forms a matrix  $D = [\mathbf{d}_1, \dots, \mathbf{d}_n] \in \mathbf{R}^{p \times n}$  where  $n$  is the number of images in the sequence. The matrix version of (1) is

$$D = AX^T + E \quad (2)$$

where  $X = [\bar{\mathbf{x}}_1^T, \dots, \bar{\mathbf{x}}_n^T]^T$ . The columns of the matrix  $X$ , i.e., vectors  $\mathbf{x}_1, \dots, \mathbf{x}_r \in \mathbf{R}^{n \times 1}$ , represent the TAC of each source image. We will use the bar symbol,  $\bar{x}_i$ , to distinguish the  $i$ th row of matrix  $X$ , while  $\mathbf{x}_i$  will be used to denote the  $i$ th column. The noise  $\mathbf{e}_t$  is assumed to be homogeneous Gaussian with zero mean and unknown precision (inverse variance)  $\omega$

$$f(D|A, X, \omega) = \prod_{t=1}^n \mathcal{N}(A \bar{\mathbf{x}}_t, \omega^{-1} I_p). \quad (3)$$

Here,  $\mathcal{N}$  denotes Gaussian distribution and  $I_p$  denotes identity matrix of the size  $p$ . The precision parameter  $\omega$  of the normal density function has a conjugate prior in the form of the Gamma distribution

$$f(\omega) = \mathcal{G}(\vartheta_0, \rho_0) \quad (4)$$

where constants  $\vartheta_0, \rho_0$  are chosen.

Assumption of the homogeneous noise (3) can be too restrictive in modalities with Poisson distributed noise, such as scintigraphy. In that case, we may perform scaling of the data [33]. The original data  $D_{\text{orig}}$  is scaled using the correspondence analysis

$$d_{ij} = \frac{d_{ij, \text{orig}}}{\sqrt{\sum_{i=1}^p d_{ij, \text{orig}} \sum_{j=1}^n d_{ij, \text{orig}}}}. \quad (5)$$

When this operation is performed, inverse scaling needs to be applied to the estimates of  $A$  and  $X$  for their presentation in the original scale. Note that scaling (5) is only asymptotically optimal for Poisson noise, and may introduce bias for low count scenarios.

### B. Model of Source Images

The only assumption on the images of the underlying compartments is that they are not present in all pixels of the observed sequence. Therefore, we assume that only some elements of the source images  $\mathbf{a}_k$  have nonzero value, all other elements are zeros. The sparse solution is achieved by priors that favor zero solution when insufficient data are available [27]. Specifically, we will use the mechanism of automatic relevance determination (ARD) which is commonly used in hierarchical models [34]. Each pixel of each underlying source image is assumed to have truncated Gaussian prior with unknown precision

$$f(a_{i,k}|\xi_{i,k}) = t\mathcal{N}(0, \xi_{i,k}^{-1}, [0, 1]) \quad (6)$$

$$f(\xi_{i,k}) = \mathcal{G}(\phi_0, \psi_0), \quad \forall i = 1, \dots, p, \forall k = 1, \dots, r \quad (7)$$

where  $t\mathcal{N}(\mu, \sigma, [0, 1])$  denotes Gaussian distribution with support on interval  $[0, 1]$  Appendix B, and  $\phi_0, \psi_0$  are chosen constants. Note that each pixel of the source images has an extra unknown parameter,  $\xi_{i,k}$  forming a matrix  $\Xi$  of the same dimensions as  $A$ .

### C. Model of Time-Activity Curves

Accumulation of the tracer in a specific tissue compartment is commonly modeled by convolution of input function  $\mathbf{b}$  (usually associated with TAC of the blood) with a compartment specific kernels  $\mathbf{u}_k$  [25]. Therefore, we model TAC of the  $k$ 'th source as

$$\mathbf{x}_k = B\mathbf{u}_k, \quad \forall k = 1, \dots, r \quad (8)$$

where matrix  $B$  is defined as

$$B = \begin{pmatrix} b_1 & 0 & 0 & 0 \\ b_2 & b_1 & 0 & 0 \\ \dots & b_2 & b_1 & 0 \\ b_n & \dots & b_2 & b_1 \end{pmatrix}. \quad (9)$$

Using matrix  $U$  for aggregating convolution kernels as  $U = [\mathbf{u}_1, \dots, \mathbf{u}_r]$ , we can rewrite the matrix  $X$  in terms of matrices  $B$  and  $U$  as  $X = BU$  and rewrite the data model (2) as

$$D = AX^T + E = AU^T B^T + E. \quad (10)$$

Note that all matrices  $A, U$ , and  $B$  have to be estimated from the data matrix  $D$ .

1) *Model of Convolution Kernels*: Many parametric forms of the convolution kernels has been proposed [25], [22]. However, these parametric forms may not fit well to the real data. Therefore, we once again use only the assumption of sparsity of the convolution kernel using ARD prior.

The prior distribution of the convolution kernels  $\mathbf{u}_k, \forall k = 1, \dots, r$ , is modeled as normally distributed truncated to positive values

$$f(u_{j,k}|v_{j,k}) = t\mathcal{N}(0, v_{j,k}^{-1}, [0, \infty]), \quad \forall j = 1, \dots, n, \quad (11)$$

$$f(v_{j,k}) = \mathcal{G}(\alpha_0, \beta_0) \quad (12)$$

where  $\alpha_0, \beta_0$  are selected constants. Each pixel of the convolution kernels  $U$  thus has its unknown variance, forming a matrix  $\Upsilon$  of the same size as  $U$ .

2) *Model of the Input Function*: The input function from (8) stored in vector  $\mathbf{b} = [b_1, \dots, b_n]$  is an arbitrary vector of positive real numbers. The prior distribution of this vector is

$$f(\mathbf{b}|\varsigma) = t\mathcal{N}(\mathbf{0}_{n,1}, \varsigma^{-1}I_n, [0, \infty]) \quad (13)$$

$$f(\varsigma) = \mathcal{G}(\zeta_0, \eta_0) \quad (14)$$

where  $\mathbf{0}_{n,1}$  denotes zeros matrix of the respected size and constants  $\zeta_0, \eta_0$  are chosen.

### D. Summary of the Model

The probabilistic model of the sparse blind source separation and deconvolution is formulated as a joint distribution

$$f(D, A, U, \mathbf{b}, \omega, \Xi, \Upsilon, \varsigma) = f(D|A, U, \mathbf{b}, \omega) f(A|\Xi) f(\Xi) f(U|\Upsilon) f(\Upsilon) f(\mathbf{b}|\varsigma) f(\varsigma) \quad (15)$$

where the elements on the right-hand side are given by (3)–(4), (6)–(7), and (11)–(14). The task is to find the posterior densities of the source images,  $f(A)$ , convolution kernels,  $f(U)$ , and input function  $f(\mathbf{b})$  from the data matrix  $D$ .

## III. ESTIMATION OF THE MODEL PARAMETERS

Following the Bayesian approach, we will estimate the unknown source images and TACs as posterior expected values. However, evaluation of the marginal distribution of (15) is analytically intractable. Therefore, we use the Variational Bayes (VB) approximation [35], [36] also known as ensemble learning method [18].

### A. Variational Bayes Method

Let  $D$  denote a multivariate observation of a parametric probabilistic model given as  $f(D|\theta)$ , where  $\theta = [\theta_1, \dots, \theta_q]^T$  is a multivariate parameter. Our information about the  $\theta$  is quantified by a prior distribution,  $f(\theta)$ . The Variational Bayes method [35], is a technique for approximate evaluation of shaping parameters of posterior distribution  $f(\theta|D)$ . The approximation is chosen in a restricted form  $\check{f}(\theta|D)$  and its shaping parameters are optimized in order to minimize the Kullback-Leibler divergence to the true posterior

$$\text{KLD} \left( \check{f}(\theta|D) || f(\theta|D) \right) = \int \check{f}(\theta|D) \ln \frac{\check{f}(\theta|D)}{f(\theta|D)} d\theta. \quad (16)$$

*Theorem 1*: Let  $f(\theta|D)$  be the posterior distribution of multivariate parameter  $\theta = [\theta_1, \dots, \theta_q]^T$ . Let  $\check{f}(\theta|D)$  be an approximate distribution restricted to the set of conditionally independent distributions as

$$\check{f}(\theta|D) = \prod_{i=1}^q \check{f}(\theta_i|D). \quad (17)$$

Then, the minimum of KLD (16)

$$\tilde{f}(\theta|D) = \arg \min_{\check{f} \in (17)} KLD \left( \check{f}(\theta|D) \| f(\theta|D) \right) \quad (18)$$

is reached for

$$\tilde{f}(\theta_i|D) \propto \exp \left( E_{\check{f}(\theta_i|D)} [\ln (f(\theta, D))] \right), \quad i = 1, \dots, q \quad (19)$$

where symbol  $\propto$  means up to normalizing constant,  $E_{\check{f}(\cdot)}$  means expected value of an argument with respect to distribution  $\check{f}$ , and  $\theta_{/i}$  denotes complement of  $\theta_i$  in  $\theta$ ; hence,  $\theta_{/i} = [\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_q]$ .

Proof of the theorem can be found, e.g., in [18]. Note that (19) forms a set of implicit functional equations. The solution involves extra steps, that has been systematized in [35]. We will need only the following steps: 1) selection of the conditional independent parameter partitioning (17), 2) identification of the standard form of the marginals (19), and evaluation of the analytical moments of (19), and 3) running of the iterative algorithm for evaluation of shaping parameters of (19).

### B. Parameter Partitioning

The first step of the VB method is selection of a partitioning of the parameters of the natural logarithm of the model (15). Some partitioning may appear naturally as a result of conditional independence of the likelihood function. Some partitioning has to be forced to obtain tractability. Specifically, for likelihood function (3) with convolution model (10), the logarithm is

$$\ln f(D|A, U, \mathbf{b}, \omega) = \frac{pn}{2} \ln \omega + \frac{1}{2} \omega \text{tr} \left[ (D - AU^T B^T) (D - AU^T B^T)^T \right] \quad (20)$$

where  $\text{tr}(\cdot)$  denotes trace of a matrix. Natural partitioning arise for source images  $A$ , since the trace (20) can be rewritten in the following form:

$$\sum_{i=1}^p \left[ \omega \bar{\mathbf{a}}_i \sum_{j=1}^n (\bar{\mathbf{x}}_j D_{i,j})^T - \frac{1}{2} \omega \bar{\mathbf{a}}_i \sum_{j=1}^n (\bar{\mathbf{x}}_j^T \bar{\mathbf{x}}_j) \bar{\mathbf{a}}_i^T \right] + c_A \quad (21)$$

where  $c_A$  is a term aggregating all elements of (20) independent of  $A$ . Note that only the elements of rows  $\bar{\mathbf{a}}_i$  of the matrix  $A$  interact with each other. Hence, the posterior  $f(A|D)$  is naturally partitioned into  $f(A|D) = \prod_{i=1}^p f(\bar{\mathbf{a}}_i|D)$ .

This does not happen for the convolution kernels  $U$ . Using matrix algebra [37]

$$\text{tr}(AU^T B^T B U A^T) = \mathbf{u}^T (A^T A \otimes B^T B) \mathbf{u}$$

where  $\mathbf{u} = \text{vec}(U) = [\mathbf{u}_1^T, \dots, \mathbf{u}_r^T]^T$ , and  $\otimes$  is the Kronecker product [37]. We note that all elements of the convolution kernels interact with each other, hence forcing conditional independence as it was proposed in [26] will introduce approximation error. In this paper, we do not impose this restriction and evaluate full posterior  $f(\mathbf{u}|D)$ .

For tractability reasons, we still need to impose conditional independence between all precision parameters  $\omega, \zeta, \Xi, \Upsilon$  and

their mean values, e.g.,  $f(A, \Xi|D) \approx f(A|D)f(\Xi|D)$ . However, under this condition, the elements of the precision parameters also naturally partition into  $f(\Xi|D) = \prod_i \prod_k f(\xi_{i,k}|D)$ .

### C. Standard Forms of the VB Marginals

The VB theorem (19) is applied to (15) and the set of VB-marginals is established. The recognized standard distributions for the model are as follows:

$$\tilde{f}(\omega|D) = \mathcal{G}(\vartheta, \rho) \quad (22)$$

$$\tilde{f}(\mathbf{u}|D) = t\mathcal{N}(\mu_{\mathbf{u}}, \Sigma_{\mathbf{u}}, [0, \infty]), \quad \tilde{f}(v_{j,k}|D) = \mathcal{G}(\alpha_{j,k}, \beta_{j,k}) \quad (23)$$

$$\tilde{f}(\mathbf{b}|D) = t\mathcal{N}(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}, [0, \infty]), \quad \tilde{f}(\zeta|D) = \mathcal{G}(\eta, \zeta), \quad (24)$$

$$\tilde{f}(\bar{\mathbf{a}}_i|D) = t\mathcal{N}(\mu_{\bar{\mathbf{a}}_i}, \Sigma_{\bar{\mathbf{a}}_i}, [0, 1]), \quad \tilde{f}(\xi_{i,k}|D) = \mathcal{G}(\phi_{i,k}, \psi_{i,k}). \quad (25)$$

The shaping parameters  $\vartheta, \rho, \mu_{\mathbf{u}}, \Sigma_{\mathbf{u}}, \alpha_{j,k}, \beta_{j,k}, \mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}, \eta, \zeta, \mu_{\bar{\mathbf{a}}_i}, \Sigma_{\bar{\mathbf{a}}_i}, \phi_{i,k}, \psi_{i,k}$  of the posterior densities (22)–(25) are given in Appendix A. They form a set of implicit equations which will be solved iteratively.

### D. Estimation of the Number of Sources

The number of sources  $r$  can be manually selected for the whole procedure as a static parameter, however, the results of separation may be sensitive to this choice. The ARD prior has been applied as sparsity criterion for selection of the number of relevant sources, e.g., in [18]. In the proposed model, the ARD prior is used as sparsity criterion for individual pixels of the source images. Thus the sparsity at source level would have to be enforced by an additional parameter, leading to multilevel or multiresolution priors. This is certainly possible but it would further complicate the algorithm.

To avoid even more complex prior structures, we propose an alternative automatic approach for estimation of the number  $r$  using the estimate of precision parameter  $\omega$ .

Specifically, we note that VB solution of the scalar version of the model (2),  $d = ax + e$ , yields nonzero signal (i.e.,  $\hat{a}\hat{x} > 0$ ) when  $d > 2\sqrt{\omega^{-1}}$  [26]. Using this inference bound, the sum of  $d_{i,j}^2$  corresponding to a pixel from the  $k$ th source, should be  $n$  times greater than the noise level

$$\mathbf{x}_k^T \mathbf{x}_k > 2n\omega^{-1}. \quad (26)$$

This observation will be used as a criterion for removal of weak sources within the iterative procedure. Since removal of a source influence all others we disallow further removal for the next 50 iterations of the algorithm. The algorithm starts from  $r_{\max}$  sources and terminates if all sources satisfy (26) or the minimum number of sources  $r_{\min}$  is reached. The interval  $[r_{\min}, r_{\max}]$  can be specified by an expert or heuristically. In this paper, we used heuristics based on differences of singular values  $\sigma_i$  of the data matrix  $D$ . Specifically, when  $\sigma_{i+1}^2$  is less than 95% of  $\sigma_i^2$ , then  $r_{\max} = i$ . We observed that this starting point for  $r$  overestimates significantly the true value of  $r$ . The same heuristics is used for selection of  $r_{\min}$  but the coefficient is set to 75%.

In general, we recommend to slightly overestimate the number of  $r_{\max}$  of the relevant sources, since the redundant source would be estimated to be weak or removed by the automatic criteria (26). If the  $r_{\max}$  is chosen lower than the true number of sources, the sources will be always mixed. Condition (26) may remove even a valid signal if the number nonzero elements in  $\mathbf{x}_k$  is much lower than  $n$ . Removal of sources becomes more aggressive with growing  $n$ .

### E. Iterative Algorithm

An iterative solution of the set of implicit equations from Appendix A is given in Algorithm 1. The algorithm will be denoted as the sparse blind source separation and vectorized deconvolution (S-BSS-vecDC). Since VB suffers from local minima [35], good initial conditions of the algorithm are required to reach convergence to an acceptable solution. The parameter most sensitive to its initial estimate is the input function,  $\mathbf{b}$ . Good results were obtained with the pulse function or an exponential. The initial shapes of the convolution kernels,  $\mathbf{u}_k$  are chosen as unit pulses with different lengths and delays.

---

#### Algorithm 1 Iterative S-BSS-vecDC algorithm

---

- 1) Initialization:
  - (a) Initialize constants  $\vartheta_0, \rho_0, \phi_0, \psi_0, \alpha_0, \beta_0, \zeta_0, \eta_0 = 10^{-10}$
  - (b) Initialize the maximum number of sources,  $r_{\max}$ , and the minimum number of sources,  $r_{\min}$ .
  - (c) Initialize the input function,  $\mathbf{b}$ , and the convolution kernels,  $U$ .
- 2) Iterate equations from Appendix A until no source is removed and maximum number of iterations is not reached:
  - (a) Compute shaping parameters:  $\mu_{\bar{\mathbf{a}}_i}, \Sigma_{\bar{\mathbf{a}}_i}$ , and  $\phi_i, \psi_i \forall i = 1, \dots, p$ .
  - (b) Compute shaping parameters:  $\mu_{\mathbf{u}}, \Sigma_{\mathbf{u}}$ , and  $\alpha, \beta$ .
  - (c) Compute shaping parameters:  $\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}$ , and  $\eta, \zeta$ .
  - (d) Compute shaping parameter:  $\vartheta, \rho$ .
  - (e) If the convergence is reached, compute criterion (26) for each source and remove source if the criterion is fulfilled.
- 3) Report estimates  $\hat{A}, \hat{X}$ .

The iterations of the shaping parameters and their moments use equations from Appendix A with the exception of moment  $\widehat{B^T B}^{-1}$ . This moment suffers from numerical instability when the maximum of the input function is not on the first element. We adopt the Moore-Penrose pseudoinverse to restore numerical stability. Specifically, we discard singular values which are smaller than 10% of the mean of all singular values.

MATLAB implementation of Algorithm 1 is available from [http://www.utia.cas.cz/AS/softwaretools/image\\_sequences](http://www.utia.cas.cz/AS/softwaretools/image_sequences) together with documentation and tutorial examples.

## IV. EXPERIMENTS AND RESULTS

The proposed algorithm is tested on simulated and clinical data sets of dynamic renal scintigraphy. The result of estimation are compared to the results of the competing methods.

### A. Competing Methods

All methods in this section provide solution to the blind source separation problem (1). However, they differ in additional assumptions (priors) on the source images and TACs. In the following paragraphs, we briefly summarize the key assumptions.

1) *BSS+*: The variational blind source separation (BSS+) model [18] is based on (3) and (4). The ARD principle is used only for determination of the number of sources, i.e., the unknown scalar variance is common to all elements of TACs  $\mathbf{x}_k$ . In effect, sparsity of the source images and TACs is not encouraged.

2) *FAROI*: The factor analysis with integrated regions of interests (FAROI) model [23] is an extension of the BSS+ method to encourage sparse estimates of the source images. An indicator variable  $\mathbf{i}$  of the same dimension as  $A$  is introduced, where each element  $i_{i,k}$  indicates if the pixel  $a_{i,k}$  belongs to the zero component or nonzero component. The estimation procedure is thus closely related to clustering algorithms. The number of relevant sources is estimated using the ARD principle on the TACs as in the BSS+ method.

3) *CAM-CM*: Interpretation of signal superposition as a mixture model is also used in the CAM-CM method [22]. However, the number of unknown sources is determined for each pixel. The TACs are modeled as a convolution of parametric input function and parametric convolution kernels. Both the input function and the kernels are assumed to have exponential shape with an unknown rate parameter.

4) *S-BSS-DC*: The sparse blind source separation and deconvolution (S-BSS-DC) method [26] is based on the same model as proposed in this paper. Evaluation of the posterior distribution was however restricted to the mutually independent convolution kernels as explained in Section III-B.

### B. Experiment With Model-Generated Data

Quality of estimation of the proposed method is first validated on a synthetic image sequence that was generated according to the model (1). We simulate the data using the additional assumptions on the curves required by the CAM-CM model and using the same images [22]. The size of each image is  $50 \times 50$  pixels ( $p = 2500$ ) and 50 time points are simulated ( $n = 50$ ). We simulate three sources:  $b = \exp(-t/3)$ ,  $u_1 = \exp(-t/10)$ ,  $u_2 = 100 \exp(-4t)$ , and  $u_3 = 1/2 \exp(t/100)$ , where  $t = 1, \dots, n$ . The standard deviation of the noise is assumed to be proportional to the signal with coefficient 0.5 such as  $d_{i,j} = d_{i,j} + 0.5e_{i,j}\sqrt{d_{i,j}}$ , where  $e_{i,j}$  is random realization from a normal distribution with mean 0 and standard deviation 1. The ground truth source data are shown in Fig. 1, left, and are available for download together with the S-BSS-vecDC algorithm.

We compare the estimation results of the CAM-CM, S-BSS-DC and S-BSS-vecDC for initial value  $r_{\max} = 4$  in order to test the ability of the methods to recover the correct number of sources. This experiment demonstrates the advantage of S-BSS-vecDC over S-BSS-DC. While the S-BSS-DC algorithm (Fig. 1, the fifth and the sixth columns) splits the third source in two almost equal estimates, the S-BSS-vecDC (Fig. 1, the seventh and the eighth columns) aggregates the

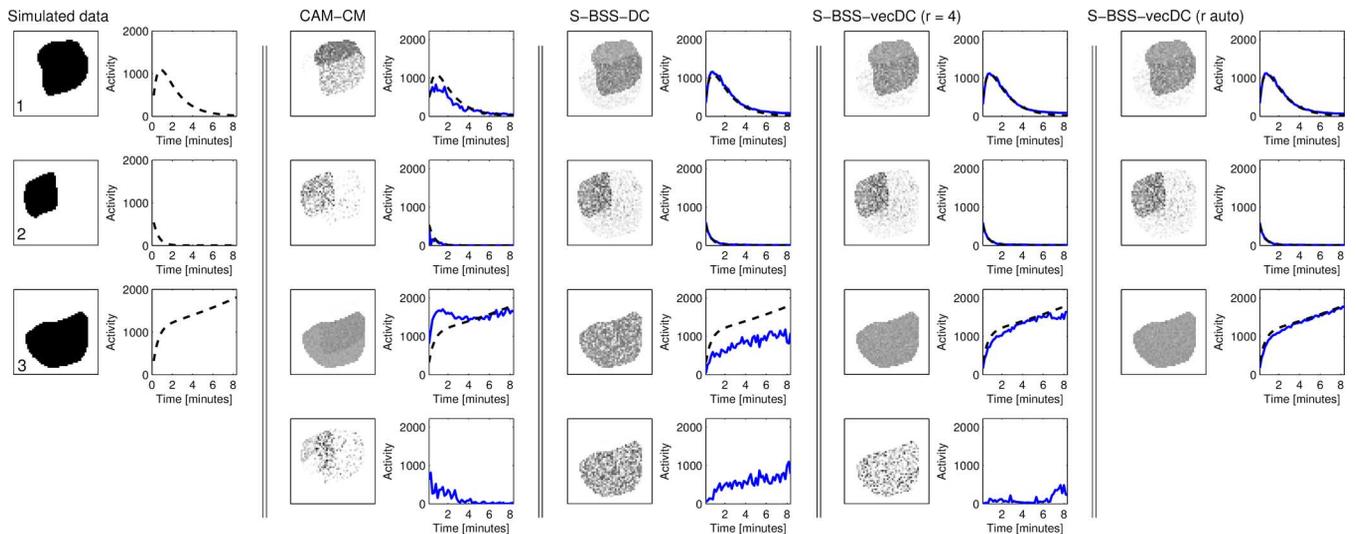


Fig. 1. Results of blind source separation of synthetic data displayed in the two column format. Source images are in the odd columns and the corresponding time-activity curves in the even columns. Simulated sources are displayed in columns 1 and 2, followed by the results of separation using the CAM-CM algorithm (columns 3 and 4), the S-BSS-DC algorithm (columns 5 and 6), and the S-BSS-vecDC (columns 7 and 8). All methods were run with the maximum number of sources,  $r_{\max} = 4$ . Results of the S-BSS-vecDC algorithm with the automatic estimation of the number of sources are displayed in columns 9 and 10.

TABLE I  
COMPARISON OF RESULTS FROM FIG. 1 USING MSE AND MAE

algorithm	$r_{\max}$	k (source)	$\mu_{\text{MSE}}^k$	$\mu_{\text{MAE}}^k$
CAM-CM	4	1	0.0440	0.1420
		2	0.0214	<b>0.0389</b>
		3	0.1150	0.2388
S-BSS-DC	4	1	0.0057	0.0684
		2	<b>0.0021</b>	0.0393
		3	0.4401	0.6465
S-BSS-vecDC	4	1	<b>0.0043</b>	<b>0.0513</b>
		2	0.0022	0.0413
		3	<b>0.0273</b>	<b>0.1428</b>
S-BSS-vecDC	auto	1	0.0030	0.0447
		2	0.0016	0.0344
		3	0.0077	0.0720

Bold numbers indicate the best result of separation obtained by a method with the number of sources set to four ( $r_{\max} = 4$ ). The sources are identified by the same numbers as in the first column of Figure 1.

activity from the third source into the third estimate and the fourth estimate is insignificant. Since the activity of the fourth source is lower than the noise level, it could be automatically removed using (26), yielding further improvement (Fig. 1, the ninth and the tenth columns). Note that such criteria would not be effective for the S-BSS-DC algorithm.

The visual comparison is accompanied with quantitative results using mean square error (MSE) and mean absolute error (MAE)

$$\mu_{\text{MSE}}^k = \frac{1}{n} \sum_{j=1}^n \left( \hat{x}_{j,k} - x_{j,k}^{\text{gt}} \right)^2, \mu_{\text{MAE}}^k = \frac{1}{n} \sum_{j=1}^n \left| \hat{x}_{j,k} - x_{j,k}^{\text{gt}} \right| \quad (27)$$

for each source in Table I,  $k = 1, \dots, r$ . Here,  $\hat{x}_j$  denotes elements of the estimated TAC and  $x_j^{\text{gt}}$  denotes the simulated TAC.

For the four sources  $r_{\max} = 4$ , the proposed S-BSS-vecDC algorithm provides the best estimates (bold numbers in Table I, middle block) of the first and the third source while the best estimate of the second source is provided by the S-BSS-DC method

(in terms of MSE) and by the CAM-CM (in terms of MAE). None of the algorithms is able to correctly determine the correct number of sources. However, the proposed extension of the S-BSS-vecDC algorithm (Section III-D) correctly detects the number of sources which results in the best MSE and MAE criteria (Table I, bottom block). For completeness, we also performed evaluation using other methods from Section IV-A. On these data, none of them provide more accurate results than the S-BSS-vecDC algorithm.

### C. Experiment With Simulated Dynamic Renal Studies

The algorithms were tested with Monte Carlo simulated data of dynamic renal study [28]. We analyzed 12 scintigraphic sequences from [29]. Each sequence consists of 120 images with the resolution of  $128 \times 128$  pixels, and is accompanied by the ground truth (GT) data without tissue background, attenuation, and noise. Reference TACs thus can be extracted easily without interference of the contaminating structures.

We use algorithms: BSS+, FAROI, S-BSS-DC, and S-BSS-vecDC. We do not use the CAM-CM algorithm since it has computational issues with data matrices of this size. The algorithms were applied to each of the two kidneys separately using rectangular ROIs, yielding 24 sequences of individual left and right kidneys to be analyzed. For comparison of respective methods, TACs of renal parenchyma were chosen because they are clinically important, often difficult to extract from real patient data, and in this case they could be reliably extracted from the GT data by an expert physician. Example estimates of renal parenchyma for all tested algorithms and one kidney sequence are presented in Fig. 2. The images of renal parenchyma are in the first row and the corresponding TACs are in the second row. The solid blue lines indicate the estimated TACs and the dashed black lines indicate the reference TACs. Statistical comparison of all 24 estimates is given in Table II using the average MSE, denoted as  $\bar{\mu}_{\text{MSE}}$ , and the standard deviation, denoted as

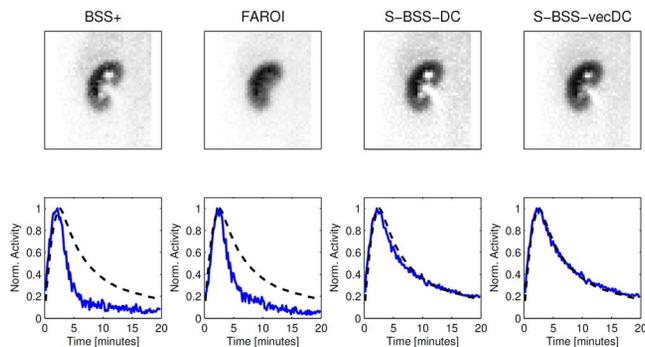


Fig. 2. Source images and time-activity curves corresponding to the renal parenchyma tissue estimated in simulated renal study using algorithms: BSS+, FAROI, S-BSS-DC, and S-BSS-vecDC. Solid blue lines indicate the estimated time-activity curves while dashed black lines indicate the reference time-activity curves.

TABLE II  
STATISTICAL COMPARISON OF ESTIMATED AND REFERENCE TACS  
OF RENAL PARENCHYMA IN SIMULATED STUDY

MSE	24 parenchyma curves		
algorithm	$\bar{\mu}_{\text{MSE}} \pm \text{std}(\mu_{\text{MSE}})$	$\bar{d} \pm \text{std}(d)$	$p$
BSS+	$0.0724 \pm 0.0481$	$0.0455 \pm 0.0503$	0.0002
FAROI	$0.0481 \pm 0.0437$	$0.0212 \pm 0.0431$	0.0243
S-BSS-DC	$0.0361 \pm 0.0209$	$0.0092 \pm 0.0152$	0.0071
S-BSS-vecDC	$0.0269 \pm 0.0129$	–	–

$\bar{\mu}_{\text{MSE}}$  and  $\text{std}(\mu_{\text{MSE}})$  are the mean and standard deviation of  $\mu_{\text{MSE}}$  for the 24 curves.

$\bar{d}$  and  $\text{std}(d)$  is the average and standard deviation of paired differences of  $\mu_{\text{MSE}}$  of the S-BSS-vecDC method and that of the competing method (positive difference means improvement).

$p$  is the  $p$ -value of the paired two-tailed t-test of paired differences between MSEs of the S-BSS-vecDC method and other respective methods (BSS+, FAROI, S-BSS-DC).

$\text{std}(\mu_{\text{MSE}})$ . The  $\bar{d}$  denotes the average difference (positive difference means improvement) of the S-BSS-vecDC's MSE from that of the competing methods, and  $p$  denotes the  $p$ -value of the statistical paired sample two-tailed t-test of MSEs from the S-BSS-vecDC method and from other methods, demonstrating that the improvement of the S-BSS-vecDC is statistically significant.

On the tested data, the proposed S-BSS-vecDC algorithm outperforms all other used algorithms in term of proximity of the estimated TACs to those obtained by the experienced physician from the GT data.

#### D. Experiment With Clinical Dynamic Renal Studies in Patients

The same type of experiment as in Section IV-C was performed with real data from dynamic renal scintigraphy. A set of 19 studies was analyzed and the same parameters as in Section IV-C were computed. Data of dynamic renal scintigraphy in patients were chosen from a large set of anonymized data considered to be included into the database [29]. The criterion of choice was clear visibility of dynamic structures. Each sequence consists of 100–180 images of resolution  $128 \times 128$  pixels. Rectangular ROI was placed around each kidney and then processed by BSS+, FAROI, S-BSS-DC, and S-BSS-vecDC algorithms. In this data, true TACs were not available and a physician experienced in analysis of dynamic

TABLE III  
STATISTICAL COMPARISON OF ESTIMATED AND REFERENCE TACS  
OF RENAL PARENCHYMA IN CLINICAL DATA

MSE	38 parenchyma curves		
algorithm	$\bar{\mu}_{\text{MSE}} \pm \text{std}(\mu_{\text{MSE}})$	$\bar{d} \pm \text{std}(d)$	$p$
BSS+	$0.0349 \pm 0.0271$	$0.0165 \pm 0.0268$	0.0005
FAROI	$0.0426 \pm 0.0323$	$0.0242 \pm 0.0329$	<0.0001
S-BSS-DC	$0.0240 \pm 0.0280$	$0.0056 \pm 0.0223$	0.1296
S-BSS-vecDC	$0.0184 \pm 0.0161$	–	–

$\bar{\mu}_{\text{MSE}}$  and  $\text{std}(\mu_{\text{MSE}})$  are the mean and standard deviation of the  $\mu_{\text{MSE}}$  from 38 parenchyma curves.

$\bar{d}$  and  $\text{std}(d)$  is the average and standard deviation of paired differences of  $\mu_{\text{MSE}}$  of the S-BSS-vecDC method and that of the competing method (positive difference means improvement).

$p$  is the  $p$ -value of the paired two-tailed t-test of MSEs from the S-BSS-vecDC method and from the competing methods.

renal studies was asked to extract the curves. The ROIs were defined manually on a computer screen with a cursor using anatomical and physiological knowledge, and currently recommended procedure guidelines [30].

The results in terms of MSE are summarized in Table III, using the same methodology as in the previous Section. In case of clinical data used in our experiment, the S-BSS-vecDC method provided better estimates than any other competing method, however, the difference from S-BSS-DC is not statistically significant.

## V. CONCLUSION

We propose a probabilistic model of dynamic image sequences that involves superposition of observed structures in the recorded images and convolution of time-activity curves with unknown input function and unknown convolution kernels. This model is used to develop a method of blind source separation that decomposes the sequence into a sum a source images and their corresponding time-activity curves. The key assumption of the method is that the images of the underlying tissues and the unknown convolution kernels are sparse. The assumption of sparsity is incorporated using Bayesian approach. Specifically, the chosen form of prior probability distributions of the source images and the convolution kernels assigns higher probability to sparse structures. Due to the concept of hierarchical priors, the method also automatically adjusts all uncertainties in the model with the exception of the maximum number of sources which remains to be the only tuning parameter. Hence, the method can be considered as a general purpose tool for source separation.

To demonstrate separation ability of the proposed method, it has been applied to 12 simulated dynamic renal studies with known reference time-activity curves and to 19 clinical dynamic renal studies in adult patients in which the reference time-activity curves were extracted by an experienced physician. With both data sets, the proposed method estimated the reference time-activity curves with significantly lower errors than other state-of-the-art methods. Notably, the results were achieved without any domain-specific assumptions. To test performance of the method and facilitate its validation in analysis of various dynamic data sets, MATLAB implementation of the algorithm is available for download.

## APPENDIX A

## VB EQUATIONS OF S-BSS-VECDC MODEL

Shaping parameters of posterior distributions (22)–(25) are given as

$$\begin{aligned}\vartheta &= \vartheta_0 + \frac{pn}{2}, \\ \rho &= \rho_0 + \frac{1}{2} \text{tr} \left( DD^T - 2\widehat{A}\widehat{X}^T D^T \right) + \frac{1}{2} \text{tr} \left( \widehat{A}^T \widehat{A} \widehat{X}^T \widehat{X} \right), \\ \Sigma_{\mathbf{u}}^{-1} &= \widehat{A}^T \widehat{A} \otimes \widehat{\omega} \widehat{B}^T \widehat{B} + \text{diag}(\text{vec}(\widehat{\Upsilon})) \\ \mu_{\mathbf{u}} &= \Sigma_{\mathbf{u}} \left( \widehat{A}^T \widehat{A} \otimes \widehat{\omega} \widehat{B}^T \widehat{B} \right) \text{vec} \left( \widehat{B}^T \widehat{B}^{-1} \widehat{B}^T D^T \widehat{A} \widehat{A}^T \widehat{A}^{-1} \right), \\ \alpha_k &= \alpha_{k,0} + \frac{1}{2} \mathbf{1}_{n,1}, \beta_k = \beta_{k,0} + \frac{1}{2} \text{diag} \left( \widehat{\mathbf{u}}_k \widehat{\mathbf{u}}_k^T \right) \\ \Sigma_{\mathbf{b}}^{-1} &= \zeta I_n + \widehat{\omega} \sum_{i,j=1}^r \left( \widehat{\mathbf{a}}_i^T \widehat{\mathbf{a}}_j \right) \left( \sum_{k,l=0}^{n-1} \Delta_k^T \Delta_l \widehat{\mathbf{u}}_{k+1,j} \widehat{\mathbf{u}}_{l+1,i} \right) \\ \mu_{\mathbf{b}} &= \Sigma_{\mathbf{b}} \widehat{\omega} \sum_{k=1}^r \left( \sum_{j=0}^{n-1} \Delta_j \widehat{\mathbf{u}}_{j+1,k} \right)^T D^T \widehat{\mathbf{a}}_k \\ \zeta &= \zeta_0 + \frac{n}{2}, \eta = \eta_0 + \frac{1}{2} \text{tr} \left( \widehat{\mathbf{b}}^T \widehat{\mathbf{b}} \right) \\ \Sigma_{\widehat{\mathbf{a}}_i}^{-1} &= \widehat{\omega} \sum_{j=1}^n \left( \widehat{\mathbf{x}}_j^T \widehat{\mathbf{x}}_j \right) + \text{diag}(\widehat{\Xi}_i), \\ \mu_{\widehat{\mathbf{a}}_i} &= \Sigma_{\widehat{\mathbf{a}}_i} \widehat{\omega} \sum_{j=1}^n \left( \widehat{\mathbf{x}}_j d_{i,j} \right) \\ \phi_i &= \phi_{i,0} + \frac{1}{2} \cdot \mathbf{1}_{r,1}, \psi_i = \psi_{i,0} + \frac{1}{2} \text{diag} \left( \widehat{\mathbf{a}}_i^T \widehat{\mathbf{a}}_i \right).\end{aligned}$$

Here,  $\widehat{x}$  denotes a moment of respective distribution,  $\text{tr}()$  denotes a trace of argument,  $\text{diag}()$  denotes a square matrix with argument vector on diagonal and zeros otherwise or a vector composed from diagonal element of argument matrix,  $\mathbf{1}_{n,1}$  denotes the matrix of ones of dimension  $n \times 1$ , the auxiliary matrix  $\Delta_k \in \mathbb{R}^{n \times n}$  is defined as  $(\Delta_k)_{i,j} = \begin{cases} 1, & \text{if } i - j = k, \\ 0, & \text{otherwise,} \end{cases}$  and standard moments of required probability distributions are given Appendix B and, e.g., in [35, Appendix].

## APPENDIX B

## TRUNCATED NORMAL DISTRIBUTION

Truncated normal distribution  $t\mathcal{N}$  of a scalar variable  $x$  on interval  $a \leq x \leq b$  is defined as

$$x \sim t\mathcal{N}(\mu, \sigma, [a, b]) = \frac{\sqrt{2} \exp((x - \mu)^2)}{\sqrt{\pi} \sigma (\text{erf}(\beta) - \text{erf}(\alpha))} \chi_{[a,b]}(x)$$

where  $\alpha = a - \mu/\sqrt{2\sigma}$ ,  $\beta = b - \mu/\sqrt{2\sigma}$ , function  $\chi_{[a,b]}(x)$  is defined as  $\chi_{[a,b]}(x) = 1$  if  $x \in [a, b]$  and  $\chi_{[a,b]}(x) = 0$  otherwise.  $\text{erf}()$  is the error function. The moments of truncated normal distribution are

$$\widehat{x} = \mu - \sqrt{\sigma} \frac{\sqrt{2} [\exp(-\beta^2) - \exp(-\alpha^2)]}{\sqrt{\pi} (\text{erf}(\beta) - \text{erf}(\alpha))} \quad (28)$$

$$\widehat{x^2} = \sigma + \mu \widehat{x} - \sqrt{\sigma} \frac{\sqrt{2} [b \exp(-\beta^2) - a \exp(-\alpha^2)]}{\sqrt{\pi} (\text{erf}(\beta) - \text{erf}(\alpha))}. \quad (29)$$

## ACKNOWLEDGMENT

The authors acknowledged with thanks cooperation of Dr. M. Samal of Charles University Prague who provided us with part of experimental and clinical data and supported us with comments on interpretation of the results and on the motivation part of the paper.

## REFERENCES

- [1] R. Di Paola, J. Bazin, F. Aubry, A. Aurengo, F. Cavailloles, J. Herry, and E. Kahn, "Handling of dynamic sequences in nuclear medicine," *IEEE Trans. Nucl. Sci.*, vol. 29, no. 4, pp. 1310–1321, Aug. 1982.
- [2] M. Margadán-Méndez, A. Juslin, S. Nesterov, K. Kalliokoski, J. Knuuti, and U. Ruotsalainen, "ICA based automatic segmentation of dynamic cardiac PET images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 3, pp. 795–802, May 2010.
- [3] J. Mateos-Pérez, M. Desco, M. Dae, C. García-Villalba, L. Cussó, and J. Vaquero, "Automatic TAC extraction from dynamic cardiac PET imaging using iterative correlation from a population template," *Comput. Methods Programs Biomed.*, vol. 111, no. 2, pp. 308–314, 2013.
- [4] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. Med. Imag.*, vol. 32, no. 5, pp. 821–837, May 2013.
- [5] J. Zaknun, H. Rajabi, A. Piepsz, I. Roca, and M. Dondi, "The international atomic energy agency software package for the analysis of scintigraphic renal dynamic studies: A tool for the clinician, teacher, and researcher," in *Proc. Semin. Nucl. Med.*, 2011, vol. 41, pp. 73–80.
- [6] M. Caglar, G. Gedik, and E. Karabulut, "Differential renal function estimation by dynamic renal scintigraphy: Influence of background definition and radiopharmaceutical," *Nucl. Med. Commun.*, vol. 29, no. 11, p. 1002, 2008.
- [7] E. Garcia, R. Folks, S. Pak, and A. Taylor, "Totally automatic definition of renal regions-of-interest from Tc-99m MAG3 renograms: Validation in patients with normal kidneys and in patients with suspected renal obstruction," *Nucl. Med. Commun.*, vol. 31, no. 5, p. 366, 2010.
- [8] M. Samal and C. Nimmon, "Automatic definition of renal regions of interest," *Nucl. Med. Commun.*, vol. 32, no. 5, pp. 419–420, 2011.
- [9] J. Fleming, "Quantitative measurements from gamma camera images," in *Mathematical Techniques in Nuclear Medicine*, C. Greaves, Ed. New York: Inst. Phys. Eng. Med., 2011, pp. 22–74.
- [10] A. Houston, "Factor analysis," in *Math. Techniques in Nucl. Medicine Inst. of Physics and Eng. in Medicine*, C. Greaves, Ed. York: , 2011, pp. 153–171.
- [11] D. Vriens, L.-F. de Geus-Oei, W. Oyen, and E. Visser, "A curve-fitting approach to estimate the arterial plasma input function for the assessment of glucose metabolic rate and response to treatment," *J. Nucl. Med.*, vol. 50, no. 12, pp. 1933–1939, 2009.
- [12] M. Alf, M. Wyss, A. Buck, B. Weber, R. Schibli, and S. Krämer, "Quantification of brain glucose metabolism by 18F-FDG PET with real-time arterial and image-derived input function in mice," *J. Nucl. Med.*, vol. 54, no. 1, pp. 132–138, 2013.
- [13] B. Lanz, C. Poitry-Yamate, and R. Gruetter, "Image-derived input function from the vena cava for 18F-FDG PET studies in rats and mice," *J. Nucl. Med.*, vol. 55, no. 8, pp. 1380–1388, Jun. 2014.
- [14] D. Sutton, "Deconvolution of the renogram," in *Mathematical Techniques in Nuclear Medicine*, C. Greaves, Ed. York: Inst. Phys. Eng. Med., 2011, pp. 106–129.
- [15] C. Patlak *et al.*, "Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data," *J. Cereb. Blood Flow Metab.*, vol. 3, no. 1, pp. 1–7, 1983.
- [16] J. Fleming and P. Kemp, "A comparison of deconvolution and the Patlak-Rutland plot in renography analysis," *J. Nucl. Med.*, vol. 40, no. 9, p. 1503, 1999.
- [17] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.
- [18] J. Miskin, "Ensemble learning for independent component analysis," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 2000.
- [19] N. Gillis and S. Vavasis, "Fast and robust recursive algorithms for separable nonnegative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 698–714, Apr. 2014.
- [20] H. Attias and C. Schreiner, "Blind source separation and deconvolution: The dynamic component analysis algorithm," *Neural Computat.*, vol. 10, no. 6, pp. 1373–1424, 1998.

- [21] D. Riabkov and E. Di Bella, "Estimation of kinetic parameters without input functions: analysis of three methods for multichannel blind identification," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 11, pp. 1318–1327, Nov. 2002.
- [22] L. Chen, P. Choyke, T.-H. Chan, C.-Y. Chi, G. Wang, and Y. Wang, "Tissue-specific compartmental analysis for dynamic contrast-enhanced MR imaging of complex tumors," *IEEE Trans. Med. Imag.*, vol. 30, no. 12, pp. 2044–2058, Dec. 2011.
- [23] V. Šmídl and O. Tichý, "Automatic regions of interest in factor analysis for dynamic medical imaging," in *Proc. IEEE Int. Symp. Biomed. Imag.*, May 2012, pp. 158–161.
- [24] O. Tichý, V. Šmídl, and M. Šámal, "Model-based extraction of input and organ functions in dynamic medical imaging," in *Proc. Conf. Computat. Vis. Med. Image Process.*, 2013, pp. 75–80.
- [25] A. Kuruc, W. Caldicott, and S. Treves, "Improved deconvolution technique for the calculation of renal retention functions," *Comp. Biomed. Res.*, vol. 15, no. 1, pp. 46–56, 1982.
- [26] V. Šmídl and O. Tichý, "Sparsity in Bayesian Blind Source Separation and Deconvolution," in *Machine Learning and Knowledge Discovery in Databases*, H. Blockeel, Ed. et al. Heidelberg, Germany: Springer, 2013, vol. 8189, Lecture Notes Comput. Sci, pp. 548–563.
- [27] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [28] G. Brodin, K. Gleisner, and M. Ljungberg, "Dynamic  $^{99m}\text{Tc}$ -MAG3 renography: images for quality control obtained by combining pharmacokinetic modelling, an anthropomorphic computer phantom and Monte Carlo simulated scintillation camera imaging," *Phys. Med. Biol.*, vol. 58, no. 10, p. 3145, 2013.
- [29] Database of dynamic renal scintigraphy [Online]. Available: <http://www.dynamicrenalstudy.org>, Jul. 02, 2014.
- [30] I. Gordon, A. Piepsz, and R. Sixt, "Guidelines for standard and diuretic renogram in children," *Eur. J. Nucl. Med. Molec. Imag.*, vol. 38, no. 6, pp. 1175–1188, 2011.
- [31] E. Durand et al., "International scientific committee of radionuclides in nephrourology (ISCORN) consensus on renal transit time measurements," *Seminars Nucl. Med.*, vol. 38, pp. 82–102, 2008.
- [32] R. Neal, "Bayesian learning for neural networks," Ph.D. dissertation, Univ. Toronto, Toronto, Canada, 1995.
- [33] H. Benali, I. Buvat, F. Frouin, J. Bazin, and R. Paola, "A statistical model for the determination of the optimal metric in factor analysis of medical image sequences (FAMIS)," *Phys. Med. Biol.*, vol. 38, p. 1065, 1993.
- [34] C. Bishop and M. Tipping, "Variational relevance vector machines," in *Proc. 16th Conf. Uncertainty Artif. Intell.*, 2000, pp. 46–53.
- [35] V. Šmídl and A. Quinn, *The Variational Bayes Method In Signal Processing*. New York: Springer, 2006.
- [36] M. Woolrich and T. Behrens, "Variational Bayes inference of spatial mixture models for segmentation," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1380–1391, Oct. 2006.
- [37] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus*. New York: Wiley, 2001.