Regularizing Direct Parametric Reconstruction for Dynamic PET with the Method of Sieves

László Szirmay-Kalos and Ágota Kacsó[1],

Abstract—This paper proposes regularization methods for direct parametric dynamic PET reconstruction, when the space-time activity function needs to be recovered from measurements. In case of high spatial and temporal resolution, the reconstruction is statistically poorly defined, requiring the inclusion of a priori information in the form of a penalty term or filtering. The method of sieves executes filtering in each iteration step, i.e. projects the actual estimate into the subspace of acceptable solutions, and has been successful in reconstructing static data. The objective of this paper is to generalize the filtering scheme for spatio-temporal reconstruction, taking into account that accurate kinetic models describing the temporal behavior are non-linear. Fast changes are impossible to distinguish from noise if only a small temporal window is examined, thus the simple extension to 4D does not provide acceptable results. We show that efficient filtering can be obtained if voxel based model parameters are modified according to the time activity functions of neighboring voxels belonging to the same anatomic region. As the dependence of the time activity function on the model parameters is non-linear for sophisticated kinetic models, the filtering step involves a non-linear parameter fitting, which can be solved analytically for the two-tissue compartment model. The presented method is built into the TeraTomo™ system.

I. INTRODUCTION

In dynamic Positron Emission Tomography (PET), we examine the dynamic nature of biological processes, like accumulation and emptying drugs in certain organs, using radiotracers emitting positrons. The positron emitted at a decay may annihilate with an electron, when two oppositely directed gamma-photons are born, which might be detected by the tomograph. The system collects the *events* of simultaneous photon incidents in detector pairs. An event is a composition of the identification of the detector pair, also called *Line Of Response* or *LOR*, and its time of occurrence. The state-of-the-art and previous work on direct estimation of kinetic parameters for dynamic PET are surveyed in review articles [15], [7].

II. DYNAMIC PET RECONSTRUCTION

Generally, we assume that the radiotracer concentration in each voxel V in time t can be expressed by a common *kinetic model* $C(\mathbf{p}_V, t)$, where spatial dependent properties of voxel V are encoded in a low dimensional vector of kinetic parameters $\mathbf{p}_V$. Such models can be defined based on the mathematical description of the biological/chemimal processes or on compartment analysis [3], [17], [16]. In this paper we consider the popular *two-tissue compartment model*

[1]: Budapest University of Technology and Economics (e-mail: szirmay@iit.bme.hu).

that is defined by the following formula of parameter vector $\mathbf{p} = (f_v, a_1, a_2, \alpha_1, \alpha_2)$:

$$C(\mathbf{p}, t) = (1 - f_v)F(a_1, a_2, \alpha_1, \alpha_2) * C_p(t) + f_v C_w(t), \quad (1)$$

where

$$F(a_1, a_2, \alpha_1, \alpha_2) = a_1\alpha_1 \exp(-\alpha_1 t) + a_2\alpha_2 \exp(-\alpha_2 t) \quad (2)$$

is the *impulse response* of the compartment model, f_v is the *fractional volume of blood*, $*$ stands for convolution, $C_p(t)$ is the known *blood activity function* serving as the input signal to the compartment system, and $C_w(t)$ is the also known *whole blood concentration function*. The objective of the reconstruction is the determination of parameter vector $\mathbf{p}$ in each voxel from the list of events.

The measurement time is decomposed into finite time intervals, called *frames*, of widths $\Delta t_1, \ldots, \Delta t_{N_T}$ and centers $t_1, \ldots, t_{N_T}$, and events are binned in frames. The expected number of radioactive decays, i.e. number of positrons generated by a unit volume of voxel V in time frame T covering $[t_T - \Delta t_T/2, t_T + \Delta t_T/2)$ is

$$\tilde{x}_T(\mathbf{p}_V) = \int_{t_T - \Delta t_T/2}^{t_T + \Delta t_T/2} C(\mathbf{p}_V, t) \exp(-\mu t)\mathrm{d}t. \quad (3)$$

where μ is the decay of the radiotracer.

During iterative *Expectation Maximization* (ML-EM) reconstruction [8], unknown coefficients are found to maximize the probability of the actually measured data. Assuming that the measured number of hits in LOR L in time interval Δt_T follows a Poisson distribution and is statistically independent of other LORs and frames, the log-likelihood of the current measurement is

$$\log \mathcal{L} = \sum_L \sum_T \left(y_{L,T} \log \tilde{y}_{L,T} - \tilde{y}_{L,T} - \log(y_{L,T}!) \right) \quad (4)$$

where $y_{L,T}$ is the number of *measured events* in LOR L and frame T, and $\tilde{y}_{L,T}$ is its expected value, which is the sum of the contributions of all N_V voxels in the volume at this time:

$$\tilde{y}_{L,T} = \sum_{V=1}^{N_V} \mathbf{A}_{L,V} \tilde{x}_T(\mathbf{p}_V) \quad (5)$$

where *system matrix* $\mathbf{A}_{L,V}$ expresses the probability that a decay in voxel V generates an event in LOR L.

The reconstruction means the maximization of the log-likelihood in Equation 4, which leads to the following non-linear equation:

$$\sum_T \frac{\partial \tilde{x}_T(\mathbf{p}_V)}{\partial \mathbf{p}_{V,P}} \left(\frac{x_{V,T}}{\tilde{x}_T(\mathbf{p}_V)} - 1 \right) = 0, \quad (6)$$

where $x_{V,T}$ is the result of a static *forward projection* evaluating equation 5, and a *back projection* taking the data from frame T

$$x_{V,T} = \tilde{x}_T(\mathbf{p}_V) \cdot \frac{\sum_L \mathbf{A}_{L,V} \frac{y_{L,T}}{\tilde{y}_{L,T}}}{\sum_L \mathbf{A}_{L,V}}.$$

In this equation expected activity $\tilde{x}_{V,T}$ depends on unknown parameter vector $\mathbf{p}_V$ of the given voxel, while $x_{V,T}$ depends on the parameter vectors of all voxels. Additionally, $x_{V,T}$ is the only factor that is affected by the elements of the system matrix. Thus, if $x_{V,T}$ were known, then the computation could be decoupled for different voxels and can be made independent of the huge system matrix.

To achieve this, a subiteration is included into the main iteration solution of this equation. In the subiteration expensive terms $x_{V,T}$ are not re-evaluated, they are updated just in the main iteration steps. Assuming that $x_{V,T}$ is constant, Equation 6 describes just a single voxel, and can thus be solved independently for all voxels. We use the *Damped Newton* and the *Levenberg-Marquardt* algorithms for the solution. These algorithms compute the parameter vector determining a time activity curve $x(\mathbf{p}_V, t)$ that fits to $x_{V,T}$.

Putting the discussed projections and curve fitting together, we obtain the following pseudo-code of the reconstruction:

> **for** $n = 1$ **to** $n_{\max}$ **do**
>   **for** $T = 1$ **to** N_T **do**
>   **foreach voxel** V *// evaluation*
> $\tilde{x}_T(\mathbf{p}_V) = $ Compute integral of Equation 3
>   **foreach LOR** L *// forward projection*
> $\tilde{y}_{L,T} = \sum_{V'} \mathbf{A}_{L,V'} \tilde{x}_T(\mathbf{p}_V)$
>   **foreach voxel** V *// back projection*
> $x_{V,T} = \tilde{x}_T(\mathbf{p}_V) \cdot \dfrac{\sum_L \mathbf{A}_{L,V} \frac{y_{L,T}}{\tilde{y}_{L,T}}}{\sum_L \mathbf{A}_{L,V}}$
>   **endfor**
>   **foreach voxel** V *// curve fitting*
> $\mathbf{p}_V = $ Solve Equation 6
> **endfor**

If fast dynamic changes are to be recovered, frames must be short and consequently the number of events in a frame is rather low. This means that reconstruction done independently in frames is either impossible or leads to very noisy data. To attack this problem, regularization is needed that enforces the smoothness both in the temporal and the spatial domains.

III. PREVIOUS WORK ON REGULARIZATION

There are various options to regularize the solution, which is essential in the case of inverse problems. Note that temporal regularization is automatically provided by the application of a prescribed kinetic model since we allow just this class of functions to be solutions. For spatial regularization, one option is the modification of the optimization target by a regularization term that penalizes unacceptable solutions having too high spatial or temporal variation. An appropriate penalty term is the *Total Variation* (TV) of the solution [6], [5]. Total variation regularization may create stair-like artifacts, which can be reduced by *Bregman iteration* [1], [18], [12]. The inclusion of the anatomic information into spatial regularization is straightforward, smoothness should be imposed only inside

anatomically homogeneous regions but not on their boundaries [2].

Another possibility for spatial regularization is the application of the *method of sieves*, where the optimization target is not modified, but the iterated approximation is filtered in each iteration step. Several authors proposed the inclusion of a voxel space filtering step in the reconstruction loop [9], [4] and it turned out that it is equivalent to the method of sieves that seeks to constrain the EM solution to a subspace of all possible solutions [10], [11], [14]. Mathematically, this approach projects the current estimate into the subspace of acceptable solutions in each iteration. Filtering can also exploit anatomic information gathered by a CT or an MR [13].

IV. SPATIAL FILTERING OF TIME ACTIVITY FUNCTIONS

Note that the reconstruction algorithm can also be imagined as a static forward projection and back projection independently for each frames, then parameter fitting in each voxel, which implicitly executes temporal filtering. Iterating these steps will establish a control loop of Fig. 1. Spatial filtering can be included at various stages of the process. It would be possible to filter $x_{V,T}$ before *Curve Fitting*, parameter values $\mathbf{p}_V$ after *Curve Fitting* and expected voxel intensities $\tilde{x}_{V,T}$ after *Evaluation* and before *Forward Projection*. As *Curve Fitting* and *Evaluation* are non-linear operations, order of filtering matters, and the results of the different options will be different. If filtering is done in the parameter space, then the EM estimation is not modified, just its proposed result is projected into acceptable solutions. On the other hand, as the number of frames is significantly larger than the number of parameters, filtering after *Curve Fitting* has to deal with much smaller parameters. Because of these reasons, we proposed the application of spatial filtering in parameter space, right after *Curve Fitting*.
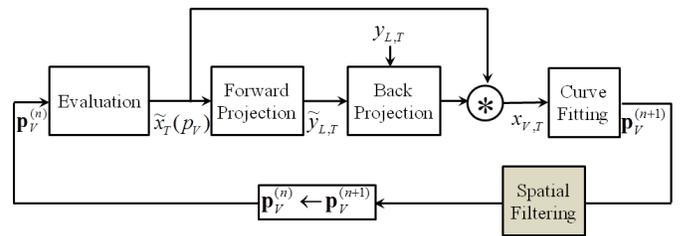

Fig. 1. Reconstruction as a control loop

For spatial filtering of the parameters, we wish to set the time activity function of each voxel to the weighted average of neighboring voxels of the same anatomic region. Weights $G(V', V)$ can be selected as a distance dependent Gaussian if V and V' belong to the same anatomic region and zero otherwise. Thus, our target filtered time activity function in voxel V is:

$$
\begin{aligned}
C(\tilde{\mathbf{p}}_V, t) &\approx \frac{\sum_{V'} C(\mathbf{p}_{V'}, t) G(V', V)}{\sum_{V'} G(V', V)} \\
&= \sum_{V'} C(\mathbf{p}_{V'}, t) w(V', V) \qquad (7)
\end{aligned}
$$

where $w(V',V)$ are the normalized weights that sum up to 1 when all neighboring V' voxels belonging to the same anatomic region are taken into account. Note that in the classical application of the method of sieves, this filtering is computed on scalars, but here we need to average functions. If the time activity curves were expressed as a linear combination of pre-defined basis functions, then the filtering of functions would be equivalent to the filtering of the parameters defining the time activity functions. However, sophisticated kinetic models, e.g. the *two-tissue compartment model* defined by equation 1 are non-linear. Clearly, filtering the parameters independently does not work necessarily for non-linear functions, because there is no guarantee that the resulting function will be "between" the filtered functions.

In our proposed filtering algorithm, f_v is handled separately and is filtered first:

$$\tilde{f}_{v,V} = \sum_{V'} f_{v,V'} w(V',V). \tag{8}$$

Having fixed its value, the remaining filtered parameters of $\tilde{\mathbf{p}} = (\tilde{f}_v, \tilde{a}_1, \tilde{a}_2, \tilde{\alpha}_1, \tilde{\alpha}_2)$ of voxel V must be obtained. As the blood input function is shared, the similarity of the time activity functions requires

$$\tilde{g}(t) = (1 - f_{v,V}) F(\tilde{a}_{1,V}, \tilde{a}_{2,V}, \tilde{\alpha}_{1,V}, \tilde{\alpha}_{2,V}, t) \approx$$

$$\sum_{V'} (1 - f_{v,V'}) F(a_{1,V'}, a_{2,V'}, \alpha_{1,V'}, \alpha_{2,V'}, t) w(V',V) = g(t). \tag{9}$$

Our objective is to make two functions $\tilde{g}(t)$ and $g(t)$ approximately equal. As we have four unknown parameters left, four constraints are needed, that can fall into two main categories.

Collocation constraints require the values or the mth derivatives be equal at fixed points of time. As F is composed of exponentials, its behavior is well defined by the value and the derivatives in $t = 0$. Thus, to find the filtered parameters, we can also require the equality of the mth derivatives at $t = 0$:

$$\left.\frac{\mathrm{d}^m}{\mathrm{d}t^m} \tilde{g}(t)\right|_{t=0} = \left.\frac{\mathrm{d}^m}{\mathrm{d}t^m} g(t)\right|_{t=0}.$$

Galerkin constraints, on the other hand, require the projections of the two functions into a subspace defined by basis functions $b_m(t)$ be equal:

$$\int_0^\infty \tilde{g}(t) b_m(t) \mathrm{d}t = \int_0^\infty g(t) b_m(t) \mathrm{d}t.$$

Galerkin constraint with basis function $b_0(t) = 1$ enforces the energy conservation of the filtering scheme, i.e. the total activity after filtering will be equal to the total unfiltered activity. Let us choose the mth basis function as $b_m(t) = t^m$, which means that the subspace defined by the basis functions is the space of polynomials. Increasing m, larger t values have more important role to define the unknown parameters. In this way, Galerkin constraints can complement collocation constraints focusing just on the initial behavior of the functions.

As F is a linear combination of exponentials of form $a\alpha \exp(-\alpha t)$, the derivatives and the projection integrals are also combinations of the derivatives and integrals of exponentials:

$$\left.\frac{\mathrm{d}^m}{\mathrm{d}t^m} a\alpha \exp(-\alpha t)\right|_{t=0} = a\alpha^{m+1},$$

$$\int_0^\infty a\alpha \exp(-\alpha t) t^m \mathrm{d}t = a\alpha^{-m} m!. \tag{10}$$

Thus, for this exponential family, collocation constraints and Galerkin constraints become similar and extend the domain of m to all integers.

Imposing these requirement on equation 9 and taking the formula of the equation 2 into account, we obtain

$$a_1 \alpha_1^m + a_2 \alpha_2^m = A_m, \tag{11}$$

where we omitted subscript V for the sake of simplicity, and A_m values are weighted averages:

$$A_m = \sum_{V'} \frac{1 - f_{v,V'}}{1 - \tilde{f}_v} \left(a_{1,V'} \alpha_{1,V'}^m + a_{2,V'} \alpha_{2,V'}^m\right) w(V',V).$$

Collocation and Galerkin constraints provide infinitely many possibilities to execute the filtering on functions, from which we should choose one based on the application dependent concept of similarity and on the complexity of the scheme since this operation should be executed for every voxel and every iteration. Let us consider four constraints defined by consecutive m values m, $m+1$, $m+2$, $m+3$.

Denoting the left side of equation 11 with a given m by l_m, it can be seen that

$$(l_{m+1} - \tilde{\alpha}_1 l_m)(l_{m+3} - \tilde{\alpha}_1 l_{m+2}) = (l_{m+2} - \tilde{\alpha}_1 l_{m+1})^2.$$

Thus, the same rule should also be applicable to the right sides of these equations:

$$(A_{m+1} - \tilde{\alpha}_1 A_m)(A_{m+3} - \tilde{\alpha}_1 A_{m+2}) = (A_{m+2} - \tilde{\alpha}_1 A_{m+1})^2.$$

This results in the following second order equation for $\tilde{\alpha}_1$:

$$(A_m A_{m+2} - A_{m+1}^2)\tilde{\alpha}_1^2 + (A_{m+1}A_{m+2} - A_m A_{m+3})\tilde{\alpha}_1 +$$

$$(A_{m+1}A_{m+3} - A_{m+2}^2) = 0,$$

which can be solved analytically. Knowing $\tilde{\alpha}_1$, parameter $\tilde{\alpha}_2$ can be computed as

$$\tilde{\alpha}_2 = \frac{A_{m+3} - A_{m+2}\tilde{\alpha}_1}{A_{m+2} - A_{m+1}\tilde{\alpha}_1},$$

or it can equivalently taken as the second root of the quadratic equation. The remaining unknown parameters can be obtained using the following substitutions:

$$\tilde{a}_2 = \frac{A_{m+1} - A_m \tilde{\alpha}_1}{\tilde{\alpha}_2^m(\tilde{\alpha}_2 - \tilde{\alpha}_1)}, \quad \tilde{a}_1 = \frac{A_m - \tilde{\alpha}_2^m \tilde{a}_2}{\tilde{\alpha}_1^m}.$$

These equations work for arbitrary m, which is worth selecting to include the zero value in $m, m+1, m+2, m+3$ to guarantee energy conservation. There are four options satisfying this requirement: $m = 0$, $m = -1$, $m = -2$, $m = -3$.

Figure 2 shows the application of the proposed non-linear averaging scheme for two functions and compares it to the parameter-wise linear averaging. Note that the proposed method really puts the average in between the two functions to be averaged, but parameter-wise averaging fails due to the non-linearity of the problem.

V. RESULTS

To examine the proposed method, we use a 2D mathematical tomograph model (Fig. 3) where the detector ring contains 90 detector crystals and each of them is of size 2.2 in voxel units and participates in 47 LORs connecting this crystal to crystals being in the opposite half circle, thus the total number of LORs is $90 \times 47/2 = 2115$. The voxel array to be reconstructed is in the middle of the ring and has 32×32 resolution, i.e. 1024 voxels. The measured data is obtained with Monte Carlo simulation of a brain model where there are three homogeneous regions, including the white matter, the gray matter, and the background. The simulation generated 16k hits in total, distributed in 100 frames covering a 10 second long interval. Note that this is a low statistic measurement where the average number of hits per frame per LOR is less than 0.08.
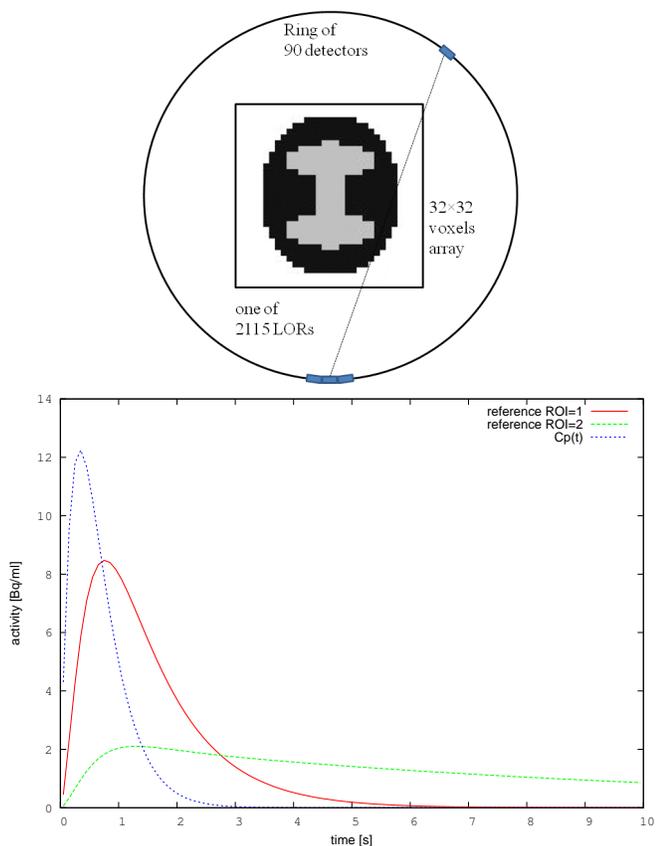

Fig. 3. 2D tomograph model

The measurements are reconstructed without any spatial regularization, with TV penalty on the activity values, and with the proposed method. The results are shown in Figs. 4–6. Note that TV penalty reduces both the bias and the variance of the reconstruction, but is poorer than the result obtained with the method of sieves. Efficient penalty based regularization algorithms use the *one-step-late option*, which has negligible computational overhead, but may prohibit convergence when the regularization parameter λ is too strong. Here we used $\lambda = 0.1$, which is found to be optimal, i.e. lower values cause higher variance reconstructions, higher values divergence and
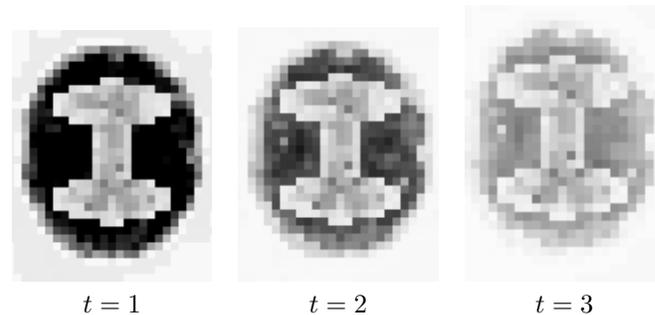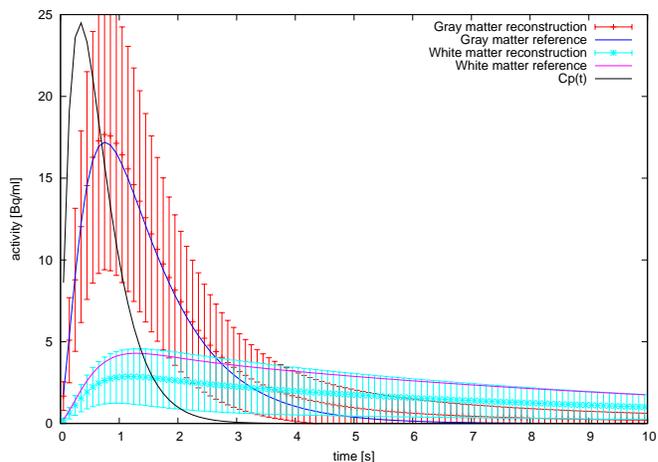

Fig. 4. Time activity curves reconstructed with *no regularization*, showing the average and the standard deviation for the anatomic regions in $t = 1$, $t = 2$, and $t = 3$.

stronger chess-pattern like artifacts. Thus, using the one-step-late option, the applicability of TV reguralization is limited. This is not the case for the method of sieves, where the homogeneity of anatomic regions can be enforced without limits by increasing the standard deviation of the position based Gaussian.

VI. CONCLUSIONS

In this paper we investigated the regularization problem of direct parametric PET reconstruction. We proposed the application of spatial averaging of the voxel-based time activity functions, i.e. the method of sieves, as a way of regularization. To implement the basic idea, we also addressed the problem of averaging non-linear functions in parameter space. The proposed method can use anatomic information about region boundaries and remains stable for aggressive filtering as well. In our fully-3D implementation all steps are implemented on the GPU where the added computational cost of filtering is negligible with respect to forward and back projection calculations.
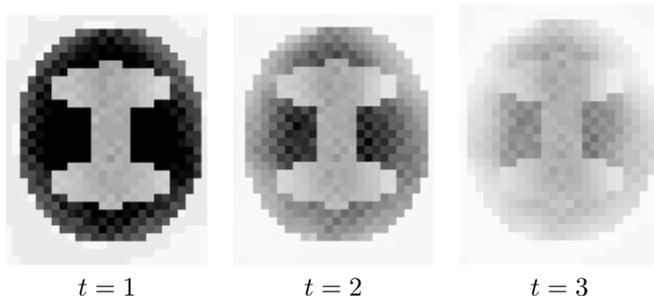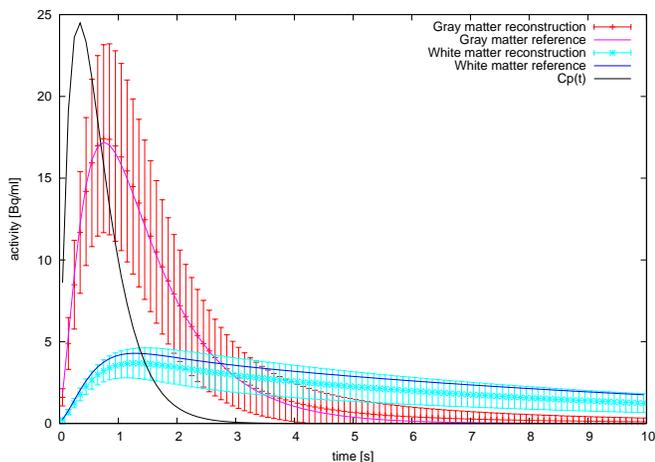
Fig. 5. Time activity curves reconstructed with *TV regularization* of regularization parameter $\lambda = 0.1$, showing the average and the standard deviation for the anatomic regions in $t = 1$, $t = 2$, and $t = 3$
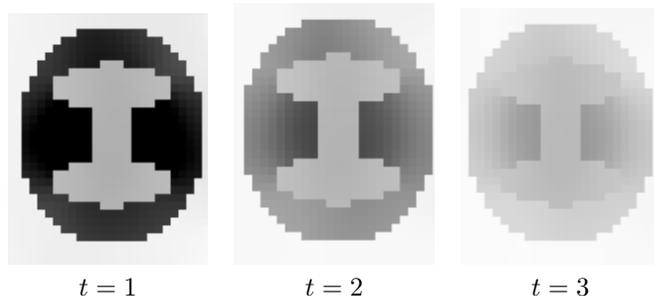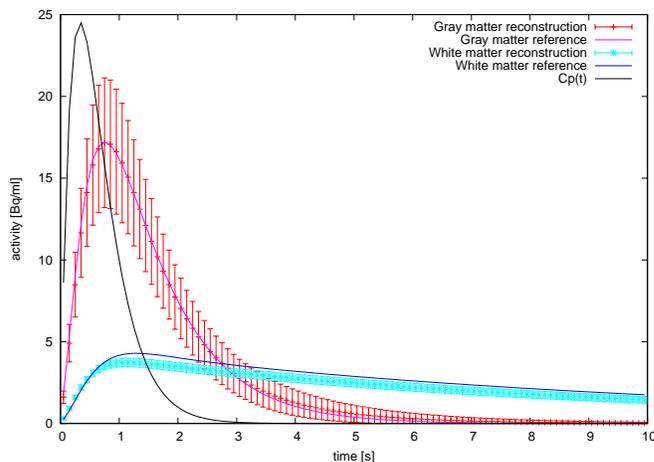

Fig. 6. Time activity curves reconstructed with the proposed *method of the sieves* setting $\sigma = 1.5$, showing the average and the standard deviation for the anatomic regions in $t = 1$, $t = 2$, and $t = 3$

REFERENCES

[1] L. Bregman. The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[2] C. Chan, R. Fulton, D. D. Feng, and S. Meikle. Regularized image reconstruction with an anatomically adaptive prior for positron emission tomography. *Physics in Medicine and Biology*, 54:7379–400, 2009.

[3] M.E. Kamasak, C.A. Bouman, E.D. Morris, and K. Sauer. Direct reconstruction of kinetic parameter images from dynamic pet data. *IEEE Transactions on Medical Imaging*, 24(5):636–650, May 2005.

[4] M. Magdics, L. Szirmay-Kalos, B. Tóth, and T. Umenhoffer. Filtered sampling for PET. In *IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2012.

[5] M. Magdics, B. Tóth, B. Kovács, and L. Szirmay-Kalos. Total variation regularization in PET reconstruction. In *KÉPAF*, pages 40–53, 2011.

[6] M. Persson, D. Bone, and H. Elmqvist. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. *Physics in Medicine and Biology*, 46(3):853–866, 2001.

[7] A.J. Reader and J. Verhaeghe. 4d image reconstruction for emission tomography. *Physics in Medicine and Biology*, 59(22):R371, 2014.

[8] L. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1:113–122, 1982.

[9] E.T.P. Slijpen and R.J. Beekman. Comparison of post-filtering and filtering between iterations for SPECT reconstruction. *IEEE Transactions on Nuclear Science*, 46(6):2233–2238, 1999.

[10] D.L. Snyder and M.I. Miller. The use of sieves to stabilize images produced with the em algorithm for emission tomography. *IEEE Transactions on Nuclear Science*, 32(5):3864–3872, Oct 1985.

[11] D.L. Snyder, M.I. Miller, L.J. Thomas, and D.G. Politte. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Transactions on Medical Imaging*, 6(3):228–238, Sept 1987.

[12] L. Szirmay-Kalos, B. Tóth, and G. Jakab. Efficient Bregman iteration in fully 3d PET. In *IEEE Nuclear science symposium and medical imaging conference*, MIC'14, 2014.

[13] L. Szirmay-Kalos, M. Magdics, and B. Tóth. Volume enhancement with externally controlled anisotropic diffusion. *The Visual Computer*, pages 1–12, 2016. doi:10.1007/s00371-015-1203-y

[14] E. Veklerov and J. Llacer. The feasibility of images reconstructed with the methods of sieves. *IEEE Transactions on Nuclear Science*, 37(2):835–841, Apr 1990.

[15] G. Wang and J. Qi. Direct estimation of kinetic parametric images for dynamic pet. *Theranostics*, 3(10):802–815, 2013.

[16] G. Wang and J. Qi. An optimization transfer algorithm for nonlinear parametric image reconstruction from dynamic pet data. *IEEE Transactions on Medical Imaging*, 31(10):1977–1988, Oct 2012.

[17] J. Yan, B. Planeta-Wilson, and R.E. Carson. Direct 4d list mode parametric reconstruction for pet with a novel em algorithm. In *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, pages 3625–3628, Oct 2008.

[18] W. Yin, S. Osher, J. Darbon, and D. Goldfarb. Bregman iterative algorithms for compressed sensing and related problems. *IAM Journal on Imaging Sciences*, 1(1):143–168, 2008.

	f_v	a_1	a_2	α_1	α_2
$F1$	0.5	0.05	2	2	0.5
$F2$	0.6	0.1	5	1.8	0.1
$\tilde{F}_{m=0}$	0.55	0.25	3.5	1.35	0.16
$\tilde{F}_{m=-1}$	0.55	0.72	2.85	0.84	0.099
$\tilde{F}_{m=-2}$	0.55	1.5	2.07	0.47	0.087
$\tilde{F}_{m=-3}$	0.55	1.26	2.32	0.08	0.21

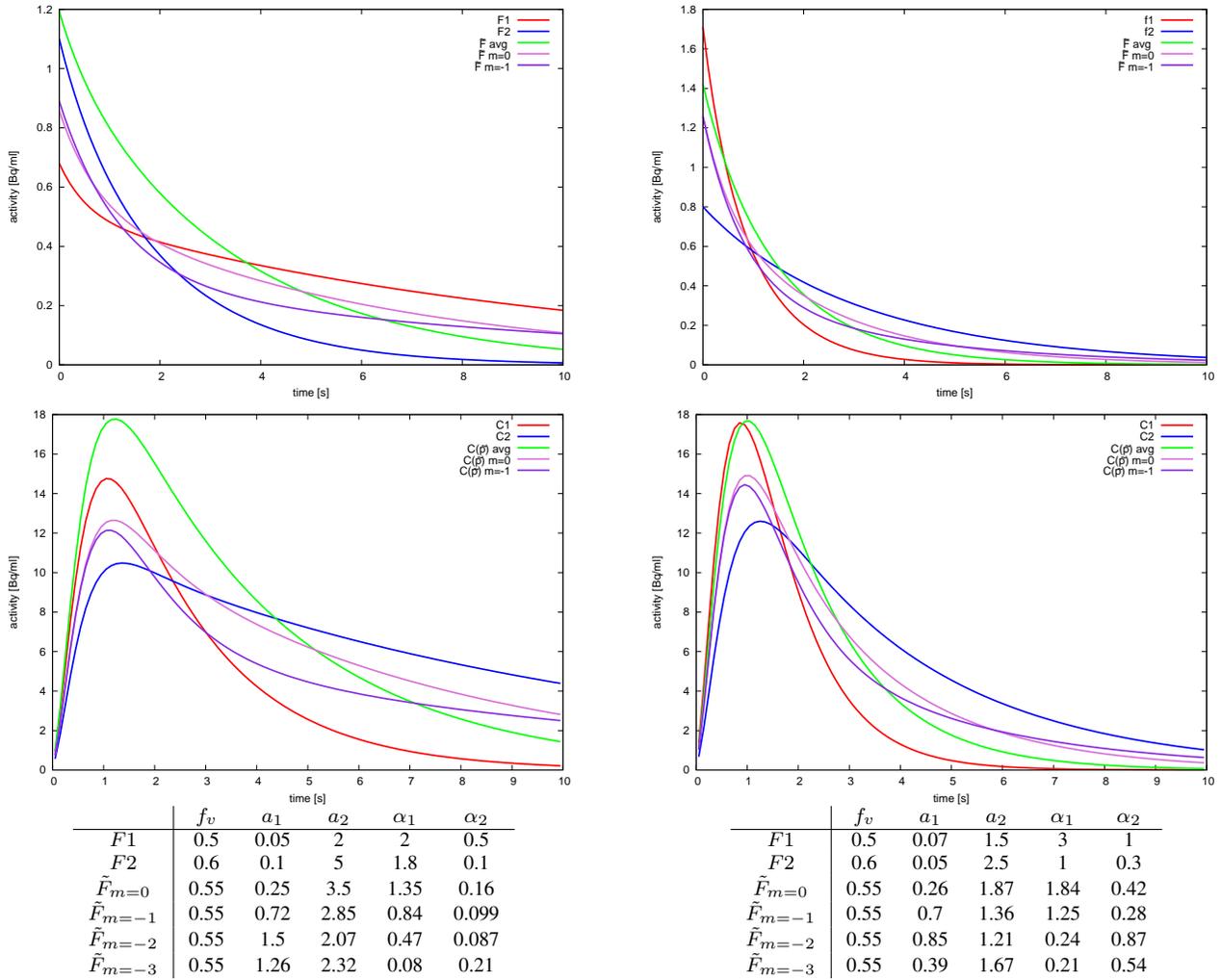|	f_v	a_1	a_2	α_1	α_2
$F1$	0.5	0.07	1.5	3	1
$F2$	0.6	0.05	2.5	1	0.3
$\tilde{F}_{m=0}$	0.55	0.26	1.87	1.84	0.42
$\tilde{F}_{m=-1}$	0.55	0.7	1.36	1.25	0.28
$\tilde{F}_{m=-2}$	0.55	0.85	1.21	0.24	0.87
$\tilde{F}_{m=-3}$	0.55	0.39	1.67	0.21	0.54

Fig. 2. Averaging of two functions with the proposed non-linear scheme and a comparison to the direct averaging of the non-linear parameters. The upper row shows the impulse response functions F, the lower row concentration functions C.