

Deconvolution-Based Partial Volume Correction in Raclopride-PET and Monte Carlo Comparison to MR Based Method

Jussi Tohka^a Anthonin Reilhac^b

^a*Institute of Signal Processing, Tampere University of Technology, Tampere, Finland, email: jussi.tohka@tut.fi*

^b*CERMEP, Bron, France*

Abstract

1

In this work, we evaluated three iterative deconvolution algorithms and compared their performance to partial volume (PV) correction based on structural imaging in brain positron emission tomography (PET) using a database of Monte Carlo simulated images. We limited our interest to quantitative radioligand PET imaging, particularly to ¹¹C Raclopride and striatal imaging. The studied deconvolution methods included Richardson-Lucy, reblurred Van Cittert, and reblurred Van Cittert with the total variation regularization. We studied the bias and variance of the regional estimates of binding potential (BP) values and the accuracy of regional TACs as a function of the applied image processing. The resolution/noise tradeoff in parametric BP images was addressed as well. The regional BP values and TACs obtained by deconvolution were almost as accurate than those by structural imaging based PV correction (GTM method) when the ideal volumes of interests (VOIs) were used to extract TACs from the images. For deconvolution methods, the ideal VOIs were slightly eroded from the exact anatomical VOI to limit the bias due to tissue fraction effect which is not corrected for by deconvolution based methods. For the GTM method, the ideal VOIs were the exact anatomical VOIs. The BP values and TACs by deconvolution were less affected by segmentation and registration errors than those with the GTM based PV correction. The BP estimates and TACs with deconvolution based PV correction were more accurate than BPs and TACs derived without PV correction. The parametric images obtained by the deconvolution based PV correction showed considerably improved resolution with only slightly increased noise level compared to the case with no PV correction. The reblurred Van Cittert method was the best of the studied deconvolution methods. We

¹ The original version published as J. Tohka and A. Reilhac. Deconvolution-Based Partial Volume Correction in Raclopride-PET and Monte Carlo Comparison to MR Based Method. *NeuroImage*, 39(4):1570-1584, 2008. <http://dx.doi.org/10.1016/j.neuroimage.2007.10.038>

conclude that the deconvolution is an interesting alternative to structural imaging-based PV correction as it leads to quantification results of similar accuracy, and it is less prone to registration and segmentation errors than structural imaging based PV correction. Moreover, PV corrected parametric images can be readily computed based on deconvolved dynamic images.

1 Introduction

The partial volume effect (PVE) is an important degrading factor in quantitative PET brain imaging. The term PVE refers to two factors contributing to blur in images: In the terminology of Aston et al. (2002), *the tissue fraction effect* is caused by several distinct types of tissue existing in a single voxel. *Spill-over* (and *spill-in*) - caused by a non-zero positron range and image reconstruction among other factors - mean that the reconstructed image is a convolution of the true image by a point spread function (PSF). The difference between the two sources of PVE for us is that the spill-over and spill-in could be corrected based on the information in the PET image itself and the measured PSF while the correction of the tissue fraction effect requires additional external information. This external information can be in the form of the segmented magnetic resonance (MR) image of the same subject. For the purposes of this paper, there is no need to differentiate between spill-in and spill-over, and therefore the term spill-over is used to mean both spill-over and spill-in.

To date, several methods to correct for PVEs have been proposed. Usually, these methods rely on the availability of a segmented (MR) image of the subject that is co-registered with the PET image. Aston et al. (2002) provided a unifying statistical framework covering many of these methods. The widely applied GTM method (Rousset et al., 1998) utilizes structural information from MRI and assumes regional homogeneity of the radioactivity levels in the PET image. This way a linear system is formed using the structural information and the scanner PSF. The solution to this system then provides the underlying true radioactivity levels based on the radioactivity levels measured from PVE corrupted data using the structural information. In addition, there exist methods, termed pixel-by-pixel methods, that relax the regional homogeneity assumption for a single (gray matter) structure (Müller-Gärtner et al., 1992; Meltzer et al., 1996; Strul and Bendriem, 1999) while assuming that the true radioactivity levels in other brain structures are known. These methods are not routinely applied for Raclopride imaging, and they cannot account for the contamination between two or more different gray matter structures. Therefore, a direct suitability of these pixel-by-pixel methods to striatal imaging with Raclopride can be doubted because there are (at least) two distinct sub-cortical gray matter structures (caudate and putamen) whose activity level is different from the activity level in cortex.

Because the MR-based PV correction methods rely on the structural information, they are sensitive to the errors in MR-PET registration and volume of interest (VOI) delineation. The effects of mis-registration and mis-segmentation in MR-based PV correction have been previously studied by Baete et al. (2004); Frouin et al. (2002); Meltzer et al. (1999); Quarantelli et al. (2004) and Zaidi et al. (2006). A majority of the above mentioned works targeted ^{18}F -FDG-PET imaging of cortical regions. Frouin et al. (2002) and Zaidi et al. (2006) studied also striatal imaging using ^{18}F -Dopa. Therefore, these studies are of the greatest interest to us. Frouin et al. (2002) studied the effect of the registration and segmentation errors to the GTM-based PV correction with a Monte Carlo simulated ^{18}F -Dopa phantom. They simulated various erroneous segmentations and registrations and concluded that both registration and segmentation errors affected the precision of the recovery of the true radioactivity levels. Especially, the variation in the recovered activity levels was found to increase when the extents of the segmentation and registration errors were increased. The segmentation and registration errors were considered separately, and no results about the pooled registration and segmentation errors were provided. Zaidi et al. (2006) studied the effect of several MR tissue classification algorithms to the PV correction with ^{18}F -Dopa and ^{18}F -FDG. The delineations of caudate and putamen were performed relying on a segmented brain template, and this template was the same for all tissue classification algorithms considered. Still, it was found that the choice of the tissue classification algorithm could lead up to 10% differences in the activity level of the caudate nuclei in the PVE corrected PET data. For the putamen, the differences were usually below 1%.

The sensitivity of MR-based PV correction to inevitable segmentation and registration errors leads us to consider purely PET based PVE correction. As mentioned, the spill-over can be corrected for by utilizing only the PET image itself and the measured PSF. The process is termed deconvolution. For general tutorials about the subject, see (Carasso, 1999; Biemond et al., 1990). These methods have not received much attention with brain PET although deconvolution problems share a similarity with the image reconstruction problems (Demoment, 1989). With SPECT (single photon emitting computed tomography), there has been some interest towards deconvolution-based PVE correction (Mignotte et al., 2002; Charalambous et al., 1992). However, the quantification aspects were not considered in the above references. Kao et al. (1997) restored sinograms before the reconstruction. Obviously, spill-over caused by reconstruction cannot be addressed this way. Several sources of the spill-over can be addressed in the iterative maximum likelihood reconstruction algorithms (Leahy and Qi, 2000). Although this would be a favorable avenue to proceed, there are practical difficulties. For example, it is not always possible to choose the reconstruction algorithm or reconstruct the images retrospectively with a better algorithm than the one applied originally. Also, these reconstruction methods call for an advanced modeling of the data acquisition

and, in practice, these models are necessarily approximative (see Leahy and Qi (2000)). Still, a recent technique generates the PVE corrected image based on the multiresolution analysis of the PET and co-registered MRI images of the subject (Boussion et al., 2007). This technique performs the correction using the MRI data but without the need to segment nor classify the data. Whereas it has successfully been validated for static FDG PET and rCBF SPECT data, this technique still suffers from limitations and is not yet applicable for multi frame PET data. Moreover, the suitability of this technique for striatal imaging is unknown.

In this study, we evaluate a few well-known deconvolution algorithms and their modifications for their suitability to PV correction in quantitative parametric PET imaging using ^{11}C Raclopride. Specific attention is given to the comparison of the deconvolution based PV correction and MR based PV correction in presence of segmentation and registration errors. The evaluation is performed by utilizing a publicly available database of Monte Carlo simulated ^{11}C -Raclopride images (Reilhac et al., 2005). An important feature of the database is that the normal anatomical variability is modeled in the database. This allows us to draw conclusions about its impact on the variability of regional physiological parameters of interest. The main criteria of the comparison are the bias and variance - with respect to the anatomical variability in the database - in regional binding potential (BP) values. This is an additional novel feature of the present study, since the previous efforts have characterized results mainly with respect of the capability of recovering the true activities rather than the underlying physiological parameters of interest. In addition to regional BP values, we consider the noise level and resolution of the parametric BP images with the deconvolution based PV corrections as well as the quality of regional time activity curves (TACs).

2 Partial volume correction

A beginning note: The BP values in dynamic ^{11}C -Raclopride are typically computed with the simplified reference region model using cerebellum as the reference region (Lammertsma and Hume, 1996; Gunn et al., 1997). Since the model is nonlinear, the PV correction must be performed separately for each frame of the dynamic image and the PV correction of the parametric image is not reasonable.

2.1 Deconvolution methods

2.1.1 Image model

The task is to restore a 3-D image t which has been deteriorated by a known PSF h . We assume that the PSF is symmetric, non-negative, integrates to one, and that $h(0) > 0$. A good model for observed image i is $i = \mathcal{N}_2(\mathcal{N}_1(t) * h)$, where $*$ denotes 3-D convolution and $\mathcal{N}_1, \mathcal{N}_2$ are noise processes. In practice, we assume a simpler image model

$$i = \mathcal{N}_2(t * h). \quad (1)$$

In other words, we ignore the noise process taking place prior to blurring by the PSF. This is a notable simplification since a part of noise in PET stems from the statistical nature of radioactive decay and the contribution of this noise takes place before resolution deterioration due to factors related to detection and reconstruction. The task is then to estimate the true image t based on the observed image i , the PSF h and the properties of the noise process \mathcal{N}_2 . This problem is, in general, ill-posed and similar with the image reconstruction problem.

2.1.2 Iterative deconvolution methods

The simplest approach to deconvolution is the inverse filtering in which the discrete Fourier transform (DFT) of the true image is estimated by the ratio $DFT(i)/DFT(h)$, where $DFT(\cdot)$ denotes 3-D DFT. This approach leads to excessive noise amplification. Another linear approach is based on the linear Wiener filtering, which we have concluded to be of no apparent interest in our application in an earlier conference paper (Tohka and Reilhac, 2006). Therefore, we concentrate on the iterative deconvolution techniques in this work. Two classes of iterative deconvolution techniques are studied. The first one is based on the Poisson noise model and the maximum likelihood principle. The second one is based on the Gaussian noise model and the least squares principle.

A class of deconvolution approaches is based on the maximum likelihood (ML) principle and the expectation maximization (EM) algorithm. These image restoration algorithms are widely used in processing of astronomical images and also in fluorescence microscopy where the image formation can be seen as a counting process as in PET. In the Richardson-Lucy (RL) algorithm (Richardson, 1972; Lucy, 1974)², we assume that each $i(x)$ (x is the 3-D coordinate

² Both Richardson and Lucy originally derived this algorithm without any explicit

of an image voxel) is drawn independently from the Poisson distribution with the parameter $(t * h)(x)$. That is, we assume that the noise process \mathcal{N}_2 can be described by Poisson distribution. This pre-supposes that $i(x) \geq 0$. We set negative values of $i(x)$ to 0 and denote $i^+(x) = \max(i(x), 0)$. This gives the log-likelihood

$$l(t) \propto \sum_x [i^+(x) \log(t * h)(x) - (t * h)(x)]. \quad (2)$$

Maximizing the log-likelihood with the EM algorithm leads to the update rule

$$\begin{aligned} t_{k+1}(x) &= (s * h)(x) t_k(x); t_0(x) = i^+(x) \\ s(x) &= \frac{i^+(x)}{(t_k * h)(x)}, \end{aligned} \quad (3)$$

where t_k is the estimate of the true image t at the iteration k . The algorithm preserves of positivity (i.e. $t_{k+1}(x) > 0$ if $t_k(x) > 0$) and the flux (i.e. $\sum_x t_{k+1}(x) = \sum_x t_k(x)$) (Carasso, 1999). Also, if $t_k(x) = 0$ then $t_{k+1}(x) = 0$. Hence, $(t_k * h)(x) > 0$ when $t_0(x) > 0$ because $h(x) \geq 0$ and $h(0) > 0$. The algorithm, when iterated until convergence, gives the inverse filtering solution (Carasso, 1999), and thus leads to excessive noise amplification. We have previously studied several variants designed to reduce the noise amplification problem (Tohka and Reilhac, 2006). The variants included early termination after a fixed number of iterations, the damped RL algorithm (White, 1993), and regularization with the maximum entropy prior (Nunez and Llacer, 1990). However, none of these variants seemed particularly interesting and hence we do not carry out more extensive experimentation reported in this work for these variants. We note that there exists a number of other regularization approaches that can be associated with the RL algorithm.

In the iterative least squares formulations, the noise process \mathcal{N}_2 is Gaussian. The image t is estimated by minimizing the least squares (LS) criterion

$$\sum_x \|i(x) - (t * h)(x)\|^2. \quad (4)$$

The LS formulation has the inverse filter as the trivial solution (Carasso, 1999). However, one can formulate an iterative scheme that is terminated before convergence. The steepest descent scheme applied to minimization of (4) leads to the following iteration rule:

$$t_{k+1}(x) = t_k(x) + \alpha(h * (i - h * t_k))(x); t_0(x) = i(x), \quad (5)$$

noise model considerations.

where $\alpha \in (0, 2)$ is the step length. (This selection guarantees the convergence.) In this work, we set $\alpha = 1.5$. We call this iterative scheme reblurred Van Cittert iteration (later on VC) according to Carasso (1999). We consider the reblurred Van Cittert method instead of the standard Van Cittert method (see e.g. Biemond et al. (1990)) because the latter is a simple error correcting procedure that is not based on the optimization of any well-known statistical criterion. Note that in this work we do not enforce the positivity of the solutions as we did in the earlier conference paper (Tohka and Reilhac, 2006).

The total variation based regularization (Rudin et al., 1992) leads to an interesting modification of the VC method. Now, we are minimizing the functional

$$\sum_x ||i(x) - (t * h)(x)||^2 + \lambda_{TV} \sum_x |\nabla t(x)|, \quad (6)$$

where λ_{TV} is a regularization parameter. To guarantee the differentiability of the above functional, $|\nabla t(x)|$ is replaced by $\sqrt{|\nabla t(x)|^2 + \beta^2}$, where β is a small positive constant (Dobson and Vogel, 1997). Under the steepest descent scheme, the update equation is written as

$$t_{k+1}(x) = t_k(x) + \alpha(h * (i - h * t_k))(x) + \alpha \lambda_{TV} \operatorname{div} \frac{\nabla t_k(x)}{\sqrt{|\nabla t_k(x)|^2 + \beta^2}}. \quad (7)$$

We implemented the computation of the divergence operator *div* as Rudin et al. (1992). We selected $\lambda_{TV} = 0.25$, which was the largest regularization parameter value observed experimentally to lead to the convergence under our implementation. The parameter β was set to one percent of the mean of the absolute image intensities: $\beta = \sum_x |i(x)| / \sum_x 1$. We note that there is a relatively large body of work concerning numerical aspects of the total variation regularization and our implementation is rather simplistic. The work has concentrated around to Newton type methods (e.g. Dobson and Vogel (1997)) requiring a solution of a linear system in each iteration of the algorithm. This is problematic in 3-D because of the large size of linear systems needed to be solved and the more complex structure of these systems in 3-D deconvolution as compared to 1-D and 2-D cases.

2.1.3 Implementation notes

We implemented the algorithms in Matlab (Mathworks, Natick, MA). The convolutions in (3), (5), and (7) were implemented in the Fourier space. The Matlab's built-in Fast Fourier Transform algorithm was applied to compute the DFTs.

For all the methods, the equal termination condition was applied:

$$\frac{\sqrt{\sum_x (t_{k+1}(x) - t_k(x))^2}}{\sqrt{\sum_x i(x)^2}} < 0.01. \quad (8)$$

This condition, with most methods, led to convergence rather rapidly, usually after 10 - 30 iterations. Hence, the convergence in most cases was premature. If the convergence did not occur during 100 iterations, the iteration was terminated. We did not study the effects of modifying the termination condition or threshold in this study although the results would certainly be different for different termination conditions.

2.2 MR based PV correction

We use an MR based PV correction method described in (Reilhac et al., 2000; Frouin et al., 2002) as the reference method. This method is rooted in the geometric transfer matrix (GTM) method originally introduced by Rousset et al. (1998). We will briefly describe this method in order to make its vulnerabilities to registration and segmentation errors more visible.

The principal assumption of the GTM method is that the brain can be parcellated into K regions of homogeneous activity D_1, \dots, D_K . Denote the true radioactivity levels within these regions by r_i , $i = 1, \dots, K$ and let $\mathbf{r} = [r_1, \dots, r_K]^T$. Further, denote by \hat{r}_i the activities measured based on the PVE corrupted data and let $\hat{\mathbf{r}} = [\hat{r}_1, \dots, \hat{r}_K]^T$. The homogeneity assumption allows the definition of the GTM $G = (g_{ij})_{K \times K}$, where the element g_{ij} describes the fractional contribution of the activity within the region i to the activity of the region j . If G is non-singular, the unknown true activities can be solved as

$$\mathbf{r} = G^{-1}\hat{\mathbf{r}}. \quad (9)$$

Both the matrix G and the vector $\hat{\mathbf{r}}$ are measured based on the information from the structural image. They are both subject to segmentation and registration errors. Registration errors affect only $\hat{\mathbf{r}}$ (Frouin et al., 2002). Segmentation errors affect both $\hat{\mathbf{r}}$ and G . If the interest lies in recovering the radioactivity concentration in a small structure of high activity level relative to its surroundings, over (resp. under) segmentation decreases (increases) \hat{r}_i and increases (decreases) g_{ii} . These have similar effects to the PV correction yielding a strongly biased estimates of the true activity. Under/over segmentation effects can be reduced by a particular implementation of the GTM method, and the union of the VOIs used to compute $\hat{\mathbf{r}}$ and G do not need to cover the whole field of view. For example, Frouin et al. (2002) defined VOIs

used to compute $\hat{\mathbf{r}}$ and G by

$$VOI_i = \{x : \frac{\int_{y \in D_i} h(x - y)}{\max_{z \in D_i} \int_{y \in D_i} h(z - y)} > 0.5\},$$

where D_1, \dots, D_K are obtained by segmenting the MR image of the subject into K regions. Although these implementational details probably have a large influence to the robustness of the GTM PV corrections, they are not always followed and their effects are not well documented. Thus, in this study, we assume that $VOI_i = D_i$, i.e. that the VOIs obtained from MR segmentation and used for PV correction with GTM are equal.

3 Materials and methods

3.1 PET Monte Carlo database

Our material consisted of the publicly available database of 16 Monte Carlo simulated dynamic ^{11}C -Raclopride PET brain images (Reilhac et al., 2005). (There are 17 images in the database, however, we reserved one of them for tuning the parameters of the methods). Realistic Monte Carlo PET simulation requires a model of physiology (TACs for each brain region) and a model of anatomy (3D labeled volume). All the images had the same underlying physiology but different underlying anatomy. The physiological model was derived from an actual ^{11}C -Raclopride study. It contained 7 TACs (putamen, caudate/nucleus accumbens, white matter, extra-striatal gray matter, cerebellum, blood pool, and skin/fat/muscle). The realistic anatomical models were derived from T1 weighted brain MR images of healthy volunteers. One model was generated by manual labeling (the Jacob phantom, Kabani et al. (1998)), and 15 others by an automatic segmentation algorithm based on label propagation (Collins et al., 1999). The caudate VOI included both head and body of the caudate, but not the tail. The striatum in the anatomical models was composed of caudate, putamen, and nucleus accumbens and we considered nucleus accumbens as a separate structure from caudate and putamen. The anatomical models had dimensions of $181 \times 217 \times 181$ voxels. The voxel size was 1mm^3 .

Multiple anatomical models allowed us to study the variation caused by the anatomical variability to the quantification results. (We call this *anatomy-induced variance*). Note that the variability of quantification results is due to both anatomy induced variance and imaging variance due to the stochastic nature of PET imaging. The latter is always present in Monte Carlo PET simulations. In order to study the relative sizes of the two sources of variability,

we simulated 7 replicates of dynamic images based on a single anatomical model (the Jacob phantom).

The database was simulated by the PET-SORTEO Monte Carlo simulation software (Reilhac et al., 2004) modeling the Exact ECAT HR+ scanner operating in 3-D mode. PET-SORTEO permits faithful reproduction of the image formation process of PET tomographs and can be reliably used for the validation of data analysis methods (Reilhac et al., 2005). The simulation model accounts for most of the phenomena encountered during PET acquisitions including the Poisson nature of the emission, positron range in tissues, annihilation photons non-collinearity, scatter, randoms and system dead-time. The following corrections were applied to raw data: randoms (online correction), normalization, scatter, arc, attenuation, dead-time and decay corrections. The data was reconstructed with the filtered backprojection (FBP) with the Hann(ing) filter and the frequency cutoff of $0.3mm^{-1}$. For the comparison purposes, we reconstructed the images also with the FBP with the ramp filter with $0.5mm^{-1}$ cutoff. These reconstruction filters represent the two extremes in terms of the resolution/noise tradeoff. Namely, the Hann filter with the cutoff of $0.3mm^{-1}$ produces images with a low noise level but with a poor resolution. The ramp filter produces noisy images with a good resolution. The resulting dynamic images consisted of 26 frames of volumes of the size $128 \times 128 \times 63$. The voxel size was $2.11 \times 2.11 \times 2.43 mm^3$.

3.2 Finding the PSF

The PSF was approximated by a 3-D Gaussian kernel. The full-width at half maximum (FWHM) parameters of the kernel were computed based on a reconstructed image of a point source that was located at 5cm from the center of the field of view of the simulated scanner. The FWHM parameters for the kernel for ^{11}C Raclopride reconstructed with the Hann filter were $9.3mm$ in the axial directions and $9.1mm$ in the z -direction. For the ramp filter, FWHM was $5mm$ in each direction. The Gaussian kernels were represented by $21 \times 21 \times 21$ arrays in our implementation.

3.3 VOIs and simulating registration/segmentation errors

The VOIs for each region and subject were defined based on the anatomical models employed for the generation of the database. We studied regional quantification with two distinct classes of VOIs for each subject and method: ideal and standard. For the MR based GTM PV correction method, the ideal VOIs are the ones based on the exact anatomy of the subject. However, the full-size anatomical VOIs are not optimal for activity recovery if the tissue

fraction effect is not compensated for. Therefore, to create ideal VOIs for the other methods than MR based PV correction, we applied morphological erosion (in the space of the anatomical models) to reduce the size of the VOIs systematically as little as possible while removing the tissue fraction effect. This method is systematic and, in our opinion, it agrees well with manual VOI delineation procedures used in PET image analysis.

The standard VOI class included the exact anatomical VOI set and four different sets of VOIs per each subject in which VOIs were deformed by segmentation and registration errors. Note that the exact anatomical VOI is as well the ideal VOI for GTM, but this does not hold for other methods. The errors in the MR-PET registration were generated by automatically co-registering the MR image, which was the basis of the anatomical model, to the simulated PET image. The time averaged PET images computed from the dynamic images were used in the registration process. The registration was implemented by the MNI's *autoreg* program (Collins et al., 1994) using mutual information as the similarity criterion (Maes et al., 1997). We started the registration from two different initializations, therefore simulating two different registration errors per subject. In the first initialization, named *register*, the MR and PET images were initially in register. In the second initialization, named *native*, the MR image was in the native space as opposed to the PET image in the stereotactic MNI-305 space.

Fifteen anatomical models in the database were created by means of the automatic segmentation algorithm. Therefore, it was not pertinent to model the segmentation errors based on automatic segmentation algorithms. Instead, we resorted to similar means in simulating segmentation errors than Frouin et al. (2002). Two different erroneous segmentations per subject were generated based on the mathematical morphology operations. We took the full resolution anatomical model and eroded the VOIs of the principal structures of interest (putamen, caudate, and cerebellum) replacing the necessary labels based on the nearest neighbor interpolation. The same procedure was repeated for dilation. The structuring element for the morphological dilation/erosion was the $3 \times 3 \times 3$ cube centered at the origin. The average volumes of striatal VOIs are given in Table 1. Our segmentation errors were larger than the extreme cases of (Frouin et al., 2002). This was mainly due to a different structuring element used in the morphological operations, but differences in anatomical models can also contribute to this difference. For each subject, combining 2 erroneous segmentations with 2 erroneous registrations yielded 4 cases of the VOIs in the PET space containing both registration and segmentation errors.

	Putamen	Caudate
Original	$10.97 \pm 0.45cm^3$	$10.16 \pm 1.07cm^3$
Eroded	$5.33 \pm 0.31cm^3$	$3.93 \pm 0.77cm^3$
Dilated	$18.14 \pm 0.63cm^3$	$18.48 \pm 1.42cm^3$

Table 1

Average volumes \pm standard deviation of volumes of putamen and caudate VOIs in original, eroded and dilated VOI sets.

3.4 Computing regional TACs and BP values

Each 3-D frame of each of the 16 dynamic PET images was corrected for PVE by deconvolution techniques and by the MR based GTM method. The PV corrections were primarily performed for the images reconstructed with the Hann filter. In addition, we performed VC-deconvolution based PV corrections for the images reconstructed with the ramp filter to study how the increase in noise level and resolution affected deconvolution; this method is abbreviated as VC-ramp. Abbreviations VC, VC-TV, and RL refer to the deconvolutions of the Hann filtered data. When highlighting the differences between the two reconstructions is necessary, VC deconvolution based on Hann filter reconstructed data is abbreviated as VC-Hann.

The MR based correction was performed separately for 5 anatomical models per subject (4 with segmentation/registration errors and 1 ideal). It yielded directly regional TACs. The regional TACs for images PVE corrected by the deconvolution algorithms as well as for the images not corrected for PVE were computed based on the labeled images by averaging the radioactivity concentration within a VOI. This yielded 6 sets of TACs per subject and algorithm (1 ideal, 1 based on the exact anatomy, and 4 with segmentation/registration errors). Each TAC set included three TACs (putamen, caudate, and cerebellum).

The regional binding potential (BP) values of the D₂ dopamine receptors were computed from the regional TACs with the simplified reference region model using the cerebellum as the reference region (Gunn et al., 1997). The cerebellum TAC was always from the same TAC set than TACs for caudate and putamen. We computed the true BP values using the physiological ground-truth TACs. These were the same for all 16 subjects and VOI models per subject: 2.7263 for putamen and 1.6397 for caudate. The BP values for the putamen and caudate are reported in this study, since the most important applications of the Raclopride imaging are within striatal imaging of D₂ dopamine receptors. Hence, the TACs for putamen, caudate and cerebellum (the reference region) were of principal interest here. The nucleus accumbens shares the TAC with the caudate in the database. Therefore, we did not study the

BP values in nucleus accumbens, or ventral striatum, although typical BP values of this structure are possibly distinct from caudate’s typical BP values (Mawlawi et al., 2001).

The parametric BP images were computed with the simplified reference region model using the cerebellum as the reference region (Gunn et al., 1997). The cerebellum TAC was extracted from each dynamic image based on the ground-truth anatomy. This TAC is separate for each PV correction method and different from the physiological ground-truth. The BP image computation is not possible for the GTM method.

3.5 *Experiments and the evaluation criteria*

We evaluated the performance of the PV correction in three different levels using the following criteria:

- (1) We computed the bias and the variance of the regional BP values computed for 16 subjects for each PV correction method³. A small bias is important if absolute quantification is required as, for example, in receptor occupancy studies (Passchier et al., 2002). A small variance is more important than a small bias in group comparisons if the bias is expected to be equal between the two groups. The square root of the variance (the standard deviation) is reported to enable easy comparison to bias values. The variance in regional BP values can be decomposed into imaging and anatomy induced variance. It is important to address the relative amount of these two sources variance if, for example, the interest is in correlating anatomical and functional findings (see Morris et al. (1999)). We estimated imaging variance using 7 realizations based on the Jacob anatomical model. Then, the proportion of the anatomy induced variance was estimated by subtracting the estimated imaging variance from the BP value variance across 16 subjects. Due to small sample sizes (16 and 7) for this experiment, the results should be viewed with some caution.
- (2) We compared the extracted TACs to the ground truth TACs using the root mean square error (RMSE) between the ground truth TACs and extracted TACs as the index of performance. RMSEs were computed for each structure of interest separately and averaged over the 16 images, i.e.

$$RMSE_{TAC} = \tag{10}$$

³ These are reported in absolute units rather than percentages of the true BP values to enable a direct calculation of the related measures of performance such as the mean square error. Percent-wise bias can be calculated by dividing absolute bias by the true BP (2.7263 for putamen, 1.6397 for caudate).

$$\frac{1}{16} \sum_{subj=1}^{16} \sqrt{\sum_{fr=1}^{26} \frac{\Delta_{fr}}{\sum_{fr} \Delta_{fr}} (TAC_{ext}^{subj}(fr) - TAC_{true}(fr))^2},$$

where TAC_{ext}^{subj} is the TAC extracted from $subj$ th dynamic image, TAC_{true} is the ground truth TAC and Δ_{fr} is the duration of the frame fr .

- (3) We studied the resolution-noise tradeoff and accuracy in parametric imaging. To address resolution-noise tradeoff, we computed the coefficient of variation (CoV) and the contrast recovery coefficient (CRC) for caudate and putamen. These were computed only for the VOIs based on the exact anatomy. CoV was computed by dividing the standard deviation of the BP values within each VOI by the mean of the BP values within the VOI for each subject and then averaging these ratios over the 16 subjects. The formula for CRC was adapted from Kessler et al. (1984):

$$CRC = \sum_{subj=1}^{16} \frac{\hat{BP}_{subj} / \hat{BP}_{subj}^B - 1}{BP / BP^B - 1}. \quad (11)$$

Above, \hat{BP}_{subj} and \hat{BP}_{subj}^B , respectively, are the average BP value in a VOI and the average BP value in the white matter region surrounding the VOI in the parametric image of the subject $subj$. \hat{BP}_{subj}^B was computed by averaging BP of $|VOI|$ white matter voxels closest to the VOI, where $|VOI|$ is the number of voxels within the VOI. The symbols BP and BP^B refer to the correct BP value for VOI and white matter. White matter was chosen to represent the background BP because it is the only structure, which has enough voxels in the immediate neighborhood of both caudate and putamen in the anatomical models. For example, CSF is not suitable for this purpose because its distance from putamen is too large.

The accuracy of parametric images was evaluated by computing the RMSE of BP values over a VOI and averaging this over subjects:

$$RMSE_{BP} = \frac{1}{16BP} \sum_{subj=1}^{16} \sqrt{\sum_{x \in VOI_{subj}} \frac{(BP_{subj}(x) - BP)^2}{|VOI|}} \quad (12)$$

where $BP_{subj}(x)$ refers to the BP value of the voxel x in the parametric image of the subject $subj$, and BP is the true BP value for the structure in question. We report these errors as percentages of the true BP values because absolute RMSE over a VOI has little meaning. We note that that $RMSE_{BP}$ can be compared to the mean absolute error of the regional BP values (MAE_{VOI}): Let BP_{subj} be the regional BP value based on TACs for the subject $subj$. Writing $BP_{subj}(x) = BP_{subj}$ for all x , and $|VOI| = 1$ in (12), we obtain MAE normalized by BP . The accuracy of parametric images can be proportioned to the accuracy of the regional BP values by comparing $RMSE_{BP}$ and MAE_{VOI} measures.

4 Results

4.1 Bias and variance in regional BP values

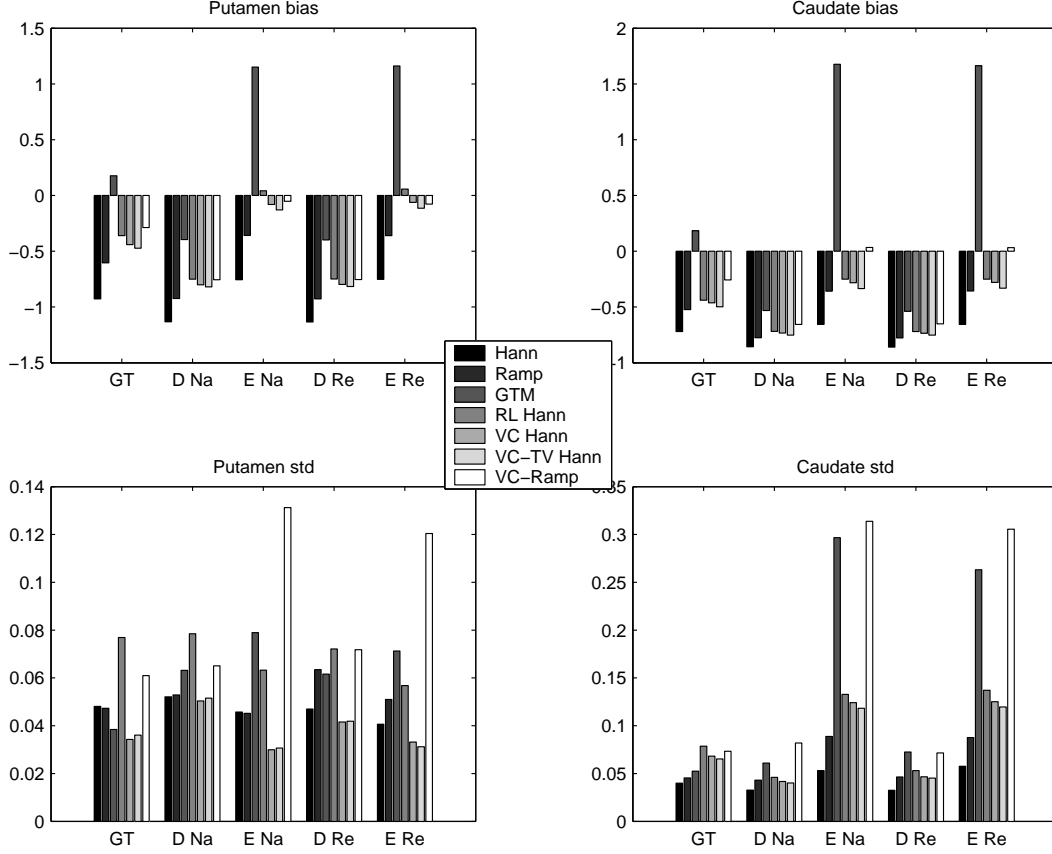


Fig. 1. The bias and standard deviation of BPs in putamen and caudate with different segmentation and PV correction methods. GT refers to ground truth segmentation. D and E refer to dilated and eroded ground truth segmentations. Na and Re refer to registration errors from the native and register initializations. For example, D Re refers to the dilated ground truth segmentation and registration error from the register initialization. Hann and Ramp refer to unprocessed images based on the Hann and Ramp filter based reconstruction. RL (VC, VC-TV) Hann refers to the Hann-filter reconstructed image corrected by RL (VC, VC-TV) deconvolution. VC Ramp refers to ramp filter-reconstructed image corrected by VC deconvolution. GTM refers to the MR-based GTM method.

The biases and standard deviations of the regional BP values in putamen and caudate with the standard VOI class are shown in Fig. 1. With the exact anatomical segmentation (ground-truth, GT), the MR-based GTM method yielded the most precise results (the smallest absolute value of the bias) as expected. The deconvolution based PV correction reduced the bias compared to the cases with no PV correction. The reconstruction with the ramp filter

reduced the bias in BP values compared to the reconstruction with the Hann filter. With the exact anatomical segmentation, the standard deviations of BP values in putamen were the smallest with the LS-based deconvolution methods coupled with Hann reconstruction. Also, with the GTM method, the standard deviation was smaller in putamen compared to the cases with no PV correction. The RL deconvolution produced the largest standard deviation of the BP value in putamen closely followed by VC-ramp. With caudate, the standard deviation with the unprocessed Hann reconstructed data was the lowest and the standard deviations with the deconvolution methods were the highest.

The standard deviations of the RL deconvolution were consistently higher than the LS based methods (VC and VC-TV). Also, while the bias was smaller for the RL deconvolution than for the LS-based deconvolutions, this difference was not large. Therefore, we consider the results of the LS based deconvolution were better than the results of the RL deconvolution in this experiment. The differences between VC-Hann and VC-TV deconvolutions were not large. Interestingly, VC-ramp reduced considerably the bias compared to VC-Hann. VC-ramp increased also the standard deviation of the regional BP values compared to VC-Hann. However, the standard deviation by VC-ramp also always lower than the standard deviation by RL with BPs computed based on the exact anatomical segmentation.

The segmentation errors affected most the results by the GTM-method. For putamen, the bias varied from -0.400 (dilated-register, -15% of the true BP value) to 0.984 (eroded-register, 36% of true BP value). For caudate, the bias varied from -0.721 (dilated-register, -44% of the true BP value) and 1.492 (eroded-native, 90% of the true BP value). In addition, the standard deviations with the GTM-method were the largest (together with VC-ramp) when anatomical information was erroneous as is visible in Fig. 1. Thus, this method appears to be very sensitive to segmentation errors in the BP quantification with Raclopride. The errors in VOIs affected the BP values obtained by deconvolution based PV correction (starting from the Hann-filter-reconstruction) and no PV correction in a similar manner. This is not surprising since the deconvolution based regional BP value computation applies the anatomical information in the same manner than the traditional BP computation. The standard deviation of the BPs with VC-ramp were considerably larger than with VC-Hann when the VOIs did not match the exact anatomy.

The differences in the registration error did not cause any notable differences to the average BP values (compare results with the native and register registrations and the same segmentation in Fig. 1). We additionally performed subject-by-subject analysis, that is, compared the BP value from a native registration to the corresponding register registration. The maximum of absolute difference divided by the average of the two BP values varied between 3.50%

and 3.93% for putamen and between 6.17% and 6.83% for caudate. There were no large differences between PV correction methods. This corroborates the earlier findings in Frouin et al. (2002) for dynamic imaging and shows that effects of registration errors are small in relation to segmentation errors in parametric imaging with Raclopride. As opposed to Frouin et al. (2002), here both rotation and translation errors were present.

The results with the ideal (see section 3.3) VOIs for each method are given in Table 2. Compared to GTM, the absolute values of bias with RL and VC deconvolutions were closer to zero in putamen and nearly as close to zero in caudate. The biases were considerably smaller with deconvolution based PV corrections than without the PV correction. The standard deviations of the BP values were the lowest for the LS based methods (VC and VC-TV) in putamen. In caudate, GTM and raw Hann reconstruction featured the lowest standard deviations. In this experiment, the biases of VC were notably smaller than the biases of VC-TV and there were no great differences in standard deviations of these two methods. VC-ramp could almost remove the bias with ideal VOIs. However, this happened at the expense of considerably increasing the standard deviations. Based on this experiment, VC and RL based deconvolution methods appeared preferable.

	Bias		Std	
	Put	Cau	Put	Cau
Hann	-0.743	-0.597	0.047	0.052
ramp	-0.368	-0.315	0.060	0.089
GTM	0.177	0.184	0.039	0.053
RL	0.046	-0.184	0.075	0.124
VC	-0.084	-0.238	0.034	0.110
VC-TV	-0.133	-0.290	0.033	0.105
VC-ramp	-0.112	0.048	0.155	0.278

Table 2

Comparison of BP statistics with ideal VOIs for each method. The ideal VOIs were the exact anatomical VOIs for GTM and the eroded VOIs for other methods.

The proportion of the anatomy induced variance for both reconstructions and the three deconvolution methods are given in Table 3. An interesting finding is that VC and VC-TV deconvolutions could almost eliminate the anatomy induced variance in putamen BP. This did not happen with caudate. The RL deconvolution based PV correction yielded similar proportions of imaging and anatomy induced variance as were obtained from uncorrected Hann-reconstructed image. The Hann reconstruction decreased the proportion of the imaging variance compared to the reconstruction with the ramp filter as was expected.

	put	cau
Hann	88.5	83.5
ramp	38.5	46.1
RL	80.7	82.3
VC	1.6	76.5
VC-TV	15.4	74.0

Table 3

The proportion of variance in BP values which is anatomy induced (in percents). This was not measured for the GTM and VC-ramp methods.

4.2 Errors in TACs

The RMSEs between ground truth TACs and extracted TACs are listed in Table 4 for the standard VOI class. The results are organized with respect to VOI sets (ground truth anatomy, eroded, and dilated) that were applied for TAC extraction. The eroded (resp. dilated) $RMSE_{TAC}$ values were computed by averaging eroded native and eroded register (resp. dilated native and dilated register) VOIs because the differences caused by different registrations were negligible. With the ground truth anatomical model based VOIs, the $RMSE_{TAC}$ values with the GTM based PV correction were the lowest except for the cerebellum, where VC-deconvolution yielded a lower $RMSE_{TAC}$. $RMSE_{TAC}$ values for VC-ramp were lower than for VC-Hann for the VOIs defined by the exact anatomy.

$RMSE_{TAC}$ values with the ideal VOIs are compared in Table 5. $RMSE_{TAC}$ values for GTM, VC-Hann, VC-TV, and RL were almost equal in putamen and in cerebellum. In caudate, $RMSE_{TAC}$ values for deconvolution methods were greatly improved by using ideal, slightly eroded VOIs instead of the exact anatomical VOIs. However, these still remained higher than the $RMSE_{TAC}$ value of GTM with the exact anatomical VOI that is ideal for GTM. These differences are visualized in Fig. 2, where average TACs across 16 subjects are compared to the ground truth TAC. The improvement by deconvolution based PV corrections compared to the case of TACs extracted from unprocessed images was clear with any set of VOIs. The TACs based on ideal VOIs were not as good for VC-ramp as they were for deconvolutions based on the Hann-reconstructed volumes. There were no significant differences between the $RMSE_{TAC}$ values for three (Hann-based) deconvolution methods. Surprisingly, when we studied frame-wise standard deviations of TACs of 16 subjects, we found almost no difference between RL, VC, and VC-TV methods in putamen and in caudate. However, in cerebellum, TACs based on RL deconvolved images were more variable across subjects than those with the LS based methods (see Fig. 3). This might explain the higher variability in regional BP

values observed with RL as compared to VC and VC-TV.

We computed activity recovery coefficients (ARC, Rousset et al. (1998)) averaged over frames and subjects for the GTM method in order to explain high bias in BP values with VOIs contaminated with segmentation and registration errors. Average ARC can be understood as a measure of bias in TACs. For putamen, these varied from 87% (dilated-register) to 131% (eroded-native). For caudate, these varied from 77% (dilated-native) to 163% (eroded-native). Thus, it appears that the bias increased in the BP computation with caudate. The variation of ARCs both over time and over subject was substantial. The bias for the eroded VOIs was larger than for the dilated ones as was the case with BP values. It is probable that this was, at least partially, caused by larger relative segmentation errors in the case of the eroded VOIs, and it appears that the bias increased superlinearly with the segmentation error. This was expected because, as explained in Section 2.2, systematic segmentation errors affect both the matrix G and the estimated radioactivity level \hat{r}_i , and both of these biases are likely to have the same sign.

	Ground truth			Eroded			Dilated		
	Put	Cau	Cere	Put	Cau	Cere	Put	Cau	Cere
Hann	86.77	66.14	5.30	68.14	58.27	3.09	102.12	76.42	5.23
ramp	60.88	50.29	5.16	36.46	34.96	3.64	85.98	70.75	5.21
GTM	11.39	14.02	2.31	121.43	162.45	5.95	43.90	51.68	4.50
RL	42.60	45.30	4.43	10.13	29.56	2.28	72.82	66.68	4.83
VC	41.07	42.00	1.89	10.22	26.51	1.51	70.60	63.99	2.52
VC-TV	43.62	44.55	1.95	12.69	29.98	1.44	72.01	65.22	2.54
VC-ramp	33.03	27.57	3.80	24.82	32.71	3.79	70.48	61.12	4.40

Table 4

RMSEs in nCi/cc different PV corrections and standard sets of VOIs. To save space, we averaged results of the two sets of dilated and eroded VOIs.

4.3 Resolution-noise tradeoff in parametric imaging

The results concerning the resolution-noise tradeoff in BP images are listed in Table 6. Examples of parametric BP images are shown in Fig. 4. The deconvolution based PV correction (started from Hann-filter-reconstructed images) yielded the better resolution and less noise than the ramp-filter-reconstruction in all the cases with a single exception: The CRC with VC-TV deconvolution in caudate was slightly lower than with the ramp-filter. Hence, the deconvolution outperformed the reconstruction with the ramp filter in this experiment.

	Put	Cau	Cere
Hann	68.97	54.49	3.79
ramp	38.51	32.11	4.11
GTM	11.39	14.02	2.31
RL	10.57	24.44	2.66
VC	10.78	22.52	1.53
VC-TV	13.40	25.93	1.46
VC-ramp	27.89	34.56	3.90

Table 5

RMSEs in nCi/cc for PV corrected TACs when the VOIs for each method were ideal. The ideal VOIs for GTM are anatomical ground truth VOIs. The ideal VOIs for other methods were reduced anatomical ground truth VOIs described in Section 3.3.

As compared to the unprocessed Hann-filter-reconstructed data, the CoVs for the deconvolution methods were slightly higher. However, this was compensated by much higher CRC values. For example, in caudate the increase of CRC was over three-fold by the deconvolution while the increase of the noise level was under 15% for the LS based deconvolution methods.

The deconvolved parametric images were more accurate than those that were not PV corrected as it is visible from $RMSE_{BP}$ values reported in Table 6. Table 6 also provides comparison between $RMSE_{BP}$ and MAE_{VOI} values. MAE_{VOIs} for the GTM method are displayed in the caption of Table 6. It is obvious from the definition of these error measures that MAE_{VOI} is smaller than $RMSE_{BP}$ for a fixed method. MAE_{VOI} was closely related to bias in regional BP values (see Section 4.1), because the the subject-wise biases had the same sign in most cases. For both caudate and putamen, $RMSE_{BP}$ values for RL, VC, and VC-TV methods were smaller than MAE_{VOI} values for the uncorrected Hann-filter-reconstructed data. This serves as an indication of the accuracy of parametric images PV-corrected by deconvolution methods.

For parametric imaging, it was important to start deconvolution from Hann-filter-reconstructed data: VC-ramp method led to excessive noise amplification and improved resolution only slightly. CoV and $RMSE_{BP}$ were over three times larger for VC-ramp than for VC- Hann. In addition, the difference between MAE_{VOI} and $RMSE_{BP}$ was considerably greater for VC-ramp than for any other method.

Excluding VC-ramp, the RL deconvolution led to the best CRCs and the worst CoVs of the deconvolution methods studied. There were not too large differences between the standard VC method and the VC method regularized by total variation. Thus, for parametric imaging the standard VC algorithm

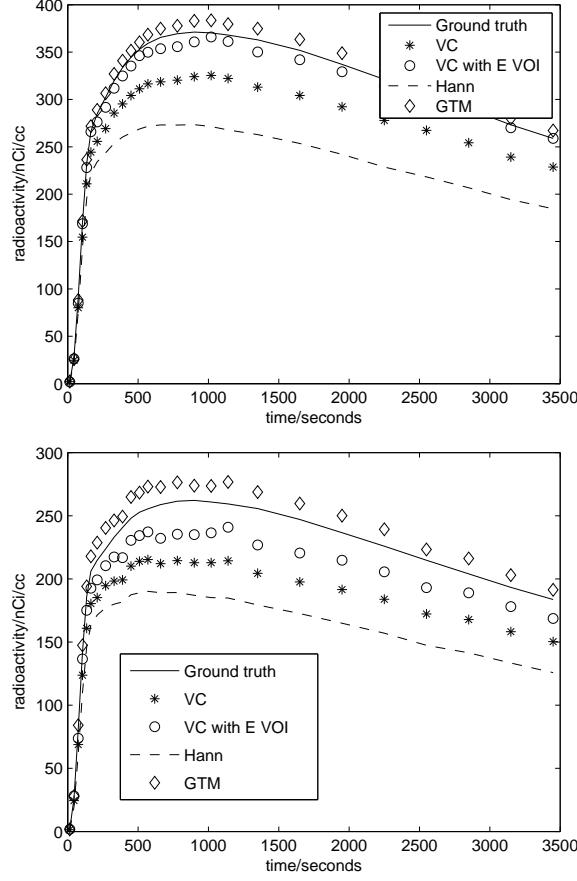


Fig. 2. The averages (over 16 subjects) of extracted TACs compared to the ground truth TAC. From top: putamen TACs, and caudate TACs. The other TACs were extracted based on ground-truth anatomy except the ones labeled VC with E VOI, which were extracted using the ideal VOI for the VC method. The average TACs with other deconvolution methods are highly similar to the VC TACs.

appeared to be the best choice because of its simplicity and good performance. However, it might be that the VC-TV method could benefit from a more powerful optimization method than the steepest descent algorithm applied in this work.

Parametric images displayed artifacts on coronal and sagittal views (see Fig. 4). These artifacts, which were most pronounced on the ramp reconstructed volume and the RL PV corrected volume, existed on the first and the last axial slices in the images. The artifacts were most pronounced on first frames of dynamic images and parametric images, and they were likely to be caused by inaccuracies in scatter correction and normalization.

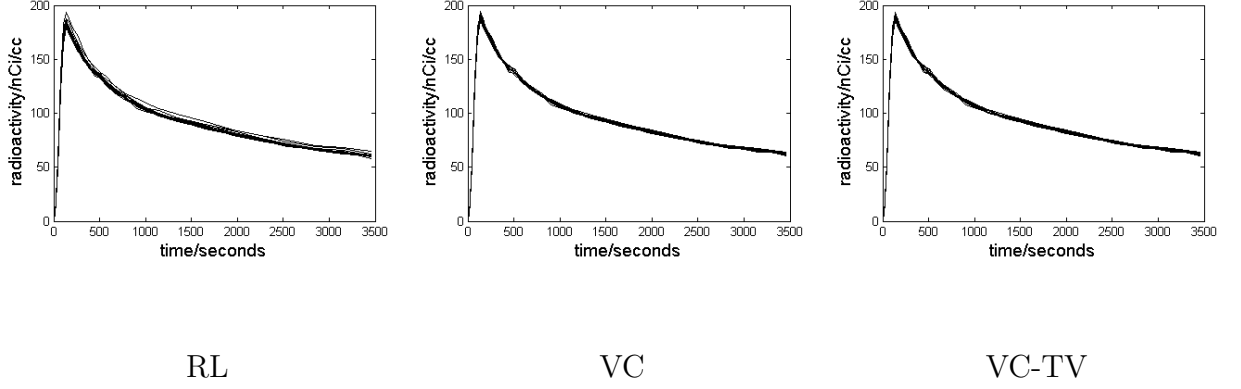


Fig. 3. The variation within cerebellum TACs. Cerebellum TACs corrected for PVE

for all 16 distinct neuroanatomies are shown.

	putamen				caudate			
	CRC	CoV	$RMSE_{BP}$	MAE_{VOI}	CRC	CoV	$RMSE_{BP}$	MAE_{VOI}
Hann	13.69	20.63	38.21	34.00	7.22	34.48	48.68	43.88
ramp	19.66	33.62	32.88	22.14	22.17	51.73	46.03	31.81
RL	25.40	29.45	29.24	13.22	26.21	44.92	42.93	26.75
VC	23.42	27.09	28.57	16.18	24.07	38.76	40.23	28.23
VC-TV	22.60	26.58	28.89	17.36	21.15	38.53	41.31	30.36
VC-ramp	27.78	90.47	94.29	10.54	35.57	136.85	148.06	15.65

Table 6

Contrast recovery coefficients (CRC)s, the coefficients of variation (CoV) and $RMSE_{BP}$ in percents. Also MAEs for regional BP values (MAE_{VOI} , in percents) are provided to enable a comparison of parametric imaging and VOI-based BP analysis. MAE_{VOI} s were computed only for the exact anatomical VOIs. MAE_{VOI} s for GTM were 6.50% for putamen and 11.23% for caudate.

4.4 Demonstration with clinical data

In addition to Monte Carlo studies, which allow for the quantitative evaluation of deconvolution, we provide a qualitative demonstration on the effects of deconvolution to a parametric BP image with data acquired using an Exact-ECAT HR+ scanner. A dynamic ^{11}C -Raclopride image of a healthy volunteer, who had given a written informed consent, was obtained from the image database of CERMEP imaging centre. The dynamic imaging protocol consisted of 42 frames with a duration between 30s and 5 minutes. The total scan duration was 100 minutes. A transmission scan to correct for the attenuation as well as a T1 weighted MR image were also available. The corrections and reconstructions were performed in the same manner as described for the

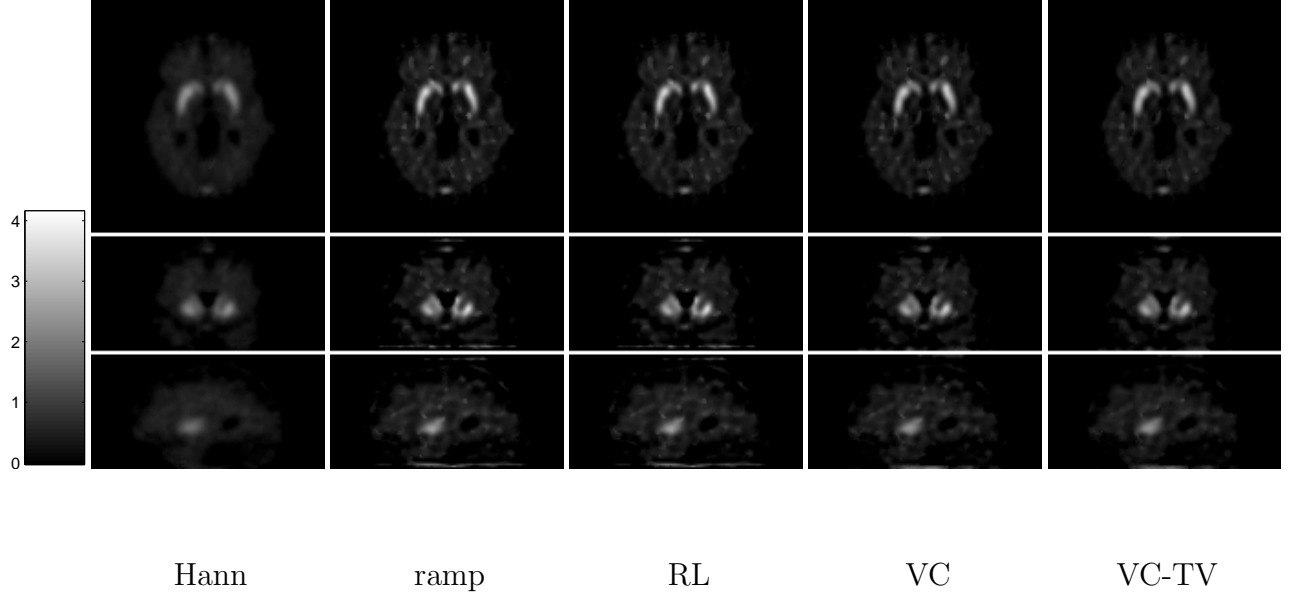


Fig. 4. Parametric BP images based on unprocessed data and data that was corrected for PVE by deconvolution algorithms. BP values below zero are not shown.

Gray-scale is shown on left.

Monte Carlo simulated data. The resulting dynamic images consisted of 42 frames of volumes of the size $128 \times 128 \times 63$. The voxel size was $2.05 \times 2.05 \times 2.43 \text{ mm}^3$. The images FBP reconstructed with the Hann filter were deconvolved frame by frame with RL, VC, and VC-TV methods. We did not try to deconvolve the ramp-filter reconstructed image due to the unsuitability of this combination noticed in the previous subsection. The BP images were computed using the cerebellum as the reference region (Gunn et al., 1997). The cerebellar VOI was obtained by automatically segmenting the MR image of the subject and registering the MR image to the time averaged PET image. The segmentation was carried out with the ANIMAL+INSECT algorithm (Collins et al., 1999) as described by Reilhac et al. (2005). The registration was performed with MNI's **autoreg** program (Collins et al., 1994) using mutual information as the cost function. Thus, the parametric image generation was completely automated.

Parametric images resulting from Hann and ramp reconstructed dynamic volumes as well as deconvolved dynamic volumes are shown in Fig. 5. The parametric images displayed the same characteristics as the parametric images based on the Monte Carlo simulated data (see Fig. 4) as was expected. The RL deconvolution produced a slightly noisier parametric image than the VC and VC-TV methods. The parametric image based on the ramp-filter-reconstructed dynamic volume had a much noisier appearance than the deconvolution-based images. Also, the resolution increase compared to the

Hann reconstructed volume was apparent (c.f. coronal cross sections in Fig. 5). The only clear difference to the Monte Carlo study was that the range of the BP values in the parametric images with the clinical study extended slightly higher. However, this was observed also with the images without deconvolution based PV correction.

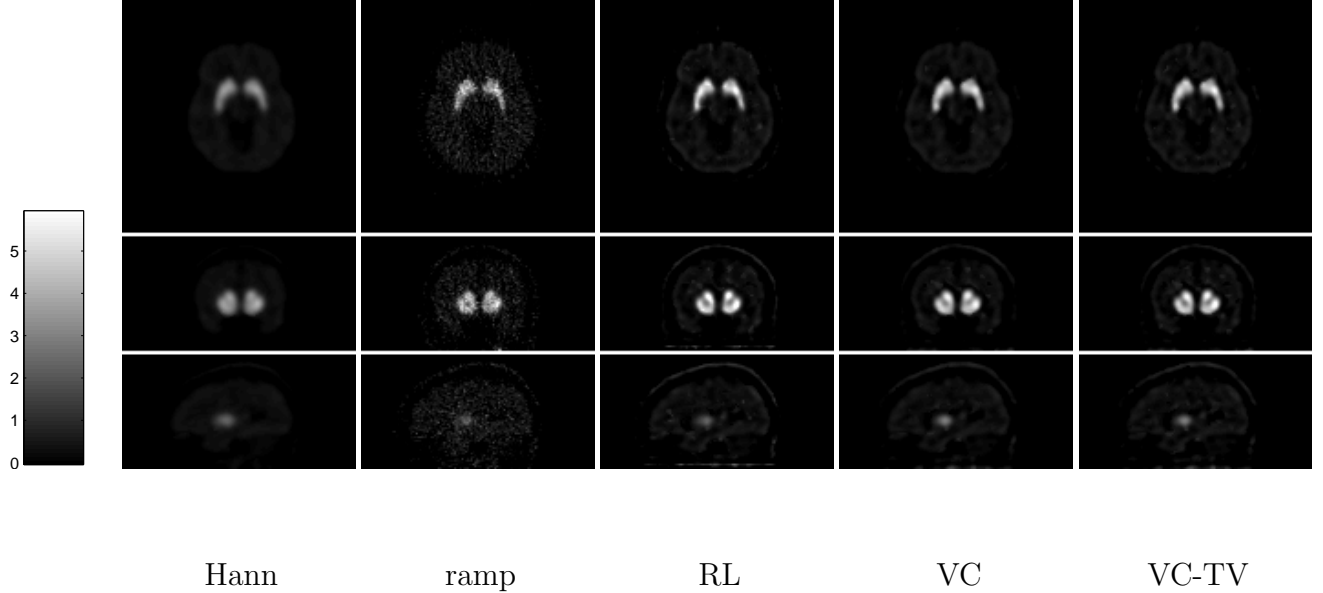


Fig. 5. Parametric BP images based on the clinical data set. BP values below zero are not shown. Gray-scale is shown on left.

5 Discussion

We have studied iterative deconvolution techniques for the PV correction in the quantitative PET brain imaging and compared these to the widely applied MR-based GTM technique. We found that the deconvolution improved the accuracy of the BP quantification with ^{11}C -Raclopride as well as provided more accurate regional TACs compared to the case when PV correction was omitted. The GTM method outperformed the deconvolution methods when the conditions for it were ideal, that is, the anatomy of the subject was perfectly known. However, when the VOIs were adjusted to correct for tissue fraction effect, deconvolution techniques provided regional BP estimates and TACs that were nearly as accurate as with the GTM method.

The anatomical information required by the GTM method is subject to errors arising from MR segmentation and MR-PET registration. The deconvolution based PV correction was more stable with respect to these inevitable

errors than the GTM method. Moreover, with deconvolution, segmentation and registration errors did not cause any excess variability to the BP values as compared to the case of no PV correction. The vulnerability of the GTM method to segmentation and registration errors in striatal imaging has been demonstrated before by Frouin et al. (2002); Zaidi et al. (2006) who studied how well GTM can recover the true activity levels when VOI definitions were not ideal. In this study, we addressed the influences of the applied image processing on the physiological parameters of interest in addition to influences to recovery of regional activity concentrations. We found GTM more sensitive to segmentation and registration errors than Frouin et al. (2002). There are several differences between the two studies that can explain this. 1) The implementations of the GTM method were different between the two studies as was explained in Sect. 2.2. 2) The segmentation errors in our case were larger than the segmentation errors in (Frouin et al., 2002). This was caused by the differences in anatomical phantoms as well as by different structuring elements applied for the mathematical morphology operations for segmentation error generation. 3) Frouin et al. (2002) studied segmentation/registration errors with ^{18}F -L-Dopa whereas we concentrated on ^{11}C -Raclopride imaging. 4) Finally, our primary interest lied in the accuracy of the BP estimation while previous studies addressed the recovery of the correct radioactivity level for each structure. As we noted in Section 4.2, the bias caused by mis-segmentation and mis-registration was more pronounced in regional BP values, especially with caudate, than in TACs.

The experiments were performed with realistic Monte Carlo simulated image data Reilhac et al. (2004, 2005). The multiple anatomical models present in the database allowed us to study the anatomy induced variability to the BP quantification as the function of the applied image processing. It was noted that the anatomy and imaging induced regional BP variation was reduced in putamen but increased in caudate by the deconvolution based PV correction. The same happened with the GTM based PV correction when the perfect knowledge of the anatomy was available. This was probably because the shape of caudate is much more complex than the shape of putamen and thus caudate is more prone to PVE. We performed additional Monte Carlo simulations to separate the anatomy induced variation from the variation due to the stochastic nature of the emission tomography. It was noted that VC and VC-TV deconvolutions almost eliminated the anatomy-induced variation from the putamen BP, but they did not considerably reduce the proportion of the anatomy induced variation in the caudate BP.

In addition to studying regional BP values and TACs, we computed the parametric images based on the image data PV corrected by the deconvolution. Notable increases in the resolution were obtained with modest increases in noise level when studied against different extremes of the FBP reconstruction. It proved to be important to deconvolve the images reconstructed using

the Hann filter: the noise amplification due to deconvolution was excessive with ramp-filter reconstructed images. Parametric imaging appears as one of the most interesting applications of deconvolution based PV correction. First, deconvolution algorithms could provide a useful preprocessing step for automated PET image segmentation algorithms such as the one in (Tohka et al., 2006). In fact, this application was one of the motivating factors for this study. Secondly, the resolution increase could be expected for time averaged PET images as well, although this was not directly addressed in this work. Since time-averaged PET images are typically applied for MR-PET co-registration, deconvolution algorithms could improve the accuracy of MR-PET co-registration. Thirdly, clustering of TACs has recently received interest in dynamic radioligand PET imaging (see e.g. Glatting et al. (2004); Liptrot et al. (2004)). A majority of clustering algorithms applied to radioligand PET do not model PVE or do it implicitly, rather than explicitly, e.g. by considering separate partial volume clusters (Glatting et al., 2004). Therefore, the TAC clustering in radioligand PET could benefit from the deconvolution based PV correction.

Iterative reconstruction techniques could improve the quality of parametric images and regional quantitative BP estimates compared to FBP reconstruction. However, iterative reconstruction techniques are not yet widely used for radioligand PET. A recent comparative study concluded that iterative reconstruction (ordered subsets expectation maximization, OSEM) could be used for human neurotransmission PET studies but OSEM was not necessarily superior to FBP (Morimoto et al., 2006).

The studied deconvolution methods included the Richardson-Lucy algorithm, the reblurred Van Cittert algorithm, and the reblurred Van Cittert algorithm regularized with the total variation functional. The former is based on the Poisson noise model and maximum likelihood principle. The latter two are based on the least squares formulation under the Gaussian noise model. In addition, we compared the VC deconvolution for the images FBP-reconstructed using Hann and ramp filters. We excluded many candidate algorithms from this study based on our previous conference article (Tohka and Reilhac, 2006). The LS based algorithms provided better results than the RL algorithm in this study. They were more immune to the anatomy induced variation than the Richardson-Lucy method with the regional BP computations. Interestingly, when studying the extracted regional TACs, only the cerebellum TAC appeared to vary more from subject to subject with the RL deconvolution. The variation in TACs of striatal structures was highly similar between RL and LS-based deconvolutions. This indicates the importance of investigating the effect of image processing to the final parameters of interest in addition to characterizing the errors in regional TACs in quantitative PET. VC-ramp produced less bias in the regional BP values than VC-Hann. However, also anatomy induced variation in regional BP values was greater and robustness

to segmentation errors was compromised with VC-ramp. The standard VC algorithm performed slightly better than the regularized version. However, we used the steepest descent method for the optimization and there might be additional gains available with the regularized version if a more advanced optimization algorithm would be adopted. Another possibility to improve the deconvolution results would be to adopt a spatially varying PSF since the resolution across the PET image is not uniform. However, for the striatal imaging this would probably have a negligible effect, since striatum is typically located near the center of the field of view and the resolution at striatum is expected to be well approximated by the measured PSF. The cerebellum, needed as the reference region with Raclopride, is not located near the center of the field of view. However, the TACs for the cerebellum were already rather accurate, and thus there was little room for improvement with cerebellum TACs.

All the deconvolution algorithms were fast, typically converging in 10 to 30 iterations per image frame when started from Hann-filter-reconstructed data. An increase in the noise level, as with VC-ramp, increased the number of required iterations. VC-ramp was often terminated because the maximum number of iterations (100) was reached before the stopping criterion (Eq. (8)) was satisfied. The average time consumption for one iteration in our Matlab implementation was 4 seconds with the RL and VC methods and 5 seconds with the VC-TV method. The time consumptions are reported for a IBM xSeries 335 Linux server with a 3.2 GHz Intel Xeon processor.

Noise level, image resolution and image size all have their influence to bias and variance of physiological parameters of interest as well as the number of iterations needed for deconvolution algorithms to converge. As these parameters vary according to the hardware and image correction/reconstruction choices, it is interesting to try to address the effects of these parameters to deconvolution performance. A low noise level appears to be more important than a high resolution as witnessed by comparisons of the VC-Hann and VC-ramp algorithms, at least when parametric imaging was considered. The higher noise level meant also increased number of iterations. Indeed, in our experience, a decreased resolution (if the image sizes and noise levels are the same) does not especially degrade the deconvolution performance. If the number of voxels is increased together with the resolution (by decreasing the voxel dimensions) while maintaining the noise level, the number of iterations needed will increase. Effects to other, more important performance indicators are harder to address without specifying a particular application. However, we expect the studied deconvolution algorithms to perform similarly as in the current study if applied to the data from a more modern scanner than the Exact Ecat HR+.

6 Conclusion

The deconvolution based PV correction is a viable alternative to MR based PV correction with quantitative ^{11}C Raclopride PET. Deconvolution based PV correction was found to provide quantification accuracy close to the MR-based GTM PV correction when a small adjustment of VOIs to compensate for tissue fraction effect was applied. Deconvolution based PV correction was also found to be less sensitive to differences in VOI delineation than the GTM method. Moreover, in comparison to the MR based methods, deconvolution based PV correction is available without individual MR images and is free on the assumptions of the structure-wise homogeneity of the receptor availability. The latter makes deconvolution based PV correction suitable for parametric imaging unlike many MR based PV correction schemes.

The deconvolution based PV correction was found to improve quantification accuracy compared to the case when PV correction was omitted. When compared against the different extremes (in terms of resolution/noise tradeoff) of the FBP reconstruction, the parametric images from the deconvolved dynamic images were slightly more noisy, but that was well compensated by a superior resolution of the parametric images.

The best deconvolution method in this study was the reblurred Van Cittert algorithm. Its quantitative performance was superior to the Richardson-Lucy algorithm and the VC with the total variation regularization.

Acknowledgments

The work of J.T. was supported by the Academy of Finland under the grants 108517, 104834, and 213462 (Finnish Centre of Excellence program (2006-2011)).

Thanks to Vesa Oikonen from Turku PET Centre for helpful discussions about binding potential computations with Raclopride.

References

- Aston, J., Cunningham, V., Asselin, M.-C., Hammers, A., Evans, A., Gunn, R., 2002. Partial volume correction: Estimation and algorithms. *J Cereb Blood Flow Metab* 22 (8), 1019 – 1034.

- Baete, K., Nuyts, J., Van Laere, J., Van Paesschen, W., Ceyssens, S., De Ceuninck, L., Gheysens, O., Kelles, A., Van den Eynden, J., Suetens, P., Dupont, P., 2004. Evaluation of anatomy based reconstruction for partial volume correction in brain fdg-pet. *NeuroImage* 23, 305 – 317.
- Biemond, J., Lagendijk, R., Mereserau, R., 1990. Iterative methods for image deblurring. *Proc. IEEE* 78, 856 – 883.
- Boussion, N., Cheze Le Rest, C., Reilhac, A., Turkheimer, F., Visvikis, D., 2007. Quantitative evaluation of multiresolution partial volume compensation in fdg pet and rcbf spect using realistic simulation and clinical brain image. Submitted to *NeuroImage*.
- Carasso, A., 1999. Linear and nonlinear image deblurring: A documented study. *SIAM J. Numer. Anal.* 36, 1659 – 1689.
- Charalambous, C., Ghaddar, F., Kouris, K., 1992. Two iterative image restoration algorithms with applications to nuclear medicine. *IEEE Trans Med Imag* 11, 2 – 8.
- Collins, D. L., Neelin, P., Peters, T. M., Evans, A. C., 1994. Automatic 3d inter-subject registration of mr volumetric data in standardized talairach space. *Journal of Computer Assisted Tomography* 18 (2), 192 – 205.
- Collins, D. L., Zijdenbos, A. P., Baaré, W. F. C., Evans, A. C., 1999. ANIMAL+INSECT: Improved cortical structure segmentation. In: *Proc. of the 16th International Conference on Information Processing in Medical Imaging (IPMI)*, vol. 1613 of LNCS. pp. 210 – 223.
- Demoment, G., 1989. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Trans on Acoustics, Speech, and Signal Processing* 37, 2024 – 2036.
- Dobson, D., Vogel, C., 1997. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.* 34, 1779 – 1791.
- Frouin, V., Comtat, C., Reilhac, A., Gregoire, M.-C., 2002. Correction of partial-volume effect for PET striatal imaging: Fast implementation and study of robustness. *J Nucl Med* 43, 1715 – 1726.
- Glatting, G., Mottaghy, F., Karitzky, J., Baune, A., Sommer, F., Landwehrmeyer, G., Raske, S., 2004. Improving binding potential analysis in [11c]raclopride pet studies using cluster analysis. *Med Phys* 31, 902 – 906.
- Gunn, R. N., Lammertsma, A. A., Hume, S., Cunningham, V., 1997. Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *Neuroimage* 6 (4), 279 – 287.
- Kabani, N. J., MacDonald, D. J., Holmes, C. J., Evans, A. C., 1998. 3d anatomical atlas of the human brain. In: *OHBM 1998*.
- Kao, C.-M., Yap, J., Mukherjee, J., Wernick, M., 1997. Image reconstruction for dynamic PET based on low-order approximation and restoration of the sinogram. *IEEE Trans. Med. Imag.* 16, 738 – 749.
- Kessler, R., Ellis, J., Eden, M., 1984. Analysis of emission tomographic scan data: Limitations imposed by resolution and background. *J Comput Assist Tomography* 8, 514 – 522.

- Lammertsma, A., Hume, S., 1996. Simplified reference tissue model for PET receptor studies. *Neuroimage* 4, 153 – 158.
- Leahy, R., Qi, J., 2000. Statistical approaches in quantitative positron emission tomography. *Statistics and Computing* 10, 147 – 165.
- Liptrot, M., Adams, K., Martiny, L., Pinborg, L., Lonsdale, M., Olsen, N., Holm, S., Svarer, C., Knudsen, G., 2004. Cluster analysis in kinetic modelling of the brain: a noninvasive alternative to arterial sampling. *NeuroImage* 21, 483 – 493.
- Lucy, L., 1974. An iterative technique for retriification of observed distributions. *Astron J* 79, 745 – 765.
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 16 (2), 187 – 198.
- Mawlawi, O., Martinez, D., Slifstein, M., Broft, A., Chatterjee, R., Hwang, D. R., Huang, Y., Simpson, N., Ngo, K., Heertum, R. V., Laruelle, M., Sep 2001. Imaging human mesolimbic dopamine transmission with positron emission tomography: I. accuracy and precision of d(2) receptor parameter measurements in ventral striatum. *J Cereb Blood Flow Metab* 21 (9), 1034–1057.
- Meltzer, C., Kinahan, P., Greer, P., Nichols, T., Comtat, C., Cantwell, M., Lin, M., Price, J., 1999. Comparative evaluation of mr-based partial-volume correction schemes for PET. *J Nucl Med* 40, 2053 – 2065.
- Meltzer, C., Zubieta, J., Links, J., Brakeman, P., Stumpf, M., Frost, J., 1996. Mr-based correction of brain pet measurements for heterogeneous gray matter radioactivity distribution. *J Cereb Blood Flow Metab* 16, 650 – 658.
- Mignotte, M., Meunier, J., Soucy, J.-P., Janicki, C., 2002. Comparison of deconvolution techniques using a distribution mixture parameter estimation: Application in single photon emission computed tomography imagery. *J. Electronic Imaging* 11 (1), 1 – 14.
- Morimoto, T., Ito, H., Takano, A., Ikoma, Y., Seki, C., Okauchi, T., Tanimoto, K., Ando, A., Shiraishi, T., Yamaya, T., Suhara, T., 2006. Effects of image reconstruction algorithm on neurotransmission pet studies in humans: comparison between filtered backprojection and ordered subsets expectation maximization. *Ann Nucl Med* 20, 237 – 243.
- Morris, E., Chefer, S., Lane, M., Muzic, R., Wong, D., Dannals, R., Matochik, J., Bonab, A., Villemagne, V., Grant, S., Ingram, D., Roth, G., London, E., 1999. Loss of D2 receptor binding with age in rhesus monkeys: importance of correction for differences in striatal size. *J Cereb Blood Flow Metab* 19, 218 – 229.
- Müller-Gärtner, H., Links, J., Prince, J., Bryan, R., McVeigh, E., Leal, J., Davatzikos, C., Frost, J., 1992. Measurement of radiotracer concentration in brain gray matter using positron emission tomography: Mri-based correction for partial volume effects. *J Cereb Blood Flow Metab* 12, 571 – 583.
- Nunez, J., Llacer, J., 1990. A fast bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images. *IEEE*

- Trans. Med. Imag. 9, 159 – 171.
- Passchier, J., Gee, A., Willemsen, A., Vaalburg, W., van Waarde, A., 2002. Measuring drug-related receptor occupancy with positron emission tomography. *Methods* 27, 278 – 286.
- Quarantelli, M., Berkouk, K., Prinster, A., Landeau, B., Svarer, C., Balkay, L., Alfano, B., Brunetti, A., Baron, J., Salvatore, M., 2004. Integrated software for the analysis of brain pet/spect studies with partial-volume-effect correction. *J Nucl Med* 45, 192 – 201.
- Reilhac, A., Batan, G., Michel, C., Grova, C., Tohka, J., Collins, D., Costes, N., Evans, A., 2005. PET-SORTEO: validation and development of database of simulated PET volumes. *IEEE Trans Nucl Sci* 52 (5), 1321 – 1328.
- Reilhac, A., Lartizien, C., Costes, N., Sans, S., Comtat, C., Gunn, R., Evans, A., 2004. PET-SORTEO: A Monte Carlo-based simulator with high count rate capabilities. *IEEE Transactions on Nuclear Science* 51 (1), 46 – 52.
- Reilhac, A., Rousset, O., Comtat, C., Frouin, V., Gregoire, M.-C., Evans, A., 2000. A correction algorithm for partial volume effects in 3D PET imaging: Principle and validation. In: *IEEE-NSS/MIC 2000 Conference Record Vol. 3*. pp. 18/62 – 18/66.
- Richardson, W., 1972. Bayesian-based iterative method for image restoration. *J Opt Soc Am* 62, 55 – 59.
- Rousset, O., Ma, Y., Evans, A., 1998. Correction for partial volume effects in PET: principle and validation. *J. Nucl. Med.* 39, 904–911.
- Rudin, L., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259 – 268.
- Strul, D., Bendriem, B., 1999. Robustness of anatomically guided pixel-by-pixel algorithms for partial volume effect correction in positron emission tomography. *J Cereb Blood Flow Metab* 19, 547 – 559.
- Tohka, J., Reilhac, A., 2006. A monte carlo study of deconvolution algorithms for partial volume correction in quantitative pet. In: *In NSS/MIC Conference Record*. pp. 3339 – 3345.
- Tohka, J., Wallius, E., Hirvonen, J., Hietala, J., Ruotsalainen, U., 2006. Automatic extraction of caudate and putamen in [^{11}C]-raclopride PET using deformable surface models and normalized cuts. *IEEE Trans Nucl Sci* 53, 220 – 228.
- White, R., 1993. Image restoration using the damped richardson-lucy method. In: *Second Workshop on the Restoration of Images and Spectra from the Hubble Space Telescope*. pp. 103 – 110.
- Zaidi, H., Ruest, T., Schoenahl, F., Montandon, M.-L., 2006. Comparative assessment of statistical brain MR image segmentation and their impact on partial volume correction in PET. *NeuroImage* 32, 1591 – 1607.