RESEARCH ARTICLE

# Quantitative Intensity Harmonization of Dopamine Transporter SPECT Images Using Gamma Mixture Models

Alberto Llera,[1] Ismael Huertas,[1,2] Pablo Mir,[2] Christian F. Beckmann[1,3,4]

[1]*Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

[2]*Unidad de Trastornos del Movimiento, Servicio de Neurología y Neurofisiología Clínica, Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Avda. Manuel Siurot s/n., 41013, Seville, Spain*

[3]*Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, The Netherlands*

[4]*Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, Oxford, UK*

## Abstract

*Purpose:* Differences in site, device, and/or settings may cause large variations in the intensity profile of dopamine transporter (DAT) single-photon emission computed tomography (SPECT) images. However, the current standard to evaluate these images, the striatal binding ratio (SBR), does not efficiently account for this heterogeneity and the assessment can be unequivalent across distinct acquisition pipelines. In this work, we present a voxel-based automated approach to intensity normalize such type of data that improves on cross-session interpretation.

*Procedures:* The normalization method consists of a reparametrization of the voxel values based on the cumulative density function (CDF) of a Gamma distribution modeling the specific region intensity. The harmonization ability was tested in 1342 SPECT images from the PPMI repository, acquired with 7 distinct gamma camera models and at 24 different sites. We compared the striatal quantification across distinct cameras for raw intensities, SBR values, and after applying the Gamma CDF (GDCF) harmonization. As a proof-of-concept, we evaluated the impact of GCDF normalization in a classification task between controls and Parkinson disease patients.

*Results:* Raw striatal intensities and SBR values presented significant differences across distinct camera models. We demonstrate that GCDF normalization efficiently alleviated these differences in striatal quantification and with values constrained to a fixed interval [0, 1]. Also, our method allowed a fully automated image assessment that provided maximal classification ability, given by an area under the curve (AUC) of AUC = 0.94 when used mean regional variables and AUC = 0.98 when used voxel-based variables.

*Conclusion:* The GCDF normalization method is useful to standardize the intensity of DAT SPECT images in an automated fashion and enables the development of unbiased algorithms using multicenter datasets. This method may constitute a key pre-processing step in the analysis of this type of images.

## Introduction

The imaging of the dopamine transporter (DAT) with nuclear medicine techniques, such as single-photon emission computed tomography (SPECT) or positron emission tomography (PET), is a very useful and widespread tool for the diagnosis of Parkinson's disease (PD) and other neurodegenerative parkinsonisms [1]. In these diseases, there is a progressive degeneration of the dopaminergic neurons in the nigrostriatal pathway, which projects from the substantia nigra to the striatum. DATs are located at the presynaptic nerve terminals and are responsible for reuptake of dopamine in the synapses of these neurons to the striatum. Hence, DAT imaging allows for *in vivo* visualization of the state of these projections and thus determining the presence of neuronal degeneration [2]. DATs can be imaged with molecular binding agents such as the cocaine derivative I-123 labeled 2β-carbomethoxy-3β-(4-iodophenyl)-N-(3-fluoropropyl) nortropane ([$^{123}$I]FP-CIT). Binding of this molecule in the brain is stable for 3–6 h after its administration and best imaged 3–4 h post-injection, which makes it ideal for clinical use. Its use was approved by the European Medicines Agency in 2000 and by the Food and Drug Administration in the USA in January 2011 [3].

The current standard method of assessment consists of the quantification of tracer binding in the main striatal regions of interest (ROI), namely putamen and caudate, through the calculation of the striatal binding ratio (SBR). In this calculation, the tracer binding in displaceable regions (*i.e.*, the striatum) is normalized to the tracer binding in non-displaceable regions. The non-displaceable region refers to a region where the DATs are poorly expressed, commonly the occipital cortex or the cerebellum [4]. The SBR of a particular ROI is then formulated as

$$SBR_{\mathrm{ROI}} = \frac{C_{\mathrm{ROI}} - C_{\mathrm{NDB}}}{C_{\mathrm{NDB}}}$$

where $C_{\mathrm{ROI}}$ is the mean count per voxel in the ROI (whole striatum, putamen, or caudate) and $C_{\mathrm{NDB}}$ represents the mean count per voxel in the non-displaceable region. Unfortunately, the SBR does not efficiently harmonize the large variability in intensity that DAT SPECT images present. One of the largest sources for this variability, apart from the subject-related physiological factors, is due to the imaging system, comprised of the acquisition hardware and the image reconstruction software [5, 6]. Indeed, sites with distinct gamma cameras but comparable subjects produced significantly different SBR values in the European multicenter database of healthy controls for [$^{123}$I]FP-CIT SPECT (ENC-DAT) [5]. This discrepancy is partly due to the fact that this ROI-based assessment makes assumptions about the intensity distribution of the image that are not valid, since the ratio between displaceable and non-displaceable ROIs may largely differ across different acquisition set-ups. This bias has negative consequences in the analysis of these images, including (i) the limitation to pool and compare images from multiple sites, (ii) the generation of non-comparable estimates in machine learning algorithms and a resulting poor generalization, and (iii) lack of reproducibility of single-site findings. There have been recent efforts to reduce this variability in ENC-DAT dataset by revisiting the reconstruction algorithm [7]. However, this approach is far from being a straightforward step since it requires dealing with detailed camera-specific factors such as correction of scatter and septal penetration for image reconstruction, which may turn the normalization into a cumbersome process.

In this work, we present a normalization method that relies on the harmonization of the intensity histogram of the reconstructed image. We have obtained exceptional harmonization metrics by only making use of basic information about the acquisition, such as site and gamma camera model, which goes in favor of practicability. More specifically, our method consists of a reparametrization of the voxel values according a mixture model of two Gamma distributions fitting a normative image histogram. Such reparametrization is based on the cumulative density function of the Gamma component modeling the specific uptake so that the new feature space is constrained to the interval [0, 1].

To test our method, we used over 1000 images obtained from the Parkinson's Progression Markers Initiative (PPMI) repository and which were acquired at multiple sites. We demonstrate that our normalization procedure (i) equalizes the striatal DAT quantification across distinct SPECT cameras and (ii) improves the performance of algorithms developed with images acquired at multiple sites. The application of our method allows to pooling and comparing images from different sites and studies. This harmonization step is crucial to guarantee the reproducibility of findings and the generalization of the algorithms using this type of scans.

## Material and Methods

### Dataset

In this work, we made use of DAT SPECT images obtained from the Parkinson's Progression Markers Initiative (PPMI)

database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI—a public-private partnership—is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners (www.ppmi-info.org/fundingpartners). This study involves more than 30 clinical sites worldwide and the DAT SPECT is one of the imaging tests that are being collected. These conditions make this multicenter repository ideal to reflect the heterogeneity of real world data and thus to test our method for intensity harmonization. We randomly selected 1342 DAT SPECT scans from which 210 corresponded to independent healthy controls and 1132 to PD patients (439 independent subjects with repeated scans over the course of the disease). The list of PPMI identifiers of the scans/subjects used in this work is provided in Table S1 in supplementary material. PPMI repository also contains data from SWEDDS subjects (scans without evidence for dopaminergic deficit from patients clinically presumed to have Parkinson's disease), but these were not included in this work. Each participating PPMI site received approval from an ethical standards committee on human experimentation before study initiation and written informed consent for research was obtained from all participants in the study.

These scans were performed at 24 different sites and were acquired with 7 different SPECT camera models from a variety of manufacturers, including (1) MARCONI 3000XP, (2) PHILIPS BrightView, (3) SIEMENS Encore2, (4) SIEMENS IP2, (5) PICKER HERMES Workstation, (6) GE MILLENNIUM MG, and (7) GE VARICAM. The details and demographics of the sample, grouped by camera, are shown in Table 1. All brain images were spatially normalized to MNI152 standard space using a linear affine transformation in FSL (http://fsl.fmrib.ox.ac.uk/fsl) [8–10] and a custom DAT SPECT template (http://www.nitrc.org/projects/spmtemplates) [11]. The spatially normalized images were subsequently masked with a whole brain mask available in FSL to include only gray and white matter areas.

## DAT SPECT Intensity Profile

Our intensity normalization procedure is based on the assumption that the voxel-wise imaging data contain voxels that belong to two distinct classes: (i) a majority of voxels that show non-displaceable dopamine uptake and (ii) a spatially confined region within the image with elevated levels of uptake. Consequently, as shown in Fig. 1a, the histogram of DAT SPECT images is characterized by a positive-valued distribution with two prominent features: (i) a broad main component with positive mode at low values, which corresponds to the majority of voxels with non-displaceable uptake (background intensity) and (ii) a right-skewed long tail at higher intensity values than the former and which corresponds to the activation voxels, namely the specific uptake region.

However, the shape and x-axis offset of this intensity histogram can largely vary depending upon the scanned subject and the imaging system. This is due to the fact that the intensity profile of this type of images heavily depends on a variety of factors including acquisition hardware and settings, reconstruction algorithm and physiological-related factors, such as age, gender, and metabolism [5, 7]. This is illustrated in Fig. 1b where it is depicted the mean intensity distribution of the raw images of the healthy controls grouped by camera model. Here, it can be noticed the large differences in terms of both shape of the distribution and quantitative offset produced by one of the major sources of variability, the imaging system. Moreover, as mentioned before, other acquisition-related factors beyond the camera (e.g., collimators, reconstruction technique) can largely contribute to this variability in intensity. This is reflected in the bulky profile of some distributions, especially in those where data from many sites were grouped (e.g., SIEMENS Encore2).

## Intensity Normalization with Gamma Distributions

We propose to intensity normalize DAT SPECT images by modeling the intensity profile with a mixture model of Gamma distributions. The Gamma distribution is a positive-supported continuous distribution characterized by two parameters: shape and scale, $\Theta = \{s, r\}$. These parameters enable the function to adopt a wide range of forms (Fig. 2a), and in fact, well-known distributions such as the exponential, the Gaussian, and chi-square are special cases of the Gamma. Thus, given the tailed intensity profile of DAT SPECT images, Gamma distributions appear to be well-

**Table 1.** Demographics and sample characteristics grouped by gamma camera model and diagnosis. Note: some sites have more than one camera

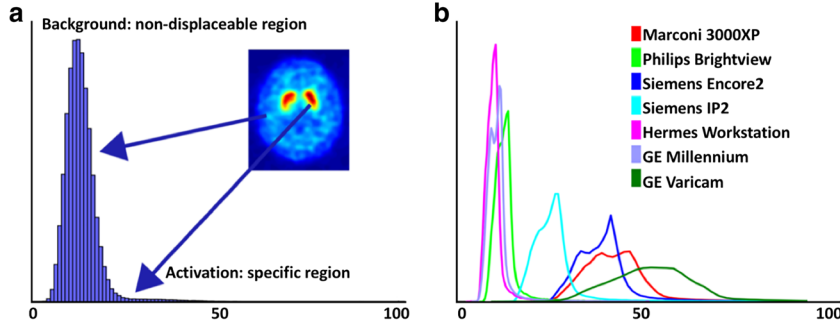| | MARCONI 3000XP | PHILIPS BrightView | SIEMENS Encore2 | SIEMENS IP2 | PICKER HERMES | GE MILLENIUM | GE VARICAM |
|---|---|---|---|---|---|---|---|
| Site (N) | 1 | 3 | 14 | 6 | 4 | 1 | 2 |
| Controls | | | | | | | |
| N | 15 | 21 | 90 | 30 | 31 | 14 | 9 |
| gender (M/F) | 9 / 6 | 15 / 6 | 60 / 30 | 15 / 15 | 23 / 8 | 10 / 4 | 6 / 3 |
| age (years) | 60 ± 11 | 62 ± 13 | 62 ± 10 | 59 ± 11 | 61 ± 13 | 55 ± 13 | 64 ± 13 |
| PD | | | | | | | |
| N | 179 | 133 | 494 | 128 | 108 | 61 | 29 |
| gender (M/F) | 128 / 51 | 65 / 68 | 309 / 185 | 97 / 31 | 68 / 40 | 44 / 17 | 20 / 9 |
| age (years) | 62 ± 10 | 68 ± 9 | 63 ± 10 | 61 ± 9 | 63 ± 9 | 60 ± 11 | 58 ± 11 |

Fig. 1. **a** Characteristic intensity histogram of a DAT SPECT image. **b** Mean camera-specific DAT SPECT image histogram for the healthy controls.

suited to model these image distributions. Specifically, we propose to model DAT SPECT image data with a mixture model of two Gamma distributions, one modeling the background and one modeling the specific uptake. The strategy consists of three conceptual steps:

1. Obtain a normative distribution using SPECT images with same/similar acquisition settings.
2. Fit a two-component Gamma mixture model to that normative distribution.
3. Reparametrize further SPECT images (either healthy or pathologic) from that particular acquisition set-up using the mixture model. Each voxel value x will be reparametrized based on its cumulative density function (CDF(x)) under the Gamma distribution modeling the specific region (*i.e.*, the second Gamma). Consequently, all voxel values will be re-casted into the interval [0, 1].

This procedure is illustrated in Fig. 2b, where the red dashed line represents the fit of the normative distribution and the orange area represents the CDF(x) under the activation component. In this work, we have considered 7

normative models, one per gamma camera model, using the images of the healthy controls available for each camera independently. Nevertheless, other alternatives to generate the normative model may be also valid. Indeed, when enough samples are available, we encourage the creation of normative models that are specific for a single acquisition set-up, *e.g.*, a site, rather than grouping per camera model. Also, the use of an anthropomorphic phantom or scans from patients may serve to approximate the normative model when data from healthy controls are not available.

## Mathematical Formulation

Let $X = \{x_1, \ldots, x_{S_1}, \ldots, x_S\}$ represent a set of vectors corresponding to DAT SPECT images from S subjects recorded using the same camera model. Consider that the first $S_1$ subjects are healthy control subjects and the remaining ones have some pathology (*e.g.*, PD). For each subject $s$, the vector $x_s = (x_1^s, \ldots, x_n^s)$ represents the measurements at the $n \in N$ spatial locations in the brain (voxels). The proposed method consists of three main steps:
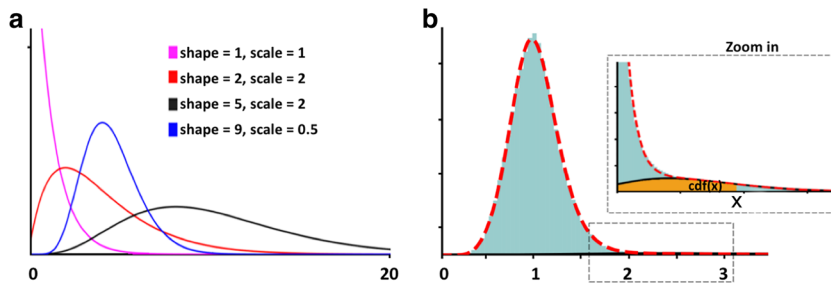


Fig. 2. **a** Gamma probability density function across different shape and scale parameters. **b** Modeling of a camera-specific normative intensity profile by a mixture of two gamma distributions. The green bars represent the mean image histogram, and the red dashed line the mixture model fit. The inset shows a zoom on the specific uptake region where the black continuous line shows the gamma distributed component modeling the specific region, and the orange colored area represents the cumulative density function (CDF) of a voxel x under such gamma component. Each voxel value x in the image will be reparametrized by its CDF(x) value.

1. Obtain a normative distribution using SPECT images with same/similar acquisition settings. This step involves a subject-wise scaling followed by the selection of a set of images:

a. Subject-wise scaling with respect to the mode: $\tilde{x}_j^s = \frac{x_j^s}{M_s}$; $j \in \{1, ..., N\}, s \in \{1, ..., S\}$, where $M_s$ denotes the mode of the subjects' intensity histogram.
b. Concatenate scaled images of a specific normative model: $x = (\tilde{x}_1, ..., \tilde{x}_{S_1})$ which corresponds to the healthy subjects of a given camera model.

2. Learn the parameters of a normative mixture model of two Gamma components of the form:

$$p(x|\pi, \Theta) = \prod_{n=1}^{NS_1} \sum_{k=1}^{2} \pi_k G_k(x_n, \Theta_k)$$

where $N$ is the number of voxels and $S_1$ is the number of subjects selected for the normative distribution, $k$ indexes the two mixture components, $G_k$ expresses the $k$-th Gamma distribution, $\Theta_k = \{s_k, r_k\}$ represents Gamma shape and scale parameters, and $\pi_k$ are the mixing proportions. We used a Variational Bayesian approach to learn the model parameters [12].

3. Substitute each scaled voxel value (1a) by its CDF (area) under the Gamma distribution modeling the specific region ($k = 2$):

$$CDF\left(\tilde{x}_j^s\right) = \frac{1}{r_2^{s_2} \Gamma(s_2)} \int_0^{\tilde{x}_j^s} t^{s_2-1} \exp\left(\frac{-t}{r_2}\right) dt$$

where $\Gamma$ represents the Gamma function [12], $s$ and $j$ the indexes of subject and brain location respectively, and $\Theta_2 = \{s_2, r_2\}$ represents Gamma parameters of the component modeling the specific region.

We will further denote this reparametrization as the Gamma CDF (GCDF) intensity normalization. Noteworthy, in step 1a, we used the mode to subject-wise scale the images. We found the mode a convenient metric since it is data-driven, avoids definitions of ROIs, and provides a reliable mapping of each subject distribution to a distribution with mode one that allows an easy, principled and robust initialization for the mixture model fit. However, we believe that, given the flexibility of the mixture model to learn the scaled data, other scaling metrics (*e.g.*, non-displaceable binding, SBR) may be also valid.

*Statistical Analyses*

In order to investigate the harmonization ability of our newly proposed method, we compared across cameras the intensity values of putamen and caudate after harmonization with GCDF. As a reference, we evaluated the differences, also across cameras, of the putamen and caudate raw intensity values and the SBR values available on PPMI website. Despite striatum quantitative data of both healthy controls and PD are available, we only made comparisons in the control group, as disease severity may largely confound the striatal uptake differences in PD. We used ANOVA and Tukey post hoc analyses and the significance threshold was set to 0.007 (0.05/7) according to Bonferroni correction.

Then, as a proof-of-concept example, we evaluate the performance of a binary classifier based on logistic regression to distinguish between healthy controls and PD after GCDF normalization when images from multiple sites were pooled together. As a reference, previous single-center studies have demonstrated that the diagnostic accuracy of visual expert reading or SBR-based classifiers can be as high as of 95 % [1, 4]. We evaluated a low dimensional classifier using striatal ROI variables (bilateral putamen and caudate) and also a higher dimensional one using all striatal voxels; in both cases we used a logistic regression classifier and tenfold cross-validation for evaluation. Feature selection in the voxel-based regression was performed by L1 regularization. We also calculated, as a reference for the ROI- and voxel-based approaches, the performance of the classifiers using the SBR and the voxel raw intensities, respectively. The mean area under curve (AUC) of the 10 classifiers obtained from cross-validation was used as a measure of classification performance. The striatum and its subregions were based on a manually delineated template (4622 voxels, https://www.nitrc.org/projects/striatalvoimap).

## Results

Figure 3 illustrates an example of a normal DAT SPECT image before (Fig. 3a) and after (Fig. 3b) the normalization step, where an enhancement of the contrast can be noticed.

Figure 4 represents the image intensity transformation over the different steps of the method. First, raw images showed a very heterogeneous profile (Fig. 4(a)). Then, subject dependent global scaling brought the intensity values to similar x-axis range, and the two-component mixture models were created (Fig. 4(b)). Finally, the CDF integral over the second component equalized the image histograms with values between 0 and 1 for all gamma cameras (Fig. 4(c)).

*Regional Quantification of the Striatum*

The descriptive values for striatal subregional quantification (raw intensities, SBR, and GCDF-harmonized values) across Gamma cameras are shown in Table 2. As expected, post hoc comparisons revealed significant differences in the
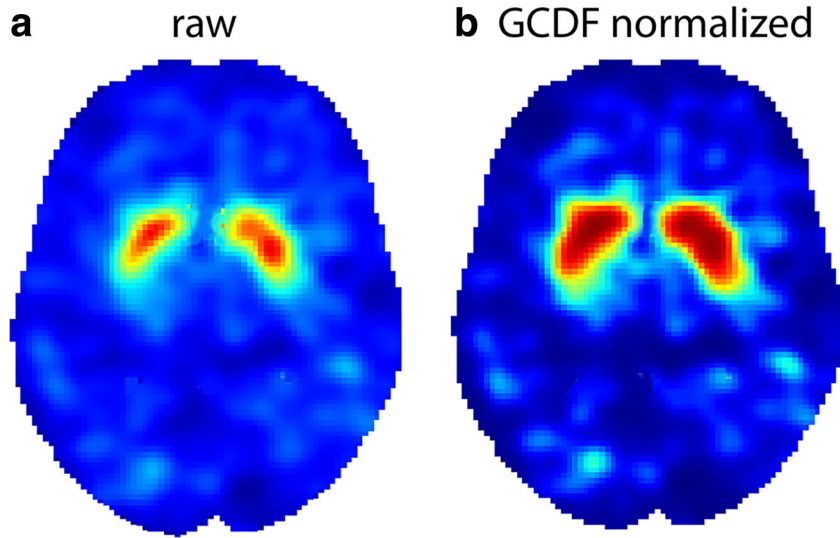
**Fig. 3.** Example of a DAT SPECT image **a** before and **b** after normalization (GCDF normalized).

mean putamen and caudate raw intensity values between multiple cameras in the sample of healthy controls. We also observed that many of these differences were still present in the SBR values quantified by the PPMI sites themselves. The proposed GCDF-based normalization remarkably produced homogeneous mean striatal intensity values and efficiently eliminated all the differences across cameras.

*Diagnostic Classifiers Using Multisite Data*

The results of the classifiers are summarized in Table 3. We found that the classifiers using the striatum data after GCDF normalization provided high diagnostic performance using images from multiple sites. When all images were GCDF-harmonized and pooled, the classifier using ROI variables
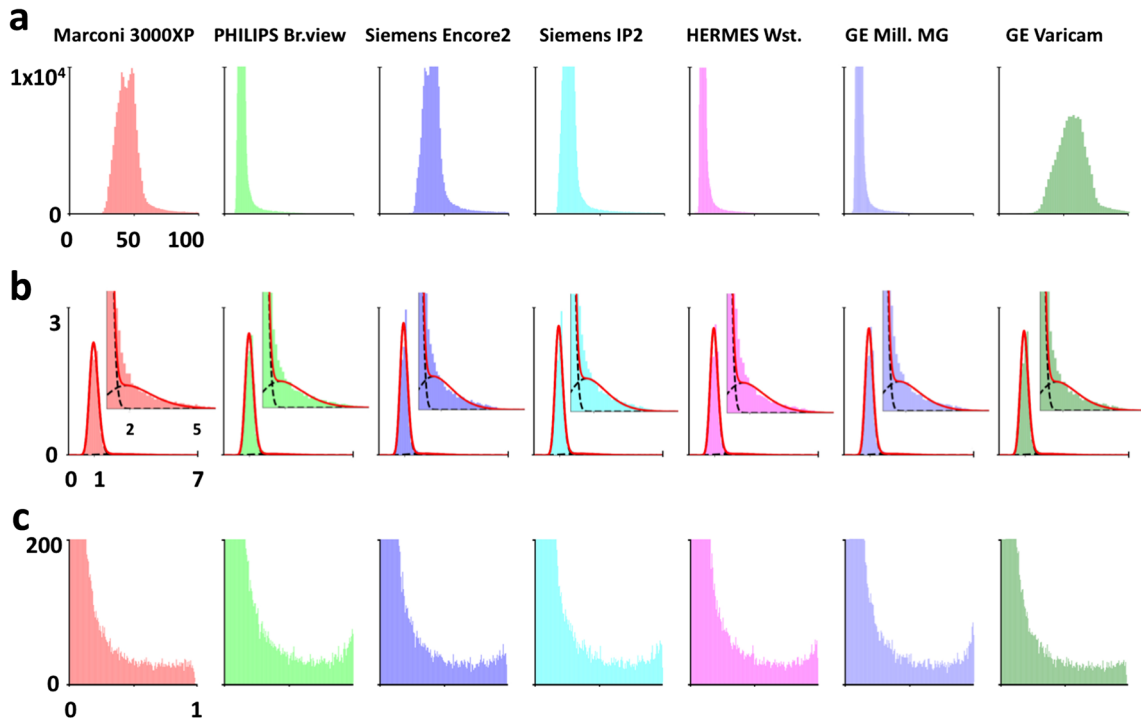


**Fig. 4.** Normalization image histograms transformation for each camera model. (a) Normative distributions given by the images of all the healthy controls. (b) Scaled data histograms and two-component gamma mixture models fits (red lines). (c) Data histograms after GCDF normalization.

**Table 2.** Mean values for striatal variables for healthy controls and PD patients using the raw values, the striatal binding ratio (SBR), and gamma cumulative density function (GCDF) normalization methods. Superscripts indicate the pairwise post hoc differences between the cameras, which are numerically labeled as shown in their names

| | | MARCONI | PHILIPS | SIEMENS | | PICKER | GE | |
| | | 3000XP[1] | BrightView[2] | Encore2[3] | IP2[4] | HERMES[5] | MILLENNIUM[6] | VARICAM[7] |
|---|---|---|---|---|---|---|---|---|
| Controls | N | 15 | 21 | 90 | 30 | 31 | 14 | 9 |
| | Raw caudate | $113 \pm 21^{2,4,5,6}$ | $33 \pm 7^{1,3,7}$ | $95 \pm 54^{2,4,5,6}$ | $62 \pm 37^{1,3,7}$ | $27 \pm 8^{1,3,7}$ | $28 \pm 5^{1,3,7}$ | $142 \pm 42^{2,4,5,6}$ |
| | Raw putamen | $101 \pm 18^{2,4,5,6}$ | $29 \pm 6^{1,3,7}$ | $87 \pm 49^{2,4,5,6}$ | $56 \pm 35^{1,3,7}$ | $24 \pm 7^{1,3,7}$ | $25 \pm 4^{1,3,7}$ | $130 \pm 38^{2,4,5,6}$ |
| | SBR caudate | $3.48 \pm 0.65^{3,4}$ | $2.96 \pm 0.50$ | $2.70 \pm 0.43^{1,5,7}$ | $2.72 \pm 0.48^{1,5,7}$ | $3.47 \pm 0.64^{3,4}$ | $3.08 \pm 0.65$ | $3.55 \pm 0.43^{3,4}$ |
| | SBR putamen | $2.59 \pm 0.72^{2,3,4}$ | $1.99 \pm 0.39^{3,5,7}$ | $1.91 \pm 0.37^{1,5,7}$ | $1.93 \pm 0.51^{1,5,7}$ | $2.52 \pm 0.54^{2,3,4}$ | $2.25 \pm 0.54$ | $2.94 \pm 0.34^{2,3,4}$ |
| | GCDF caudate | $0.65 \pm 0.06$ | $0.65 \pm 0.08$ | $0.66 \pm 0.08$ | $0.66 \pm 0.09$ | $0.64 \pm 0.08$ | $0.58 \pm 0.08$ | $0.57 \pm 0.10$ |
| | GCDF putamen | $0.54 \pm 0.05$ | $0.53 \pm 0.06$ | $0.55 \pm 0.08$ | $0.53 \pm 0.08$ | $0.52 \pm 0.07$ | $0.48 \pm 0.07$ | $0.48 \pm 0.06$ |
| PD | N | 179 | 133 | 494 | 128 | 108 | 61 | 29 |
| | Raw caudate | $93 \pm 23$ | $29 \pm 12$ | $75 \pm 43$ | $58 \pm 32$ | $27 \pm 11$ | $22 \pm 6$ | $108 \pm 31$ |
| | Raw putamen | $82 \pm 20$ | $26 \pm 11$ | $67 \pm 38$ | $51 \pm 27$ | $23 \pm 9$ | $19 \pm 4$ | $92 \pm 25$ |
| | SBR caudate | $1.99 \pm 0.50$ | $1.72 \pm 0.50$ | $1.70 \pm 0.47$ | $1.66 \pm 0.51$ | $2.20 \pm 0.59$ | $1.94 \pm 0.69$ | $2.63 \pm 0.59$ |
| | SBR putamen | $0.76 \pm 0.24$ | $0.67 \pm 0.27$ | $0.70 \pm 0.26$ | $0.69 \pm 0.35$ | $0.84 \pm 0.38$ | $0.71 \pm 0.27$ | $1.16 \pm 0.31$ |
| | GCDF caudate | $0.43 \pm 0.13$ | $0.42 \pm 0.16$ | $0.44 \pm 0.15$ | $0.39 \pm 0.15$ | $0.43 \pm 0.13$ | $0.44 \pm 0.16$ | $0.38 \pm 0.15$ |
| | GCDF putamen | $0.26 \pm 0.09$ | $0.27 \pm 0.12$ | $0.28 \pm 0.11$ | $0.24 \pm 0.11$ | $0.25 \pm 0.10$ | $0.27 \pm 0.10$ | $0.20 \pm 0.08$ |

Post hoc significance determined by $p < 0.007$

achieved AUC = 0.94 and the classifier using voxel data AUC = 0.98.

In general, the performance of the classifiers trained with GCDF-harmonized data was superior in the majority of the 16 classifiers as compared with the reference approaches (11 superior, 3 equal, and 2 inferior). Importantly, the performance of the GCDF voxel-based classifiers was in general superior to the performance of the SBR-based classifiers (AUC 0.98 *vs.* 0.92), which constitutes a clinically relevant improvement with respect to current standards. Although this may come at the expense of an increased complexity and computation time, our method allows to automating the analysis and benefiting from the higher spatial dimension of the voxel feature space. This opens the doors to more sophisticated analysis methods and challenging applications.

## Discussion

In this work, we have introduced a probabilistic normalization schema valid for DAT SPECT images based on a mixture of Gamma distributions. We showed that the proposed methodology provides a reparametrization of the

voxel data in the interval [0, 1] which effectively equalize the intensity of images from different cameras and centers. Furthermore, the proposed normalization step allowed an automated assessment of the images which provided high discrimination ability between controls and PD. Our method is also extensible to other modalities of functional imaging following the activation/background paradigm such as other SPECT radioligands or PET.

Data harmonization is a key element in the generation of reliable and reproducible results in science. In fact, recent years have seen an increase in alarming signals regarding the lack of replicability in neuroimaging research [13, 14]. There are many factors influencing this problem, some subject-related such as the variability in the collection of phenotypic data and the inherent heterogeneity in disease, but also the acquisition hardware and data analysis methods can play a major role. Unfortunately, this lack of standardization may generate site- or study-specific neuroimaging findings and estimates and therefore cause irreproducible and non-generalizable results. We have observed this phenomenon in DAT SPECT data from PPMI repository, where the mean SBR values from distinct cameras presented significant

**Table 3.** Mean (± standard deviation) tenfold classification performance (controls *vs.* Parkinson) given by the area under the curve (AUC). The classifiers were built using ROI striatal variables (SBR and after GCDF normalization) and all striatal voxels (raw and after GCDF normalization)

| | | ALL | MARCONI | PHILIPS | SIEMENS | | PICKER | GE | |
| | | | 3000XP | BrightView | Encore2 | IP2 | HERMES | MILLENIUM | VARICAM |
|---|---|---|---|---|---|---|---|---|---|
| Sites (N) | | 24 | 1 | 3 | 14 | 6 | 4 | 1 | 2 |
| RIO-based | SBR | $0.92 \pm 0.02$ | $0.96 \pm 0.01$ | $0.93 \pm 0.02$ | $0.93 \pm 0.02$ | $0.90 \pm 0.02$ | $0.91 \pm 0.02$ | $0.93 \pm 0.01$ | $1 \pm 0.00$ |
| | GCDF | $0.94 \pm 0.02$ | $1 \pm 0.00$ | $0.94 \pm 0.01$ | $0.91 \pm 0.03$ | $0.95 \pm 0.02$ | $0.96 \pm 0.01$ | $0.91 \pm 0.02$ | $1 \pm 0.00$ |
| Voxel-based | Raw | $0.93 \pm 0.02$ | $1 \pm 0.00$ | $0.94 \pm 0.02$ | $0.91 \pm 0.03$ | $0.89 \pm 0.01$ | $0.92 \pm 0.02$ | $0.96 \pm 0.01$ | $0.94 \pm 0.01$ |
| | GCDF | $0.98 \pm 0.01$ | $1 \pm 0.00$ | $0.97 \pm 0.01$ | $0.96 \pm 0.02$ | $0.97 \pm 0.02$ | $1 \pm 0.00$ | $0.96 \pm 0.01$ | $1 \pm 0.00$ |

differences. These differences can affect the interpretation and image pooling in multicenter studies, and eventually bias algorithms trained with images from different acquisition pipelines.

In our study, we have found that the GCDF normalization method was effective to homogenize the mean regional quantification of the striatum across different cameras, thus outperforming in this task to the classical SBR calculation. SBR values are commonly in the range [0.5, 4], and in line with our observations with PPMI data, previous work has shown that these values can differ substantially across sites [5, 6]. We have demonstrated that our method restricts the quantification to values in the interval [0, 1] independently of the camera and diminishes the intensity differences across SPECT devices. These new normalized values represent the CDF under the camera-specific normative Gamma distribution modeling the specific uptake region. Thanks to this new reparametrization, mean ROI values (*i.e.*, CDF values in putamen and caudate) could be unbiasedly compared across studies and pooled for meta-analysis purposes.

To probe this, we found that the GCDF normalization allowed deploying diagnostic classifiers that reached maximal classification accuracies using multisite data (AUC: SBR = 0.92 *vs* GCDF = 0.94). The performances using ROI variables were similar to those using the SBR, which are all in line with the reported performances in the literature (AUC = 0.90–0.95), and, reflect no significant performance gain or loss when using the GCDF normalization in single camera ROI-based comparisons. Positively, the performances using voxel variables were higher than the ROI-based classifiers and therefore above the current state-of-the-art levels (AUC = 0.95–1), although this could also be related to the higher complexity of the classifiers. Nevertheless, the key aspect is the fact that GCDF normalization harmonizes at the voxel level, thus allowing the use of automated voxel-based analyses and algorithms. This could maximize objectivity, practicability, and reproducibility against traditional specialist-dependent calculations. Indeed, there have been other recent proposals to intensity normalize DAT SPECT images with voxel-based methods [15–17]. However, in these works, the authors only included images from a single center and the multisite problem was not explicitly tackled. Moreover, our method has yielded higher classification performances [15–17].

Our method is flexible and the strategy to generate the normative distributions and fit the mixture models can be adapted to the sample and acquisition characteristics of a particular multicenter scenario. In this work, we created normative distributions using data from healthy controls, but other alternatives may be also valid when SPECT data from healthy subjects are not available. These alternatives may include (i) an anthropomorphic phantom, (ii) DAT SPECT images from patients without evidence of degeneration and reported as normals (*e.g.*, SWEDDs, non-degenerative parkinsonisms), or (iii) DAT SPECT images from PD patients: under the assumption of sparseness and using an adequate number of images, focal deficits may be compensated across, and the mean image may be a valid approximation. To illustrate this, we have also calculated the mixture models using only images from PD patients (Fig. S1 in Electronic Supplementary Material). Not surprisingly, the mixture models represent slightly shorter tailed distributions than the ones obtained using the healthy controls, modeling then lower values, presumably due to a lower DAT binding. Consequently, the CDF functions of the activation components are sharper sigmoid-like functions than the ones obtained when using the healthy controls, directly translating into a less smooth data transformation. Although in this suboptimal case the intensity harmonization was also achieved for all the cameras, classification performance was slightly lower than the one obtained using healthy controls (AUC = 0.92). Hence, we would like to be cautious and still encourage the use of non-degenerated SPECT data whenever possible; since being able to use patients for normative modeling would be a very relevant advantage, further research will be conducted to investigate possible biases induced by uptake deficits, sample sizes, or disease severity imbalance across centers that could affect the quality of the reparametrization.

As final remarks, it is important to point out the potential bias in our calculations owing to grouping images per camera model. As briefly mentioned in the methods section, a better approach would have been to fit models to normative distributions that are specific of single acquisition settings (e.g., a site). We decided to grouped them because, given the absence/scarcity of healthy cases in some sites, further stratification could have led to poorly reliable normative distributions. However, we noticed heterogeneous intensity profiles in images acquired with the same camera model (see Fig. 1b), probably caused by the variability in scanning parameters at different sites (e.g., collimators, reconstruction algorithm technique/version). Indeed, we believe that this may explain why a pair of SBR-based classifiers outperformed GCDF classifiers (Siemens Encore2 (14 sites) and GE Millennium (1 site)). We suspect that this effect could be related to the choice of suboptimal normative distributions, probably by the inclusion in the same normative model of data recorded under significantly different set-ups. The harmonization ability of our method relies on the accuracy of the normative distribution, and in theory, acquisition-specific models may provide more accurate results and control for the effect of other acquisition factors. Moreover, we also believe that biological factors such as age and gender should be also accounted for. In future work, it would be interesting to conduct a more thorough characterization of the distinct sources of variability, including data from poorly harmonized scenarios, and to evaluate the performance of our method also under different normative distributions scenarios. The simplicity and flexibility of the proposed data harmonization will allow in the near future the fusion of large datasets to perform multivariate modeling approaches that will provide detailed spatially interpretable information unavailable to date.

## Conclusion

This work presents an objective and automated intensity normalization method for DAT SPECT images. The method allows reparametrizing the voxel values of DAT SPECT images into a new normalized feature space (between 0 and 1) by modeling a normative intensity histogram as a mixture model of Gamma distributions. We have demonstrated that this normalization step alleviates the differences in striatal quantification produced by different gamma cameras and improves the accuracy of diagnostic algorithms trained with multicenter data. This harmonization step enables the reproducibility and generalization of clinical applications and therefore it may constitute a key pre-processing step in the analysis of this type of images.

**Compliance with Ethical Standards**

*Conflict of Interest*

CFB is director and shareholder in SBGNeuroLtd. The other authors declare no conflicts of interest.

*Ethical Approval and Inform Consent*

Each participating PPMI site received approval from an ethical standards committee on human experimentation before study initiation and written informed consent for research was obtained from all participants in the study.

## References

1. Tatsch K, Poepperl G (2013) Nigrostriatal dopamine terminal imaging with dopamine transporter SPECT: an update. J Nucl Med 54:1331–1338

2. Booij J, Teune LK, Verberne HJ (2012) The role of molecular imaging in the differential diagnosis of parkinsonism. Q J Nucl Med Mol Imaging 56:17–26

3. Djang DS, Janssen MJ, Bohnen N et al (2012) SNM practice guideline for dopamine transporter imaging with 123I-ioflupane SPECT 1.0. J Nucl Med 53:154–163

4. Badiavas K, Molyvda E, Iakovou I, Tsolaki M, Psarrakos K, Karatzas N (2011) SPECT imaging evaluation in movement disorders: far beyond visual assessment. Eur J Nucl Med Mol Imaging 38:764–773

5. Varrone A, Dickson JC, Tossici-Bolt L, Sera T, Asenbaum S, Booij J, Kapucu OL, Kluge A, Knudsen GM, Koulibaly PM, Nobili F, Pagani M, Sabri O, Vander Borght T, van Laere K, Tatsch K (2013) European multicentre database of healthy controls for [123I]FP-CIT SPECT (ENC-DAT): age-related effects, gender differences and evaluation of different methods of analysis. Eur J Nucl Med Mol Imaging 40:213–227

6. Tossici-Bolt L, Dickson JC, Sera T, de Nijs R, Bagnara MC, Jonsson C, Scheepers E, Zito F, Seese A, Koulibaly PM, Kapucu OL, Koole M, Raith M, George J, Lonsdale MN, Münzing W, Tatsch K, Varrone A (2011) Calibration of gamma camera systems for a multicentre European [123I]FP-CIT SPECT normal database. Eur J Nucl Med Mol Imaging 38:1529–1540

7. Buchert R, Kluge A, Tossici-Bolt L, Dickson J, Bronzel M, Lange C, Asenbaum S, Booij J, Atay Kapucu LÖ, Svarer C, Koulibaly PM, Nobili F, Pagani M, Sabri O, Sera T, Tatsch K, Vander Borght T, van Laere K, Varrone A, Iida H (2016) Reduction in camera-specific variability in [123I]FP-CIT SPECT outcome measures by image reconstruction optimized for multisite settings: impact on age-dependence of the specific binding ratio in the ENC-DAT database of healthy controls. Eur J Nucl Med Mol Imaging 43:1323–1336

8. Jenkinson M, Beckmann CF, Behrens TE et al (2012) FSL. NeuroImage 62:782–790

9. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, de Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, de Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23(Suppl 1):S208–S219

10. Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM (2009) Bayesian analysis of neuroimaging data in FSL. NeuroImage 45:S173–S186

11. Garcia-Gomez FJ, Garcia-Solis D, Luis-Simon FJ et al (2013) Elaboration of the SPM template for the standardization of SPECT images with Jenkinson M, Beckmann CF, Behrens TE -Ioflupane. Rev Esp Med Nucl Imagen Mol 32:350–356

12. Llera A, Vidaurre D, Pruim RHR, Beckmann CF (2016) Variational mixture models with gamma or inverse-gamma components. ArXiv e-prints Volume 16072016

13. Gorgolewski KJ, Poldrack RA (2016) A practical guide for improving transparency and reproducibility in neuroimaging research. PLoS Biol 14:e1002506

14. Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, Nichols TE, Poline JB, Vul E, Yarkoni T (2017) Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat Rev Neurosci 18:115–126

15. Brahim A, Gorriz JM, Ramirez J, Khedher L (2014) Linear intensity normalization of DaTSCAN images using mean square error and a model-based clustering approach. Stud Health Technol Inform 207:251–260

16. Brahim A, Ramirez J, Gorriz JM et al (2015) Comparison between different intensity normalization methods in variational mixture models with gamma or inverse-gamma components—ioflupane imaging for the automatic detection of parkinsonism. PLoS One 10:e0130274

17. Salas-Gonzalez D, Gorriz JM, Ramirez J, Illan IA, Lang EW (2013) Linear intensity normalization of FP-CIT SPECT brain images using the alpha-stable distribution. NeuroImage 65:449–455