



**HAL**  
open science

## **SPEQTACLE: An automated generalized fuzzy C-means algorithm for tumor delineation in PET**

Jérôme Lapuyade-Lahorgue, Dimitris Visvikis, Olivier Pradier, Catherine Cheze Le Rest, Mathieu Hatt

### ► To cite this version:

Jérôme Lapuyade-Lahorgue, Dimitris Visvikis, Olivier Pradier, Catherine Cheze Le Rest, Mathieu Hatt. SPEQTACLE: An automated generalized fuzzy C-means algorithm for tumor delineation in PET. *Journal of Medical Physics*, 2015, 42 (10), 10.1118/1.4929561 . inserm-01203004

**HAL Id: inserm-01203004**

**<https://inserm.hal.science/inserm-01203004>**

Submitted on 22 Sep 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SPEQTACLE: an automated generalized fuzzy C-means algorithm for tumor delineation in PET

---

Jérôme Lapuyade-Lahorgue<sup>1</sup>, Dimitris Visvikis<sup>1</sup>, Olivier Pradier<sup>1,2</sup>, Catherine Cheze Le Rest<sup>3</sup>,  
Mathieu Hatt<sup>1</sup>

<sup>1</sup>LaTIM, INSERM, UMR 1101, Brest, France

<sup>2</sup> Radiotherapy department, CHRU Morvan, Brest, France

<sup>3</sup>DACTIM, Nuclear medicine department, CHU Milétrie, Poitiers, France

1

2 **Corresponding author:** M. Hatt

3 INSERM, UMR 1101,LaTIM

4 CHRU Morvan, 2 avenue Foch

5 29609 Cedex, Brest, France

6 Tel: +33(0)2.98.01.81.11

7 Fax: +33(0)2.98.01.81.24

8 E-mail: [hatt@univ-brest.fr](mailto:hatt@univ-brest.fr)

9

10 **Wordcount:** ~9400 (8900 for main body, 540 for appendix, 255 for abstract)

11 **Disclosure of Conflicts of Interest:** No potential conflicts of interest were disclosed.

12 **Funding:** No specific funding.

13 **Abstract**

14 Purpose: accurate tumordelineation in PET images is crucial in oncology. Although recent  
15 methods achieved good results, there is still room for improvement regarding tumors with  
16 complex shapes, low signal-to-noise ratio and high levels of uptake heterogeneity.

17 Methods: We developed and evaluated an original clustering-based method called  
18 SPEQTACLE (Spatial Positron Emission Quantification of Tumor - AutomatiCLp-norm  
19 Estimation), based onthe fuzzy C-means (FCM)algorithm with a generalization exploiting a  
20 Hilbertian norm to more accurately account for the fuzzy and non-Gaussian distributions of  
21 PET images.An automatic and reproducibleastimation scheme of the norm on an image-by-  
22 image basis wasdeveloped.Robustness was assessed by studying the consistency of results  
23 obtained on multiple acquisitions of the NEMA phantom on three different scanners with  
24 varying acquisitions parameters. Accuracy was evaluatedusing classification errors (CE)  
25 onsimulated and clinical images. SPEQTACLE was compared to another FCM  
26 implementation (FLICM)and FLAB.

27 Results: SPEQTACLE demonstrateda level of robustness similar to FLAB (variability of  
28  $14\pm 9\%$  vs.  $14\pm 7\%$ ,  $p=0.15$ ) and higher than FLICM ( $45\pm 18\%$ ,  $p<0.0001$ ), and improved  
29 accuracy with lower CE( $14\pm 11\%$ ) over bothFLICM ( $29\pm 29\%$ ) and FLAB ( $22\pm 20\%$ ) on  
30 simulated images. Improvement was significant for the more challenging cases with CE of  
31  $17\pm 11\%$  for SPEQTACLEvs.  $28\pm 22\%$  for FLAB ( $p=0.009$ ) and  $40\pm 35\%$  for FLICM  
32 ( $p<0.0001$ ). For the clinical cases, SPEQTACLE outperformed FLAB and FLICM ( $15\pm 6\%$  vs.  
33  $37\pm 14\%$  and  $30\pm 17\%$ ,  $p<0.004$ ).

34 Conclusions: SPEQTACLE benefitted from the fully automatic estimation of the norm on a  
35 case-by-case basis.This promising approach will be extended to multimodal imagesand  
36 multi-class estimation in future developments.

37 **Keywords:** PET segmentation - clustering methods -Fuzzy C-means-Hilbertian norm.

38

## 39 **Introduction**

40 Positron Emission Tomography (PET) is established as a powerful tool in numerous  
41 oncology applications<sup>1</sup>, including target definition in radiotherapy planning<sup>2</sup>, and therapy  
42 monitoring<sup>3, 4</sup>, two applications for which tumor delineation is an important step, allowing for  
43 instance further quantification of PET images such as the extraction of image based  
44 biomarkers<sup>5-7</sup>. Within this context, automatic 3D functional volume delineation presents a  
45 number of advantages relative to manual delineation which is tedious, time-consuming and  
46 suffers from low reproducibility<sup>8</sup>. PET imaging is characterized by lower spatial resolution (~4-  
47 5mm 3D full width at half maximum (FWHM)) and signal-to-noise ratio (SNR) compared to  
48 other medical imaging techniques such as Magnetic Resonance Imaging (MRI) or Computed  
49 Tomography (CT). In addition, the existing large variability in scanner models and associated  
50 reconstruction algorithms (and their parameterization) leads to PET images with varying  
51 properties of textured noise, contrast, resolution and definition in clinical routine practice,  
52 which becomes a critical issue in multi-centric clinical trials<sup>9</sup>. Thus, automatic, repeatable and  
53 accurate, but also robust segmentation of tumor volumes is still challenging. Many methods  
54 based on various image segmentation paradigms, including but not limited to fixed and  
55 adaptive thresholding, active contours and deformable models, region growing, statistical  
56 and Markovian models, watershed transform and gradient, textural features classification,  
57 and fuzzy clustering, have been already proposed<sup>10, 11</sup>. Despite the recent improvements and  
58 the high level of accuracy and robustness achieved by some of these state-of-the-art  
59 methods, there is still room for improvement, especially regarding the delineation of tumors  
60 with complex shapes, high level of uptake heterogeneity, and/or low image SNR.

61 Methods including clustering and Bayesian estimation have demonstrated promising  
62 performance in PET tumor volume segmentation.

63 On the one hand, in Bayesian segmentation methods, statistical distributions (also called  
64 noise distributions) of the intensities are modeled by summarizing the histogram of the

65 images considering a reduced number of parameters to estimate. These methods provide  
66 automatic algorithms allowing noise modeling and *prior* solution selection, which allows them  
67 in turn to be less sensitive to noise than other segmentation approaches due to their  
68 statistical modeling<sup>12</sup>. Bayesian segmentation methods can be viewed as regularized “blind”  
69 statistical approaches in which the prior probability constraints the solution. This prior  
70 distribution can be defined in different ways according to the targeted application, for  
71 instance using hidden Markov field or chain models where the *prior* distribution is a Markov  
72 field distribution<sup>13</sup>. Relatively recent examples of such methods specifically developed for  
73 PET include Fuzzy Hidden Markov Chains (FHMC) and the Fuzzy Locally Adaptive Bayesian  
74 (FLAB) methods. In FHMC, the prior distribution was modeled using fuzzy hidden Markov  
75 chains<sup>14</sup>, whereas in FLAB the 3D neighborhood of a given voxel was used to locally  
76 estimate the fuzzy measure for each voxel<sup>15, 16</sup>, leading to a more accurate segmentation of  
77 small structures. FLAB can be considered to be one of the state-of-the-art methods for PET,  
78 according to its wide success due to its robustness, its repeatability and its overall accuracy  
79 demonstrated on both simulated and various clinical datasets including radiotracers of  
80 hypoxia and cellular proliferation<sup>8, 16–23</sup>.

81 On the other hand, clustering methods aim at partitioning the images into clusters depending  
82 on the statistical properties of the voxel intensities. The main interest of clustering methods  
83 compared to Bayesian methods lies in their low computational cost, as well as easier  
84 parameters estimation and overall implementation. The most known and used clustering  
85 method is the K-means clustering which has been extended to Fuzzy C-means clustering  
86 (FCM) by considering a fuzzy instead of a deterministic measure on the cluster’s  
87 membership. The Fuzzy C-means (FCM) algorithm has several advantages including  
88 flexibility and low computational cost. However, it fails to correctly address non-Gaussian  
89 noise, geometrical differences between clusters, spatial dependency between voxels, as well  
90 as the variability of fuzziness and noise properties or textures of the PET images that arise

91 from the large range of PET image reconstruction algorithms and post-reconstruction filtering  
92 schemes currently used in clinical practice.

93 Regarding FCM more specifically, amongst the other different generalizations of FCM, some  
94 incorporate a more accurate description of the clusters' geometry in the data model, for  
95 example by replacing the Euclidian norm by the Mahalanobis distance<sup>24</sup>. This method  
96 requires estimating the covariance matrices of each cluster additionally to the centers of the  
97 clusters and therefore takes into account that the clusters are not necessarily of identical  
98 sizes. Another version uses the Lebesgue  $l^1$  and  $l^\infty$  norms instead of the Euclidian norm<sup>25</sup>.  
99 Other authors have proposed to replace this Euclidian norm by a Hilbertian kernel<sup>26</sup>, which is  
100 more reliable in cases where the data does not follow a Gaussian mixture model. Finally,  
101 other authors have replaced the probability measure by evidential measure as in the  
102 "possibilistic" FCM<sup>27</sup>. This last approach is interesting within the context of evidential theory,  
103 however the way hard decision is carried out is heuristic and difficult to justify<sup>28</sup>. Amongst the  
104 methods exploiting the spatial information, it was proposed to generalize FCM by introducing  
105 spatial constraints to regularize it<sup>29</sup>. Other methods, such as the Fuzzy Local Information C-  
106 Means (FLICM) algorithm, incorporate in the minimization criteria the distance between  
107 voxels<sup>30</sup>.

108 The goal of this work was to focus on FCM and to propose a novel generalization in order to  
109 improve on the accuracy without sacrificing on robustness of PET tumor segmentation  
110 results compared to current state-of-the-art techniques, for challenging heterogeneous  
111 tumours. We have chosen to generalize FCM using a Hilbertian kernel, with the norm  
112 parameter not set empirically or *a priori* but rather estimated on an image-by-image basis,  
113 using a fully automatic scheme based on a likelihood maximization algorithm. The new  
114 algorithm was compared to FLICM and FLAB in terms of robustness and accuracy on real  
115 and simulated PET image datasets.

## 116 **Materials and methods**

### 117 **A. FCM algorithm and its extensions**

#### 118 *Classical FCM algorithm*

119 The FCM algorithm consists in finding for each class  $i \in \{1, \dots, C\}$ , where  $C$  is the number of  
120 classes, and for each voxel  $u \in V$  of the finite set of voxels  $V \subset \mathfrak{R}^3$ , the centers  $\mu_i \in \mathfrak{R}$  and  
121 the degrees of belief  $p_{u,i} \in [0,1]$  minimizing the criterion:

$$122 \sum_{u \in V} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^2 \quad (1)$$

123 under the constraint:  $\sum_{i=1}^C p_{u,i}^m = 1$ ,

124 where  $y_u$  is the observed intensity for the voxel  $u$  and the parameter  $m > 1$  controls the  
125 fuzzy behavior and is usually chosen as  $m = 2$ .

126 The details regarding this minimization are provided in [appendix A](#).

127 Regarding the segmentation, for each voxel  $u \in V$ , the class  $i \in \{1, \dots, C\}$  maximizing the  
128 probability  $p_{u,i}$  is chosen. This decision step is the same for the generalized FCM (GFCM).

#### 129 *FCM as a Bayesian inference method*

130 The traditional “hard” K-means clustering is equivalent to a Bayesian method where the  
131 observations are modeled as a Gaussian mixture. FCM clustering can also be rewritten in  
132 order to highlight a *prior* distribution regarding the parameters  $p_{u,i}$  and  $\mu_i$ , and a likelihood  
133 associating observations with the parameters. This idea has already been exploited by  
134 choosing *prior* distributions to optimize the estimation<sup>31</sup>. The minimization of eq. (1) is  
135 equivalent to the maximization of:

136  $P = f\left(\sum_{u \in V} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^2\right)$ , where  $f$  is a positive function such that  $P$  is a probability

137 density according to the observed variables  $(y_u)_{u \in V}$  called "likelihood". From statistics, the

138 maximization of  $P$  is equivalent to a likelihood maximization and is exhaustive (*i.e.* uses the

139 entire information of the sample) if the density of  $(y_u)_{u \in V}$  maximizes the Shannon entropy.

140 Moreover, one can show that a distribution whose form is given by

141  $P = f\left(\sum_{u \in V} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^2\right)$  is an elliptical distribution (*i.e.* isodensities are ellipsoid) with

142 center  $\frac{\sum_{i=1}^C p_{u,i}^m \mu_i}{\sum_{i=1}^C p_{u,i}^m}$  and dispersion given by  $\frac{1}{\sqrt{\sum_{i=1}^C p_{u,i}^m}}$ . An elliptical distribution is entirely

143 determined by its functional parameter  $f$ , its center and its dispersion. Amongst the elliptical

144 distributions with the same center and dispersion, one can show that the maximum entropy is

145 reached if  $f$  is an exponential function.

146 Consequently, in this case, the minimization of eq. (1) is equivalent to the maximization of:

$$\begin{aligned}
 147 \quad P &= \exp\left(-\frac{1}{2} \sum_{u \in V} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^2\right) \\
 &= \prod_{u \in V} \exp\left(-\frac{1}{2} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^2\right) \quad (2)
 \end{aligned}$$

148 Also:

$$\begin{aligned}
 &\sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^2 \\
 149 \quad &= \left(\sum_{i=1}^C p_{u,i}^m\right) \times \left(y_u^2 - 2 \frac{\sum_{i=1}^C p_{u,i}^m \mu_i y_u}{\sum_{i=1}^C p_{u,i}^m} + \frac{\sum_{i=1}^C p_{u,i}^m \mu_i^2}{\sum_{i=1}^C p_{u,i}^m}\right).
 \end{aligned}$$

150 Consequently, conditionally to the parameters, the observations  $y_u$  are independent and  
 151 Gaussian distributed as:

$$152 \quad p(y_u | (\mu_i)_{1 \leq i \leq C}, (p_{u,i})_{1 \leq i \leq C}) \\
 := N \left( \frac{\sum_{i=1}^C p_{u,i}^m \mu_i}{\sum_{i=1}^C p_{u,i}^m}, \frac{1}{\sqrt{\sum_{i=1}^C p_{u,i}^m}} \right) \quad (3)$$

153 Whereas the *prior* distribution for parameters is given by:

$$154 \quad p((\mu_i)_{1 \leq i \leq C}, (p_{u,i})_{1 \leq i \leq C, u \in V}) \propto \\
 \prod_{u \in V} \frac{1}{\sqrt{\sum_{i=1}^C p_{u,i}^m}} \exp \left[ -\frac{1}{2} \left( \sum_{i=1}^C p_{u,i}^m \mu_i^2 - \frac{\left( \sum_{i=1}^C p_{u,i}^m \mu_i \right)^2}{\sum_{i=1}^C p_{u,i}^m} \right) \right]$$

### 155 *Drawbacks of the classical FCM*

156 The previous theory results in two major drawbacks:

- 157 (a). FCM clustering is equivalent to a maximum *posterior* estimation when the  
 158 observations follow a Gaussian distribution conditionally to the parameters.  
 159 Consequently, FCM leads to inaccurate estimation when the data are not Gaussian.
- 160 (b). Similarly, FCM clustering assumes that the observations are independent  
 161 conditionally to the parameters, leading to inaccurate segmentation in the presence of  
 162 spatial dependencies.

### 163 **B. SPEQTACLE algorithm: an automatic Generalized FCM algorithm (GFCM)**

164 In this work we investigated the advantage of generalizing FCM by considering the Hilbertian  
 165  $l^p$ -norm instead of the Euclidian norm and providing an associated scheme that enables a  
 166 fully automated estimation of the norm parameter for optimal delineation on a case-by-case  
 167 basis, in order to reduce user interaction and avoid empirical optimization. Indeed, a user-

168 defined choice of the norm parameter based on visual analysis seems challenging because  
 169 of its non-intuitive nature, and would suffer from low reproducibility. An alternative would be  
 170 to optimize empirically the norm value on a training dataset, although it is unlikely that a  
 171 single norm value would be appropriate for all cases. We have consequently developed an  
 172 approach to automatically estimate the norm value for each image.

173 The proposed algorithm is called Spatial Positron Emission Quantification of Tumor  
 174 volume:AutomatiCL<sup>P</sup>-norm Estimation (SPEQTACLE).

175 *Principle of GFCM algorithm*

176 In the GFCM algorithm, the minimization criterion becomes:

$$177 \sum_{u \in V} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^\alpha \quad (4)$$

178 where, the norm parameter  $\alpha > 1$  and with no solution for  $\alpha = 1$ . Moreover, the cluster centers  
 179  $\mu_i$  cannot be estimated explicitly when  $\alpha \neq 2$ , whereas  $\alpha = 2$  corresponds to the standard  
 180 FCM. When  $\alpha > 2$  and  $\alpha < 2$ , the centers are computed using the Newton-Raphson  
 181 algorithm and gradient descent respectively (for details we refer the reader to [Appendix B.](#)  
 182 [and C.](#)).

183 *Generalized Gaussian distribution*

184 We assume that conditionally on the parameters  $(\mu_i)_{1 \leq i \leq C}$  and  $(p_{u,i})_{1 \leq i \leq C}$  the observation is  
 185 approximately distributed as a generalized Gaussian distribution whose density is

186  $y \rightarrow \frac{\alpha}{2\sigma\Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\frac{|y-\mu|^\alpha}{\sigma^\alpha}\right)$  parameterized by a center  $\mu = \frac{\sum_{i=1}^c p_{u,i}^m \mu_i}{\sum_{i=1}^c p_{u,i}^m}$ , a dispersion  $\sigma =$

187  $\frac{1}{\left(\sum_{i=1}^c p_{u,i}^m\right)^{\frac{1}{\alpha}}}$  and a shape  $\alpha$ .

188 *Estimation of the norm*

189 The estimation technique presented in the next section is based on the above generalized  
 190 Gaussian distribution. Contrary to the Gaussian case, it is only an approximation; indeed the  
 191 expression (4) can be expressed as a product of  $\sum_{i=1}^c p_{u,i}^m$  and a term of form  $|y_u - \beta|^\alpha$  only in  
 192 the Gaussian case, which corresponds to  $\alpha=2$ . However, it becomes a generalized  
 193 Gaussian distribution if  $p_{u,i} = 1$  holds for only one class. This approximation is valid as long  
 194 as the probabilities ( $p_{u,i}$ ) are not too far from the configuration  $p_{u,i} = 1$ . Consequently, the  
 195 norm parameter has to be estimated from an area for which one can consider that  $p_{u,i} = 1$   
 196 holds. In practice, this area was automatically selected using a background subtraction  
 197 method in order to provide a first guess of the tumor region, as recently proposed<sup>32</sup>. In order  
 198 to simplify the estimation task, we have chosen to estimate the norm for this background-  
 199 subtracted region, which is likely to correspond to a first estimation of the tumor region, and  
 200 set a single norm parameter value for all classes.

201 The next step involves the estimation of the different parameters using likelihood  
 202 maximization.

203 Let us denote  $\mu = \frac{\sum_{i=1}^c p_{u,i}^m \mu_i}{\sum_{i=1}^c p_{u,i}^m}$  and  $\sigma = \frac{1}{\left(\sum_{i=1}^c p_{u,i}^m\right)^{\frac{1}{\alpha}}}$ .

204 First, one can assume that these values do not depend on  $u$  and secondly, that the  
 205 distribution of the observations  $y_u$  in the selected area is approximately the generalized  
 206 Gaussian distribution. Let  $(y_u)_{u \in W}$  be the sample from the selected area  $W$ , the maximum  
 207 likelihood estimators of  $\mu$ ,  $\sigma$  and  $\alpha$ , denoted  $\hat{\mu}_{ML}$ ,  $\hat{\sigma}_{ML}$  and  $\hat{\alpha}_{ML}$  are solutions of the system:

208 a. 
$$\sum_{u \in W} \text{sgn}(y_u - \hat{\mu}_{ML}) |y_u - \hat{\mu}_{ML}|^{\hat{\alpha}_{ML}-1} = 0 ;$$

209 b. 
$$\hat{\sigma}_{ML}^{\hat{\alpha}_{ML}} = \hat{\alpha}_{ML} \times \frac{1}{|W|} \sum_{u \in W} |y_u - \hat{\mu}_{ML}|^{\hat{\alpha}_{ML}} ;$$

210 c. 
$$\frac{\hat{\alpha}_{ML} + \psi\left(\frac{1}{\hat{\alpha}_{ML}}\right)}{\hat{\alpha}_{ML}^2} = \frac{1}{|W|} \sum_{u \in W} \log\left(\frac{|y_u - \hat{\mu}_{ML}|}{\hat{\sigma}_{ML}}\right) \frac{|y_u - \hat{\mu}_{ML}|^{\hat{\alpha}_{ML}}}{\hat{\sigma}_{ML}^{\hat{\alpha}_{ML}}},$$

211 where,  $|W|$  is the cardinality of  $W$  and  $\psi$  is the log-derivative of the Eulerian function (see  
 212 [AppendixD](#)).

213 These equations are not linear and cannot be solved independently. Consequently, the  
 214 solution is estimated by using a combination of a variational method and the Newton-  
 215 Raphson algorithm as outlined below:

216 1. Let  $\mu^{(0)}$ ,  $\sigma^{(0)}$  and  $\alpha^{(0)}$  be the initial values ;

217 2. From  $\alpha^{(p)}$ , compute  $\mu^{(p+1)}$  by solving  $\sum_{u \in W} \text{sgn}(y_u - \mu^{(p+1)}) |y_u - \mu^{(p+1)}|^{\alpha^{(p)}} = 0$  using the

218 Newton-Raphson algorithm;

219 3. From  $\alpha^{(p)}$  and  $\mu^{(p+1)}$ , compute  $\sigma^{(p+1)} = \alpha^{(p)} \times \frac{1}{|W|} \sum_{u \in W} |y_u - \mu^{(p+1)}|^{\alpha^{(p)}} ;$

220 4. From  $\mu^{(p+1)}$  and  $\sigma^{(p+1)}$ , compute  $\alpha^{(p+1)}$  by solving

$$\frac{\alpha^{(p+1)} + \psi\left(\frac{1}{\alpha^{(p+1)}}\right)}{\alpha^{(p+1)}} =$$

$$\frac{1}{|W|} \sum_{u \in W} \log \left( \frac{|y_u - \mu^{(p+1)}|}{\sigma^{(p+1)}} \right) \frac{|y_u - \mu^{(p+1)}|^{\alpha^{(p+1)}}}{(\sigma^{(p+1)})^{\alpha^{(p+1)}}}$$

222 5. Repeat steps 2, 3 and 4 until convergence.

223 Although such generalized Gaussian distributions have properties that allow for convergence  
 224 of the maximum likelihood estimation, the stopping criteria has to be defined. One could  
 225 assume that the estimation can be stopped when the successive values of  $\alpha^{(p)}$  (resp.  $\mu^{(p)}$   
 226 and  $\sigma^{(p)}$ ) are sufficiently close to each other, using the absolute distances as stopping  
 227 criteria. However, the values of the parameters can be close, whereas the distance between  
 228 the resulting distributions may be large. Indeed, the smaller  $\sigma$  is, the more sensitive to the  
 229 value of  $\mu$  is the resulting density. To overcome this drawback, we used a more appropriate  
 230 distance; namely the distance between distributions rather than the distance between  
 231 parameters' values. This distance is defined from the Fisher information matrix ([Appendix E](#)).

232 It has been previously shown that the set of given parameterized distributions is a  
 233 Riemannian manifold whose metric tensor is given by the Fisher information matrix<sup>33</sup>. More  
 234 precisely, let  $\Lambda = \{y \rightarrow p(y|\theta) : \theta \in \Theta\}$  be a smooth manifold of statistical distributions  
 235 parameterized by an open set  $\Theta \subset \mathfrak{R}^k$ , the distance between "close" distributions  
 236  $y \rightarrow p(y|\theta)$  and  $y \rightarrow p(y|\theta + d\theta)$  is given by:

237  $dl = \sqrt{(d\theta)^* I(\theta) d\theta}$ , where  $I(\theta)$  is the Fisher information matrix and  $(d\theta)^*$  is the transpose  
 238 of the vector  $d\theta$ .

239 For the generalized Gaussian random variables that we use in SPEQTACLE, the Fisher  
 240 information relative to the position parameter  $\mu$ , the dispersion parameter  $\sigma$  and the norm  
 241 parameter  $\alpha$  are given respectively by:

$$242 \quad I(\mu) = \alpha(\alpha - 1) \frac{\Gamma\left(\frac{\alpha - 1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)}$$

$$I(\alpha) = 1 + \frac{1}{\alpha^2}$$

$$243 \quad I(\sigma) = \frac{\alpha}{\sigma^2} \text{ and } + \frac{2(1 + \alpha)}{\alpha^3} \psi\left(\frac{1}{\alpha}\right) + \frac{1}{\alpha^3} \psi\left(\frac{1}{\alpha}\right)^2, \\ + \frac{1 + \alpha}{\alpha^4} \psi\left(\frac{1}{\alpha}\right)$$

244 where,  $\Gamma$  and  $\psi$  are the Eulerian function and its log-derivative respectively. In the norm  
 245 estimation algorithm, we evaluate the distance between distributions twice; namely when  
 246  $\sigma^{(p)}$  and  $\alpha^{(p)}$  are recomputed. It maybe also possible to evaluate the distance when  $\mu^{(p)}$  is  
 247 recomputed. However, if the two other parameter sequences  $\sigma^{(p)}$  and  $\alpha^{(p)}$  do not vary, one  
 248 can reliably assume that the parameter sequence  $\mu^{(p)}$  does not vary either. As the Fisher  
 249 information relative to  $\sigma$  is given by  $I(\sigma) = \frac{\alpha}{\sigma^2}$ , for fixed values of  $\alpha$  and  $\mu$  the infinitesimal

250 distance between two generalized Gaussian distributions  $p(y|\alpha, \mu, \sigma)$  and

251  $p(y|\alpha, \mu, \sigma + d\sigma)$  is  $\frac{\sqrt{\alpha}}{\sigma} d\sigma$  and the distance between  $p(y|\alpha, \mu, \sigma^{(p)})$  and

252  $p(y|\alpha, \mu, \sigma^{(p+1)})$  is given by:

$$253 \quad D(\sigma^{(p)}, \sigma^{(p+1)}) = \sqrt{\alpha} \left| \int_{\sigma^{(p)}}^{\sigma^{(p+1)}} \frac{d\sigma}{\sigma} \right| \\ = \sqrt{\alpha} \left| \log\left(\frac{\sigma^{(p+1)}}{\sigma^{(p)}}\right) \right|$$

254 Regarding the parameter  $\alpha$ , the integration of the Fisher metric is not explicit and requires  
 255 time consuming numerical methods. We have used the Kullback-information “metric” instead,  
 256 as a good approximation of the Fisher metric when the consecutive values of  $\alpha^{(p)}$  are close  
 257 (Appendix E). When  $\mu$  and  $\sigma$  are set, the Kullback information from  $p(y|\alpha^{(p)}, \mu, \sigma)$  to  
 258  $p(y|\alpha^{(p+1)}, \mu, \sigma)$  is given by:

$$259 \quad K(\alpha^{(p+1)} : \alpha^{(p)}) = \log \left( \frac{\alpha^{(p+1)} \Gamma\left(\frac{1}{\alpha^{(p)}}\right)}{\alpha^{(p)} \Gamma\left(\frac{1}{\alpha^{(p+1)}}\right)} \right) + \frac{\Gamma\left(\frac{1+\alpha^{(p)}}{\alpha^{(p+1)}}\right)}{\Gamma\left(\frac{1}{\alpha^{(p+1)}}\right)} - \frac{1}{\alpha^{(p+1)}}.$$

260 Finally, in the maximum likelihood estimation algorithm,  $D(\sigma^{(p)}, \sigma^{(p+1)})$  and  $K(\alpha^{(p+1)} : \alpha^{(p)})$   
 261 are evaluated when the value of  $\sigma^{(p+1)}$  and  $\alpha^{(p+1)}$  are respectively computed. The stopping  
 262 rule is a fixed threshold value  $\varepsilon = 10^{-7}$  small enough to ensure convergence.

### 263 **C. Algorithm evaluation methodology**

#### 264 *Repeatability and dependency of the norm estimation on initial tumor region*

265 In order to evaluate the repeatability of SPEQTACLE, the whole process (background-  
 266 subtracted area definition used to estimate the norm, followed by the iterative estimation of  
 267 the norm and the modified FCM clustering) was applied 20 times to the same tumor images.

268 In order to investigate the dependency of the estimated norm value on the background-  
 269 subtracted region, we made smaller or larger the result of this fully automated procedure<sup>32</sup>  
 270 by one to three voxels in all directions and relaunched the estimation procedure on the new  
 271 area.

#### 272 *Robustness assessment*

273 We first evaluated the robustness of the SPEQTACLE algorithm. Robustness was defined as  
 274 the ability of the automatic algorithm to provide consistent results for a given known object of

275 interest, considering varying image properties such as spatial sampling (voxel size), SNR,  
276 contrast, texture, filtering, etc. This evaluation was carried out using a dataset of phantoms  
277 containing homogeneous spheres on homogeneous background that were acquired in  
278 different PET/CT scanners, each with varying acquisition and reconstruction parameters (see  
279 section D. Datasets). Homogeneous spheres on homogeneous background are not  
280 appropriate for the evaluation of absolute accuracy since they represent a simplistic set-up  
281 and because of the bias due to cold sphere walls<sup>34, 35</sup>. On the other hand, they are well suited  
282 for the task of robustness estimations since any present bias present is the same for all  
283 acquisitions and they can provide a wide range in imaging settings for a given known object.  
284 The four spheres with largest diameters (37, 28, 22 and 17 mm) were segmented  
285 individually. The 13 and 10mm spheres were not included in the analysis because they were  
286 not filled in all acquisitions and are often too small with respect to the reconstructed voxel  
287 size to provide meaningful results.

#### 288 *Accuracy assessment*

289 To evaluate the accuracy of the new algorithm relative to that of current state-of-the-art  
290 methods more challenging cases such as relatively large, complex-shaped and/or  
291 heterogeneous tumors were used considering both simulated realistic tumors and clinical  
292 tumor cases (see section D. datasets).

#### 293 *Evaluation metrics*

294 For the robustness assessment, since the objects used are simple homogeneous spheres  
295 and the goal is to assess the consistency of results over various acquisitions of the same  
296 object and not absolute accuracy, the standard deviation of the determined volumes for a  
297 given sphere across the entire dataset (all scanners, all configurations) was reported as a  
298 measure of robustness.

299 For the accuracy evaluation, the classification errors (CE) were used. In the simulated  
300 dataset, CE were calculated relatively to the known ground truth. In the clinical datasets, CE

301 were calculated relatively to a surrogate of truth obtained through a statistical consensus  
302 using the STAPLE (Simultaneous Truth And Performance Level Estimation) algorithm<sup>36</sup>  
303 applied to three manual delineations performed by experts with similar training and  
304 experience. CE may result from two contributions: the false negatives, the number of  
305 misclassified voxels within the ground truth, and the false positives, the number of  
306 misclassified voxels outside of the ground truth. CE as a percentage is then calculated as the  
307 sum of positive and negative misclassified voxels, divided by the number of voxels defining  
308 the ground truth<sup>15</sup>. CE were reported as mean $\pm$ SD as well as with box-and-whisker plots in  
309 the figures.

### 310 *Comparison with other methods*

311 Within this evaluation framework, the proposed algorithm SPEQTACLE was compared to a  
312 couple of state-of-the-art methods which are improvement of the classical FCM: the Fuzzy  
313 Locally Adaptive Bayesian (FLAB)<sup>16</sup> and the Fuzzy Local Information C-means  
314 (FLICM)<sup>30</sup>. Because the standard FCM has already been extensively evaluated and  
315 compared to these extensions or other previous segmentation approaches, including on PET  
316 images<sup>15, 16, 37</sup>, it was not included in the present analysis.

317 FLAB combines a fuzzy measure with a Gaussian mixture model, and a stochastic  
318 estimation of the parameters from a FCM-based initialization. This method was developed  
319 initially for PET and thoroughly validated on both simulated and clinical datasets<sup>16, 17,</sup>  
320<sup>23</sup>. FLICM is a recent FCM algorithm with a weighted norm taking into account outliers due to  
321 the noise<sup>30</sup>. This method uses two parameters: a regularization parameter and the size of the  
322 surrounding kernel. In the present work, we have set the parameter regularization equal to 1  
323 and the kernel radius equal to 3 voxels, which are the recommended values<sup>30</sup> although they  
324 have not been optimized specifically for PET.

325 For all methods, the object of interest is first isolated in a 3D region of interest (ROI)  
326 containing the tumor, similarly as previously detailed for FLAB<sup>15</sup>. The number of

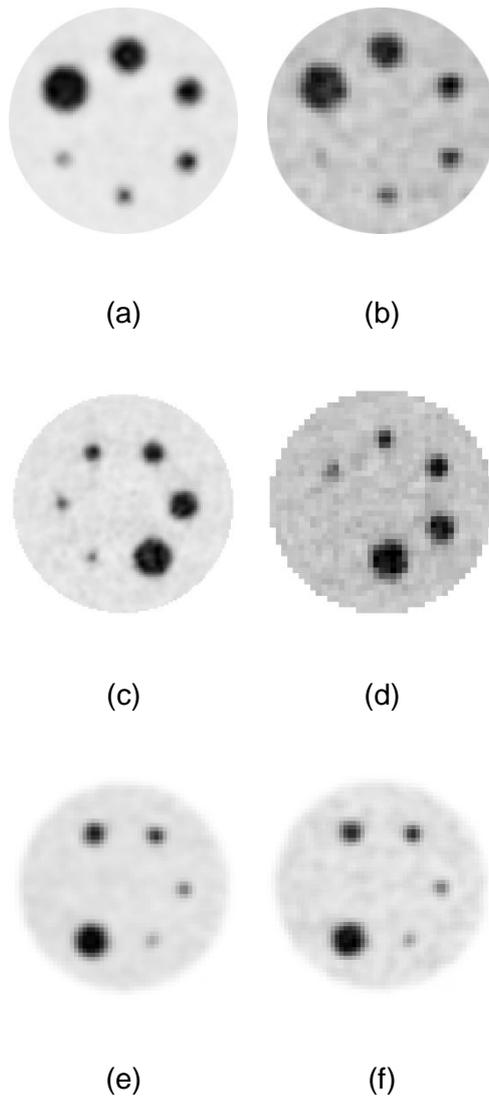
327 classes/clusters used was 2 for the robustness evaluation (homogeneous spheres) and 3 for  
328 the accuracy evaluation, in order to take into account potential tumor uptake heterogeneity.  
329 The two tumor classes were then unified for the error calculation with respect to the binary  
330 ground-truth (tumor/background). Thus, all algorithms were applied considering the same  
331 number of classes/clusters for a given image.

332 The Wilcoxon rank sum test was used to compare the results between methods. P-values  
333 below 0.05 were considered significant.

#### 334 **D. Datasets**

##### 335 *Homogeneous spheres phantoms*

336 The dataset used for the robustness evaluation consists of NEMA phantoms containing  
337 spheres of various sizes (37, 28, 22, 17, 13, 10 mm) and filled with  $^{18}\text{F}$ -FDG,  
338 that were acquired in three different PET/CT scanners: two PHILIPS scanners (a standard  
339 GEMINI and a time-of-flight (TOF) GEMINI), and a SIEMENS Biograph 16 scanner<sup>8</sup>. The  
340 standard iterative reconstruction algorithms associated with each scanner were used with  
341 their usual parameters: Time-of-Flight Maximum Likelihood-Expectation Maximization (TF  
342 ML-EM) for the GEMINI TOF, 3D Row Action Maximum Likelihood Algorithm (RAMLA) (2  
343 iterations, relaxation parameter 0.05, Gaussian post-filtering with 5mm FWHM) for the  
344 GEMINI, and Fourier rebinning (FORE) followed by Ordered Subsets Expectation  
345 Maximization (OSEM) (4 iterations, 8 subsets, Gaussian post-filtering with 5mm FWHM) for  
346 the Biograph 16. All PET images were reconstructed using CT-based attenuation correction,  
347 as well as scatter and random coincidences. For each scanner, two different values for the  
348 following acquisition parameters and reconstruction settings were considered: the contrast  
349 between the sphere and the background (4:1 and 8:1), the voxel size in the reconstruction  
350 matrix ( $2 \times 2 \times 2$  and  $4 \times 4 \times 4$  or  $5.33 \times 5.33 \times 2$  mm<sup>3</sup>) and the noise level (2 and 5 min of list mode  
351 data). Note that for the GEMINI acquisitions, the 28mm sphere was missing in the physical  
352 phantom. Figure 1 illustrates the images obtained for some of the acquisitions.



353 Fig 1. Examples of phantoms acquisitions: (a-b) the PHILIPS GEMINI TOF scanner with  
 354 5min acquisitions and (a) ratio 8:1, voxels  $2 \times 2 \times 2 \text{ mm}^3$ , (b) ratio 4:1,  $4 \times 4 \times 4 \text{ mm}^3$ . (c-d) the  
 355 SIEMENS scanner with 5min acquisitions and (c) ratio 8:1, voxels  $2 \times 2 \times 2 \text{ mm}^3$ , (d) ratio 4:1,  
 356  $5.33 \times 5.33 \times 2 \text{ mm}^3$ . (e-f) the PHILIPS GEMINI scanner with ratio 8:1, voxels  $4 \times 4 \times 4 \text{ mm}^3$ , and  
 357 (e) 5min acquisition, (f) 2 min acquisition.

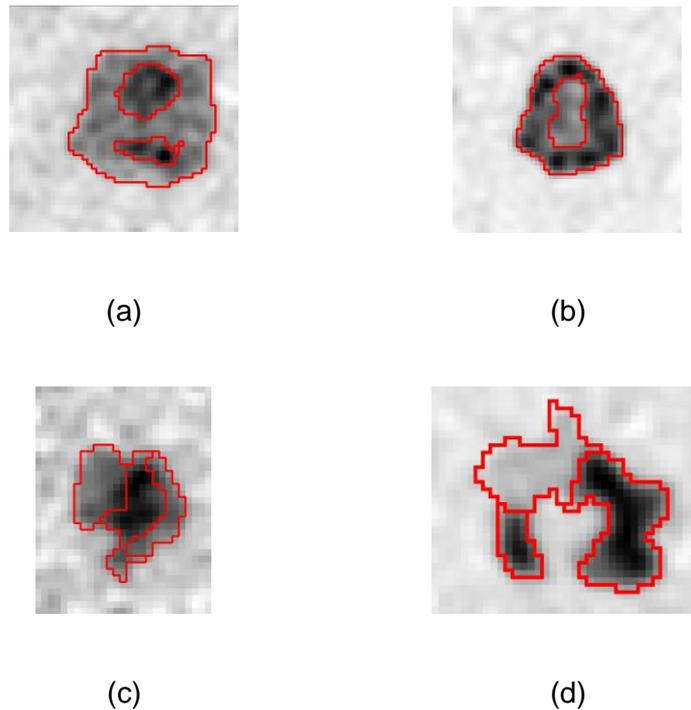
358

### 359 *Simulated PET images*

360 A set of 34 simulated PET tumor images with a wide range of contrast, noise levels, uptake  
 361 heterogeneity and shape complexity was generated following a previously described  
 362 methodology to obtain realistic complex shapes and uptake distributions of tumors for which

363 the exact ground-truth on a voxel-by-voxel basis is known<sup>38, 39</sup>. This dataset was built with  
364 relatively more challenging cases compared to previously conducted evaluations<sup>16</sup>, in order  
365 to provide more complex tumor cases with combination of low SNR, high levels of  
366 heterogeneities and complex shapes. The important steps of the procedure used to generate  
367 these images is outlined below, and the reader is referred to<sup>38, 39</sup> for more details.

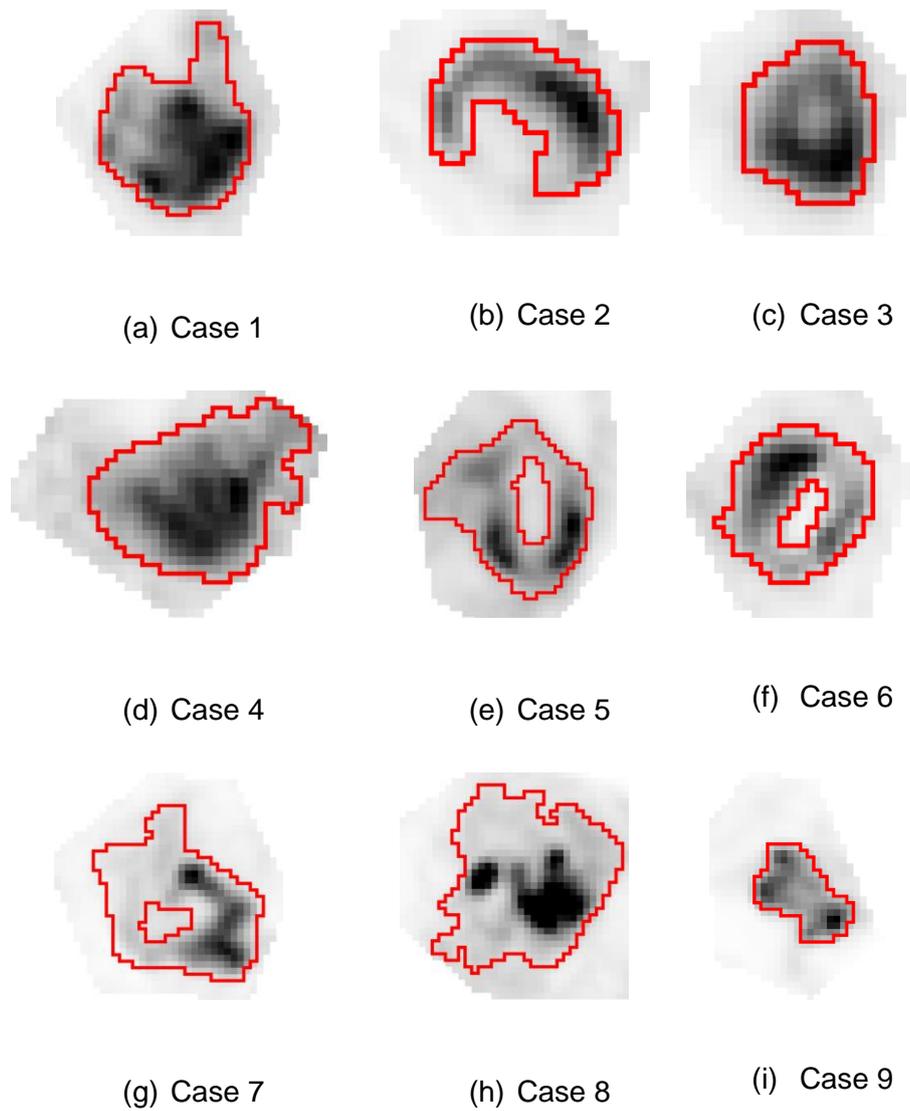
368 Each clinical tumor was first manually delineated on a clinical PET image by a nuclear  
369 medicine expert, thus creating a voxelized volume that represents the ground-truth of the  
370 tumor model used in the simulation. The activity levels attributed to each of the tumor parts  
371 were derived from the activity measured in the same areas of the tumor in the corresponding  
372 patient images. This ground-truth tumor structure was subsequently transformed into a Non-  
373 Uniform Rational B-Splines (NURBS) volume via Rhinoceros<sup>TM</sup> (CADLINK software), for  
374 insertion into the NCAT phantom<sup>40</sup> attenuation maps at the same approximate position as  
375 located in the patient. No respiratory or cardiac motions were considered. Simulations using  
376 a model of the Philips PET/CT scanner previously validated with GATE (Geant4 Application  
377 for Tomography Emission)<sup>41</sup> were carried out. A total of 45 million coincidences were  
378 simulated corresponding to the statistics of a clinical acquisition over a single axial 18 cm  
379 field of view. Images were subsequently reconstructed using the One-Pass List mode  
380 Expectation Maximization (OPL-EM) (7 iterations, 1 subset). In some cases, the same 3D  
381 tumor shape was produced with different levels of contrast and heterogeneity, voxel sizes  
382 ( $4 \times 4 \times 4$  and  $2 \times 2 \times 2$  mm<sup>3</sup>) and/or a different number of coincidences (45M or 20M) for different  
383 SNR realizations. Figure 2 illustrates some of the simulated tumors. The first two cases (fig.  
384 2a-b) present relatively simpler shapes, higher contrast and SNR, whereas fig. 2c and 2d  
385 present more complex shapes and higher levels of noise and uptake heterogeneity.



386 Fig 2. Four examples of simulated tumors. Red contours correspond to the simulation ground  
387 truth showing both external contours and sub-volumes heterogeneity.

### 388 *Clinical PET images*

389 Nine non-Small Cell Lung Cancer (NSCLC) tumors were chosen for their challenging nature  
390 with complex shapes and uptake heterogeneity. Patients fasted for at least 6 hours before 3D  
391 PET data was acquired on a Philips GEMINI PET/CT scanner without motion correction,  
392  $60 \pm 4$  min after injection of 5MBq/kg of  $^{18}\text{F}$ -FDG. Images were reconstructed with the 3D  
393 RAMLA algorithm (2 iterations, relaxation parameter 0.05, post-filtering with a Gaussian of 5  
394 mm FWHM) and a voxel size of  $4 \times 4 \times 4 \text{ mm}^3$ , using CT-based attenuation correction, scatter  
395 and random correction<sup>42</sup>. In the absence of ground-truth for these volumes, 3 different  
396 experts delineated each tumor slice-by-slice with free display settings. A statistical  
397 consensus of the segmentations was then derived using the STAPLE algorithm to generate  
398 one surrogate of truth (fig. 3).



399 Fig 3. (a-i) Clinical images of 9 NSCLC tumors. Red contours correspond to the statistical  
 400 consensus of 3 different manual delineations.

## 401 Results

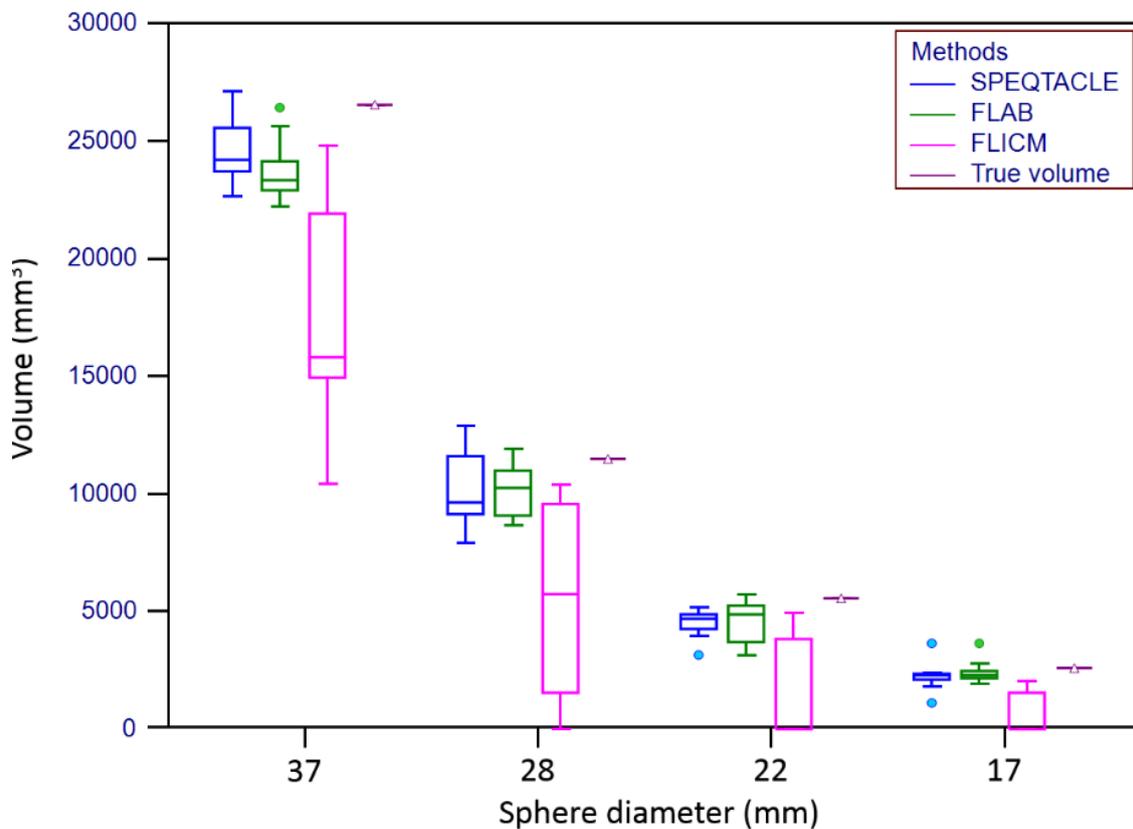
### 402 *Repeatability and dependency on initially selected tumor region*

403 The procedure was found perfectly repeatable with no variations in the resulting  
 404 segmentations on repeated applications to the same (previously defined) region of interest.  
 405 In addition, enlarging or reducing the size of the initial background-subtracted area by 1 to 3  
 406 voxels in all directions (equivalent to shrinking or increasing of the size of the region used to  
 407 estimate the norm by 5 to 15%) resulted in only minor variations in the estimated norm value

408 (3±11%, range -10% to +16%), and even smaller variations in the resulting segmentation  
409 (2±5%, range -4% to +7%). A substantial degradation of the segmentation results (20%  
410 difference) was observed when the reduction(area not covering sufficiently the tumor) or  
411 enlargement (too much background incorporated) of the initially estimated area exceeded  
412 50%.

### 413 **Robustness**

414 The robustness of FLAB and standard FCM has already been reported extensively<sup>8</sup>. In the  
415 current work we focused on three scanners and the 4 largest spheres, comparing  
416 SPEQTACLE to FLAB and FLICM. Figure 4 presents the robustness of each method,  
417 quantified by the distributions of resulting volumes for each sphere as box-and-whisker  
418 plots across the entire dataset (3 scanners, all acquisition and reconstruction parameters).  
419 Although the accuracy was not under evaluation here, the true volume was also plotted for  
420 reference.



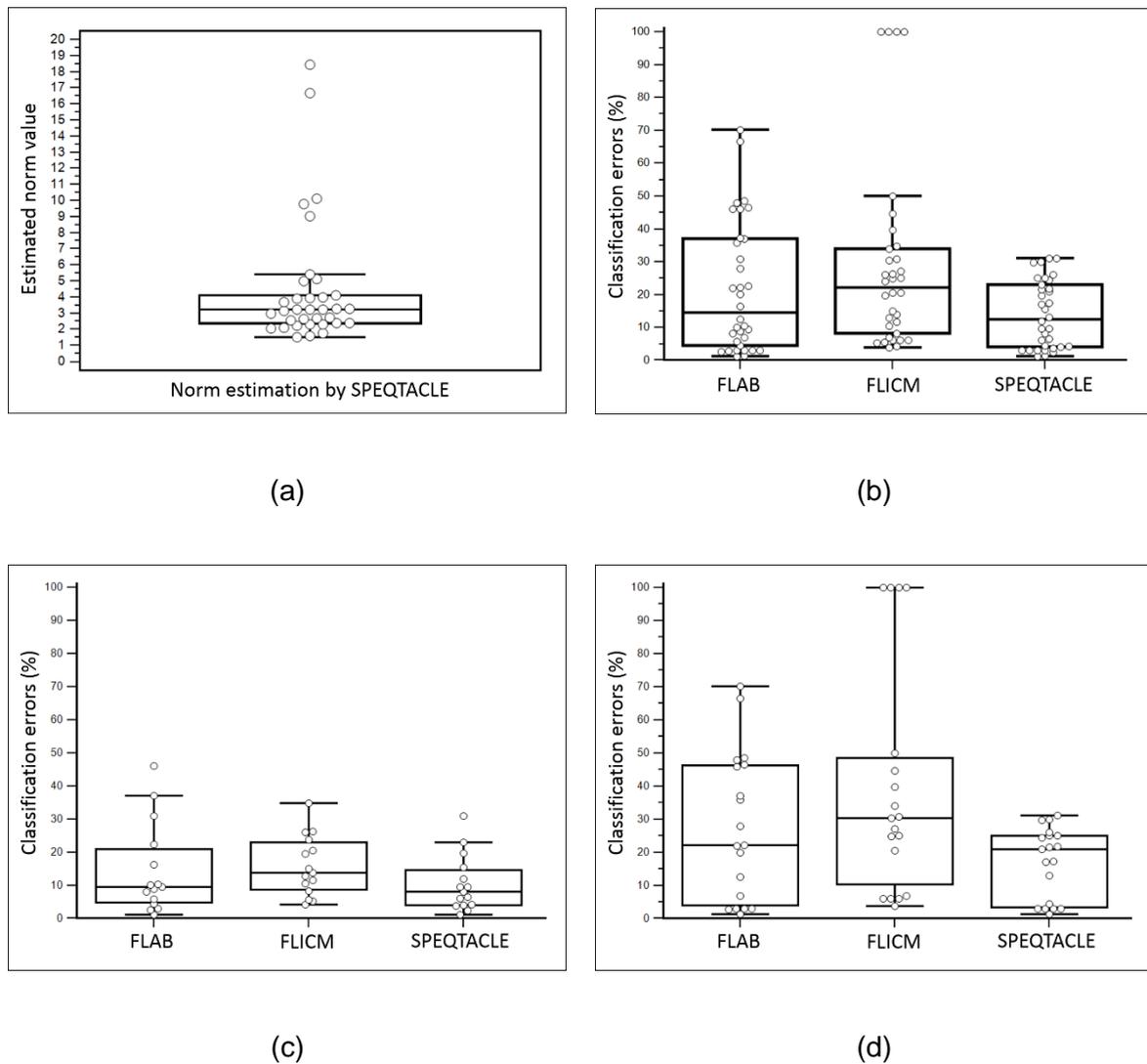
421

422 Fig 4. Distributions of volumes determined by the three methods under comparison for the  
423 four spheres of 37, 28, 22 and 17 mm in diameter across the entire robustness dataset. Box-  
424 and-whisker plots provide lower to upper quartile (25 to 75 percentile, central box), the  
425 median (middle line of the box) and the minimum to the maximum value, excluding "outlier  
426 values" which are displayed as separate dots.

427

428 The robustness performance of SPEQTACLE was satisfactory given the very large range of  
429 image characteristics. It was very similar and not statistically different ( $p=0.15$ ) from FLAB  
430 with standard deviations of 5.4%, 16.9%, 12.7% and 26.6% for SPEQTACLE vs. 5.4%,  
431 11.5%, 20.3%, and 19.3% for FLAB (for the 37, 28, 22 and 17mm spheres respectively). It  
432 should be emphasized that there were 2 outliers for the 17mm sphere and 1 for the 22 mm  
433 sphere (fig. 4). These were associated with images of some of the acquisitions for which the  
434 spheres were barely visible and spatially sampled with large voxels (see fig. 1b for an  
435 example), which explains the substantial deviation observed for these specific cases. When  
436 excluding these outliers, the robustness of SPEQTACLE increased with lower standard  
437 deviations of 7.9% and 18.8% for the 22 and 17mm sphere respectively.

438 FLICM exhibited significantly lower robustness ( $p<0.0001$ ) than FLAB and SPEQTACLE. For  
439 the spheres 28, 22 and 17 mm, this was mostly due to segmentation failures in several cases  
440 for sphere diameters  $\leq 28$  mm, with the segmentation filling the entire ROI leading to  
441 extremely large volumes. For these complete failures, we limited the resulting volume to  
442 twice the expected volume of the sphere, leading to standard deviations of 68.9%, 40.9%  
443 and 43.7% for the spheres of 28, 22 and 17mm respectively. However for the largest sphere  
444 (37 mm in diameter), the standard deviation was also higher (26.8%) than SPEQTACLE and  
445 FLAB, without an associated segmentation failure, but rather very different results depending  
446 on the different image characteristics considered.



448 Fig. 5 (a) Box-and-whisker plot of the norm parameter estimated by SPEQTACLE for the  
 449 entire set of simulated PET images. (b-d) Comparison of error rates for the three methods  
 450 with box-and-whisker plots, for (b) the 34 simulated tumors PET images, (c) the subset of  
 451 cases with estimated norm < 3 and (d) cases with norm > 3.

452 Figure 5a shows the distribution of the values for the norm parameter as estimated by  
 453 SPEQTACLE. We recall that a value of 2 corresponds to the standard FCM case. Almost half  
 454 the cases considered had an estimated norm between 3 and 6. Five cases led to estimated  
 455 norm values of 9 to 19. Given this distribution, we report the accuracy for the entire dataset,  
 456 then for the subset of cases with norm < 3 (15 cases) and finally for > 3 (19 cases), as we can

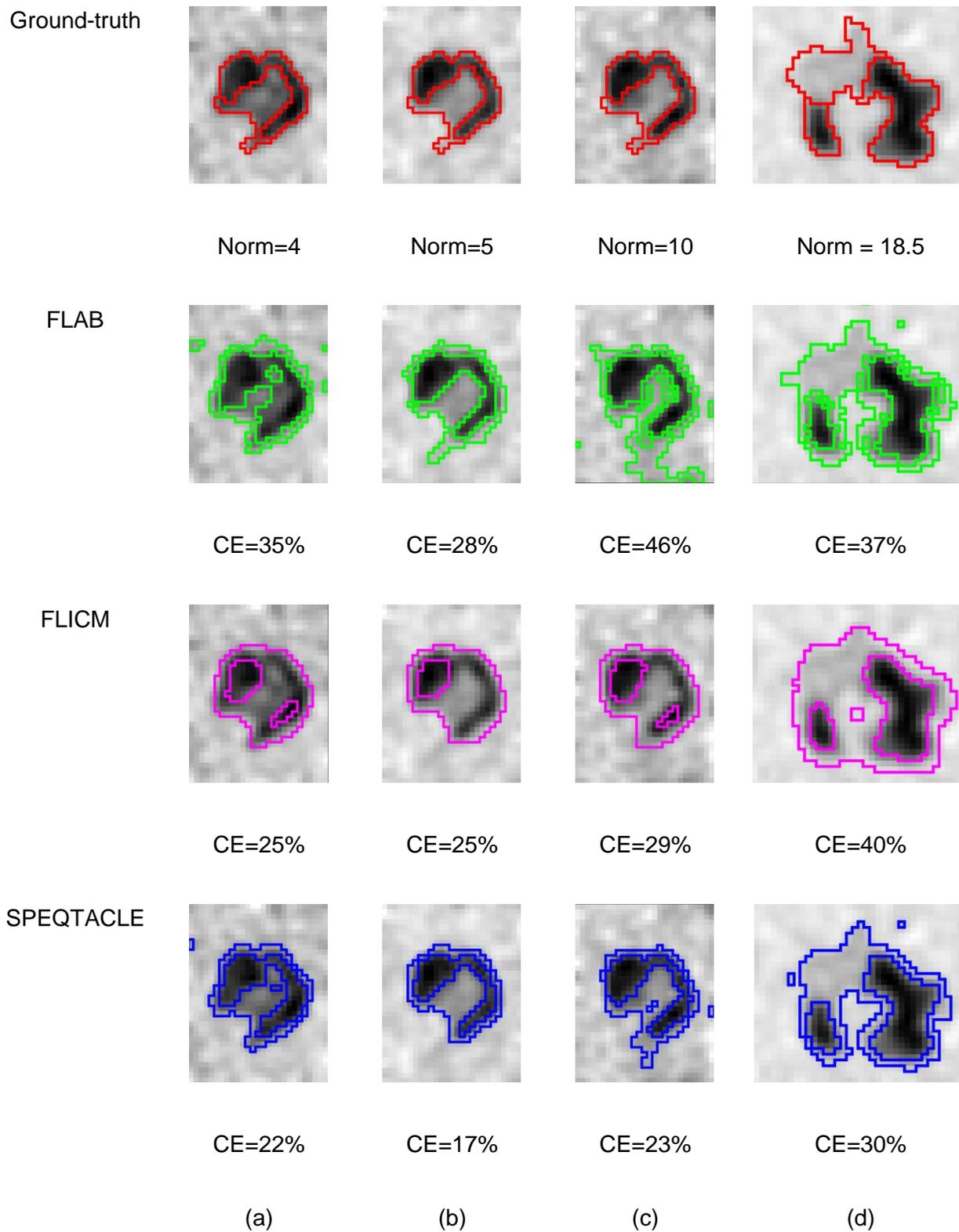
457 reasonably expect a larger improvement using SPEQTACLE over the two other algorithms  
458 for higher norm values.

459 Figure 5b shows the classification error results obtained by the three methods under  
460 comparison, for the entire set of 34 images. SPEQTACLE was found to provide lower CE  
461 than FLAB ( $p=0.0044$ ) and FLICM ( $p<0.0001$ ). FLAB, FLICM and SPEQTACLE led to CE of  
462  $21.8\pm 19.8\%$  (median 14.5%, range 1.2 – 70.2%),  $29\pm 29\%$  (median 22.3%, range 3.9 –  
463 100.0%) and  $14.4\pm 10.6\%$  (median 12.5%, range 1.3 – 37.9%) respectively. No errors above  
464 40% were observed for SPEQTACLE contrary to FLAB (up to 50-70% errors) and FLICM  
465 that even had four cases with >100% errors (complete failure of the segmentation, CE limited  
466 to 100%). SPEQTACLE had more cases with errors below 10% and between 10% and 20%  
467 than FLAB and FLICM, and fewer cases with errors between 20% and 50%.

468 Figure 5c provides the classification errors for the 15 images for which the estimated norm  
469 was  $<3$ . In this first subset, although SPEQTACLE led to the best results ( $10.5\pm 8.5\%$ , median  
470 8.3%, range 1.3 – 31%) with significantly lower errors than FLICM ( $15.3\pm 9.1\%$ , median  
471 12.9%, range 4.2 – 34.8%,  $p=0.0215$ ), no significant differences were found between  
472 SPEQTACLE and FLAB ( $14.5\pm 13.6\%$ , median 9.5%, range 1.2 – 46.1%,  $p=0.22$ ). No errors  
473 above 50% were observed for any method. It should be emphasized that despite differences  
474 between the three methods, all three achieved high accuracy performance with  $<20\%$  CE for  
475 the majority of cases.

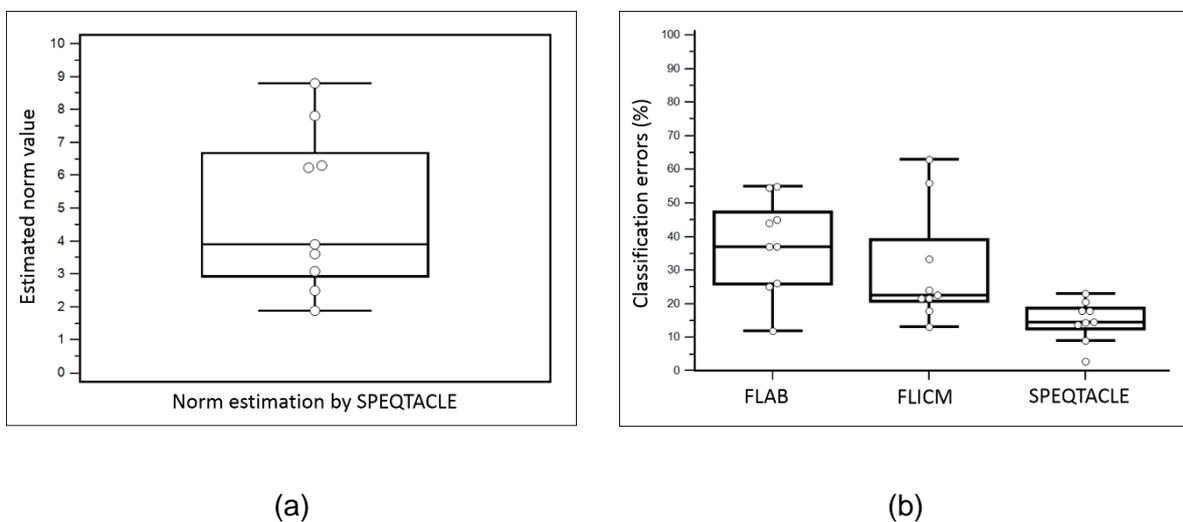
476 Figure 5d provides the classification errors for the second subset of 19 images for which the  
477 estimated norm was  $>3$ . In this dataset of clearly more challenging cases, with an error rate of  
478  $17.4\pm 11.3\%$  (median 21%, range 1.4 – 37.9%), SPEQTACLE significantly outperformed all  
479 other methods: FLAB with  $27.6\pm 22.2\%$  (median 22.2%, range 1.4 – 70.2%) ( $p=0.0092$ ) and  
480 FLICM with  $39.9\pm 34.6\%$  (median 30.5%, range 3.9 – 100.0%) ( $p<0.0001$ ). No errors above  
481 50% were observed for SPEQTACLE contrary to FLAB and FLICM, and there were less  
482 errors between 20 and 50% for SPEQTACLE than for FLAB and FLICM. Overall, the

483 accuracy achieved by SPEQTACLE in this dataset of very challenging cases was  
 484 satisfactory, with a maximum CE below 38% and a mean of 17%. Figure 6 provides some  
 485 visual examples of segmentation results for the simulated tumors.



486 Fig 6. Segmentation results for (a-c) the same simulated tumor with increasing complexity:  
 487 combinations of noise levels and heterogeneity both within the tumor (contrast between the  
 488 various sub-volumes of the tumor) or in terms of overall contrast between the tumor and the  
 489 background. These configurations were found to correspond to increasing estimated norm  
 490 values: (a) 4, (b) 5 and (c) 10. (d) presents a tumor with complex shape and high levels of  
 491 heterogeneity for which the norm was estimated at 18.45. First row is ground-truth (red)  
 492 whereas second, third and fourth rows are results from FLAB (green), FLICM (magenta) and  
 493 SPEQTACLE (blue).

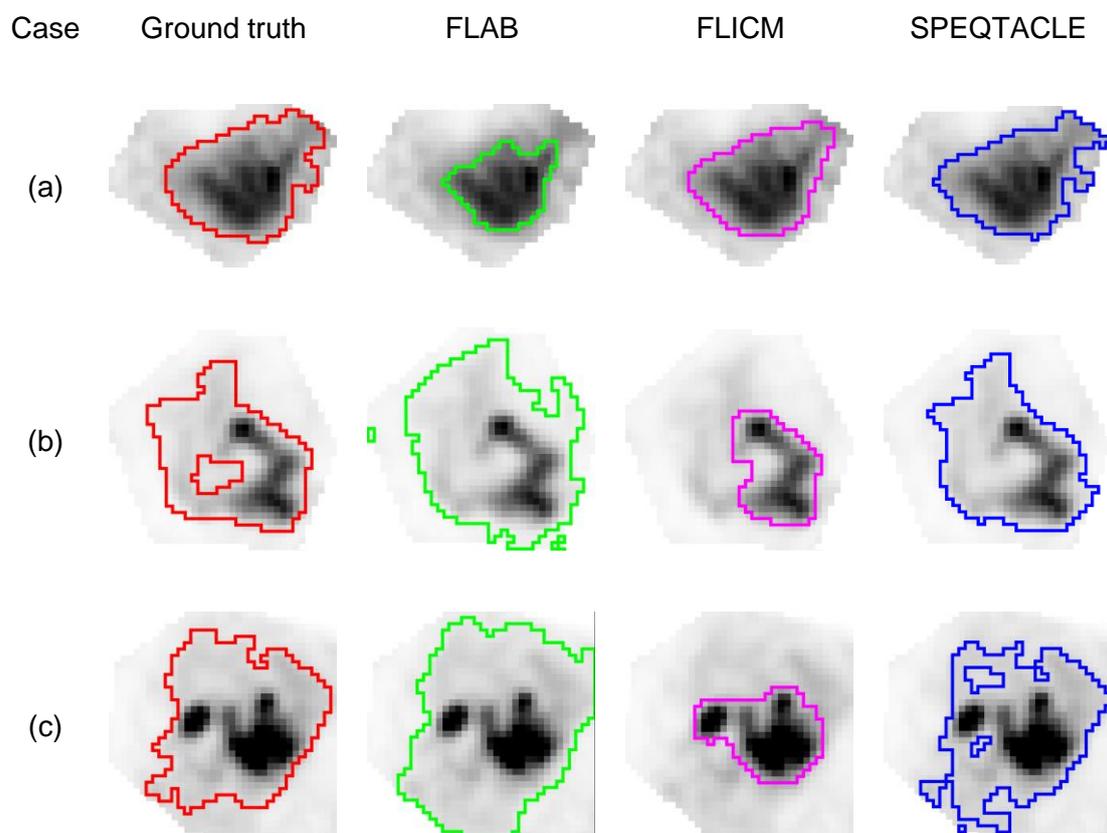
494 Figure 7 shows the estimated norm values (fig. 7a) and the classification errors (fig. 7b) for the  
 495 nine clinical images. Norm values estimated by SPEQTACLE were between 2 and 9, with  
 496 most of them being  $>3$  (7 out of 9 cases). The best performance was obtained with  
 497 SPEQTACLE with significantly ( $p < 0.004$ ) lower errors (mean  $14.9 \pm 6.1\%$ , range 2.9 – 23%)  
 498 with respect to the STAPLE-derived consensus of manual delineations, compared to FLAB  
 499 (mean  $37.3 \pm 14.3\%$ , range 12 – 55%) and FLICM ( $30.4 \pm 17.4\%$ , range 13.2 – 63%).



500 Fig 7. (a) Box-and-whisker plot of the norm parameter estimated by SPEQTACLE and (b) CE  
 501 for the three methods, for the clinical dataset.

502 Figure 8 shows the results of segmentation for all 9 clinical cases. For cases 3 and 9, the  
 503 three methods led to similar results, as the level of heterogeneity is relatively lower with

504 respect to the high overall contrast between the tumor and the surrounding background. On  
 505 the one hand, for cases 1, 4, 5 and 6, it was observed that FLAB underestimated the spatial  
 506 extent selected by the experts, by focusing on the high intensity uptake region, whereas  
 507 FLICM led to results closer to the manual contours. On the other hand, for cases 2, 7 and 8,  
 508 on the contrary FLAB slightly overestimated the manual contours, whereas FLICM  
 509 underestimated it, missing the large areas with lower uptake. In all cases, SPEQTACLE  
 510 demonstrated higher accuracy with results closer to the manual delineations.



511 Fig 8. Examples of delineations for clinical cases (a) 4, (b) 7 and (c) 8 from Fig. 3 (d), (g) and  
 512 (h): consensus of manual (red), FLAB (green), FLICM (magenta) and SPEQTACLE (blue).

### 513 Discussion

514 Although promising results for PET tumor delineation in a realistic setting beyond the  
 515 validation using simple cases (spherical and/or homogeneous uptakes) have been recently  
 516 achieved by several methods<sup>11</sup> there is still room for improvement, particularly in the case of

517 highly heterogeneous and complex shapes. The use of the fuzzy C-means clustering  
518 algorithm for delineation of PET tumors has been considered previously showing a limited  
519 performance both in accuracy<sup>15, 43, 44</sup> and robustness<sup>8</sup>. Among the recent methods dedicated  
520 to PET that demonstrated promising accuracy, the fuzzy C-means algorithm was improved  
521 using a rather complex pipeline combining spatial correlation modeling and pre-processing in  
522 the wavelet domain<sup>44</sup>. In the presented work, we rather focused on the generalization and  
523 full automation of the FCM approach to improve its accuracy and its ability to deal with  
524 challenging and complex PET tumor images, by implementing an estimation of the norm on a  
525 case-by-case basis. The improved accuracy results that we obtained on the validation  
526 datasets suggest that the optimal norm parameter can indeed be different for each PET  
527 tumor image and can vary substantially across cases, making an automatic estimation  
528 essential in the accuracy of the FCM segmentation results.

529 It should be emphasized that SPEQTACLE did not undergo any pre-processing or pre-  
530 optimization and that no parameter was set or chosen to optimize the obtained results on the  
531 evaluation datasets (either phantoms, realistic simulated or clinical tumors). The improved  
532 accuracy that SPEQTACLE achieved is therefore entirely due to its automatic estimation  
533 framework and its associated ability to adapt its norm parameter to varying properties of the  
534 image. The advantage of SPEQTACLE compared to other fuzzy clustering-based methods  
535 such as FLAB or FLICM thus lies on its ability to estimate reliably the norm parameter value  
536 on a case-by-case basis. In addition, the proposed norm estimation scheme is deterministic  
537 and convergent, therefore the repeatability of the algorithm was found to be perfect with zero  
538 variability in the results on repeated segmentations of the same image, which is an important  
539 point to ensure clinical acceptance for use by the physicians. In addition, the estimation of  
540 the norm was also found to be robust with respect to slightly larger or smaller initial  
541 determination of the tumor class using a background-subtraction approach<sup>32</sup>. In order to  
542 reach substantial differences in the segmentation results, this area had to be enlarged or

543 shrunk by more than 50%, which is very unlikely to occur unless highly inaccurate methods  
544 are used to define the initial region.

545 We showed that SPEQTACLE led to significantly higher accuracy in delineating tumor  
546 volumes with higher complexity (either in terms of shape, heterogeneity, noise levels and/or  
547 contrast), associated with a norm value higher than 3, on both simulated and clinical  
548 datasets. On the other hand, for simpler objects of interest (norm value below 3), we found  
549 that SPEQTACLE provided similar (although slightly improved) accuracy as FLAB and  
550 FLICM. Given the improved accuracy obtained with respect to FLAB on a dataset with a large  
551 range of contrast and noise levels as well as heterogeneity and shape, we expected that the  
552 robustness of SPEQTACLE should be at least similar as the one of FLAB. We indeed  
553 confirmed through a robustness analysis that the proposed automatic norm estimation  
554 scheme does not lead to decreased robustness with respect to varying image properties  
555 associated with the use of different PET/CT scanner models, reconstruction algorithms, or  
556 acquisition and reconstruction settings. Indeed, the level of robustness exhibited by  
557 SPEQTACLE was found to be similar to the one of FLAB, which had already been  
558 demonstrated as substantially more robust than standard FCM<sup>8</sup>. FLICM however was found  
559 to be much less robust, with segmentation failures for some of the configurations in the  
560 dataset. Given the fact that FLICM performed reasonably well on the accuracy dataset, its  
561 failure on the robustness evaluation might be due to the two parameters (the regularization  
562 parameter and the size of the surrounding kernel) that were set *a priori* in this study using  
563 recommended values that might not be appropriate for some of the PET images of the  
564 robustness dataset. The overall performance of FLICM might therefore be improved by  
565 optimizing these two parameters for each phantom acquisition, which is however out of the  
566 scope of the present work.

567 From a clinical point of view, our method might be easier than most of the previously  
568 proposed ones to implement in a clinical setting because it is fully automatic and perfectly  
569 repeatable, with no user intervention for parameterization beyond the localization of the

570 tumor in the whole-body image and its isolation in a 3D ROI. It is also very fast due to its low  
571 computational cost; thesegmentation of thelargest tumor (55×55×25 voxels) requires less  
572 than 1 min on a standard computer (CPU E5520 2.27 GHz×8), which could be easily  
573 shortened through algorithmic optimization and parallel computing or GPU implementation.  
574 Moreover, the algorithm itselfuses a negligible amount of memory.

575 The present work has a few limitations. It should be reminded that the proposed algorithm  
576 aims at the accurate delineation of a single pathological uptake previously detected and  
577 isolated in a ROI, similarly as FLAB. It was therefore not evaluated within the context of the  
578 simultaneous segmentation of multiple tumors (as each tumor should be processed  
579 independently when using SPEQTACLE), the detection of tumors and/or lymph nodes in a  
580 whole-body image<sup>45</sup>, nor the segmentation of diffuse and multifocal uptakes such as in  
581 pulmonary infection<sup>46</sup>. Also, we did not investigate the impact on the resulting segmentation  
582 of theinitial ROI selection, which is a first step as in most of published methods for PET tumor  
583 delineation<sup>10, 11</sup>. However, we already showed that this step has a very limited impact on the  
584 results for FLAB, as long as the ROI selection is made without incorporating nearby non-  
585 relevant uptake that would bias the estimation process<sup>15</sup>. Given that SPEQTACLE  
586 demonstrated similar robustness as FLAB, the impact of this step should be similarly low.  
587 Second, we did not include a large number of methods to compare SPEQTACLE with. Given  
588 its previous validation and demonstrated performance, FLAB can be considered a state-of-  
589 the-art method and our primary goal was to improve on that approach for challenging cases.  
590 A full comparison with numerous other methods was out of the scope of this work and might  
591 be conducted in the future using the benchmark currently being developed by the AAPM  
592 taskgroup 211<sup>147</sup>. Second, the robustness analysis was carried out on a smaller dataset than  
593 for the previously reported analysis for FLAB, FCM and thresholding methods<sup>8</sup>, however the  
594 dataset is certainly representative enough to provide a clear picture. Third, we did not  
595 evaluate the algorithms on clinical datasets with histopathology associated measurements.

---

<sup>1</sup>[http://aapm.org/org/structure/default.asp?committee\\_code=TG211](http://aapm.org/org/structure/default.asp?committee_code=TG211)

596 The one dataset available to us consists of maximum diameter measurements only<sup>17</sup>, which  
 597 might not be sufficient to highlight differences between the advanced algorithms under  
 598 comparison. On the other hand, a benchmark developed by the AAPM Taskgroup 211 is  
 599 expected to contain several clinical datasets with histopathological volumes<sup>47</sup>, and could be  
 600 used for future comparison studies. Finally, in the present implementation, the norm  
 601 parameter was estimated from an automatically pre-segmented estimation of the tumor  
 602 region, using a background-subtraction approach<sup>32</sup> in order to obtain a first guess of the  
 603 tumor class. The estimated norm was then used for all classes in the segmentation. In future  
 604 work, it would therefore be possible to potentially improve the algorithm performance by  
 605 estimating a norm parameter for each class in the ROI. In this case, the minimized criterion  
 606 in GFCM becomes:

$$607 \quad \sum_{u \in V} \sum_{i=1}^C p_{u,i}^m |y_u - \mu_i|^{\alpha_i} .$$

608 The norm parameter  $\alpha_i$  cannot be estimated by using the Newton-Raphson algorithm on the  
 609 minimized criterion of equation (4). Indeed, the norm parameter is essentially dependent on  
 610 the statistical behavior of the data and generally there is no solution  $\alpha_i \neq 1$  which minimizes  
 611 equation (4). Thus minimizing equation (4) according to  $\alpha_i$  is equivalent to solving:

$$612 \quad \sum_{u \in V} \sum_{i=1}^C p_{u,i}^m \log(|y_u - \mu_i|) |y_u - \mu_i|^{\alpha_i} = 0 .$$

613 Consequently, it depends on how data are scaled and the presence of  $y_u$  such that  
 614  $|y_u - \mu_i| > 1$  contributes to making this derivative  $> 0$ . Amongst other possible extensions, it  
 615 will be interesting to estimate a variance parameter additionally to the center parameter  $\mu_i$   
 616 and the norm parameter. Such a method would be able to fit more completely the statistical  
 617 distribution of the intensities. Indeed,  $\mu_i$  controls the mean of intensities for each cluster,  $\alpha_i$   
 618 controls the shape of the distribution whereas the variance parameter controls the disparity

619 of each cluster. Another future work will consist in extending SPEQTACLE to the multimodal  
620 situation for which each voxel becomes a vector whose components represent intensities  
621 taken from each image modality, for instance PET, CT and MRI. In the multimodal version, a  
622 norm parameter has to be estimated for each modality. The minimized criterion will thus have  
623 the same form by replacing the absolute value by a sum of absolute values.

## 624 **Conclusions**

625 In this paper, we have presented a fully automatic method for estimating the norm parameter  
626 in a generalized fuzzy C-means framework. We have developed and validated this new  
627 method for PET tumor delineation, and named it SPEQTACLE for *Spatial Positron Emission*  
628 *Quantification of Tumor: Automatic L<sub>p</sub>-norm Estimation*. The proposed approach is fully  
629 automated and perfectly repeatable. It provides improved accuracy with respect to state-of-  
630 the-art methods for realistic challenging delineation cases. This was demonstrated on both  
631 simulated and clinical datasets with complex shapes, high levels of uptake heterogeneity.  
632 The improvement in accuracy was achieved without sacrificing robustness vs. varying image  
633 properties in a multi-centric setting, which is crucial if the method is to be widely applicable in  
634 clinical practice. Future extensions of SPEQTACLE will include a multimodal version of the  
635 algorithm for PET/CT, PET/MRI and other multimodal medical imaging applications, as well  
636 as a multi-class norm estimation scheme to improve the algorithm performance.

637

638 **References**

639 <sup>1</sup> S. Hess, B.A. Blomberg, H.J. Zhu, P.F. Højlund-Carlsen, and A. Alavi, “The Pivotal Role of FDG-  
640 PET/CT in Modern Medicine,” *Acad. Radiol.* **21**(2), 232–249 (2014).

641 <sup>2</sup> G.C. Pereira, M. Traugber, and R.F. Muzic, “The role of imaging in radiation therapy planning:  
642 past, present, and future,” *BioMed Res. Int.* **2014**, 231090 (2014).

643 <sup>3</sup> K. Herrmann, M.R. Benz, B.J. Krause, K.L. Pomykala, A.K. Buck, and J. Czernin, “(18)F-FDG-PET/CT  
644 in evaluating response to therapy in solid tumors: where we are and where we can go,” *Q J Nucl  
645 Med Mol Imaging* **55**(6), 620–32 (2011).

646 <sup>4</sup> T. Carlier and C. Bailly, “State-Of-The-Art and Recent Advances in Quantification for Therapeutic  
647 Follow-Up in Oncology Using PET,” *Front. Med.* **2**, 18 (2015).

648 <sup>5</sup> M.K. Rahim, S.E. Kim, H. So, H.J. Kim, G.J. Cheon, E.S. Lee, K.W. Kang, and D.S. Lee, “Recent  
649 Trends in PET Image Interpretations Using Volumetric and Texture-based Quantification  
650 Methods in Nuclear Oncology,” *Nucl. Med. Mol. Imaging* **48**(1), 1–15 (2014).

651 <sup>6</sup> J.P.B. O’Connor, C.J. Rose, J.C. Waterton, R.A.D. Carano, G.J.M. Parker, and A. Jackson, “Imaging  
652 Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome,” *Clin.  
653 Cancer Res. Off. J. Am. Assoc. Cancer Res.* **21**(2), 249–257 (2015).

654 <sup>7</sup> S. Houshmand, A. Salavati, S. Hess, T.J. Werner, A. Alavi, and H. Zaidi, “An update on novel  
655 quantitative techniques in the context of evolving whole-body PET imaging,” *PET Clin.* **10**(1), 45–  
656 58 (2015).

657 <sup>8</sup> M. Hatt, C. Cheze Le Rest, N. Albarghach, O. Pradier, and D. Visvikis, “PET functional volume  
658 delineation: a robustness and repeatability study,” *Eur J Nucl Med Mol Imaging* **38**(4), 663–72  
659 (2011).

660 <sup>9</sup> R. Boellaard, M.J. O’Doherty, W.A. Weber, F.M. Mottaghy, M.N. Lonsdale, S.G. Stroobants, W.J.  
661 Oyen, J. Kotzerke, O.S. Hoekstra, J. Pruim, P.K. Marsden, K. Tatsch, C.J. Hoekstra, E.P. Visser, B.  
662 Arends, F.J. Verzijlbergen, J.M. Zijlstra, E.F. Comans, A.A. Lammertsma, A.M. Paans, A.T.  
663 Willemsen, T. Beyer, A. Bockisch, C. Schaefer-Prokop, D. Delbeke, R.P. Baum, A. Chiti, and B.J.  
664 Krause, “FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0,”  
665 *Eur J Nucl Med Mol Imaging* **37**(1), 181–200 (2010).

666 <sup>10</sup> M. Hatt, N. Bousson, C. Cheze-Le Rest, D. Visvikis, and O. Pradier, “[Metabolically active  
667 volumes automatic delineation methodologies in PET imaging: review and perspectives],” *Cancer  
668 Radiother* **16**(1), 70–81; quiz 82, 84 (2012).

669 <sup>11</sup> B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D.J. Mollura, “A review on segmentation of positron  
670 emission tomography images,” *Comput. Biol. Med.* **50**, 76–96 (2014).

671 <sup>12</sup> B. Braathen, W. Pieczynski, and P. Masson, *Global and local methods of unsupervised Bayesian  
672 segmentations of images*, *Mach. Graph. Vis.* 39–52 (1993).

673 <sup>13</sup> D. Benboudjema and W. Pieczynski, “Unsupervised statistical segmentation of nonstationary  
674 images using triplet Markov fields,” *IEEE Trans Pattern Anal Mach Intell* **29**(8), 1367–78 (2007).

675 <sup>14</sup> M. Hatt, F. Lamare, N. Bousson, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson,  
676 C. Cheze-Le Rest, and D. Visvikis, “Fuzzy hidden Markov chains segmentation for volume  
677 determination and quantitation in PET,” *Phys Med Biol* **52**(12), 3467–91 (2007).

678 <sup>15</sup> M. Hatt, C. Cheze le Rest, A. Turzo, C. Roux, and D. Visvikis, “A fuzzy locally adaptive Bayesian  
679 segmentation approach for volume determination in PET,” *IEEE Trans Med Imaging* **28**(6), 881–  
680 93 (2009).

681 <sup>16</sup> M. Hatt, C. Cheze le Rest, P. Descourt, A. Dekker, D. De Ruyscher, M. Oellers, P. Lambin, O.  
682 Pradier, and D. Visvikis, “Accurate automatic delineation of heterogeneous functional volumes in  
683 positron emission tomography for oncology applications,” *Int J Radiat Oncol Biol Phys* **77**(1),  
684 301–8 (2010).

685 <sup>17</sup> M. Hatt, C. Cheze-le Rest, A. van Baardwijk, P. Lambin, O. Pradier, and D. Visvikis, "Impact of  
686 tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer  
687 tumor delineation," *J Nucl Med* **52**(11), 1690–7 (2011).

688 <sup>18</sup> M. Hatt, C. Cheze-Le Rest, E.O. Aboagye, L.M. Kenny, L. Rosso, F.E. Turkheimer, N.M.  
689 Albarghach, J.P. Metges, O. Pradier, and D. Visvikis, "Reproducibility of 18F-FDG and 3'-deoxy-3'-  
690 18F-fluorothymidine PET tumor volume measurements," *J Nucl Med* **51**(9), 1368–76 (2010).

691 <sup>19</sup> B.H. de Figueiredo, M. Antoine, R. Trouette, P. Lagarde, A. Petit, F. Lamare, M. Hatt, and P.  
692 Fernandez, "Use of FDG-PET to guide dose prescription heterogeneity in stereotactic body  
693 radiation therapy for lung cancers with volumetric modulated arc therapy: a feasibility study,"  
694 *Radiat. Oncol. Lond. Engl.* **9**, 300 (2014).

695 <sup>20</sup> B. Henriques de Figueiredo, C. Zacharatou, S. Galland-Girodet, J. Benech, H. De Clermont-  
696 Gallerande, F. Lamare, M. Hatt, L. Digue, E. De Mones Del Pujol, and P. Fernandez, "Hypoxia  
697 imaging with [18F]-FMISO-PET for guided dose escalation with intensity-modulated radiotherapy  
698 in head-and-neck cancers," *Strahlenther. Onkol. Organ Dtsch. Rontgengesellschaft AI* (2014).

699 <sup>21</sup> A.I.J. Arens, E.G.C. Troost, B.A.W. Hoeben, W. Grootjans, J.A. Lee, V. Grégoire, M. Hatt, D.  
700 Visvikis, J. Bussink, W.J.G. Oyen, J.H.A.M. Kaanders, and E.P. Visser, "Semiautomatic methods for  
701 segmentation of the proliferative tumour volume on sequential FLT PET/CT images in head and  
702 neck carcinomas and their relation to clinical outcome," *Eur. J. Nucl. Med. Mol. Imaging* **41**(5),  
703 915–924 (2014).

704 <sup>22</sup> M. Hatt, A.L. Maitre, D. Wallach, H. Fayad, and D. Visvikis, "Comparison of different methods of  
705 incorporating respiratory motion for lung cancer tumor volume delineation on PET images: a  
706 simulation study," *Phys Med Biol* **57**(22), 7409–30 (2012).

707 <sup>23</sup> A. Le Maitre, M. Hatt, O. Pradier, C. Cheze-le Rest, and D. Visvikis, "Impact of the accuracy of  
708 automatic tumour functional volume delineation on radiotherapy treatment planning," *Phys  
709 Med Biol* **57**(17), 5381–97 (2012).

710 <sup>24</sup> D.E. Gustafson and W.C. Kessel, "Fuzzy clustering with a fuzzy covariance matrix," in *1978 IEEE  
711 Conf. Decis. Control 17th Symp. Adapt. Process.* (1978), pp. 761–766.

712 <sup>25</sup> R.J. Hathaway, J.C. Bezdek, and Y. Hu, "Generalized Fuzzy C-means Clustering Strategies Using Lp  
713 Norm Distances," *Trans Fuz Sys* **8**(5), 576–582 (2000).

714 <sup>26</sup> S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on  
715 new kernel-induced distance measure," *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **34**(4),  
716 1907–1916 (2004).

717 <sup>27</sup> M.-S. Yang and K.-L. Wu, "Unsupervised Possibilistic Clustering," *Pattern Recogn* **39**(1), 5–21  
718 (2006).

719 <sup>28</sup> M. Daniel, "Belief Functions: A Revision of Plausibility Conflict and Pignistic Conflict.," in *SUM*,  
720 edited by W. Liu, V.S. Subrahmanian and J. Wijsen (Springer, 2013), pp. 190–203.

721 <sup>29</sup> M.N. Ahmed, S.M. Yamany, N. Mohamed, A.A. Farag, and T. Moriarty, "A modified fuzzy C-  
722 means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med.  
723 Imaging* **21**(3), 193–199 (2002).

724 <sup>30</sup> S. Krinidis and V. Chatzis, "A robust fuzzy local information C-Means clustering algorithm," *IEEE  
725 Trans. Image Process. Publ. IEEE Signal Process. Soc.* **19**(5), 1328–1337 (2010).

726 <sup>31</sup> I. Gath and A. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Mach.  
727 Intell.* **11**(7), 773–780 (1989).

728 <sup>32</sup> I.A. Burger, H.A. Vargas, B.J. Beattie, D.A. Goldman, J. Zheng, S.M. Larson, J.L. Humm, and C.R.  
729 Schmidlein, "How to assess background activity: introducing a histogram-based analysis as a  
730 first step for accurate one-step PET quantification," *Nucl. Med. Commun.* **35**(3), 316–324 (2014).

731 <sup>33</sup> S.-I. Amari and H. Nagaoka, *Methods of Information Geometry* (American Mathematical Society,  
732 2007).

733 <sup>34</sup> B. Berthon, C. Marshall, A. Edwards, M. Evans, and E. Spezi, "Influence of cold walls on PET  
734 image quantification and volume segmentation: a phantom study," *Med. Phys.* **40**(8), 082505  
735 (2013).

736 <sup>35</sup> F. Hofheinz, S. Dittrich, C. Pöttsch, and J. van den Hoff, "Effects of cold sphere walls in PET  
737 phantom measurements on the volume reproducing threshold," *Phys. Med. Biol.* **55**(4), 1099–  
738 1113 (2010).

739 <sup>36</sup> S.K. Warfield, K.H. Zou, and W.M. Wells, "Simultaneous truth and performance level estimation  
740 (STAPLE): an algorithm for the validation of image segmentation," *IEEE Trans Med Imaging* **23**(7),  
741 903–21 (2004).

742 <sup>37</sup> H. Zaidi, M. Abdoli, C.L. Fuentes, and I.M. El Naqa, "Comparative methods for PET image  
743 segmentation in pharyngolaryngeal squamous cell carcinoma," *Eur. J. Nucl. Med. Mol. Imaging*  
744 **39**(5), 881–891 (2012).

745 <sup>38</sup> A. Le Maitre, W. Segars, S. Marache, A. Reilhac, M. Hatt, S. Tomei, C. Lartizien, and D. Visvikis,  
746 "Incorporating Patient-Specific Variability in the Simulation of Realistic Whole-Body 18F-FDG  
747 Distributions for Oncology Applications," *Proc. IEEE* **9**(12), 2026–2038 (2009).

748 <sup>39</sup> P. Papadimitroulas, G. Loudos, A. Le Maitre, M. Hatt, F. Tixier, N. Efthimiou, G.C. Nikiforidis, D.  
749 Visvikis, and G.C. Kagadis, "Investigation of realistic PET simulations incorporating tumor  
750 patient's specificity using anthropomorphic models: creation of an oncology database," *Med.*  
751 *Phys.* **40**(11), 112506 (2013).

752 <sup>40</sup> W. Segars, Development and Application of the New Dynamic NURBS-based Cardiac-Torso  
753 (NCAT) phantom (2001).

754 <sup>41</sup> F. Lamare, A. Turzo, Y. Bizais, C.C. Le Rest, and D. Visvikis, "Validation of a Monte Carlo  
755 simulation of the Philips Allegro/GEMINI PET systems using GATE," *Phys Med Biol* **51**(4), 943–62  
756 (2006).

757 <sup>42</sup> D. Visvikis, A. Turzo, A. Gouret, P. Damine, F. Lamare, Y. Bizais, and C. Cheze Le Rest,  
758 "Characterisation of SUV accuracy in FDG PET using 3-D RAMLA and the Philips Allegro PET  
759 scanner," *J. Nucl. Med.* **45**(5), 103 (2004).

760 <sup>43</sup> D.C. Weber, H. Wang, L. Cozzi, G. Dipasquale, H.G. Khan, O. Ratib, M. Rouzaud, H. Veas, H. Zaidi,  
761 and R. Miralbell, "RapidArc, intensity modulated photon and proton techniques for recurrent  
762 prostate cancer in previously irradiated patients: a treatment planning comparison study,"  
763 *Radiat Oncol* **4**, 34 (2009).

764 <sup>44</sup> S. Belhassen and H. Zaidi, "A novel fuzzy C-means algorithm for unsupervised heterogeneous  
765 tumor quantification in PET," *Med Phys* **37**(3), 1309–24 (2010).

766 <sup>45</sup> L. Bi, J. Kim, L. Wen, and D.D. Feng, "Automated and robust PERCIST-based thresholding  
767 framework for whole body PET-CT studies," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*  
768 *IEEE Eng. Med. Biol. Soc. Conf.* **2012**, 5335–5338 (2012).

769 <sup>46</sup> B. Foster, U. Bagci, null Ziyue Xu, B. Dey, B. Luna, W. Bishai, S. Jain, and D.J. Mollura,  
770 "Segmentation of PET images for computer-aided functional quantification of tuberculosis in  
771 small animal models," *IEEE Trans. Biomed. Eng.* **61**(3), 711–724 (2014).

772 <sup>47</sup> T. Shepherd, B. Berthon, P. Galavis, E. Spezi, A. Apte, J. Lee, D. Visvikis, M. Hatt, E. de Bernardi, S.  
773 Das, I. El Naqa, U. Nestle, C. Schmidlein, H. Zaidi, and A. Kirov, "Design of a benchmark platform  
774 for evaluating PET-based contouring accuracy in oncology applications," *Eur. J. Nucl. Med. Mol.*  
775 *Imaging* **39**, S264–S264 (2012).

776

777 **Appendices**

778 **A. FCM minimization step**

779 The minimization process for FCM is achieved recursively until convergence:

780 1. Let  $p_{u,i}^{(0)}$  and  $\mu_i^{(0)}$  be initial values;

781 2. From  $p_{u,i}^{(q)}$  compute:  $\mu_i^{(q+1)} = \frac{\sum_{u \in V} (p_{u,i}^{(q)})^m y_u}{\sum_{u \in V} (p_{u,i}^{(q)})^m}$

782 3. From  $\mu_i^{(q+1)}$  compute:  $p_{u,i}^{(q+1)} = \frac{1}{\sum_{k=1}^C \left( \frac{|y_u - \mu_i^{(q+1)}|}{|y_u - \mu_k^{(q+1)}|} \right)^{\frac{2}{m-1}}}$ .

783 **B. Newton-Raphson algorithm**

784 Let  $f$  be a derivable function from  $\mathfrak{R}$  to  $\mathfrak{R}$ , the Newton-Raphson algorithm is an algorithm to  
785 find the solution  $a$  such that  $f(a) = 0$ . The Newton-Raphson works as following:

786 1. Set  $a_0$  an initial value;

787 2.  $a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)}$ .

788 **C. GFCM minimization step (norm parameter is known)**

789 For fixed norm parameter  $\alpha$  and weight parameters  $(p_{u,i})_{1 \leq i \leq C}$ , the center  $\mu_j$  is estimated by  
790 minimizing:

791  $\sum_{u \in V} p_{u,j}^m |y_u - \mu_j|^\alpha$ , which is equivalent to solve the equation  $f'(\mu) = 0$  with

792  $f(\mu) = \sum_{u \in V} p_{u,j}^m |y_u - \mu|^\alpha$ .

793 We have  $f'(\mu) = -\alpha \sum_{u \in V} (p_{u,j})^m \operatorname{sgn}(y_u - \mu) |y_u - \mu|^{\alpha-1}$  and  $f''(\mu) = \alpha(\alpha-1) \sum_{u \in V} (p_{u,j})^m |y_u - \mu|^{\alpha-2}$ .

794 One can easily show that the Newton-Raphson algorithm does not converge when  $\alpha < 2$ .

795 Consequently, the minimization step of GFCM with fixed norm parameter works as following:

796 1. Let  $p_{u,i}^{(0)}$  and  $\mu_i^{(0)}$  be initial values;

797 2. If  $\alpha > 2$  compute  $\mu_j^{(q+1)}$  by the Newton-Raphson algorithm:

798 a) Let  $\mu_j^{(q+1,0)}$  be an initial value;

$$\mu_j^{(q+1,k+1)} = \mu_j^{(q+1,k)}$$

799 b) Do: 
$$\sum_{u \in V} (p_{u,j}^{(q)})^m \operatorname{sgn}(y_u - \mu_j^{(q+1,k)}) |y_u - \mu_j^{(q+1,k)}|^{\alpha-1}$$
 until convergence.  $\mu_j^{(q+1)}$  is the limit 
$$+ \frac{(\alpha-1) \sum_{u \in V} (p_{u,j}^{(q)})^m |y_u - \mu_j^{(q+1,k)}|^{\alpha-2}}{}$$

800 of this sequence.

801 3. If  $\alpha < 2$ , compute  $\mu_j^{(q+1)}$  by Gradient descent algorithm:

802 a) Let  $\mu_j^{(q+1,0)}$  be an initial value;

$$\mu_j^{(q+1,k+1)} = \mu_j^{(q+1,k)}$$

803 b) Do: 
$$+ \varepsilon \sum_{u \in V} (p_{u,j}^{(q)})^m \operatorname{sgn}(y_u - \mu_j^{(q+1,k)}) |y_u - \mu_j^{(q+1,k)}|^{\alpha-1}$$
 where  $\varepsilon$  is a fixed temporal step.

804  $\mu_j^{(q+1)}$  is the limit of this sequence.

805 4. Compute 
$$p_{u,j}^{(q+1)} = \frac{1}{\sum_{k=1}^C \left( \frac{|y_u - \mu_j^{(q+1)}|}{|y_u - \mu_k^{(q+1)}|} \right)^{\frac{\alpha}{m-1}}}$$

#### 806 D. Eulerian functions

807 The Eulerian function is defined as an integral for any complex number which real part is

808 strictly positive as:

809 
$$\Gamma(z) = \int_0^{+\infty} t^{z-1} \exp(-t) dt .$$

810 For any strictly positive integer, we have  $\Gamma(n) = (n-1)!$  and for any complex  $z$  such that  
 811  $\text{Re}(z) > 0$ , we have  $\Gamma(z) = \frac{\Gamma(z+1)}{z}$ . Consequently,  $\Gamma$  admits a meromorphic extension to  
 812 the complex plane whose singularities are negative or null integers. The infinite product of  $\Gamma$   
 813 is given by:

$$814 \quad \Gamma(z) = \frac{e^{-\gamma z}}{z} \times \prod_{n=1}^{\infty} \frac{e^{\frac{z}{n}}}{1 + \frac{z}{n}}, \text{ where } \gamma = \lim_{n \rightarrow +\infty} \left[ \sum_{k=1}^n \frac{1}{k} - \log(n) \right].$$

815 We define the di-gamma function as  $\psi = \frac{\Gamma'}{\Gamma}$ . It is also a meromorphic function which Laurent  
 816 development is given by:

$$817 \quad \psi(z) = -\gamma - \frac{1}{z} + \sum_{n=1}^{+\infty} \frac{z}{n(z+n)}. \text{ The consecutive derivatives of } \psi \text{ are given by the Laurent}$$

818 developments:

$$819 \quad \psi^{(k)}(z) = (-1)^{k+1} k! \sum_{n=0}^{+\infty} \frac{1}{(z+n)^{k+1}}.$$

## 820 E. Fisher information matrix, Kullback information divergence and related results

821 Let  $\Lambda = \{y \rightarrow p(y|\theta) : \theta \in \Theta\}$  be a smooth manifold of statistical distribution parameterized  
 822 by an open set  $\Theta \subset \mathfrak{R}^k$ , the Fisher information matrix for the value  $\theta$  of the parameter is  
 823 given by:

$$824 \quad I_{i,j}(\theta) = -E \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(Y|\theta) \right],$$

825 which is the negative of the mean of the Hessian of the log-likelihood. Under good conditions  
 826 (reversibility of integration and derivation), this matrix is strictly positive and symmetric.

827 Each statistical distribution  $y \rightarrow p(y|\theta)$  lies on an embedding set of infinite dimension;  
 828 indeed to represent the entire graph of such a function, we need an infinite number of values  
 829 for  $y$ . However, it is parameterized by a finite number of real numbers; consequently,  $\Lambda$  has  
 830 an intrinsic dimension equal to the number  $k$  of real parameters. A Riemannian manifold is  
 831 provided with an infinitesimal distance which, in our case, is given by the Fisher information  
 832 matrix. Without giving all the details regarding the differential geometry, one can say  
 833 colloquially that the distance between “close” distributions  $y \rightarrow p(y|\theta)$  and  $y \rightarrow p(y|\theta + d\theta)$   
 834 is given by:

835  $dl = \sqrt{(d\theta)^* I(\theta) d\theta}$ . Let  $\theta_1 \in \Theta$  and  $\theta_2 \in \Theta$  be two values of the parameter, the length of  
 836 the curve  $t \in [t_1, t_2] \rightarrow \theta(t)$  where  $\theta(t_1) = \theta_1$  and  $\theta(t_2) = \theta_2$  in the space of distributions is  
 837 given by:

$$838 \quad L(\theta) = \int_{t_1}^{t_2} \sqrt{(\theta'(t))^* I(\theta(t)) \theta'(t)} dt,$$

839 where  $t \rightarrow \theta'(t)$  is the derivative of  $t \rightarrow \theta(t)$  along  $t$ . The distance between the distributions  
 840  $y \rightarrow p(y|\theta_1)$  and  $y \rightarrow p(y|\theta_2)$  is the length of the smallest curve  $t \rightarrow \theta(t)$ .

841 The Kullback divergence between two probability densities  $p$  (target probability) and  $q$   
 842 (instrumental probability) is defined as:

$$843 \quad K(p:q) = \int_{\mathfrak{R}} \log \left( \frac{p(y)}{q(y)} \right) p(y) dy.$$

The Kullback divergence is not a metric and

844  $K(p:q) \neq K(q:p)$ . However, if  $p$  and  $q$  are in the same parametrical set, denoting  
 845  $K(\theta_2: \theta_1)$  the Kullback divergence for  $q = p(\theta_1)$  and  $p = p(\theta_2)$ , the Kullback divergence  
 846 satisfies the asymptotic equation:

847  $K(\theta + d\theta : \theta) = (d\theta)^* I(\theta) d\theta + o(\|d\theta\|^2),$

848 when  $d\theta$  tends to 0.