

# Image change detection using paradoxical theory for patient follow-up quantitation and therapy assessment

Simon David<sup>1</sup> david.simon@univ-brest.fr , Dimitris Visvikis<sup>1</sup> dimitris@univ-brest.fr , Gwenole Quellec<sup>1</sup> quellec@univ-brest.fr , Catherine Cheze Le Rest<sup>1</sup> catherine.cheze-lerest@chu-brest.fr , Philippe Fernandez<sup>2</sup> philippe.fernandez@u-bordeaux2.fr , Michèle Allard<sup>2</sup> michele.allard@chu-bordeaux.fr , Christian Roux<sup>1</sup> Christian.Roux@telecom-bretagne.eu , Mathieu Hatt<sup>1\*</sup> hatt@univ-brest.fr

<sup>1</sup> *Optimisation Continue des Actions Thérapeutiques par l'Intégration d'Informations INSERM : U1101 , Université de Bretagne Occidentale - Brest , CHU Brest, 5 Avenue Foch 29609 Brest Cedex 2, FR*

<sup>2</sup> *INCIA, Institut de Neurosciences Cognitives et Intégratives d'Aquitaine CNRS : UMR5287 , Université Victor Segalen - Bordeaux II , Université Sciences et Technologies - Bordeaux I , SFR Bordeaux Neurosciences , 33076 Bordeaux, FR*

\* Correspondence should be addressed to: Mathieu Hatt <hatt@univ-brest.fr >

## Abstract

**In clinical oncology, Positron Emission Tomography (PET) imaging can be used to assess therapeutic response by quantifying the evolution of semi-quantitative values such as SUV, early during treatment or after treatment. Current guidelines do not include metabolically active tumor volume (MATV) measurements and derived parameters such as total lesion glycolysis (TLG) to characterize the response to the treatment. To achieve automatic MATV variation estimation during treatment, we propose an approach based on the change detection principle using the recent paradoxical theory, which models imprecision, uncertainty and conflict between sources. It was applied here simultaneously to pre and after treatment PET scans. The proposed method was applied to both simulated and clinical datasets, and its performance was compared to adaptive thresholding applied separately on pre and post treatment PET scans. On simulated datasets, the adaptive threshold was associated with significantly higher classification errors than the developed approach. On clinical datasets, the proposed method led to results more consistent with the known partial responder status of these patients. The method requires accurate rigid registration of both scans which can be obtained only in specific body regions and does not explicitly model uptake heterogeneity. In further investigations the change detection of intra-MATV tracer uptake heterogeneity will be developed by incorporating textural features into the proposed approach.**

**Author Keywords** Change detection ; paradoxical theory ; oncology ; patient monitoring ; PET ; therapy response ; unsupervised segmentation

## Introduction

In clinical oncology, Positron Emission Tomography (PET) imaging is increasingly used for diagnosis [1], patient monitoring studies [2] and the estimation of the metabolically active tumor volume (MATV) for radiotherapy planning purposes [3]. Within the context of patient follow-up and therapy assessment, the tumor metabolic changes induced by chemotherapy and/or radiotherapy may arise prior to anatomic changes characterized using morphological imaging such as computed tomography (CT) or magnetic resonance imaging (MRI) [4]. The response to therapy may be assessed by comparing sequential PET scans acquired before (“baseline”), during (“early assessment”) and after (“late assessment”) treatment [5–7]. Both qualitative and quantitative approaches have been proposed [8]. Qualitative methods have been considered significantly less accurate and reproducible than quantitative approaches for several applications [9,10]. Quantification in PET may be carried out either on dynamic or static acquisitions, with similar reproducibility [11]. Requiring longer acquisitions and arterial sampling, the approaches based on dynamic PET imaging are currently not extensively used in clinical practice. The standardized uptake value (SUV) computed on static PET scans is still the most widely used semi-quantitative index in oncology clinical practice. Two indexes, namely the maximum SUV ( $SUV_{max}$ ) and the mean SUV ( $SUV_{mean}$ ) within a previously defined region of interest (ROI) are currently used in routine for the quantification and will be included in the present study. The reproducibility of both parameters has been assessed in different studies [12, 13], and compared to the reproducibility of MATV measurements with various automated and semi-automated methodologies [14]. In order to better characterize MATV, others indexes such as the total lesion glycolysis (TLG) were defined as the product of the MATV and the associated mean SUV [15].

In order to standardize therapy response assessment in oncology, recommendations such as WHO [16] and RECIST [17] have been established and are based on measurements of anatomic tumor dimensions on CT scans. More recently, it has been suggested in PERCIST [18] to characterize the response according to the metabolic activity quantified with PET imaging using SUV. Based on the reproducibility of such indices as previously investigated [12–14], confidence intervals were derived in order to characterize response, partial response, absence of response and progressive disease [18]. Due to the lack of appropriate automated tools, response assessment in clinical practice is usually carried out using simple SUV measurements and do not consider the overall or intra MATV variations.

All the approaches mentioned above are based on the use of global parameters such as SUV or MATV. By contrast with these “global” methods, few approaches have been developed to consider a voxel by voxel analysis. A method comparing the variation of the apparent diffusion coefficient (ADC) of each tumor voxel between the pre-treatment and mid-treatment scans has been proposed for functional MRI imaging (fMRI) [19]. Measuring the cellular density, the ADC variation of a tumor voxel can be associated with therapy response. The ADC variation measurements led to the creation of a functional diffusion map (fDM), in which each tumor voxel was classified as responder, stable or progressive. Another recent method proposed assessing the response by comparing two follow-up PET images previously co-registered using associated CT datasets [20]. A biparametric map was generated and analyzed in order to identify the tracer uptake variations within the tumor volumes.

The aim of this study was to present a novel method to characterize MATV response to therapy, based on the change detection principle previously proposed within the context of satellite and sonar imaging [21] [22]. In the proposed approach, the MATV are first delineated in each PET scan individually using a robust automatic algorithm [23]. Subsequently, signatures characterizing changes between the scans are defined by two parameters, one global and one local. The change detection process is based on the recent Dezert-Smarandache (DSm) theory, combining uncertain and paradoxical information [24]. By modeling uncertainty, imprecision and conflicting data between sources, such a process using the paradoxical theory may add valuable information to characterize the tumor evolution than the simple difference of binary segmented maps obtained on each image independently. The local voxel by voxel therapy response is finally deduced from the original follow-up PET images and the change detection map. In addition to the overall volume and activity evolution between two or more PET scans, the goal of the proposed method was to provide a quantitative accurate voxel-by-voxel response map using color coding.

The proposed method was applied on simulated and clinical follow-up  $^{18}$ F-FDG PET scans of patients with esophageal cancer treated by radio-chemotherapy. The change detection approach was compared with the current quantitative methods used in patient monitoring (SUV variations) and MATV threshold-based delineation applied independently to each of the follow-up scans.

## Materials and methods

Figure 1 presents a flowchart of the proposed method, each step being detailed in the following sections. In the following, the term *noise* denotes the statistical variations of voxels values due to the signal acquisition and image reconstruction processes, whereas the term *heterogeneity* denotes variations of uptake in small or larger groups of voxels within the MATV that may be characterized as spatial patterns associated with subjacent biological processes such as hypoxia or vascularization [25].

### Pre processing

#### *Local-based analysis*

Our method is aimed at characterizing local evolutions of volumes of interest at a voxel level in follow-up PET imaging. In this work, we therefore assume that the lesions of interest can be first isolated by a clinician in a 3D bounding box, which is determined on co-registered PET scans (figure 2). This bounding box should be large enough to encompass the entire tumor volume on all considered scans. Its shape, not necessarily cubic, should allow in most cases excluding physiological uptake close to the tumor which may disturb the change detection process, although depending on the type of malignancy, extent and location of the lesions, the process may be more or less complex and time consuming.

#### *Image deconvolution*

Tumor volumes observed in PET images are affected by the partial volume effects (PVE), due to the limited spatial resolution of PET scanners, which is about 4–6 mm full width at half maximum (FWHM) in state-of-the art PET/CT scanners [26]. The PVE lead to activity cross contamination between adjacent structures characterized by different uptake levels and might significantly impact measured activity in structures with sizes below three times the FWHM. In the co-registered PET scans, voxels most affected by PVE (mostly on the borders between structures of interest) may not have the same spatial coordinates, which can further disturb the change detection method. Moreover, the PVE affect the absolute SUV values extracted from each scan. These measurements may be significantly biased and this would lead to significant under or over estimation of the tracer uptake variation between the follow-up PET scans. Consequently, a PVE correction (PVC) was applied to each follow-up scan. The chosen PVC approach is based on 3D voxel-wise correction using an iterative deconvolution improved by wavelet and curvelet-based optimal denoising [27] [28]. This pre-processing step allows reducing the size of the blurred frontiers and extracting corrected SUVs for the quantitative characterization of the evolution of the tracer uptake.

#### *PET/PET registration*

As mentioned above, the follow-up PET scans are assumed to be registered at a voxel level before the change detection method can be applied. Within the context of patient monitoring studies, the PET/CT images may be acquired at several weeks interval. In our study, the PET/PET registration was carried out according to a previously published approach [29]. PET scans were first registered with their

corresponding CT acquired in the same bed position. Then, having more landmarks than the functional images, the two CT scans were registered with the MIPAV software using a rigid transformation, by optimizing the least square criterion. Finally, the computed CT/CT transformation matrix was applied to the PET scans leading to their registration at the voxel level. A rigid registration was used in our study for two reasons. First, the analysis was carried out on small 3D volumes of interest centered in the mediastinum. Second, elastic registration would deform the tumor shapes and volumes when registering one to the other, which would bias the comparison analysis on the voxel level of the MATV evolution between the two scans. This approach however limits the applicability of the method to body regions less impacted by respiratory motion and registration issues such as patient morphological changes between the two scans.

In addition, the local change signature of each voxel was performed within a sliding cube of  $3 \times 3 \times 3$  voxels centered on the voxel of interest, which is expected to reduce the impact of small (one or two voxels maximum) registration errors (see end of section II.B.2).

### Change Detection method

This approach is based upon change detection principle using the DSM theory to model the evolution of the tumor. The change detection has been widely used for satellite imaging [21,22] within the context of environmental (forest, agriculture) monitoring or in urban studies. However, the change detection method has not been applied to PET imaging up to now. The proposed approach was adapted from a method applied to SAR (Synthetic Aperture Imaging) images used for the detection of abrupt changes due to natural environmental damages [30].

### Image Mapping

Let us consider  $I_1$  and  $I_2$ , two 3D PET volumes of follow-up scans acquired respectively before and during (or after) treatment. Both images are assumed to be co-registered as described in section II.A.3. The first step consists in delineating the MATV in each image in order to obtain a first estimate of the tumor volumes that will be subsequently compared in the change detection approach. The algorithm used in this step (Fuzzy Local Adaptive Bayesian, FLAB) has been previously validated [23]. Using a fuzzy Bayesian model, the approach is based on a local adaptive prior model and an observation model assumed to be a mixture of Gaussian distributions. The parameters of the mixture are estimated with the stochastic expectation maximization (SEM) algorithm [31]. The voxels classification into tumor or background class is carried out according to the maximum *likelihood* criterion, based on the posterior probability. Illustrated in figure 1, black voxels are those classified as belonging to the class 'MATV' by the FLAB algorithm, whereas the other voxels have been classified as 'background'. This step provides some parameters that are subsequently used in the change signature (see sections II.B.2):

The Gaussian distributions, noted  $f$ , are defined by their mean and variance  $(\mu, \sigma_i^2)$ , where  $i$  denotes the MATV ( $i = MATV$ ) or background ( $i = BD$ ) and  $j$  denotes the affiliation to image 1 or 2.

### Change signature

The second step of the process consists in characterizing the change between the two maps obtained at the previous step (section II.B.1), considering two levels of observation: one global and one local.

The global evolution was characterized using a measure of distance between the distributions. The Bhattacharyya criterion ( $BC$ ) [32] was used to measure the overlap between two probability distributions  $p$  and  $q$ :

$$BC(p, q) = \int \sqrt{p(x)q(x)} dx \quad (1)$$

A global change criterion  $D(p, q)$  can be deduced from the  $BC$ :

$$D(p, q) = 1 - \int \sqrt{p(x)q(x)} dx \quad (2)$$

On the one hand, for differentiated distributions,  $BC$  tends to 0 and  $D$  tends to 1. On the other hand, when  $p$  and  $q$  are similar,  $BC$  tends to 1 and  $D$  tends to 0. In the proposed method, four distances were computed by considering for  $p$  and  $q$ , the distributions  $f$  associated with the PET follow-up images  $I_1$  and  $I_2$ . For example, the distance between the MATV distribution in the pre-treatment scan  $f_{MATV}^1$  and the distribution, corresponding to the BD in the post-treatment scan  $f_{BD}^2$  was defined as:

$$D(f_{MATV}^1, f_{BD}^2) = 1 - \int \sqrt{f_{MATV}^1(x)f_{BD}^2(x)} dx \quad (3)$$

The local evolution was designed based on a contrast measurement between the two images. At the local level, a given voxel is characterized by its value and the probability density function of the class it belongs to, as identified in the mapping step. For each voxel, the membership of the class would differ when the probability density function (pdf) of its class has changed, although its value could be similar in both images. The membership would also differ if the voxel belongs to a class with similar pdf while its value has changed

significantly. The membership change of a voxel  $t$  between the image  $I_1$  and  $I_2$  was therefore defined as a contrast measurement between the two pdfs  $p$  and  $q$ :

$$C_{(p, q)}(t) = 1 - \min\left(\frac{p(i(t))}{q(i(t))}, \frac{q(i(t))}{p(i(t))}\right)$$

(4)

where  $i(t)$  is the intensity of the voxel  $t$  in the image. The contrast associated to the MATV class is estimated for the voxel  $t$  with the distributions  $f_{MATV}^1$  and  $f_{MATV}^2$  associated with the pre and post-treatment scans:

$$C_{(f_{MATV}^1, f_{MATV}^2)}(t) = 1 - \min\left(\frac{f_{MATV}^1(i(t))}{f_{MATV}^2(i(t))}, \frac{f_{MATV}^2(i(t))}{f_{MATV}^1(i(t))}\right)$$

(5)

This measure allows characterizing the evolution of the classes' pdf and the voxel variation within its class and assumes that the follow-up images are registered voxel by voxel. In order to reduce the impact of potential misregistration, the contrast for each voxel  $t$  was computed within a  $3 \times 3 \times 3$  cube centered on the voxel  $t$  by averaging the values of the 27 voxels in the cube and assigning the resulting value to the center voxel.

### Change detection with paradoxical theory

In our method, the change detection process is carried out using the Dezert-Smarandache (DSm) theory of plausible and paradoxical reasoning [24]. The DSm theory is a generalization of the classical Dempster-Shafer (DS) theory. Modeling data using basic belief functions, the DSm theory allows combining different types of information sources and dealing with their inherent imprecision and uncertainties, with a better management of conflicting sources than the DS theory.

Within the context of change detection, the goal is not to achieve fusion of information provided by the different sources, but rather to directly model these changes using specific basic belief functions. In DS and DSm theories a "frame of discernment" defining the hypotheses has to be defined. Usually in change detection, the two elementary hypotheses are "change" and "no-change" with a frame of discernment defined as  $\Theta = \{\theta, \bar{\theta}\}$  [33]. The "hyper-power set"  $D^\Theta$  is derived from the frame of discernment  $\Theta$  and contains single and composite hypotheses as follows:

$$D^\Theta = \left\{ \emptyset, \theta, \bar{\theta}, \theta \cup \bar{\theta}, \theta \cap \bar{\theta}, \theta \cup \bar{\theta} \cap \theta, \theta \cap \bar{\theta} \cup \theta \right\}$$

(6)

In DSm theory, data are modeled using basic belief functions  $m$  defined in  $D^\Theta$  and verifying:

$$m : D^\Theta \rightarrow [0, 1] \left\{ \begin{array}{l} m(\emptyset) = 0 \\ \sum_{A \in D^\Theta} m(A) = 1 \end{array} \right.$$

(7)

where  $A$  is a proposition, either single or composite of  $\Theta$ . By analogy with the DS theory, the traditional belief functions of credibility and plausibility characterizing proposition  $A$  can be derived from  $m(A)$  as follows:

$$Cre(A) = \sum_{\substack{B \subseteq A \\ B \in D^\Theta}} m(B) \quad Pl(A) = \sum_{\substack{A \cap N \neq \emptyset \\ B \in D^\Theta}} m(B)$$

(8)

In order to apply the DSm theory within the context of change detection using the global and local change signature, the basic belief assignments were defined as provided in table 1 with the distance  $D(p, q)$  (eq. 3) and the contrast between membership change  $C(p, q)$  (eq. 4) where  $p$  and  $q$  are the probability density functions  $f(t)$  associated with images  $I$  and  $i$  is the class assigned to voxel  $t$  after the mapping step (section II.B.1).

A tumor evolution, modeled by the basic belief assignment  $m(\theta)$ , can be characterized by a high contrast between membership change and a significant distance between the pdfs. On the contrary, small values of contrast and distance between pdfs describe stable uptakes between the two scans and are modeled by the basic belief assignment  $m(\bar{\theta})$ . The two others basic belief assignments corresponding to the union and the intersection of  $\theta$  and  $\bar{\theta}$  characterize respectively the uncertain and the conflict. They are defined with respect to the belief mass function definition in (eq. 7). Figure 3 illustrates the pdfs corresponding to the tumor and background classes in the pre and post-treatment images. Each term in table 1 can be described and explained by a specific arrangement of the pdfs. The notions of uncertainty and paradox are stressed in the definition of  $m(\theta \cup \bar{\theta})$  and  $m(\theta \cap \bar{\theta})$ , respectively.

As illustrated in figure 1, in the map corresponding to the basic belief assignments of the intersection, voxels different between the pre and post-treatment scans are emphasized, which illustrates the conflict between the sources. By contrast, voxels similar in both scans are emphasized in the map corresponding to the basic belief assignments of the union, representing the uncertainty between the sources.

After the computation of the basic belief functions, a 'change map' (see figure 1) can be created through the maximum of credibility between the elementary hypotheses according to the credibility function defined in (eq. 8). This 'change map' provides a description of the evolution between the PET follow-up scans  $I_1$  and  $I_2$  by classifying the voxels as "change" or "no-change". According to (eq.8), the plausibility associated to the elementary hypotheses  $\theta$  and  $\bar{\theta}$  are always equal to 1. Therefore, the maximum of plausibility criterion was not used in our approach. The basic belief assignments of table 1 were defined in the general case in which several sources of information can be considered. In this paper, the change detection approach has been applied in the particular case where only one source of information is considered: the mean value in a cube of  $3 \times 3 \times 3$  voxels. In that case, the maximum of credibility criterion compares  $C(p, q)$  and  $[1 - D(p, q)]$ : if  $C(p, q) > [1 - D(p, q)]$ , then the voxel has changed, otherwise the voxel has not changed.

### **Classification**

The final step of the process is the definition of the voxel by voxel response status using the PET follow-up scans  $I_1$  and  $I_2$  in combination with the computed 'change map' (section II.B.3). The voxels were here classified according to four different statuses: responders, stable or progressive for the tumoral voxels, and those belonging to the background (or physiological) in both scans. On the one hand, the voxels identified as 'no change' in the change map could belong either to the background or to the tumor in both images. Thanks to the background and tumor means  $\mu_i^j$  computed in the mapping step, the 'no-change' voxels were classified as either background or tumor. This classification was achieved by the computation of a mean square error (MSE) between the voxel's value and the respective mean values of background and tumor in the pre treatment scan. Voxels classified as tumors and 'no change' are therefore designated as 'stable' (green voxels in the fusion map) On the other hand, the voxels identified as changing could be progressive (background evolving to tumor, red voxels in the fusion map) or responding (tumor evolving to background, blue voxels in the fusion map) and they were classified depending on the relative SUV variation between scans (decrease for responding, increase for progressive). It is then possible to deduce individual segmented maps from the fusion map for the pre and post-treatment scans. The segmentation map on the pre-treatment is deduced by the combination of voxels of the fusion map classified as responding and stable, whereas for the segmentation map on the post-treatment scan progressive and stable voxels are combined.  $SUV_{mean}$  and MATV are then extracted by copying these delineations onto each scan. The overall tumor response is then deduced from the measured  $\Delta SUV$  and  $\Delta MATV$ .

### **Validation**

The validation of the change detection method has been carried out on both simulated and clinical datasets.

#### **Simulated datasets**

The considered simulations were realistic MATVs designed based on real head and neck and esophageal lesions observed in clinical datasets, to generate realistic shapes and activity distributions as well as locations in the body. The simulated dataset used in our work was generated according to a procedure previously described using the Monte Carlo simulations toolbox GATE (Geant4 Application for Tomography Emission) [34]. This procedure combines a numerical phantom describing thorax (NCAT) [34] or head and neck (Zubal) [34] anatomical regions and a model of tumors manually created from real clinical images. Phantoms and tumors are both defined as Non-uniform rational B-splines. The voxelized phantom associated with attenuation coefficients and activity labels is then combined with a highly detailed model of the Philips Allegro PET scanner previously validated [35], and simulated in GATE in order to generate PET listmode data that is then reconstructed as 3D simulated images with voxel size  $4 \times 4 \times 4$  mm<sup>3</sup> (same as the clinical images), for which the ground-truth is known in the simulated phantoms (see examples in figure 4).

Each case contains two follow-up 3D PET scans, one pre and one post treatment. The original clinical cases used to design the simulated cases were classified as partial responders or progressive disease after radio-chemotherapy.

In this study, 25 simulated follow-up cases were considered, with various MATV shapes and tumor-to-background contrast ratios. Half of the simulated datasets, representing 13 out of 25 cases, were generated from patients' classified as partial responders. The 12 other cases were designed based on patients classified as progressive disease.

In order to assess the robustness of our method, three noise levels were considered for each case, by selecting 100, 80 and 60% of the simulated lines of response for the iterative reconstruction respectively. Consequently, 75 simulated follow-up cases were generated. Similarly to the observed clinical datasets, both homogenous and heterogeneous tracer uptake distributions within the MATV were simulated. The current implementation of our approach does not however allow finer characterization of the evolution of radiotracer uptake heterogeneities within the MATV since it does not explicitly model heterogeneity parameters, although depending on the delineations and appearance of pre and post treatment scans, such variations might still be highlighted in the generated maps. Consequently, in all simulated scans the voxels were assumed to belong either to the background (BD) or to the tumor (MATV). The table 2 provides the mean tumor volume and tumor-to-background ratios simulated for pre and post treatment scans.

#### **Evaluation metrics for simulated datasets**

For simulated datasets, the ground-truth (GT) is available. The assessment of the change detection method was achieved by computing relative volume errors (VE) and classification errors (CE) for each simulated tumor, as well as the errors in estimating the volume change. CE are defined as the sum of positive and negative classification errors (background voxels classified as tumor or tumor voxels classified as background respectively) [23]. As illustrated in the figure 4, for each simulated case, segmented maps of the pre and post treatment scans were obtained with the adaptive threshold (see section II.C.5) and the change detection method. VE and CE were larger than 100% in some cases for which adaptive threshold delineation completely failed, and were limited to 100% in such cases.

### ***Clinical datasets***

The proposed method was also applied to real clinical datasets. Seven patients with esophageal or head and neck cancer undergoing concomitant radiochemotherapy between 2005 and 2008 were considered. Two  $^{18}\text{F}$ -FDG PET scans were acquired on a PET/CT Philips Gemini scanner following standard clinical protocol before and after treatment. Images were reconstructed with voxel size  $4\times 4\times 4\text{ mm}^3$  (same as the simulated dataset). Four patients were classified as partial responders according to the RECIST criteria, one month after the completion of the therapy. The three others patients were classified as complete responders. According to the recent PERCIST recommendations including PET measurements, a response may be characterized by a SUV relative decrease above 30%, with no volume increase or apparition of new lesions [18]. This 30% value has been chosen because of the reproducibility of SUV measurements which has been determined to be within this  $\pm 30\%$  range. This reproducibility range was evaluated on double baseline (also called "test-retest") PET images, which consisted in repeated imaging of the patients at a few days interval without treatment in between. It has been recently demonstrated that using the FLAB method to delineate MATVs led to similar reproducibility performance of about  $\pm 30\%$  [14]. The authors therefore concluded that future PERCIST guidelines should consider MATV variations larger than 30% to characterize response or progression. Visual illustrations of two clinical cases are provided in the figure 6.

### ***Quantitative variation for clinical datasets***

Contrary to the simulated validation, no ground-truths were available for the clinical follow-up cases. Therefore, the evaluation of the methods was performed by comparing the measured variation of several quantitative indices with respect to the known response to therapy of the patients. This response was evaluated 1 month after the completion of the concomitant radio-chemotherapy using conventional thoraco-abdominal CT and endoscopy. Patients were classified as non responders (NR) including stable and progressive disease, partial responders (PR) or complete responders (CR). Response evaluation was based on CT evolution between pre-treatment and post-treatment scans using RECIST (Response Evaluation Criteria in Solid Tumours). Patients also underwent fibroscopy in case of partial or complete response. Complete response was confirmed by the absence of visible disease in the endoscopy and no viable tumor on biopsy. Partial CT response was confirmed by macroscopic residual ( $>10\%$  viable) on biopsy. No discordance was observed between pathological and CT evaluation. The variation of the MATV ( $\Delta\text{MATV}$ ) and mean SUV ( $\Delta\text{SUV}_{\text{mean}}$ ) between the pre-treatment and post-treatment PET scans were computed for both approaches using delineations on both scans deduced from the fusion maps (see figure 1 and section II.B.4) or directly obtained with the adaptive thresholding used by two different clinicians (see section II.C.5). These measurements were compared to the variation of maximum SUV ( $\Delta\text{SUV}_{\text{max}}$ ) with or without PVE correction. In order to measure the  $\text{SUV}_{\text{mean}}$  variations for complete responders patients, the VOI defined in the first follow-up scan were automatically copied onto the second scan, where no pathological uptake could be detected.

### ***Alternative approaches used for comparison***

In both simulated and clinical datasets, the change detection method was compared to semi-automated approaches that have been proposed for patient monitoring studies. Only the  $\text{SUV}_{\text{max}}$  variation is currently considered in clinical routine to characterize response. In order to take into account the mean SUV and MATV variations, adaptive threshold-based methods were used and applied to each scan independently. Several studies have demonstrated that the fixed threshold methods were not accurate and robust enough [36,37] and were therefore not considered here. The adaptive threshold method takes into account the background uptake and the applied threshold value is computed from the estimated contrast between the tumor activity and the background activity, and selected based on previously carried out optimization for a given scanner using phantom acquisitions of spheres [36].

However, the use of adaptive thresholding requires the manual placement of a ROI in the background for an estimation of its uptake. For the simulated datasets, the background uptake was simulated as homogeneous and the background value was automatically determined. Only one result for the adaptive thresholding is therefore provided for simulated datasets. On the other hand, clinical datasets often exhibit much more heterogeneous physiological background uptake, especially in the mediastinum and head and neck regions. The manual placement of the background region of interest may therefore lead to significant inter observer variability as previously demonstrated [38]. In our study, two nuclear medicine physicians with similar training and experience manually placed the background ROI on each PET scans of the clinical datasets, therefore leading to two different results denoted tb1 and tb2 in the tables and figures. They followed a specific protocol as they were instructed to place the ROI at least a few cm away from the lesions. They were however free to choose the actual size, shape and placement of the ROI.

## Results

### Results on simulated datasets

Illustrations of fusion maps obtained by applying the change detection method on simulated data are provided for two cases in figure 4 (e). For an easier interpretation of the maps, color codes have been associated with each type of response. Blue areas depict the responding tumor voxels. Voxels appearing in both scans (stable voxels) are characterized in green. Tumor voxels representing progressive disease are red. In both illustrated cases classified as partial responders, only green and blue areas appear. On each map, the voxel intensity is automatically set as the SUV relative variation ( $\Delta\text{SUV}$ ) between the first and the second follow-up scan. In the first case, no significant differences were observed on the segmentation of the pre and post treatment scans for both approaches. The second case illustrates a heterogeneous uptake in the tumor of the pre treatment scan. The lesion partially responded to the treatment and exhibited a homogenous uptake in the second scan. The obtained segmentations of the MATV in the post treatment scan are similar for both approaches. By contrast, the adaptive method clearly underestimated the MATV in the pre-treatment scan, contrary to the change detection method.

Figure 5 provides the VE and CE obtained by the adaptive threshold and change detection methods. The threshold-based method was applied independently on each scan. For both approaches,  $\text{VE}_1$  and  $\text{VE}_2$ ,  $\text{CE}_1$  and  $\text{CE}_2$  are the volume and classifications errors respectively assessed for the pre-treatment and post treatment scans. For both scans, the VE and CE associated to the threshold method ( $\text{VE}_1 : -25\pm 20\%$ ,  $\text{VE}_2 : -24\pm 29\%$ ,  $\text{CE}_1 : 27\pm 16\%$ ,  $\text{CE}_2 : 26\pm 25\%$ ) were significantly ( $p < 0.01$ ) higher than the ones associated with the use of the change detection approach ( $\text{VE}_1 : +2\pm 26\%$ ,  $\text{VE}_2 : -2\pm 31\%$ ,  $\text{CE}_1 : 17\pm 6\%$ ,  $\text{CE}_2 : 19\pm 7\%$ ). Standard deviations of VE for the proposed method were higher than the adaptive thresholding (26% vs. 20% and 31% vs. 29%) although the difference was not significant ( $p > 0.05$ ), whereas the standard deviations of CE were significantly ( $p < 0.001$ ) smaller (6% and 7% vs. 16% and 25%). Adaptive threshold led to a systematic under estimation in most of the cases, with some errors above 100%, whereas the change detection approach led to a mix of over and under evaluations of the true volumes (as demonstrated by the low mean VE and associated standard deviations), with no absolute errors (either VE or CE) above 30%. In these simulated cases, both homogenous and heterogeneous uptakes were simulated in the MATV according to the observed clinical scan. The adaptive threshold consistently failed to delineate heterogeneous ones, contrary to our change detection approach that prove much more robust in such configurations. Both methods provided similar errors in estimating the MATV evolution between the two scans ( $23\pm 45\%$  for change detection approach and  $25\pm 47\%$  for the adaptive threshold).

The impact of the registration error has been measured for a sub-group of the simulated data, by shifting the second follow-up scan from one and two voxels in a random direction. For one voxel shift (7 mm), volume errors of the first and the second follow-up scan increased to  $+11.4\pm 21\%$  and  $+5.3\pm 12\%$ . By contrast, a shift of two voxels (14mm) led to significantly ( $p < 0.01$ ) higher VE for both follow-up scans, respectively  $+27.5\pm 28\%$  and  $+19.6\pm 26\%$ .

### Results on clinical datasets

The results obtained on clinical datasets are illustrated in figure 6 for two patients, with the same set of color and voxels intensities as described in the simulated images validation.

The quantitative measurements for each patient are provided in table 3. According to the response status, mean measurements were computed and presented in table 4 and 5 for the patients classified as complete and partial responders respectively. Contrary to the simulated datasets, the clinical images have a noisier and more heterogeneous background making the change detection more challenging.

The patients 1 to 3 were classified as complete responders to the therapy. Therefore, the volume variation was set at  $\Delta V = -100\%$ . Only the variation of  $\text{SUV}_{\text{mean}}$  computed for both methodologies were compared with the  $\Delta\text{SUV}_{\text{max}}$  with or without PVE correction. For all the CR cases, the  $\Delta\text{SUV}_{\text{mean}}$  were close ( $-74\pm 10\%$  for  $T_1$ ,  $-73\pm 11\%$  for  $T_2$  and  $-64\pm 14\%$  for the change detection method). These measures are close to those obtained with the  $\Delta\text{SUV}_{\text{max}}$  ( $-56\pm 14\%$  without the PVC and  $-66\pm 20\%$  with PVC). As expected, PVC led to a higher variation (and associated standard deviation) of the  $\Delta\text{SUV}_{\text{max}}$ .

Regarding patients 4 to 7 classified as partial responders (PR), the lesions were expected to exhibit a significant decrease in activity and volume between the first and the second follow-up scan. In the recent PERCIST criteria, PR is characterized by a  $\Delta\text{SUV}_{\text{peak}}$  decrease higher than 30%. Moreover, a decrease of the MATV is also expected between the follow-up scans. Considered as the gold standard in the clinical routine, the  $\text{SUV}_{\text{max}}$  variation was assessed as higher than  $-30\%$  ( $-39\pm 16\%$  without PVC and  $-48\pm 14\%$  with PVC). These measurements are similar to the  $\text{SUV}_{\text{mean}}$  variations computed with both approaches. No significant differences were observed in the mean variation of  $\text{SUV}_{\text{mean}}$  with both clinicians ( $-39\pm 14\%$  for  $T_1$ ,  $-46\pm 14\%$  for  $T_2$ ), which are slightly lower than the measurements computed with the change detection method ( $-49\pm 18\%$ ). On the contrary, the variations of MATV were found to be significantly different between both clinicians ( $-67\pm 22\%$  for  $T_1$ ,  $-44\pm 27\%$  for  $T_2$ ,  $p < 0.01$ ) as well as with the change detection method ( $-36\pm 8\%$ ) ( $p < 0.001$ ). The standard deviations associated with the adaptive threshold method (22 and 27%) were significantly ( $p < 0.001$ ) higher than the one estimated with the change detection method (8%).

In order to illustrate the respective behaviors of the adaptive threshold and change detection methods, two clinical follow-up cases corresponding to patients 4 and 6 are provided in the figure 6.

For patient 4, whereas the variation of  $SUV_{mean}$  are close for  $T_1$  and  $T_2$  (-39% and -51% respectively), the MATV variations are much more different (-83% and -55% respectively). The measurements obtained with the change detection method are in line with the PR status of the patient ( $\Delta SUV_{mean} = -57\%$ ,  $\Delta MATV = -28\%$ ). The measurements obtained for the patient 6 illustrate the poor reproducibility of the adaptive threshold method. Whereas the segmentations obtained for the post treatment scan are similar for both clinicians, the segmentations obtained for the first pre-treatment differ. These results can be explained by the heterogeneous distribution of tracer uptake within the background of the mediastinum combined with the fact that each clinician placed the background region of interest in a different position, leading to large differences in the estimation of this background uptake used by the adaptive threshold to compute a solution. This is consistent with previous observations in esophageal cancer [38], whereas in lung cancer the difference was not significant due to more homogeneous uptake of FDG in the lungs [39]. This led to high differences between  $\Delta MATV$  measured by the two clinicians (-38% for  $T_1$ , -5% for  $T_2$ ).  $\Delta SUV_{mean}$  variation however, were quite close (-27% for  $T_1$ , -34% for  $T_2$ ). By contrast to the adaptive method, the change detection approach succeeded in identifying both MATVs, as it can be seen in the segmented maps. Hence, the quantitative variation of SUV and volume measured with the change detection approach are coherent with the partial response ( $\Delta SUV_{mean} = -31\%$ ,  $\Delta MATV = -37\%$ ).

For all PR patients,  $SUV_{mean}$  and MATV decrease obtained with the change detection approach were below -30%, whereas several results obtained using the adaptive threshold were not (-30%  $SUV_{mean}$  for patient 5 (Tb 1), -25%  $SUV_{mean}$  for patient 6 using (Tb 1) and -5% MATV (Tb 2).

## Discussion

In clinical oncology, one application of PET imaging is the assessment of early response to therapy by analyzing sequential scans during treatment [5–7]. Up to now, most of proposed studies characterizing the therapeutic response have been based on the measurement of the variation of a single quantitative parameter, usually the maximum SUV.

The proposed method is aimed at assessing the response using follow-up PET scans, through a change detection approach. The method is divided in subsequent steps of individual images mapping, change signature of global and local parameters and change detection using the Dezert-Smarandache (DSm) theory (see figure 1). Allowing the combination of multiple sources of information and modeling the uncertainty or conflict between these sources, the change detection with DSm theory applied on follow-up scans was expected to produce more reliable results than independent segmentations performed on each scan separately. However, the method requires accurate rigid registration of both images, which can limit the applicability of the approach to body regions and types of lesions less impacted by such issues. To reliably apply the method in more complex cases, clinical acquisitions protocols would require additional standardization to minimize changes, and respiratory motion as well as potential additional corrections would be required on each image before applying this approach. The last step of the change detection method consists in classifying the local evolution of the MATV using recently proposed criteria [18].

The change detection method was applied on both simulated and clinical follow-up datasets, and compared with adaptive threshold-based delineations performed separately on each scan. With simulated datasets, the ground-truths associated to each scan allowed the computation of volume and classifications errors. 25 follow-up cases, composed of two scans acquired before and during the treatment, were simulated with various tumor-to-background ratios, tumor shapes, levels of noise and with homogeneous or heterogeneous tracer activity within the MATV. In both pre and post treatment scans, the VE and CE associated with the adaptive threshold method were significantly higher than those measured with the change detection method. These high errors obtained with adaptive threshold can be explained by its inability to accurately take into account heterogeneous distributions of radiotracer within the MATV. The change detection method appeared more resilient to heterogeneous uptake within the tumor. Although CE associated with the proposed method were measured around 20%, this performance needs to be assessed within the context of the size of the tumors relatively to the voxel size ( $4 \times 4 \times 4 \text{ mm}^3$ ), especially for post treatment lesions that have shrunk. Although both methods provided similar accuracy in estimating the volume evolution between two scans ( $\sim 25 \pm 45\%$ ), a correct volume evolution can be also reached by two incorrect delineations in the two scans. In several cases concerning the adaptive thresholding approach, complete failures with errors  $>100\%$  were obtained but were limited to 100% to avoid excessive bias. Classification errors were significantly lower for both pre- and post-treatment with the proposed approach. This has to be considered since the goal of the proposed method is not only to provide an estimation of the volume evolution between the two scans (the only metric that could be used in the case of clinical datasets), but also to provide more accurate delineations in both pre- and post-treatment images, as well as a quantitative fusion map describing voxel-by-voxel evolutions with color coding.

In clinical datasets, the adaptive threshold and change detection methods were compared with the variation of maximum SUV in the follow-up images. The  $SUV_{mean}$  measurements were also computed for both methods. Finally, in order to more accurately characterize the response, MATV variations were computed for each approach.



For patients classified as complete responders, similar measurements were observed between  $SUV_{max}$  and  $SUV_{mean}$  variations estimated with the adaptive threshold and the change detection method. Regarding the four patients classified as partial responders, there were no significant differences between the variation of  $SUV_{max}$  and  $SUV_{mean}$  assessed by both approaches. By contrast, the MATV variations were significantly different for the adaptive method results obtained by the two clinicians and the developed approach. The MATV variations measured with the change detection approach were less variable and more consistent with the status of partial response. It also led to visually more satisfying delineations than adaptive threshold that prove more sensitive to noise and uptake heterogeneities in both MATV and background.

In this study, only the evolution of the tumor volume as a whole was characterized, although since the changes are characterized voxel by voxel, local changes in intra-tumor activity can be highlighted by the current implementation of the method, and may for instance highlight appearance of necrosis or disappearance of high uptake regions within the lesions. In further developments, the finer characterization of the evolution of tracer uptake heterogeneity within the MATV will be considered. The use of textural features analysis to characterize tracer uptake heterogeneity within MATV in PET images [40] could provide a certain number of additional features characterizing local and regional changes and could be combined with those already implemented in the use of the DSM theory presented here, especially regarding the combination of several sources of information, for example by considering two sizes of the estimation cube ( $3 \times 3 \times 3$  and  $5 \times 5 \times 5$  for example), or additional textural features analysis.

## Conclusion

A novel method based on the change detection principle has been proposed to estimate a local response to therapy in oncology. Using the recent paradoxical theory, modeling the conflict and imprecision between the fused sources, the change detection approach aims at detecting the tumor evolutions more accurately than independent segmentations performed on each PET image separately. In this study the proposed method was validated on realistic simulated images and applied to clinical datasets. It was compared favorably to threshold-based method applied separately on each follow-up scans. In the simulated validation, the adaptive threshold approach led in both pre and post treatment images to higher errors than the change detection method. Regarding the clinical datasets, whereas no difference was observed in the case of patients classified as complete responders, the adaptive threshold was much less resilient to the noise and tracer uptake heterogeneity than our proposed method in the case of partial responders. For these cases, the proposed approach led to measurements more consistent with the response status of the patients. Further work will consist in applying the change detection method on more extensive clinical datasets in the context of therapy assessment in oncology, although some limitations of the applicability of the methods have been identified for lesions in areas more impacted by respiratory motion or other inter-scans registration issues. Moreover, the method will be extended to explicitly model tracer uptake heterogeneity evolution using textural features characterization and quantification.

## Acknowledgements:

This work was partly funded by GE Healthcare and the Brittany region.

## Footnotes:

Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

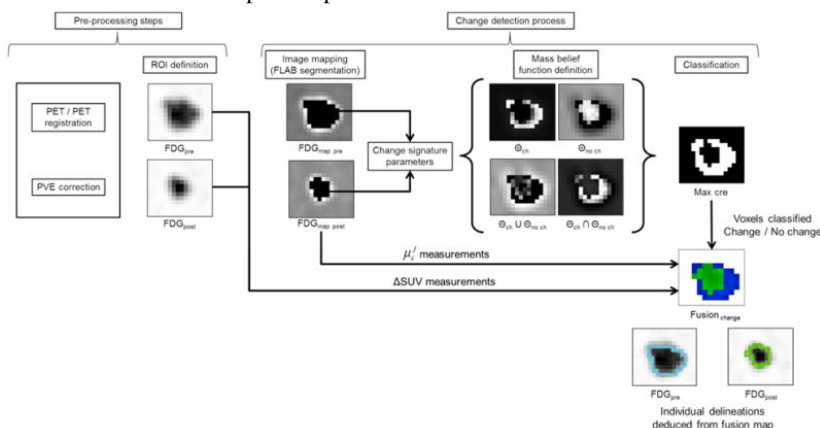
## References:

1. Jerusalem G, Withofs N, Rorive A, Hustinx R. Positron emission tomography in oncology: an update. *La Revue Du Praticien*. 2007; 57: (17) 1864 - 1870
2. Römer W, Hanauske A-R, Ziegler S, Thödtmann R, Weber W, Fuchs C, Enne W, Herz M, Nerl C, Garbrecht M, Schwaiger M. Positron emission tomography in Non-Hodgkin's lymphoma: Assessment of chemotherapy with fluorodeoxyglucose. *Blood*. 1998; 91: (12) 4464 - 4471
3. Jarritt H, Carson K, Hounsfield AR, Visvikis D. The role of PET/CT scanning in radiotherapy planning. *British Journal of Radiology*. 2006; 79: S27 - S35
4. Weber W. Assessing tumor response to therapy. *J Nucl Med: Official Publication, Society of Nuclear Medicine*. 2009; 50: (Suppl 1) 1S - 10S
5. Day FL, Link E, Ngan S, Leong T, Moodie K, Lynch C, Michael M, Winton E, Hogg A, Hicks RJ, Heriot A. FDG-PET metabolic response predicts outcomes in anal cancer managed with chemoradiotherapy. *Br J Cancer*. 2011; 105: (4) 498 - 504
6. Fei B, Wang H, Wu C, Chiu SM. Choline PET for monitoring early tumor response to photodynamic therapy. *J Nucl Med*. 2010; 51: (1) 130 - 8
7. Huang W, Zhou T, Ma L, Sun H, Gong H, Wang J, Yu J, Li B. Standard uptake value and metabolic tumor volume of (18)F-FDG PET/CT predict short-term outcome early in the course of chemoradiotherapy in advanced non-small cell lung cancer. *Eur J Nucl Med Mol Imaging*. 2011; 38: (9) 1628 - 35
8. Weber W. <sup>18</sup>F-FDG PET in Non-Hodgkin's Lymphoma: Qualitative or Quantitative?. *J Nucl Med*. 2007; 48: 1580 - 1582
9. Bentzen SM. Theragnostic imaging for radiation oncology: dose-painting by numbers. *Lancet Oncol*. 2005; 6: 112 - 117
10. Lin C, Itti E, Haioun C. Early <sup>18</sup>F-FDG PET for prediction of prognosis in patient with diffuse large B-cell lymphoma: SUV-based assessment versus visual analysis. *J Nucl Med*. 2007; 48: 1626 - 1632
11. Weber W, Ziegler S, Thodtmann R, Hanauske A, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med*. 1999; 40: 1771 - 1777 158 - 159 -
12. Nahmias C, Wahl L. Reproducibility of Standardized Uptake Value Measurements Determined by 18F-FDG PET in Malignant Tumors. *J Nucl Med*. 2008; 49: 1804 - 1808
13. Krak NC, Boellaard R. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *European Journal of Nuclear Medicine and Molecular Imaging*. 2005; 32: 294 - 301
14. Hatt M, Cheze Le Rest C, Aboagye EO, Kenny LM, Rosso L, Turkheimer FE, Albarghach NM, Pradier O, Visvikis D. Reproducibility of 18F-FDG and 18F-FLT PET tumor volume measurements. *Journal of Nuclear Med*. 2010; 51: (9) 1368 - 1376

- 15 . Larson S , Erdi Y , Akhurst T , Mazumdar M , Macapinlac H , Finn R , Casilla C , Fazzari M , Srivastava N , Yeung H . Tumor Treatment Response Based on Visual and Quantitative Changes in Globals Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis . Clin Positron Imaging . 1999 ; 2 : 159 - 171.160
- 16 . World Health Organization, . WHO handbook for reporting results of cancer treatment . WHO offset publication . 1979 ; (48 )
- 17 . Therasse P , Arbuck S , Eisenhauer E , Wanders J , Kaplan R , Rubinstein L , Verweij J , Van Glabbeke M , van Oosterom A , Christian M . New guidelines to evaluate the response to treatment in solid tumors . Journal of National Cancer Institute . 2000 ; 92 : 205 - 216
- 18 . Wahl R , Jacene H , Kasamon Y , Lodge M . From RECIST to PERCIST : Evolving Considerations for PET Response Criteria in Solid Tumors . 2009 ; J Nucl Med . 50 : 122 - 150
- 19 . Moffat B . Functional diffusion map: a non invasive MRI biomarker for early stratification of clinical brain tumor response . Proceedings of the National Academy of Sciences of the United States of America . 2005 ; 102 : (15 ) 5524 - 5529
- 20 . Necib H , Dusart M , Tylski P , Vanderlinden B , Buvat I . Detection the tumor changes between two FDG PET scans using parametric imaging . J Nucl Med Meeting Abstracts . 2008 ; 49 : 121 - 161 -
- 21 . Bruzzone L , Prieto DF . Automatic analysis of the difference image for unsupervised change detection . IEEE Transaction on Geoscience and Remote Sensing . 2000 ; 38 : (3 ) 1171 - 1182
- 22 . Le Hégarat-Masclé S , Seltz R . Automatic change detection by evidential fusion of change indices . Remote Sensing of Environment . 2004 ; 91 : (3-4 ) 390 - 404
- 23 . Hatt M , Cheze le Rest C , Turzo A , Roux C , Visvikis D . A fuzzy locally adaptive bayesian segmentation approach for volume determination in PET . IEEE Transactions on Medical Imaging . 2009 ; 28 : (6 ) 881 - 893
- 24 . Smarandache F , Dezert J . Advances and applications of DSMT for information fusion : collected works . American Research Press ; 2004 ;
- 25 . Rajendran JG , Schwartz DL , O'Sullivan J . Tumour hypoxia imaging with 18F fluoromisonidazole positron emission tomography in head and neck cancer . Clin Cancer Res . 2006 ; 12 : 5435 - 5441
- 26 . Soret M , Bacharach SL , Buvat I . Partial-volume effect in PET tumor imaging . J Nucl Med . 2007 ; 48 : 932 - 945
- 27 . Boussion N , Cheze Le Rest C , Hatt M , Visvikis D . Incorporation of wavelet based denoising in iterative deconvolution for partial volume correction in whole body PET imaging . European Journal of Nuclear Medicine and Molecular Imaging . 2008 ; 36 : (7 ) 1064 - 75
- 28 . Le Pogam A , Descourt P , Hatt M , Boussion N , Visvikis D . A combined 3-D wavelet and curvelet approach for edge preserving denoising in emission tomography . Journal of Nuclear Medicine . 2009 ; 50 : (S2 ) 533 -
- 29 . Ouksili Z . Accurate PET/PET registration of serial to assess lung tumor evolution . 4th IEEE International Symposium on Biomedical Imaging 2007 ; 732 - 735
- 30 . Mercier G , Derrode S . SAR image change detection using distance between distributions of classes . Proceedings of Geoscience and Remote Sensing Symposium . 2004 ; 6 : 3872 - 3875
- 31 . Celeux G , Diebolt J . L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités . Revue de statistique appliquée . 1986 ; 34 : (2 )
- 32 . Bhattacharyya A . On a measure of divergence between two statistical populations defined by their probability distributions . Bull Calcutta Math Soc . 35 : 1943 ; 99 - 109
- 33 . Coppin P , Jonckheere I , Nackaerts K , Muys B , Lambin E . Digital change detection methods in ecosystem monitoring: a review . International Journal of Remote Sensing . 2004 ; 25 : (9 ) 1565 - 1596
- 34 . Le Maître A , Segars WP , Reilhac A , Hatt M , Tomei S , Lartizien C , Visvikis D . Incorporating patient specific variability in the simulation of realistic whole body 18F-FDG distributions for oncology applications . Proceedings of the IEEE . 2009 ; 97 : (12 ) 2026 - 2038
- 35 . Lamare F , Turzo A , Bizais Y , Le Rest CC , Visvikis D . Validation of a Monte Carlo simulation of the Philips Allegro/GEMINI PET systems using GATE . Phys Med Biol . 2006 ; 51 : (4 ) 943 - 62
- 36 . Nestle U , Kremp S , Schaefer-Schuler A , Sebastian-Welch C , Hellwig D , Rube C , Kirsch CM . Comparison of Different Methods for Delineation of 18F-FDG PET-Positive Tissue for Target Volume Definition in Radiotherapy of Patients with Non-Small Cell Lung Cancer . Journal of Nuclear Medicine . 2005 ; 46 : (8 ) 1342 - 8
- 37 . Tylski P , Stute S , Grotu N , Doyeux K , Hapdey S , Gardi I , Vanderlinden B , Buvat I . Comparative assessment of methods for estimating tumor volume and standardized Uptake Value in FDG PET . J Nucl Med . 2010 ; 51 : 268 - 276
- 38 . Hatt M , Visvikis D , Albarghach NM , Tixier F , Pradier O , Cheze-le Rest C . Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumor delineation methodology . European Journal of Nuclear Medicine and Molecular Imaging . 2011 ; 38 : (7 ) 1191 - 1202
- 39 . Hatt M , Cheze-le Rest C , van Baardwijk A , Lambin P , Pradier O , Visvikis D . Impact of tumor size and tracer uptake heterogeneity in <sup>18</sup>F-FDG PET and CT non-small cell lung cancer tumor delineation . Journal of Nuclear Medicine . 2011 ; Nov 52 : (11 ) 1690 - 7
- 40 . Tixier F , Cheze-Le Rest C , Hatt M , Albargach NM , Pradier O , Metges J-P , Corcos L , Visvikis D . Intra-tumor heterogeneity characterized by textural features on baseline <sup>18</sup>F-FDG PET images predicts response to concomitant radio-chemotherapy in esophageal cancer . Journal of Nuclear Medicine . 2011 ; 52 : (3 ) 369 - 378 O,M,GB

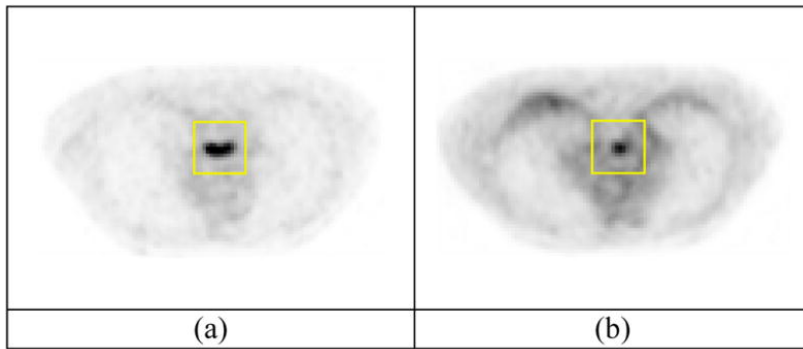
**Fig 1**

Flowchart of the proposed approach. In the image mapping step, black voxels are those classified as belonging to the MATV by the FLAB algorithm. The remaining voxels are classified as belonging to the background. In the mass function assignment definition, for all images, the voxels intensities represent values between 0 and 1. The maximum of credibility generates a binary 'change map' (white voxels 'change', black voxels 'no change'). In the classification step, using this 'change map' in combination with SUV differences between pre and post treatment as well as the mean measurements of both pre and post treatment scans, the fusion map is deduced in which green voxels are stable ('change' with less than 30% evolution or 'no change' classified as tumors depending on mean measurements), blue voxels are responding ('change' with more than 30% reduction), red voxels are progressive ('change' with more than 30% increase), and the voxel intensity is proportional to the SUV difference between pre and post treatment.

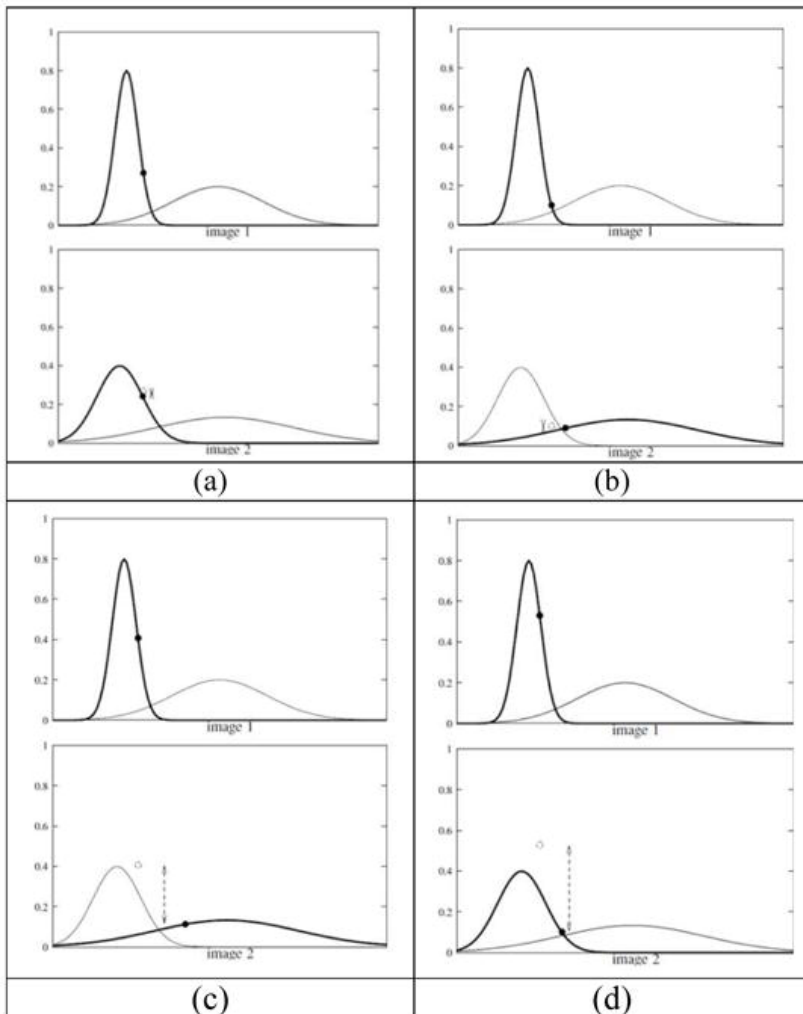


**Fig. 2**

Illustration of a VOI definition in the pre-treatment scan (a) and automatically reported on the registered mid-treatment scan (b).

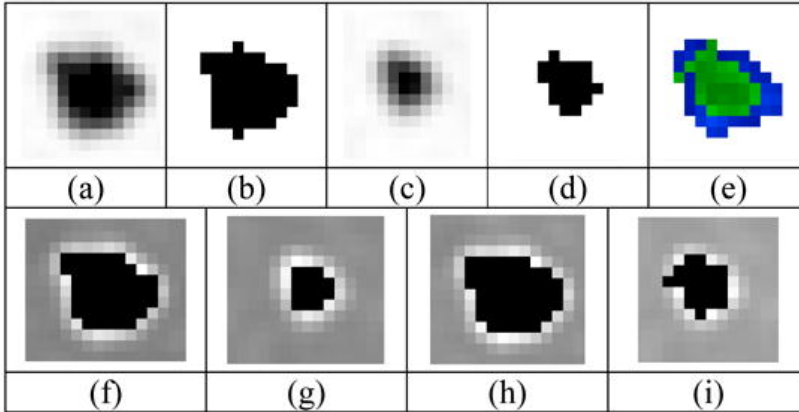
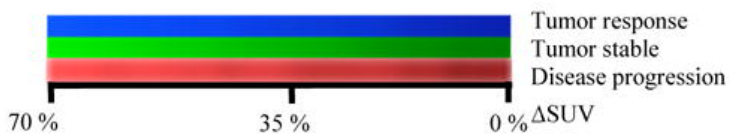
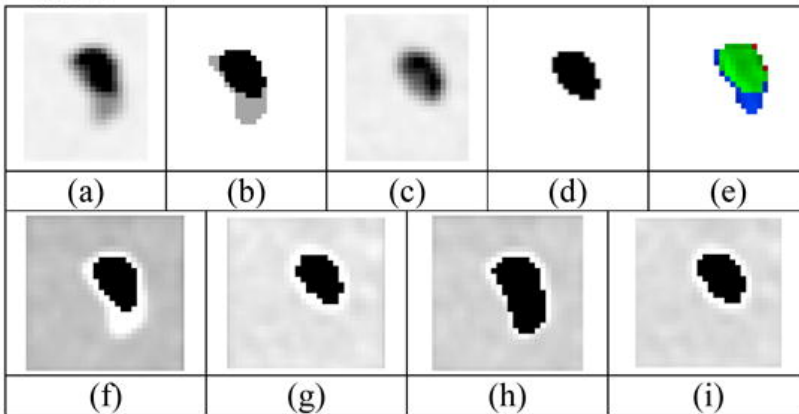
**Fig. 3**

The pdf of both classes, before treatment (image 1) and after treatment (image 2), is represented along the y-axis, as a function of voxel intensity (along the x-axis). In each figure, the intensity of one voxel ( $t$ ) is represented, before and after treatment, by a black circle. The pdf of the class assigned to voxel  $t$  is represented by the bold line. (a) ('no ch'). Voxel  $t$  was assigned to the same class before and after treatment ( $D(p,q)$  is low) and the membership degree did not change much ( $C(p,q)$  is low): the confidence that the voxel did not change is high. (b) ('ch\_or\_noch'). Voxel  $t$  was assigned to a different class before and after treatment ( $D(p,q)$  is high) with however a small change in membership degree ( $C(p,q)$  is low). The fact that  $C(p,q)$  is low suggests that pixel  $t$  may be close to the frontier between classes. Consequently the class 'change' may just be due to chance and the uncertainty is high. (c) ('ch'). Voxel  $t$  was assigned to a different class before and after treatment ( $D(p,q)$  is high). Because the membership degree change was high ( $C(p,q)$  is high), voxel  $t$  cannot be close to the frontier between classes at least in one image. Consequently, the class change is not likely to be due to chance. The confidence is therefore high that the voxel changed. (d) ('ch\_and\_noch'). Voxel  $t$  was assigned to the same class before and after treatment ( $D(p,q)$  is low) with however a high change in the membership degree ( $C(p,q)$  is high), which might be considered paradoxical, and might suggest the voxel evolution is not complete between the two scans.



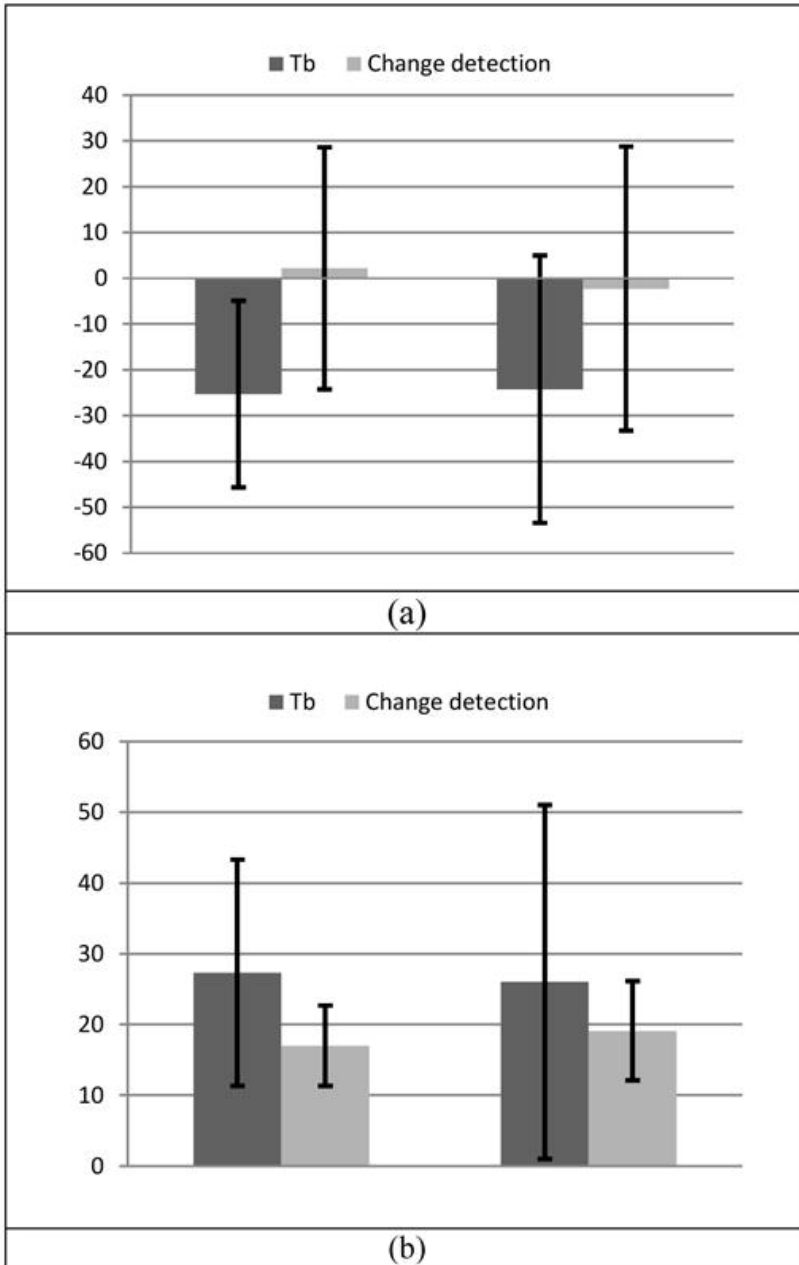
**Fig 4**

Illustration of two simulated cases (only one central 2D slice of the 3D volume is shown), with (a) the simulated pre treatment PET scan and (b) its binary ground-truth, (c) the simulated post treatment PET scan and (d) its binary ground-truth. (e) is the fusion map obtained with the proposed method. (f) and (g) are individual delineations on pre and post treatment PET scans respectively, obtained with the adaptive threshold. (h) and (i) are individual delineations obtained on the pre and post treatment PET scans respectively, by selecting voxels using the fusion map (e): for the pre treatment scan, tumor volume is obtained using responding (blue) and stable (green) voxels of the fusion map (e) whereas for the post treatment scan, tumor volume is obtained using stable and progressive (red) voxels. In this example there are no progressive voxels.

**Case 1****Case 2**

**Fig 5**

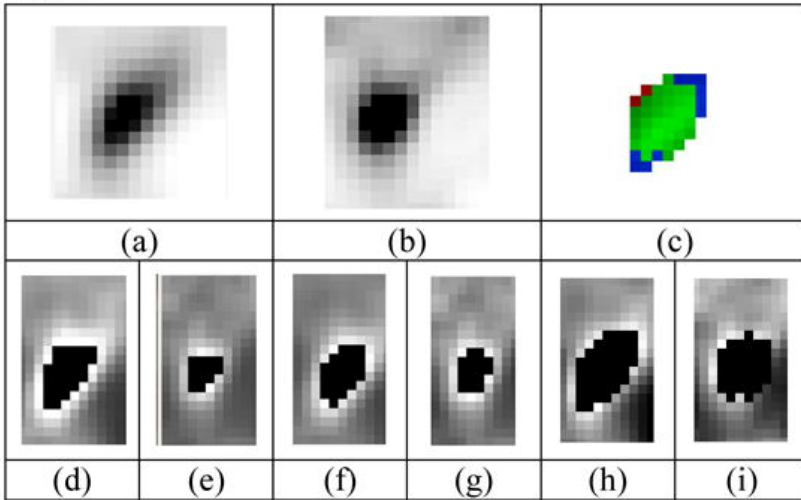
Mean and standard deviation for (a) VE and (b) CE, for the pre ( $EV_1, EC_1$ ) and post ( $EV_2, EC_2$ ) treatment scans, for adaptive threshold (tb) and change detection methods applied to the simulated cases.



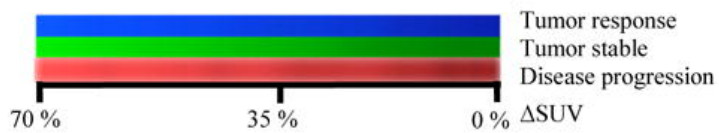
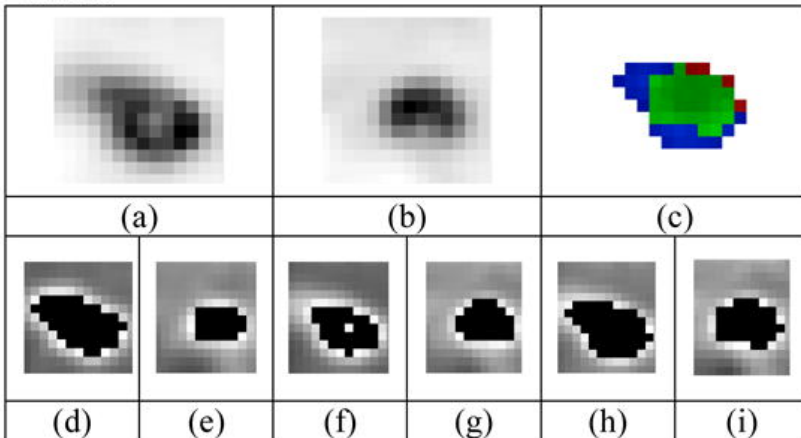
**Fig 6**

Illustration of the results for the clinical cases of patients 4 and 6, with (a) and (b) the pre and post treatment PET images respectively, (c) the classification map, (d–e) and (f–g) individual segmented maps for pre and post treatment obtained with the adaptive thresholding for the two clinicians, and (h–i) the segmentations obtained through the change detection method.

Case 1



Case 2



**TABLE I**

Hypothesis	Mass
$\emptyset$	0
$\theta$	$C(p, q) \times D(p, q)$
$\bar{\theta}$	$(1 - C(p, q)) \times (1 - D(p, q))$
$\theta \cup \bar{\theta}$	$(1 - C(p, q)) \times D(p, q)$
$\theta \cap \bar{\theta}$	$C(p, q) \times (1 - D(p, q))$

Definition of the mass function assignment with the distance  $D(p, q)$  and the contrast between membership change  $C(p, q)$  where  $p$  and  $q$  are the probability density function  $f(t)$  associated to images  $I$  and  $i$  is the class affected to voxel  $t$ .

**TABLE II**

	Tumour volume (cm <sup>3</sup> )		T/B ratio	
PET 1	27.5 ± 19	(3.8 – 90)	5.3 ± 2	(2.7 – 9.8)
PET 2	26.3 ± 23	(1.9 – 100.9)	4.4 ± 1.7	(2.0 – 8)

Mean tumor to background ratios and mean tumor volumes computed for all the simulated follow-up cases.

**TABLE III**

Patient	Method	$\Delta\text{SUV}_{\text{mean}}$ (%)	$\Delta\text{MATV}$ (%)	$\Delta\text{SUV}_{\text{max}}$ without PVC (%)	$\Delta\text{SUV}_{\text{max}}$ with PVC (%)
1	Tb 1	-73	-100	-63	-77
	Tb 2	-74	-100		
	Change	-67	-100		
2-1	Tb 1	-71	-100	-57	-69
	Tb 2	-70	-100		
	Change	-53	-100		
2-2	Tb 1	-64	-100	-37	-36
	Tb 2	-60	-100		
	Change	-54	-100		
3	Tb 1	-87	-100	-69	-81
	Tb 2	-87	-100		
	Change	-82	-100		
4	Tb 1	-39	-83	-36	-49
	Tb 2	-51	-55		
	Change	-57	-34		
5	Tb 1	-30	-62	-26	-36
	Tb 2	-35	-47		
	Change	-38	-33		
6	Tb 1	-27	-38	-33	-41
	Tb 2	-34	-5		
	Change	-31	-37		
7	Tb 1	-59	-84	-62	-67
	Tb 2	-65	-68		
	Change	-70	-46		

Measurements of MATV,  $\text{SUV}_{\text{mean}}$  and  $\text{SUV}_{\text{max}}$  evolution, computed with the adaptive threshold and the change detection methods for each patients.

**TABLE IV**

Patient	Method	$\Delta\text{SUV}_{\text{mean}}$ (%)	$\Delta\text{MATV}$ (%)	$\Delta\text{SUV}_{\text{max}}$ without PVC (%)	$\Delta\text{SUV}_{\text{max}}$ with PVC (%)
1-3	Tb 1	$-73.6 \pm 9.5$	-100	$-56.3 \pm 13.9$	$-65.6 \pm 20.3$
	Tb 2	$-72.7 \pm 11.1$	-100		
	Change	$-64.2 \pm 13.7$	-100		

Mean measurements of MATV,  $\text{SUV}_{\text{mean}}$  and  $\text{SUV}_{\text{max}}$  evolution for patients classified complete responders, computed with the adaptive threshold and the change detection methods for all the patients.

**TABLE V**

Patient	Method	$\Delta\text{SUV}_{\text{mean}}$ (%)	$\Delta\text{MATV}$ (%)	$\Delta\text{SUV}_{\text{max}}$ without PVC (%)	$\Delta\text{SUV}_{\text{max}}$ with PVC (%)
4-7	Tb 1	$-38.6 \pm 14.2$	$-66.8 \pm 21.6$	$-38.9 \pm 15.6$	$-47.9 \pm 13.7$
	Tb 2	$-46.1 \pm 14.4$	$-43.5 \pm 27.4$		
	Change	$-49.1 \pm 17.7$	$-36.6 \pm 7.5$		

Mean measurements of MATV,  $\text{SUV}_{\text{mean}}$  and  $\text{SUV}_{\text{max}}$  evolution for patients classified partial responders, computed with the adaptive threshold and the change detection methods for all the patients.