**EJNMMI Research**
a SpringerOpen Journal

**RESEARCH ARTICLE**　　　　　　　　　　　　　　　　　　　**Open Access**

# Metabolically active tumour volume segmentation from dynamic [18F]FLT PET studies in non-small cell lung cancer

Lieke L Hoyng[1], Virginie Frings[1], Otto S Hoekstra[1], Laura M Kenny[2], Eric O Aboagye[2] and Ronald Boellaard[1*]

## Abstract

**Background:** Positron emission tomography (PET) with 18F-3′-deoxy-3′-fluorothymidine ([18F]FLT) can be used to assess tumour proliferation. A kinetic-filtering (KF) classification algorithm has been suggested for segmentation of tumours in dynamic [18F]FLT PET data. The aim of the present study was to evaluate KF segmentation and its test-retest performance in [18F]FLT PET in non-small cell lung cancer (NSCLC) patients.

**Methods:** Nine NSCLC patients underwent two 60-min dynamic [18F]FLT PET scans within 7 days prior to treatment. Dynamic scans were reconstructed with filtered back projection (FBP) as well as with ordered subsets expectation maximisation (OSEM). Twenty-eight lesions were identified by an experienced physician. Segmentation was performed using KF applied to the dynamic data set and a source-to-background corrected 50% threshold (A50%) was applied to the sum image of the last three frames (45- to 60-min p.i.). Furthermore, several adaptations of KF were tested. Both for KF and A50% test-retest (TRT) variability of metabolically active tumour volume and standard uptake value (SUV) were evaluated.

**Results:** KF performed better on OSEM- than on FBP-reconstructed PET images. The original KF implementation segmented 15 out of 28 lesions, whereas A50% segmented each lesion. Adapted KF versions, however, were able to segment 26 out of 28 lesions. In the best performing adapted versions, metabolically active tumour volume and SUV TRT variability was similar to those of A50%. KF misclassified certain tumour areas as vertebrae or liver tissue, which was shown to be related to heterogeneous [18F]FLT uptake areas within the tumour.

**Conclusions:** For [18F]FLT PET studies in NSCLC patients, KF and A50% show comparable tumour volume segmentation performance. The KF method needs, however, a site-specific optimisation. The A50% is therefore a good alternative for tumour segmentation in NSCLC [18F]FLT PET studies in multicentre studies. Yet, it was observed that KF has the potential to subsegment lesions in high and low proliferative areas.

**Keywords:** 18F-FLT; Dynamic; PET; NSCLC; Kinetic filtering; Segmentation

## Background

In recent years, several studies demonstrated 18F-3′-deoxy-3′-fluorothymidine ([18F]FLT) as a useful positron emission tomography (PET) tracer for the prediction and monitoring of tumour response to chemotherapy [1-3]. Tumour tissues that show changes in [18F]FLT uptake within a week of therapy have a high likelihood of responding to treatment [4,5]. To evaluate changes in tracer uptake and viable lesion volume quantitatively, analysis requires tumour segmentation. Various manual and (semi-)automatic methods for tumour delineation have been proposed [6-8]. As manual tumour delineation is time consuming, requires expertise and is prone to observer variation, automatic tools are preferable. The performance of a (semi-)automatic tumour segmentation method in PET may vary per tracer, cancer type and lesion location. Physiological biodistribution and kinetic characteristics differ per tracer and per cancer type. The location of the lesion determines the contrast in tracer uptake between tumour and surrounding tissue: the target-to-background ratio [9]. [18F]FLT shows

* Correspondence: r.boellaard@vumc.nl
[1]Department of Radiology & Nuclear Medicine, VU University Medical Center (VUmc), P.O. Box 70571007 MB Amsterdam, The Netherlands
Full list of author information is available at the end of the article

Hoyng et al. EJNMMI Research (2015) 5:26

Page 2 of 10

lower overall uptake in tumours and higher uptake in certain healthy tissues, such as liver and bone marrow than, for example, [18F]FDG (2-fluoro-2-deoxy-D-glucose). Therefore, [18F]FLT PET may require different (semi-)automatic tumour segmentation methods for tumour delineation than [18F]FDG PET.

Gray et al. [10] proposed a supervised classification method based on kinetic filtering (KF) [11]. In the KF method, the time activity curve (TAC) of each voxel is compared to the typical reference TAC of several representative tissues (liver, bone marrow, soft tissue, etc.). In breast cancer, Gray et al. reported significantly different results of standard uptake values (SUVs) based on KF and manual delineation.

For response-monitoring purposes, the repeatability of quantitative measures, which are used to characterize tracer uptake changes, needs to be known. Therefore, it is required to determine the test-retest (TRT) performance of the KF method in [18F]FLT PET studies and to compare it with other (existing) segmentation methods. Furthermore, it is of interest to explore the sensitivity of the supervised classification method to varying imaging protocols. Finally, the method has not yet been tested for the segmentation of lung tumours.

The current study examines the TRT performance of several implementations of the KF method for segmentation of lesions in non-small cell lung cancer (NSCLC) [18F]FLT PET scans. Apart from segmentation performance, the TRT variability of SUVs and metabolically active tumour volumes as defined with KF will be compared with those based on a (static) 50% background corrected relative threshold method.

Please note that segmented tumour volume in this article refers to the metabolically active tumour volume as derived with PET, not the anatomical volume.

## Methods
### Segmentation methods
In this paper, we compare the segmentation and repeatability of several adapted versions of the KF algorithm and a 50% background corrected isocontour method (A50%), as used by Frings et al. [12]. A short introduction of the original KF method and of the A50% method is given below in order to introduce some basic terminology used in the remainder of the paper.

### Kinetic-filtering algorithm
The supervised segmentation method is described in Gray et al. [10]. The method compares for each voxel its TAC to several typical TACs of the most characteristic tissues (liver, vertebrae, tumour, heart, lung and background). These predefined TACs for various tissue types are called *kinetic classes* (KCs). KCs are derived from averaging tissue TACs from several dynamic [18F]FLT

PET studies. These are then used to segment lesions in (other) dynamic [18F]FLT studies based on calculation of the Mahalanobis distance [13]. This is the distance between the voxel TAC and tissue class, weighted by the kinetic class standard deviations:

$$D_{\mathrm{M}} = \sqrt{\sum_{t=1}^{N} \left( \frac{p_t - \mu_t}{\sigma_t} \right)^2} \qquad (1)$$

where $D_{\mathrm{M}}$ is the Mahalanobis distance of voxel $p$ to tissue class $M$, $N$ is the number of time frames, $p_t$ is the activity of the voxel at time frame $t$, and $\mu_t$ and $\sigma_t$ are the activity of the class and the standard deviation at that time frame, respectively.

In order to obtain the same temporal sampling (frames) as that of the PET study, the KCs are resampled to those of the PET scan to be analysed. Gray et al. found that the algorithm did not distinguish properly between vertebrae and tumour tissue. Therefore, the algorithm reclassifies voxels assigned to the vertebra class as tumour tissue.

### Segmentation based on A50%
A50% is a semi-automatic threshold technique. The threshold method locates the voxel with the highest uptake value near a user-indicated starting point in the tumour. Next, a region-growing algorithm using the location of the maximum tumour voxel value continues until all voxels above a certain threshold are included. In the case of the A50% method, this threshold is 50% of the sum of the maximum uptake value and the local background value. The local background value is calculated by averaging the values of a 1-voxel-thick shell at 1.5-cm distance from the boundary of an initial 70% of the maximum value isocontour. More details can be found in [14-16]. Frings et al. compared various threshold settings [12] in [18F]FLT PET and found that the 50% background adapted threshold, A50%, had the best trade-off between success rate and repeatability. Here, success rate is defined as the feasibility to define the tumour volume of interest (VOI).

### Clinical data sets and PET reconstruction
The present study involved three different KF reference data sets, each of which has been described in previous publications. The patients were included in the clinical studies after providing written informed consent in accordance with institutional review board approval. The first data set originates from the Hammersmith and Charing Cross Hospitals, London, United Kingdom, and was provided only in the form of tissue KC [10]. The other two data sets were collected at the VU University Medical Centre (VUmc), Amsterdam, The Netherlands

Hoyng *et al. EJNMMI Research* (2015) 5:26

Page 3 of 10

[1,4]. Details of the data acquisition that are directly relevant for this paper are given below; further information can be found in the original publications of these studies.

Gray et al. [10] created tissue KC from the Hammersmith data. For the main part of our study, the KF method was implemented using these KCs. The Hammersmith data consisted of 13 [18F]FLT PET scans of patients with histologically proven stage II-IV breast cancer based on American Joint Committee on Cancer stage (AJCC) criteria. A median dose of 338 MBq [18F]FLT (range 151 to 381 MBq) was injected intravenously. Simultaneously with the injection, the emission scan was started using an ECAT962/HR+ PET scanner (CTI/Siemens, Knoxville, Tennessee, USA). The patients were scanned for 95 min, and the data were binned into 31 discrete time frames of varying duration ($10 \times 30$ s, $5 \times 60$ s, $5 \times 120$ s, $5 \times 180$ s, $6 \times 600$ s). A subsequent transmission scan was used for attenuation correction. The images were reconstructed using the ordered subsets expectation maximisation (OSEM) method. The KC based on these data will be referred to as *Hammersmith KC*.

Alternative KCs were derived from one of the VUmc datasets to test whether different scanning protocols could affect the classes and the KF segmentation performance. The data consisted of nine [18F]FLT PET scans of patients with histological-confirmed NSCLC adenocarcinoma [1]. [18F]FLT (248 MBq, range 226 to 270 MBq) was injected intravenously 30 s after starting a dynamic emission scan in 3D setting on a ECAT EXACT HR+ scanner (Siemens/CTI, Knoxville, TN, USA). The total scan time was 60.5 min with the following frame lengths: $1 \times 30$ s, $6 \times 5$ s, $6 \times 10$ s, $3 \times 20$ s, $5 \times 30$ s, $5 \times 60$ s, $8 \times 150$ s and $6 \times 300$ s. All emission scans were reconstructed with $128 \times 128$ matrices using filtered back projection (FBP) with a Hanning filter (cutoff, 0.5 cycles per pixel). In addition, the images were reconstructed using OSEM, with 4 iterations and 16 subsets followed by post-smoothing of the reconstructed images using a 5-mm full width at half maximum (FWHM) Hanning filter. In the remainder of this paper, the kinetic classes based on this data set will be referred to as the *VUmc KC*.

The third data set was used to evaluate tumour segmentation performance of KF and A50%. The TRT study was first described in a previously published study [1]. The data consist of [18F]FLT PET scans of nine patients with histologically proven NSCLC. The patients were scanned twice within 7 days (mean 1.9 days, median 1 day) before any treatment. The patients were given an intravenous bolus injection with a median of 364 MBq [18F]FLT (range 252 to 397 MBq). Simultaneously, a 60-min PET scan was started using an ECAT962/HR+ PET scanner (CTI/Siemens, Knoxville, TN, USA). The 39 discrete time frames had varying durations ($6 \times 5$ s, $6 \times 10$ s, $3 \times 20$ s, $5 \times 30$ s, $5 \times 60$ s, $8 \times 150$ s and $6 \times 300$ s). The scans were reconstructed both with FBP and OSEM in the same way as applied for obtaining the VUmc KC. From here on, this data set will be referred to by the *test-retest data* or *TRT data*.

## Kinetic classes of kinetic filters

KCs are derived as described in Gray et al. [10]. For every scan of the Hammersmith OSEM-reconstructed data, the average of the last three frames was created, and tissue VOI was defined manually by an experienced physician on the several tissue types. Next, dose-normalized TACs were extracted for every tissue type. The TACs were averaged per tissue type for all scans, resulting in a mean dose-normalized TAC or KC. The standard deviations for all time frames were obtained and used as weighting factors during KF segmentation, as described previously in the 'Kinetic-filtering algorithm' section (Equation 1). As indicated before, these reference tissue classes were provided to us and are called the *Hammersmith kinetic classes* (*Hammersmith KC*).

For the VUmc KF development dataset, the same procedure was applied to create VUmc-specific KC, both from the FBP- and the OSEM-reconstructed images. These KCs are called the *VUmc kinetic classes* (*VUmc KC*).

The TRT data were averaged over the last three frames (45- to 60-min post injection (p.i.)) for both FBP- and OSEM-reconstructed images. As no anatomical imaging data was available, VOIs were delineated manually from the averaged image data. A total of 28 tumour lesions were identified by an experienced physician. The VOIs were defined including the entire tumour plus a rim of a few voxels with soft tissue surrounding the lesions. These VOIs are used as *lesion masks*. As will be explained later, the lesion masks are used to generate some adapted versions of the KF method and to evaluate the performances of the different implementations of the KF method.

## KF method adaptations

During the evaluation of the KF methods using the VUmc TRT dataset, several adaptations of the KF method were explored. The adapted KF versions were kept identical to the original method as described in Gray et al. [10], except for the following changes:

- KF of FBP- versus OSEM-reconstructed TRT data.
- Use of different sets of KC: the Hammersmith classes and FBP-reconstructed and OSEM-reconstructed VUmc kinetic classes.
- Adjustment of weighting factors in calculating the Mahalanobis distance. The standard deviations were scaled with several values (0.5, 1.0 and 2.0).
- TRT data scans were resampled to better resemble the frame duration and distribution of the

Hoyng et al. EJNMMI Research (2015) 5:26

Page 4 of 10

Hammersmith data. The total number of frames was reduced from 40 to 31 by reduction of the first 10-min data from 26 to 16 time frames. This was performed by temporally interpolating the TRT data to the Hammersmith frame times.

- Reclassification of voxels that are incorrectly classified as liver and/or vertebrae within the lesion mask into tumour voxels.
- Reduction of the set of KCs, by leaving out the liver KC or both the liver and vertebrae KC.
- Temporal smoothing of voxel TACs before application of the KF method. Every frame was averaged with the previous and the next time frames.
- Limitation of the time interval of the PET data. Tested time intervals are: 0 to 5, 9 to 60, 15 to 60 and 30 to 60 min. Frames outside the time interval were disregarded during classification.

Combinations of these adaptations were tested leading to over 50 different KF versions. In Table 1, a small subset of the most successful implementations is summarized.

### Performance evaluations

For each test and retest scan of the TRT dataset, (combinations of) the above-mentioned adapted versions of the KF method were applied to automatically segment the tumour. In addition, the A50% method was applied as reference semi-automated lesion delineation. The metabolically active tumour volume and SUV TRT (percentage difference and absolute differences) for both measures were compared using those lesions that were detected by both methods. The absolute difference and relative TRT variability for volume and SUV between test and retest data were calculated for both the KF and A50% methods for all identified lesions (within the manually predefined lesion masks) and averaged. The TRT variability was calculated as the absolute value of difference between the test and retest value, divided by the mean of both measurements and multiplied by 100. In order to further compare

the A50% and KF segmentation performance, the number of lesions that could be reliably segmented by each method was reported.

### Results

The TACs of the Hammersmith and VUmc KC are shown in Figure 1. Though the shape of the KCs appear to be similar, differences can be observed in amplitude, standard deviations, frame distribution and total scan time (90 vs. 60 min for Hammersmith and VUmc scans, respectively).
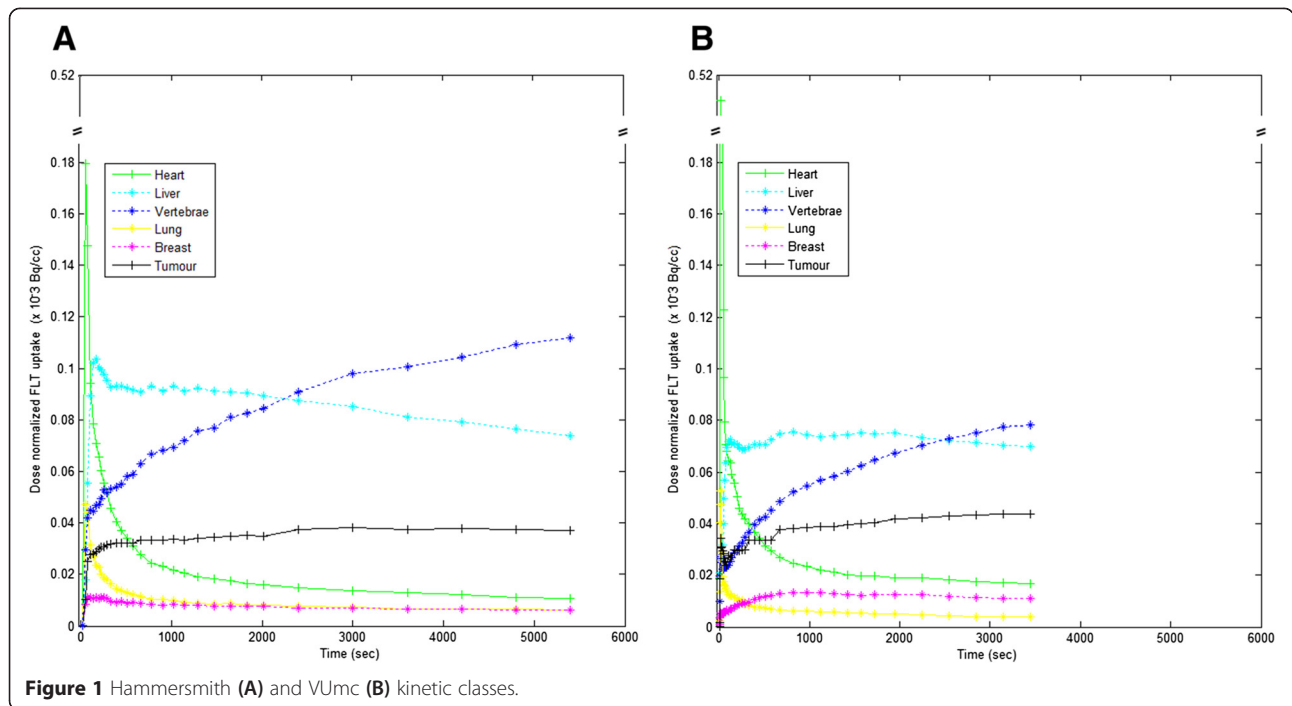
The original as well as various modified versions of the KF method provided very poor segmentation results when applied to FBP-reconstructed images due to image artefacts. Figure 2A is an example of the result of post-processing a FBP-reconstructed image with the kinetic-filtering method. The artefacts that are visible on the original FBP image (Figure 3A) are clearly affecting the classification of the kinetic-filtering method (Figure 2A). Therefore, we only considered OSEM-reconstructed images of both KF and A50% (Figures 3B, C and 2B).

The median A50%-defined lesion size (PET volume) was 8.6 cm$^3$ (range 0.7 to 51.9 cm$^3$). The original implementation of KF segmented 15 out of 28 lesions (KF1, Table 1), whereas A50% segmented all lesions. All undetected lesions were metastases and had a A50%-defined metabolic volume smaller than 2.5 cm$^3$. All primary lesions and metastases larger than 2.5 cm$^3$ were identified. On average, the original implementation of KF (using the Hammersmith kinetic filters) provided 97% larger volumes and 16% lower SUV than measured with A50%. The lower SUV obtained with KF compared to A50% was caused by the larger tumour segmentation as well as by KF misclassification of highly active areas in the lung tumour as being liver, which were excluded from the obtained tumour segmentation. Median metabolically active tumour volume and SUV TRT variability of KF and A50% were 20.7% vs. 10.0% and 4.7% vs. 4.3%, respectively, as also illustrated in the boxplot in Figure 4A and C, respectively.

**Table 1 Overview of various KF methods and their corresponding settings**

| KF version | KC | Description | Undetected lesions |
|---|---|---|---|
| 1 | H'smith | Original KF algorithm | 13 |
| 2 | H'smith | Temporal smoothing was added | 11 |
| 3 | H'smith | Frame times of the scan data adjusted | 2 |
| 4 | H'smith | Liver KC reclassified as tumour KC, classification of 9- to 60-min p.i. | 4 |
| 5 | H'smith | Liver KC reclassified as tumour KC, classification of 20- to 60-min p.i. | 2 |
| 6 | H'smith | Liver KC omitted | 6 |
| 7 | H'smith | Liver and vertebrae KS omitted | 1 |
| 8 | VUmc | Original KF algorithm | 10 |
| 9 | VUmc | Liver KC reclassified as tumour KC, temporal smoothing, classification of 15- to 60-min p.i. | 2 |

Only the results of the original KF and the eight most successful adaptations are provided. KC, kinetic class; KF, kinetic filtering; p.i., post injection.

Hoyng *et al. EJNMMI Research* (2015) 5:26

Page 5 of 10



**Figure 1** Hammersmith **(A)** and VUmc **(B)** kinetic classes.

Given the high number of either undetected lesions or partly misclassified lesions by the original implementation of KF, our aim was primarily to optimize the KF method performance and test its feasibility for the VUmc TRT data. Due to the large amount of KF versions, only the results obtained from the original and the best performing implementations are shown in Figure 4 and listed in Table 1.

The best performing KF implementations in combination with the Hammersmith KC were those that were obtained using (in random order) the following: temporal smoothing (KF2); frame times of the TRT data PET scans interpolated to the frame times of the Hammersmith KCs

(i.e. not vice versa, KF3); KC liver being reclassified as KC tumour in combination with analysis over a limited time interval from 9- to 60-min p.i. (KF4); KC liver being reclassified as KC tumour, analysis over a limited interval from 20- to 60-min p.i. (KF5); use of a reduced set of KC, i.e. the liver KC was omitted (KF6); use of reduced set of KC, i.e. both liver and vertebrae KC were omitted (KF7); and the original KF implementation (KF1) in combination with the VUmc KC is indicated by KF8. The best performing KF implementation in combination with the VUmc KC was the one that reclassified liver and vertebrae as tumour KC, used temporal smoothing and restricted the analysis time interval from 15- to 60-min p.i. (KF9).
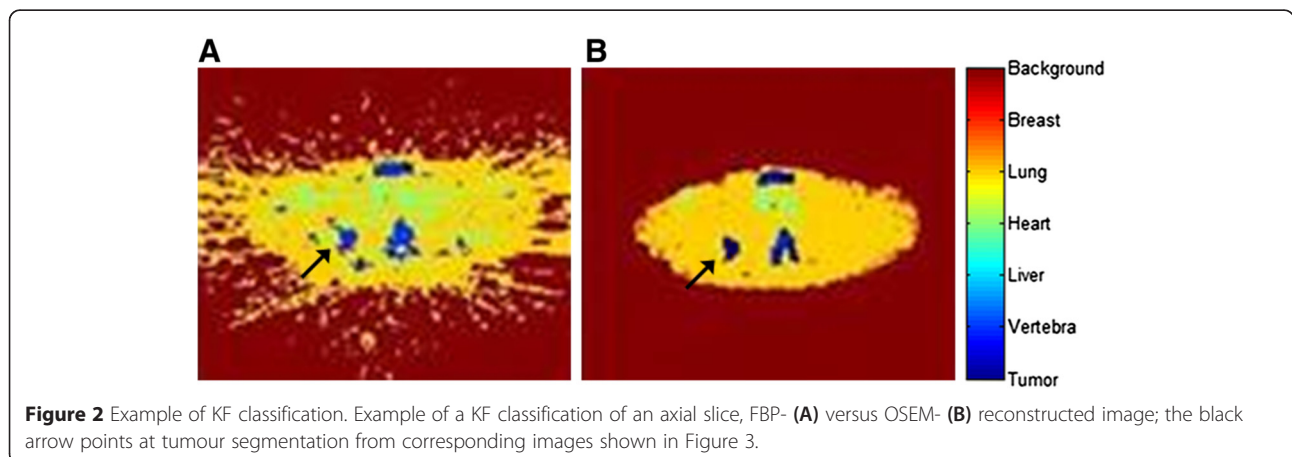


**Figure 2** Example of KF classification. Example of a KF classification of an axial slice, FBP- **(A)** versus OSEM- **(B)** reconstructed image; the black arrow points at tumour segmentation from corresponding images shown in Figure 3.

Hoyng *et al. EJNMMI Research* (2015) 5:26

Page 6 of 10



**Figure 3** Example [$^{18}$F]FLT PET axial slice, FBP and OSEM reconstructed. Axial slice showing [$^{18}$F]FLT uptake, using FBP and OSEM reconstructions: FBP, 60-min p.i. **(A)**; OSEM, 5-min p.i. **(B)**; OSEM, 60-min p.i. The tumour has a coloured overlay of the lesion mask **(C)**.

The adapted KF method versions (as listed above and in Table 1) reduced the amount of undetected lesions and improved the percentage repeatability performance as compared to the original KF implementation (*KF1*). The best performing KF methods in combination with the Hammersmith KC were *KF3*, *KF4* and *KF5*, i.e. when using temporally resampled PET images, reclassifying liver as tumour or when leaving out liver as a kinetic class.

KF with temporal smoothing and a reduced set of KC (no liver and vertebral KC) (*KF7*) showed a good TRT variability compared to other KF implementations and was able to reliably segment all but one lesion. Figure 5 shows that the method classified too many voxels as tumour, even outside the lesion mask. This version of the KF method is therefore not successful. The red colour in the lung area in the segmentation pictures of
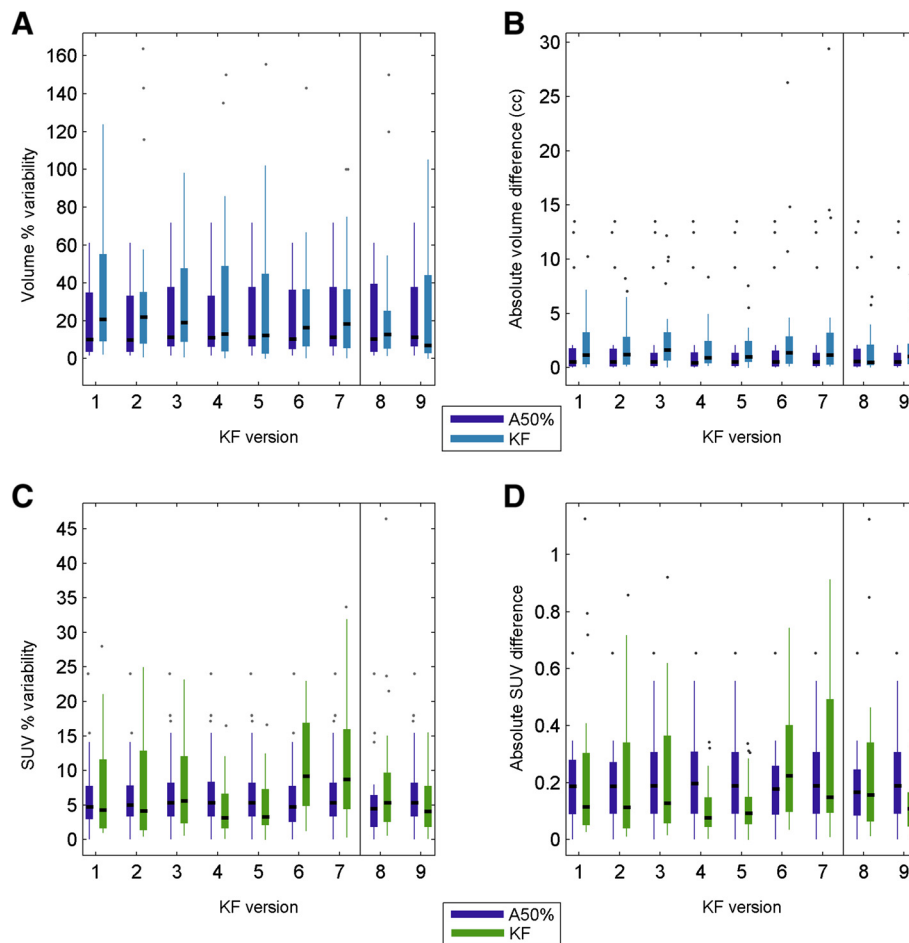


**Figure 4** Test-retest performance of the most successful KF versions and of A50%. Percentage metabolically active tumour volume variability **(A)**, absolute metabolically active tumour volume difference **(B)**, percentage SUV variability **(C)** and absolute SUV difference **(D)**. Per KF version, the A50% TRT performance is shown for those lesions that were detected by the specific KF version. KF versions are listed in Table 1. KF, kinetic filtering; SUV, standard uptake value.
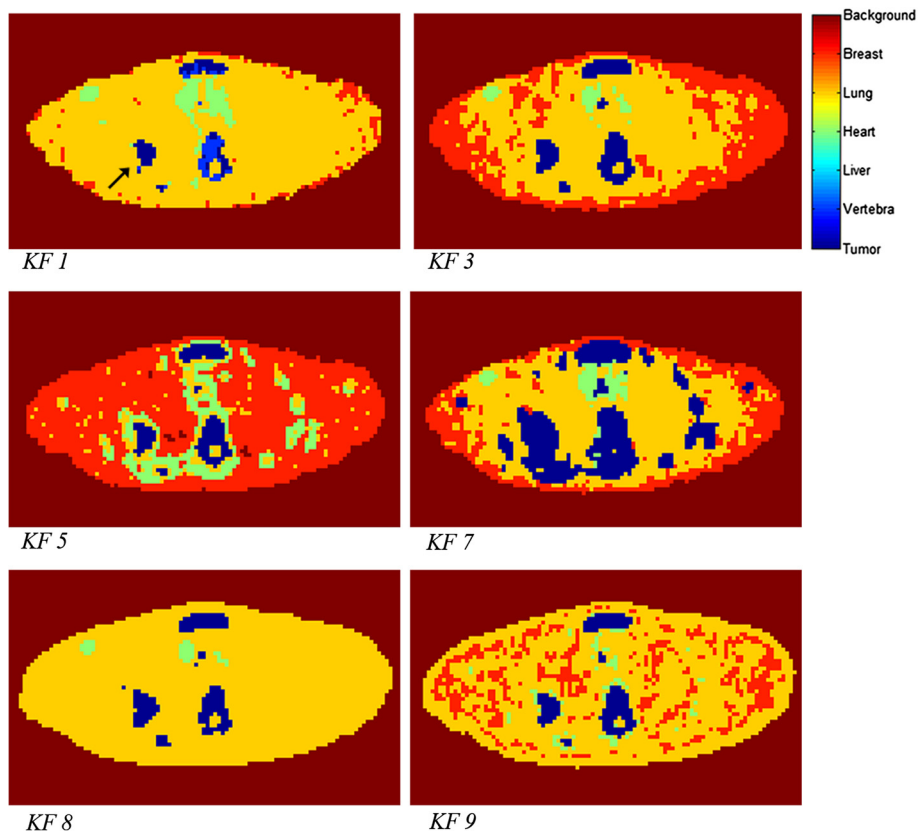
Hoyng *et al. EJNMMI Research* (2015) 5:26

Page 7 of 10



**Figure 5** Segmentation examples obtained from several KF implementations. The KF numbering corresponds to the KF numbering as listed in Table 1. The black arrow in the image of KF1 points at the tumour. Every image represents the same axial image plane in the same patient. KF, kinetic filtering.
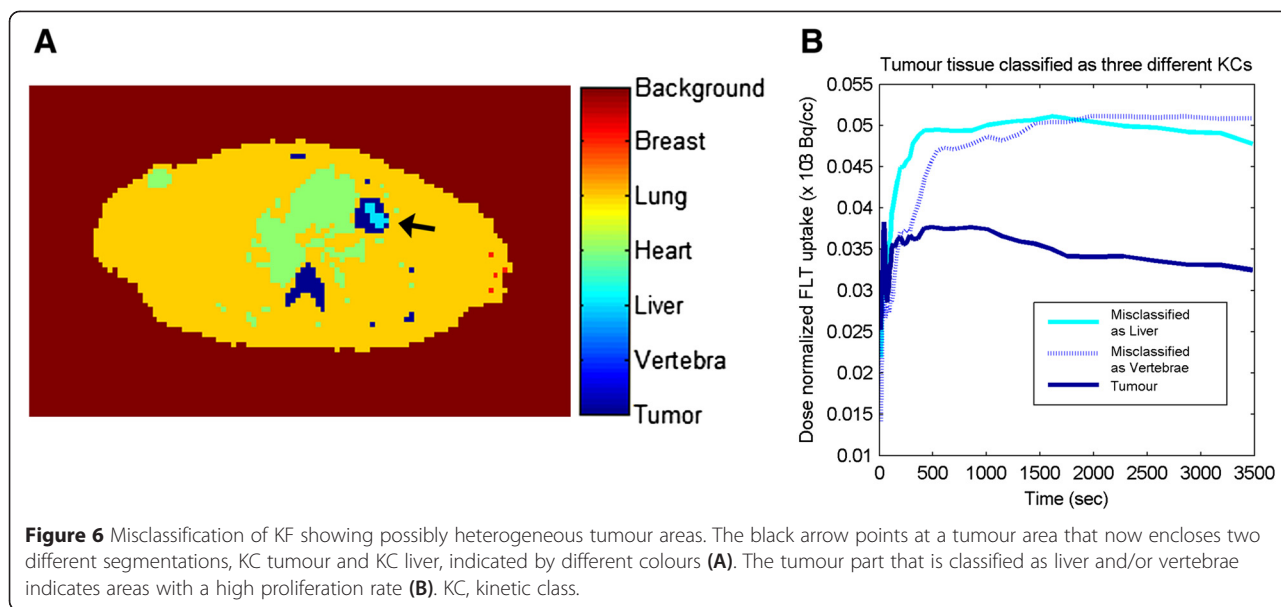
*KF3, KF5* and *KF9* shows misclassification of heart and lung tissue as breast tissue. However, repeatability of metabolically active tumour volume was similar to that of A50% and fewer lesions were undetected in these versions as compared to the original KF. However, the spread in these methods is higher than with A50%.

Misclassification of tumour voxels as liver or vertebrae was observed in the original KF implementation. After further analysis, these misclassifications seem to reflect a heterogeneous kinetic behaviour within the tumour as shown by the high-uptake trend of the averaged TACs of the tumour voxels classified as liver or vertebrae (Figure 6B).

## Discussion

KF allows automated segmentation of lesions from dynamic [18F]FLT PET studies. In this study, various modifications of the KF method of Gray et al. [10] are examined. Since none of the tested methods reliably distinguished vertebrae and tumour, we introduced lesion masks. Combined with the KF method, these masks enable the segmentation of tumour as well as redefinition of voxels classified as liver and/or vertebrae into tumour (as there can be no liver or vertebrae within the lesion masks).

Comparison of the Hammersmith KC with the VUmc KC shows that the TACs of several KCs are very similar in shape and in relative uptake intensity (Figure 1 Hammersmith (A) and VUmc (B) kinetic classes). Tumour tissue and bone marrow show relatively high uptake due to rapid proliferation. The average [18F]FLT PET bone marrow TAC demonstrates irreversible tissue kinetics. The liver metabolizes [18F]FLT to [18F]FLT-glucuronide causing an even higher uptake signal in liver tissue but with explicit reversible kinetics, i.e. a decreasing trend over time. Recent studies suggest that most NSCLC tissue TACs are best fitted by a two-tissue reversible model in nonlinear regression analysis when applying a 90-min dynamic scan [17-19]. Small amplitude and shape discrepancies between the two KCs might be caused by a difference in tracer injection and/or imaging procedure. Differences in administration procedures, i.e. a fast or slow bolus injection, have a direct effect on the shape of any TAC for all tissues which may also affect the performance of the KF methods. As the use of Hammersmith KC versus VUmc KC has a large effect on the observed TRT performances, use of optimized or site-specific KC seems to be important for obtaining reliable results. Consequently,

Hoyng *et al. EJNMMI Research* (2015) 5:26

Page 8 of 10



**Figure 6** Misclassification of KF showing possibly heterogeneous tumour areas. The black arrow points at a tumour area that now encloses two different segmentations, KC tumour and KC liver, indicated by different colours **(A)**. The tumour part that is classified as liver and/or vertebrae indicates areas with a high proliferation rate **(B)**. KC, kinetic class.

application of KF to multicentre studies requires highly standardized imaging procedures to apply KC derived elsewhere, or alternatively, the KCs need to be defined per imaging site. Furthermore, the specific framing of the dynamic scan affects KF performance. Best results with Hammersmith KC were obtained by either adjusting the frame times to a more even distribution of the time frames over the course of the PET scan or by disregarding the first 10 min of the scan (KF3-KF5). The 30-min difference in total scan duration between the two sites might be relevant in this context as well. Better TRT performance was obtained by allowing a more prominent role of the 'tails' of the TACs in the distance calculations. Possibly, use of longer than 60-min scan durations could have further improved performance; however, this was not possible with the currently available datasets. Optimisation of administration procedure in combination with adjustment of frame times of the dynamic scans might improve the performance of the tested KF methods. Moreover, we found a difference in the standard deviations of the Hammersmith KC and VUmc KC. Since these act as the weighting factor when calculating the Mahalanobis distance, this difference can influence the classification results. Although we tested various levels of rescaling of the amplitude of the weighting factors, we did not find an optimal setting. Reclassification of voxels that were initially classified as vertebrae or liver to tumour provided the best TRT results (KF4 and KF5) with Hammersmith KC. Obviously, when studying liver metastases, this would not be appropriate. Similarly, we expect the A50% isocontour method to be less reliable in hepatic metastases due to the high physiological [18F]FLT uptake value of the liver. KF with VUmc KC needed adjustment as well due to the required order of the Mahalanobis distances between a voxel and the KCs

to allow for KC reclassification of KC vertebrae. Reduced VUmc KC performed best with only 2 missing lesions out of 28 and good metabolic active tumour volume and SUV TRT variability. Currently, patients with liver metastases are undergoing [18F]FLT-PET scans in order to further assess the performance of KF in areas with high physiological [18F]FLT uptake in a future study.

Scan and administration procedures were identical for both the VUmc development dataset and the TRT dataset, and therefore, obviously KF performed better in combination with the VUmc KC than in combination with the Hammersmith KC. If KC were not adapted to the study-specific and centre-specific tracer administration and scanning settings, then 50% of the lesions were not properly segmented. Yet, after carefully adjusting the method to align KC to scan conditions of the TRT data set or applying KC derived from scans performed on the same scanner and obtained using the same imaging procedure, the KF method performance improved and was able to reliably delineate 93% of the lesions. At the same time, metabolically active tumour volume and SUV TRT performances improved to similar levels observed with the A50% method (around 10%, on average). The improved performance after local (study-specific) optimisation of KF again illustrates that for a multicentre study standardized administration and scanning protocols are essential. Yet, once KF is optimized to the specific scan conditions at hand, the method can be used as an initial automated visualisation to identify lesions in dynamic [18F]FLT PET studies.

An additional advantage of KF classifications could be their potential use as a segmentation method for kinetically heterogeneous tumours. Detection of tumour heterogeneity could be used as a prognostic biomarker, to

Hoyng et al. EJNMMI Research (2015) 5:26

Page 9 of 10

asses treatment response and for dose escalation in radiation therapy [20-22]. Therefore, it is necessary to study (automatic) delineation of intratumour subvolumes. To date, only few such studies have been performed to address this issue [9,23]. In our study, the observed intratumour (mis-)classification of voxels into liver or vertebrae may indicate that the method can differentiate between metabolically high and low active tumour tissues within a predefined lesion mask. Some preliminary results are shown in Figure 6, together with TACs, for these differently classified intratumour regions (now still denoted as liver and vertebrae). Further adaptations of this method might improve its use for intratumour tracer uptake heterogeneity assessment or segmentations and will be part of future research.

## Conclusions

For dynamic [$^{18}$F]FLT PET studies in NSCLC patients, kinetic filtering and an adaptive contrast-oriented isocontour method (A50%) have similar repeatability performance to define metabolically active tumour volume but only after careful optimisation of the kinetic-filtering method settings. The kinetic classes and/or the kinetic-filtering method therefore requires a protocol-specific optimisation. In multicentre studies, A50% might be a good alternative segmentation method for [$^{18}$F]FLT PET in NSCLC. Yet, KF has the potential to identify low and high proliferative areas within heterogeneous lesions.

**Authors' contributions**
LH carried out the evaluation and adaptation of the kinetic-filtering algorithm for the [$^{18}$F]FLT PET data of NSCLC and drafted the manuscript. VF carried out the manual tumour delineation of the VUmc data sets. OH participated in the clinical interpretation of the results. LK and EA provided the original kinetic-filtering algorithm and participated in discussions on the interpretation of the results of the adapted algorithms. RB conceived of the study and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

**Author's information**
Lieke L Hoyng is the first author.

**Author details**
$^1$Department of Radiology & Nuclear Medicine, VU University Medical Center (VUmc), P.O. Box 70571007 MB Amsterdam, The Netherlands. $^2$Comprehensive Cancer Imaging Centre, Imperial College London, Hammersmith Hospital, W12 0NN London, UK.

## References

1. De Langen AJ, Klabbers B, Lubberink M, Boellaard R, Spreeuwenberg MD, Slotman BJ, et al. Reproducibility Of quantitative $^{18}$F-3-deoxy-3-fluorothymidine measurements using positron emission tomography. Eur J Nucl Med Mol Imaging. 2009;36:389–95.
2. Hoeben BAW, Troost EGC, Span PN, Van Herpen CML, Bussink J, Oyen WJG, et al. $^{18}$F-FLT PET during radiotherapy or chemoradiotherapy in head and neck squamous cell carcinoma is an early predictor of outcome. J Nucl Med. 2013;54:532–40.
3. Contractor KB, Kenny LM, Stebbing J, Rosso L, Ahmad R, Jacob J, et al. [$^{18}$F]-3'-deoxy-3'-fluorothymidine positron emission tomography and breast cancer response to docetaxel. Clin Cancer Res. 2011;17:7664–72.
4. Frings V, Van Der Veldt A, Boellaard R, Herder GJM, Giovannetti E, Honeywell R, et al. Pemetrexed induced thymidylate synthase inhibition in non-small cell lung cancer in vivo: a pilot study with 3'-deoxy-3'-[$^{18}$F]fluorothymidine positron emission tomography. Plos One. 2013;8:E63705.
5. Kenny LM, Coombes CR, Vigushin DM, Al-Nahhas A, Shousha S, Aboagye EO. Imaging early changes in proliferation at 1 week post chemotherapy: a pilot study in breast cancer patients with 3'-deoxy-3'-[$^{18}$F]fluorothymidine positron emission tomography. Eur J Nucl Med Mol Imaging. 2007;34:1339–47.
6. Cheebsumon P, Van Velden FH, Yaqub M, Frings V, Lammertsma AA, Boellaard R. Effects of image characteristics on performance of tumor delineation methods: a test–retest assessment. J Nucl Med. 2011;52:1550–8.
7. Shepherd T, Teräs M, Beichel RR, Boellaard R, Bruynooghe M, Dicken V, et al. Comparative study with new accuracy metrics for target volume contouring in pet image guided radiation therapy. IEEE Trans Med Imag. 2012;31:2006–24.
8. Hatt M, Cheze-Le Rest C, Aboagye EO, Kenny LM, Rosso L, Turkheimer FE, et al. Reproducibility of $^{18}$F-FDG and 39-deoxy-39-$^{18}$F-fluorothymidine PET tumor volume measurements. J Nucl Med. 2010;51:1368–76.
9. Tomasi G, Shepherd T, Turkheimer FE, Visvikis D, Aboagye EO. Comparative assessment of segmentation algorithms for tumor delineation on a test-retest [$^{11}$C]choline dataset. Med Phys. 2012;39:7571–9.
10. Gray KR, Contractor KB, Kenny LM, Al-Nahhas A, Shousha S, Stebbing J, et al. Kinetic filtering Of [$^{18}$F]fluorothymidine in positron emission tomography studies. Phys Med Biol. 2010;55:695–709.
11. Turkheimer F, Edison P, Pavese N, Roncaroli F, Anderson A, Hammers A, et al. Reference and target region modeling of [$^{11}$C]-(R)-PK11195 brain studies. J Nucl Med. 2007;48:158–67.
12. Frings V, De Langen AJ, Smit E, Van Velden FH, Hoekstra O, Van Tinteren H, et al. Repeatability of metabolically active volume measurements with $^{18}$F-FDG and $^{18}$F-FLT PET in non-small cell lung cancer. J Nucl Med. 2010;51:1870–7.
13. Mahalanobis PC. On the generalized distance in statistics. Proc Nat Inst Sci. 1936;2:49–55.
14. Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. J Nucl Med. 2004;45:1519–27.
15. Krak NC, Boellaard R, Hoekstra OS, Twisk JWR, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. Eur J Nucl Med Mol Imaging. 2005;32:249–301.
16. Cheebsumon P, Van Velden FH, Yaqub M, Hoekstra CJ, Velasquez LM, Hayes W, et al. Measurement of metabolic tumor volume: static versus dynamic FDG scans. EJNMMI Res. 2011;14:1–35.
17. Frings V, De Langen AJ, Yaqub M, Schuit RC, Van Der Veldt A, Hoekstra O, et al. Methodological considerations in quantification of 3'-deoxy-3'-

[$^{18}$F]fluorothymidine uptake measured with positron emission tomography in patients with non-small cell lung cancer. Mol Imaging Biol. 2014;16:136–45.

18. Frings V, Yaqub M, Hoyng LL, Golla SV, Windhorst A, Schuit RC, et al. Assessment of simplified methods to measure 3'deoxy-3'-[$^{18}$F]fluorothymidine uptake changes in EGFR mutated non-small cell lung cancer patients undergoing EGFR tyrosine kinase inhibitor treatment. J Nucl Med. 2014;55:1417–23.

19. Muzi M, Mankoff DA, Grierson JR, Wells JM, Vesselle H, Krohn KA. Kinetic modeling Of 3'-deoxy-3'-fluorothymidine in somatic tumors: mathematical studies. J Nucl Med. 2005;46:371–80.

20. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2013;144:646–74.

21. Thorwarth D, Geets X, Paiusco M. Physical radiotherapy treatment planning based on functional PET/CT data. Radiother And Onc. 2010;96:317–24.

22. Gottesman MM, Foho T, Bates SE. Multidrug resistance in cancer: role of ATP–dependent transporters. Nature Rev Cancer. 2002;2:48–58.

23. Willaime JMY, Turkheimer FE, Kenny LM, Aboagye EO. Quantification of intra-tumour cell proliferation heterogeneity using imaging descriptors of $^{18}$F fluorothymidine-positron emission tomography. Phys Med Biol. 2013;58:187–203.