D5.4 Validation report on software performance under experimental phantom conditions and under clinical conditions

WP n° and title: 5 Automated analysis of PET based in-vivo monitoring in ion beam therapy and feedback to treatment planning

WP leader: D. Georg (MUW), B. Jones (UOXF)

Author(s): S. Helmbrecht, P. Kuess

Contributor(s):

Deliverable due: M54

Report date: July 31st, 2014

Dissemination Level

<table>
<thead>
<tr>
<th>PU</th>
<th>Public</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>Restricted to other programme participants (including the Commission Services)</td>
<td></td>
</tr>
<tr>
<td>RE</td>
<td>Restricted to a group specified by the consortium (including the Commission Services)</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>Confidential, only for members of the consortium (including the Commission Services)</td>
<td></td>
</tr>
</tbody>
</table>
TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS ................................................................. 3
PUBLISHABLE SUMMARY ................................................................................................. 4
INTRODUCTION .................................................................................................................. 5
PHANTOMS ........................................................................................................................... 6
EVALUATION ALGORITHMS ............................................................................................. 10
  Method 1 -- Range comparison and cavity detection algorithm: ........................................ 10
  Method 2 -- Pearson correlation coefficient: ..................................................................... 11
RESULT .................................................................................................................................. 13
  Phantom A ........................................................................................................................... 13
  Phantom B ........................................................................................................................... 15
DISCUSSION ......................................................................................................................... 16
REFERENCES ......................................................................................................................... 18
## LIST OF ABBREVIATIONS AND DEFINITIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>PET</td>
<td>Positron Emission Tomography</td>
</tr>
<tr>
<td>ibPET</td>
<td>In-beam Positron Emission Tomography</td>
</tr>
<tr>
<td>PMMA</td>
<td>Polymethylmethacrylate, ([C_5H_8O_2]_n)</td>
</tr>
<tr>
<td>CT</td>
<td>X-ray computed tomography</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>PTV</td>
<td>Planning Target Volume</td>
</tr>
<tr>
<td>MATLAB</td>
<td>The MathWorks, Natick, MA, USA</td>
</tr>
<tr>
<td>IDL</td>
<td>Interactive Data Language, ITT Visual Information Solutions, Boulder, CO, USA</td>
</tr>
<tr>
<td>DICOM</td>
<td>Digital Imaging and Communications in Medicine</td>
</tr>
<tr>
<td>PYTHON</td>
<td>Python Software Foundation</td>
</tr>
<tr>
<td>TUD</td>
<td>Technical University Dresden</td>
</tr>
<tr>
<td>PT-PET</td>
<td>Particle Therapy PET</td>
</tr>
<tr>
<td>CTV</td>
<td>Clinical Target Volume</td>
</tr>
<tr>
<td>MUW</td>
<td>Medical University Vienna</td>
</tr>
<tr>
<td>PCC</td>
<td>Pearson Correlation Coefficient</td>
</tr>
</tbody>
</table>
PUBLISHABLE SUMMARY

Two types of phantoms were designed to produce well defined deviations in the activity distributions. Pure range differences were simulated using the first phantom type while the other emulated cavity structures. The phantoms were irradiated with $^{12}$C-ions. PT-PET measurements were performed by means of a camera system installed at the beamline. Different measurement time scenarios were investigated, assuming a PET scanner directly at the irradiation site or placed within the treatment room. The images were analyzed by means of the Pearson Correlation Coefficient (PCC), a range calculation algorithm and a dedicated cavity filling detection method.

Range differences could be measured with an error of less than 2 mm. The range comparison algorithm yielded slightly better results than the PCC method. The filling of a cavity structure could be safely detected if its inner diameter was at least 5 mm.

Both approaches evaluate the PT-PET data in an objective way and deliver promising results for in-beam and in-room PET for clinical realistic dose rates.
INTRODUCTION

With the increase of new particle therapy centers all over the world, Particle Therapy - Positron Emission Tomography (PT-PET) gains more and more relevance for treatment verification [Frey2014]. Up to now PT-PET is the only clinically applied method for in vivo verification of ion-beam radiotherapy during or close in time to treatment.

During $^{12}$C-ion irradiation, a relatively low $\beta$-activity of $1.6 \text{ kBq cm}^{-3}$ is generated per Gray therapeutic dose by auto activation of the tissue by the beam and the beam itself [Priegnitz2008]. This activity concentration is about 1-2 orders of magnitude less compared to typical nuclear medicine applications of PET imaging. Therefore, the low number of events is one of the major challenges for PT-PET. For treatment monitoring, the measured activity distribution is compared to a reference activity [Poenisch2004,Nishio2010]. Up to now no consensus exists in the radiotherapy community on the ideal realization of PT-PET in clinical routine. Shakirin et al. [Shakirin2011] described three different techniques: in-beam PET (PET cameras with limited angle geometry, fully integrated into the treatment site) [Enghardt2004], off-line PET (the usage of a standard PET-CT scanner outside the treatment room) [Parodi2007] and in-room PET (full ring PET scanner inside the treatment room). In-room PET may also utilize a portable PET scanner [Zhu2011,Min2013].

Regarding off-line PET the most important limitation is the extensive influence of the patient metabolism [Parodi2007,Fiedler2008,Helmbrecht2013]. The chemistry of the created radioactive nuclides is very complex and the metabolic processes depend strongly on the anatomic region and the respective physiology of the patient. Therefore, these processes cannot be simulated in a phantom or easily modelled in a $\beta^+$-activity prediction tool. Thus off-line PET was not considered in this work.

The reference activity distribution can be calculated by means of a Monte Carlo simulation using the planning CT and the treatment plan as outlined in detail in [Poenisch2004].

A visual comparison of the $\beta^+$-activity distributions This procedure demands a great expertise from the evaluator. Furthermore, this task adds a massive workload to the clinical workflow [Fiedler2010]. Two approaches were presented to automate this comparison in Kueess et al and Helmbrecht et al [Kueess2012,Helmbrecht2012,Kueess2013] (see also Deliverable 5.1, 5.2 and 5.3). These studies showed that the detection of ion beam range differences of 4 mm and more is possible in head and neck patients.

In general, one of the main reason for range uncertainties during particle therapy treatment is a change in the tissue along the beam path. This can be due to changes in tumor size, changes in cavity fillings or due to weight loss or gain. So far the investigations of the automated evaluation approaches were primarily relying on artificially altered patient data [Helmbrecht2012,Kueess2012,Kueess2013]. An objective reference data set to test the algorithms was so far not available.

Thus, specially built phantoms were irradiated with $^{12}$C beams and respective PT-PET images were recorded. By means of these experiments the detection of range shifts and cavity fillings was investigated.
PHANTOMS

Two types of phantoms were developed in the framework of this study (cf. Fig. 1). The feasibility and accuracy of the measurement of range differences for the complete PT-PET imaging chain was evaluated by means of phantom type A (cf. Fig. 1a), which was designed to simulate simple range differences. The analysis was performed for eight values of preset range differences, three different numbers of annihilation signals and four time scenarios. The phantom was built from a polymethyl metacrylate (PMMA) block with the dimensions of 9x9x20 cm³ with an end-to-end hole in longitudinal direction. Two moveable PMMA cylinders with a length of 8 cm and 15 cm and a diameter of 2.5 cm allowed the generation of range differences from -20 mm to +20 mm. An air gap between the two cylinders created a range extension. If the cylinders were in touch and protruded the frontal plane -- with respect to the incident beam -- an underrange could be created.

Cavity structures as e.g. found in the human head were modeled by the type B phantoms, that were built in two versions (Fig. 1b). They were made of a 3x12x20 cm³ sized PMMA block with three holes in the transversal direction. The holes had diameters of 9 mm, 13 mm, and 18 mm (type B1) and 11 mm, 15 mm, and 23 mm (type B2), respectively. For each hole, one hollow cylinder with a wall thickness of 3 mm was constructed from bone equivalent material (SB3 Bone, Cortical RMI 450, Gammex-RMI GmbH, Giessen-Allendorf, Germany) as well as from PMMA. The SB3 material had a density of 1.81 g cm³ and consisted of 3.4% hydrogen, 31.4% carbon, 1.8% nitrogen, 36.5% oxygen, <0.1% chlorine, and 26.8% calcium. The values were given in mass fractions. The hole and a respective hollow cylinder formed the cavity. PMMA plugs were inserted into the hollow cylinder to simulate a filled cavity, as illustrated in Figure 1b. The difference in density between SB3 bone and PMMA (1.81 g cm⁻³ 1.19 g cm⁻³ = 0.62 g cm⁻³) was in good agreement with the density difference between cranium and brain (1.61 g cm⁻³ 1.04 g cm⁻³ = 0.57 g cm⁻³).
Both phantoms were irradiated with 12C ions with six energies between 244.2 to 261.6 AMeV at the former medical beam line at the GSI Helmholtzzentrum für Schwerionenforschung GmbH. The beam was created by a heavy-ion synchrotron. A total dose of 5 Gy was delivered to the target. The planned dose distribution is illustrated in Figure 2. A pause of at least 60 min between consecutive irradiations guaranteed that potential residual activity did not distort the respective experiment. According to a respective model the relative activity decreased to less than 5% after 60 min. PT-PET measurements were carried out with the camera system BASTEI, installed at the beam line. The camera consisted of two heads with a sensitive area of 42 x 21 cm². Each head consisted of 8x4 position sensitive bismuth germanate (BGO) block detectors, each divided in 8x8 pixels with an edge length of 6.75 mm, resulting in 2048 pixels per head. The resolution of the system after image reconstruction was about 5.0 mm. The camera was described in detail by Enghardt et al. [Enghardt2004, Enghardt1999]. This system was used to monitor the irradiation of approximately 440 patient treatments during the German Heavy Ion Tumor Therapy Project between 1997 and 2008. The acquisition was started prior to the first beam extraction (spill) of the synchrotron and was stopped 20 min after the end of irradiation. The coincidence data were recorded in list mode, accompanied by time stamps in 10 ms intervals. The type A phantoms were irradiated with cylinder positions for generating range differences between -6 and +10 mm in steps of 2 mm. The cavity phantoms (type B) were irradiated with integrated hollow cylinders made of PMMA or SB3 with and without PMMA plugs. The reference situation was a preset range difference of 0 mm for the type A phantom and unfilled cavities for type B. Thus, a simulation of the β⁺-activity distribution described in [Poenisch2004] was not necessary in order to obtain a reference image.

Due to the higher dose compared to clinical PT-PET acquisition, the number of annihilation signals was reduced by means of a Monte Carlo method to 20% and 30% of the initial counts. Additionally, data sets containing all coincidences were investigated as well. Since many random coincidences occurred during the beam application due to prompt gamma rays [Parodi2005b], only data captured during the spill pauses of approximately 3 s and during the follow-up measurement were used for image reconstruction.

As already mentioned above, in this work in-beam and in-room PET scenarios were investigated and compared by changing the time intervals for image reconstruction, as illustrated in Figure 3. The in-beam scenario implies a PET measurement that starts simultaneously with the irradiation and ends 40 s after its termination. An exact scenario for an in-room PET measurement is clinically not determined. Thus, the investigation considered three different time parameters [Kuess2013].

---

Figure 1: Phantoms built for the experimental verification of automated evaluation algorithms. Type A is designed to create well-defined range differences from -20 to +20 mm using a fixed beam energy $E_0$. The position of the cylinders for creation of an under- or overrange is depicted in the scheme in the upper left. Type B simulates filled cavities in bony structures, it is constructed from (1) a PMMA cube with longitudinal holes. Hollow cylinders (2) made of bone equivalent material or PMMA are placed in the holes. Cavity filling is simulated by means of PMMA plugs (3).
The transfer time was assumed to be one or two minutes, the measurement time three or five minutes. Figure 3 shows the time intervals for the scenarios in-room 1, in-room 2 and in-room 3. The number of events in the scenarios selected was $7.6 \cdot 10^5$, $1.0 \cdot 10^6$, $8.7 \cdot 10^5$ and $6.9 \cdot 10^5$ in the case of the type A phantom and $8.1 \cdot 10^5$, $1.1 \cdot 10^6$, $9.6 \cdot 10^5$ and $7.5 \cdot 10^5$ in the case of the type A phantom at 5 Gy. The image reconstruction was performed by means of a dedicated maximum likelihood expectation maximization (MLEM) algorithm [Shepp1982, Lauckner1999, Poenisch2003b].

![Diagram](image)

Fig. 2 Planned dose distribution in the phantoms shown in frontal view for the isocentric plane. The beam enters from the top. In case of the type B phantom the position of the holes and the cylinders is shown schematically.
Fig. 3: Selection of coincidences for the different PET monitoring scenarios. The graph shows the number of registered events per second. The spikes between 20 s and 180 s originate from the increased count rate during the beam application.
EVALUATION ALGORITHMS

The algorithms that were developed for the automatic evaluation of PT-PET image are described in preceding deliverables (D5.1, 5.2, 5.3) as well as in [Helmbrecht2012, Kuess2012]. Therefore the will only described here very briefly:

Method 1 -- Range comparison and cavity detection algorithm:

From the reconstructed $\beta^+$-activity distributions one dimensional subsets, referred to as profiles, along the beam direction are extracted and scaled to the maximum value. At the end of the particle path, the activity shows a characteristic decreasing behaviour, the so called distal edge. The difference of the position of the distal fall-off of these profiles equals the difference in range of the primary particles. These distal edges are extracted by applying thresholds and demanding a decreasing behaviour of the activity in the selected interval. For the phantom experiments a lower threshold of 0.2 and a higher threshold of 0.8 was applied. By means of shifting the distal edges against each other, the range difference is obtained. This algorithm is applied to every point in a plane perpendicular to the beam direction referred to as beam's eye view. The result is a two dimensional matrix $M$ of range differences. This matrix can be used for visual inspection directly [Unholtz2011] or be used as basis for further processing. In the latter case it has to be cut to a reasonable region of interest, in the clinical case the projection of the planning target to the beam's eye view.

An unexpected filling of cavities (e.g. for paranasal sinus of the patient) due to tissue swelling or deposition of mucus, causes severe differences in particle range and consequently in dose deposition. For the detection of cavity fillings, in a first step a segmentation of the planning CT is performed. Voxels with Hounsfield units $1024 \leq H \leq -200$ are marked as air, $-199 \leq H \leq 99$ as soft tissue (ST) and $100 \leq H \leq 3071$ as bone [SchlegelBille2004]. A matrix $G$ is calculated with the same dimensions as the planning CT. If a voxel in the CT meets the condition for containing air, the associated voxel in the segmentation matrix $G$ is set to one, otherwise to zero. The two matrices $A^{(1)}$ and $A^{(2)}$ describe the activity distributions from 1) the measurement and 2) the reference, which is obtained from a measurement of another fraction or a simulation. They are rotated in a way that the beam direction equals a principal direction $k$. The variables $k$, $m$ and $n$ denominate the axes of the matrix. The difference of activity in the normally air-filled voxels is calculated as follows:

$$
\Delta A_{m,n} = \sum_{i \in [k_S, k_E]} \left( A^{(1)}_{k,m,n} \cdot G_{k,m,n} \right) - \sum_{i \in [k_S, k_E]} \left( A^{(2)}_{k,m,n} \cdot G_{k,m,n} \right)
$$

$k_S$ denotes the entrance point of the beam into the patient, $k_E$ the exit point of the extension line of the beam path and $i$ an voxel between $k_S$ and $k_E$. $\Delta A$ is a two dimensional matrix, describing the difference in activity along the beam path in the voxels that are filled with air in the planning CT. It has the same dimensions as the matrix of range differences $M$. For cavity filling detection range differences greater than zero are not relevant, the same applies to the case of lower activity in the
volume of the cavity than in the reference. Therefore, positive elements in \( M \) and negative elements in \( \Delta A \) are set to zero. Thus cavity filling is represented by the matrix \( K \) where the symbol \( \odot \) represents the element wise multiplication of two matrices:

\[
K = \Delta A \odot M.
\] (2)

For an identification of a filled cavity, the relative value in comparison to the values in the proximity is important. To give an absolute value is complex since it depends on the patients anatomy and the applied double head camera system generally does not yield absolute activity values. Three possible outcomes were identified. If the value in \( K \) is zero in all pixels of the projection of the cavity volume to the beam direction, the detection has failed. If the value is greater than zero in only a part of the named projection, the situation is called partial detection. If the value is greater than zero in the complete projection the detection was successful.

The range comparison and the cavity filling detection algorithm were implemented in Python 2.7 applying the Numpy package for scientific computing.

**Method 2 -- Pearson correlation coefficient:**

The second automatic evaluation method investigated utilizes the Pearson correlation coefficient (PCC) as outlined in Kuess et al. [Kuess2012,Kuess2013] to compare the reference \( \beta^+ \)-activity distribution (no range modification or hollow cavities) to the respective \( \beta^+ \) measurement. The PCC, as depicted in Eq. 3, equals 1 if the compared images are identical and 0 if there is no correlation at all. In Eq. 3 \( A^{(2)} \) represents the reference image and \( A^{(1)} \) the measurement. The indices \( k,m,n \) denote the current voxel in the corresponding images. \( \bar{A}^{(2)} \) and \( \bar{A}^{(1)} \) represent the mean intensities in the VOI of the respective images [Birkfellner2010].

\[
PCC = \frac{\sum_{k,m,n} (A^{(2)}_{k,m,n} - \bar{A}^{(2)}) (A^{(1)}_{k,m,n} - \bar{A}^{(1)})}{\sqrt{\sum_{k,m,n} (A^{(2)}_{k,m,n} - \bar{A}^{(2)})^2} \sqrt{\sum_{k,m,n} (A^{(1)}_{k,m,n} - \bar{A}^{(1)})^2}}
\] (3)

The PCC based algorithms were written in the programming language MATLAB (R2009, 64 bit), MathWorks, Natick, MA (USA).

For both phantom types the volume of interest (VOI) was defined as the whole target volume, but for the type B phantoms the VOI was split in three parts of equal size to compare the different PCC results between holes of different diameter. The results of the range differences (type A phantom) are shown as \( \Delta PCC \) values (\( \Delta PCC=(1-PCC)\cdot100 \)) [Kuess2013]. Thus, high \( \Delta PCC \) values correlate to high deviations between the investigated PT-PET images. For the purpose of this study a dedicated threshold of \( \Delta PCC=1 \) was set. A \( \Delta PCC \) below this threshold implies that it remains uncertain if the recorded deviations are due to beam range uncertainties or based on minor changes in the experimental setup without dosimetric effect or caused by fluctuations in a potential simulation. Image reconstruction artifacts can also produce minor mismatches between two otherwise identical images. The threshold value was introduced for the following two reasons. Firstly, preceding studies on patient data sets showed that this value considers fluctuations within
the PCC calculations that are caused by a potential Monte Carlo simulation of the activity distribution [Kuess2013].

Secondly, the differences within the PCC of two similar phantom irradiation experiments yielded also PCC differences of $\Delta$PCC=[0.5-1]. Thus, the chosen threshold is a good choice for the presented phantom data sets and can also be applied in clinical use of PT-PET. This detection threshold was also accounted for when evaluating the type B phantoms. Therefore, the detection of a filled cavity implies that the calculated $\Delta$PCC was above this threshold.

The use of an additional median filter on the PT-PET images prior to comparison showed some promising results during a previous pilot study, where regarding five patient test cases the detection capability of 6 mm range modified data could be improved by 10% by using a median filter, with a kernel size of 11x11x11 voxels [Kuess2012]. Therefore, the evaluation of phantom A was also used to evaluate quantitative differences regarding the PCC evaluation between images that were pre-filtered and images that were not filtered before evaluation. Thus this investigation will show if the use of a median filter will also enhance the detection capability of measured phantom data.

Assuming a positive effect on the detection efficiency the $\Delta$PCC value should be higher if filtered images are processed than non-filtered ones. The kernel size of the used median filter was 11x11x11 voxels.
RESULT

Phantom A

Since the lateral shape of the cylinder is not perfectly reproduced by the PET measurement, only the value in the center of the cylinder was used for the analysis. The measured values showed a proper accordance to the preset range differences. To obtain a quantitative statement on the accuracy of the range measurement, the detected values were subtracted from the preset ones. Table 1 shows the mean values of all investigated range differences within each time scenario for each level of coincidence rate (20%, 30% and 100%).

The given uncertainty of the values is the standard deviation of the distribution of the differences between measured and preset range difference within one scenario and one count level. The maximum error in range difference measurement occurred at 20% coincidences in the in-room scenario with (1.4±1.1) mm. An increased number of events contributing to the image reconstruction reduced the detection error as expected. The maximum error at 30% coincidences was found in the in-beam scenario (1.1±0.4) mm. At 100% coincidences, the lowest error values were achieved, except for the in-beam scenario. The minimum value of (0.6±0.6) mm was obtained in scenario in-room 3. The mean values of the determined range differences at 30% was found inside the standard deviation at 100% coincidences and reversely. At this point, it should again be highlighted, that biological effects that blur the activity distribution over time were not taken into account in the experiments, as they are not present in phantom materials.

The results obtained with the PCC for detecting range differences in the type A phantom are presented in Figure 4 for a 30% coincidence rate. The presented data were based on median filtered images.
Figure 4: Calculated ΔPCC values for generated range differences from -6 to +10 mm in the type A phantom. The recorded data were thinned out to 30% to get a clinical realistic particle number. The color scheme corresponds to the different monitoring scenarios, as explained in the text. The semitransparent blue plain depicts the threshold (ΔPCC=1) above which a range difference qualifies as detected. All images were pretreated with a median filter using a 11x11x11 voxel kernel size.

The PCC evaluation enabled a detection of all irradiated range differences, but -2 mm which was on the edge of the defined threshold of ΔPCC=1 (blue surface in Figure 4). A linear fit through the obtained ΔPCC data points and determining the intersection at ΔPCC =1, revealed that range shifts of 1.3 mm and 1.2 to 1.6 mm in positive direction could be detected for in-beam PET and in-room scenarios, respectively. Regarding shifts in negative directions, the lowest range difference which could be detected was -2.3 mm and -2.0 to -2.3 mm for in-beam and in-room PET, respectively. Very similar results were obtained for 20% and 100% coincidence rate. In both cases the detection limit was between -2.0 to -2.6 mm in negative direction and 1.6 to 1.9 mm in positive direction, except the in-beam scenario (20% coincidence rate) where the fit resulted in a lower limit of -2.9 mm. The in-beam scenario never yielded the best result. The latter finding is in agreement with the results of the range comparison algorithm (c.f. Table 1). In-room 3 was observed to achieve the lowest detection limit in most cases (for 100% and 30% coincidence rate in positive direction and for 30% and 20% in negative direction). Considering all results given above, the automated PCC based evaluation algorithms
suggest a detection of range difference down to +2 mm and -2.5 mm in positive and negative direction, respectively. 
The effect of a median filter did not vary with respect to the different monitoring techniques. The median filter improved the detection efficiency of all investigated range differences, but only for range differences lower than -6 mm and greater than +4 mm the improvement was above 1%. The influence of such a median filter increased with increasing range difference (e.g. 3% for +8 mm range difference).

<table>
<thead>
<tr>
<th>Coincidence levels</th>
<th>Mean error in scenario / mm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>in-beam</td>
</tr>
<tr>
<td>20%</td>
<td>1.4 ± 0.8</td>
</tr>
<tr>
<td>30%</td>
<td>1.1 ± 0.4</td>
</tr>
<tr>
<td>100%</td>
<td>1.1 ± 0.6</td>
</tr>
</tbody>
</table>

Table 1: Mean absolute differences between the preset and the determined range differences obtained for coincidence level and time scenario. The given uncertainties are the standard deviations of the distributions.

**Phantom B**

An illustration of the representative activity distributions in a type B phantom is shown in Figure 6. If the cavities are filled with air (Fig. 6a) the activity distribution shows spikes distal of the heterogeneities. With the cavities filled, the spikes disappear. Only the influence of the higher density of the bone equivalent material remains visible (Fig. 6b). The range differences are displayed in Figure 6c using color coded bars for the extend of the range differences. The length of the bars indicate the position of the distal edge in the planned situation, i.e. the relevant information lies in the color. The indication of the distal edge position serves primarily to ensure that the correct position in the activity distribution was selected by the algorithm. The corresponding cavity detection matrix K is depicted in Figure 6d. Cavities with diameter ≥ 5 mm were detected in all cases (5, 7, 9, 12, 17 mm). The smallest cavity with a diameter of 3 mm could not be completely detected in all cases by means of the dedicated algorithm. The results are shown in detail in table 2. As mentioned earlier, for the PCC evaluation an altered cavity status was marked as detected if the comparison of two images yielded a ΔPCC≥1. According to this definition all cavities greater than 3 mm were detected. The detection of cavities with 3 mm diameter remained challenging with a successful detection only in one case (cf. table 2)
D 5.4 Validation report on software performance under experimental phantom conditions and under clinical condition

Dissemination

ENVISION -GA n°241851

Table 2: Result of the detection of filled cavities for the smallest cavity (diameter 3mm) of the type B phantom. n denotes the number of coincidences at the given dose (in parenthesis). A ✓ labels a successful detection, • a partial detection, × a failure. The columns P show the result for hollow cylinders made of PMMA, B for bone equivalent material, respectively. ∆R refers to the cavity detection algorithm and PCC to the imaging based evaluation approach. The investigated cavities of 5 to 17 mm were detected with both algorithms for all setups and coincidence rates.

<table>
<thead>
<tr>
<th>Coincidences</th>
<th>Method</th>
<th>Scenario P</th>
<th>Scenario B</th>
<th>Scenario in-room 1</th>
<th>Scenario in-room 2</th>
<th>Scenario in-room 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>20% · n(5 Gy)</td>
<td>∆R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>PCC</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>30% · n(5 Gy)</td>
<td>∆R</td>
<td>✓</td>
<td>•</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>PCC</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>100% · n(5 Gy)</td>
<td>∆R</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>PCC</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

Figure 6: Measured activity distribution in type B phantom in frontal view, in case of (a) open and (b) filled cavities. The image shows the isocenter plane. The beam impinges from the left side. (c) Range differences obtained by means of the range comparison algorithm displayed as colored bars. (d) Cavity detection values in individual scaling displayed in beam’s eye view.

DISCUSSION

In a survey, carried out during the 2012 annual meeting of the AAPM, 33% of the voters said, that range uncertainties are the main issue when using particle beams for radiotherapy [Freeman2012]. PT-PET addresses this issue and furthermore automated evaluation techniques will help to incorporate this monitoring method into clinical routine. In this work the achievable accuracy of a PET based dose verification was analyzed by means of phantoms using two different evaluation
techniques for PT-PET. Although the camera system has an intrinsic resolution of about 5 mm, beam range differences of approximately 2 mm within the phantoms could be detected. Since this holds true for all investigated time scenarios, one can assume that there is no physically caused deterioration of the image quality within the first 6-7 min after the irradiation. It has to be highlighted that metabolic processes also influence the PET image taken with in-room PET, even though the effect is less pronounced compared to off-line PET [Helmbrecht2013]. Biological caused effects could not be considered in these phantom experiments. The lower detection limit of in-room techniques were slightly better than for in-beam PET although the same double head camera system was used. Hence, limitations concerning image quality in an in-room PET scenario are caused only by biological, not by physical effects.

In general, both approaches presented and evaluated in this paper yielded comparable results. For the pure detection of range uncertainties the range comparison algorithm seems superior over the PCC based image evaluation. However, the reasons for deviations between measured and the expected activity distributions can be manifold. While the range comparison algorithm focuses on measuring the beam range, the PCC based algorithm considers the whole of the measured PT-PET image. Furthermore, the outcome of the PCC based evaluation strongly varies with the homogeneity of the target [Kuess2013]. Another difference between the methods is that the range comparison algorithm calculates a number that can be directly converted to the range differences, while the PCC based approach yields a value that defines the magnitude of the deviation but not the direction of the shift. Both algorithms are comparable regarding the execution times. For the pure calculation of the respective values that decide if a range difference is present or not less than 1 min is needed on a standard office computer. If using a more powerful computer or a different programming language a considerable reduction of calculation time is expected.

By evaluating the PT-PET images recorded during the irradiation of the type A phantom, it was possible to obtain a lower limit for the automated analysis of PT-PET. For a 30% (or more) coincidence rate, the detection of range differences of 1.5 to 2 mm can be guaranteed by means of the range comparison algorithm. For a 20% coincidence rate (= 1 Gy) the limit with this method could be found at approximately 2 mm, depending on the treatment scenario. If a full ring scanner would be used, a conservative estimate is a factor of 3 for the total number of coincidences in comparison to the BASTEI system utilized in this study. The detection limit of the PCC based algorithm has to be stated with 2 mm for positive range modifications and 2.5 mm for negative range modifications, independently from the investigated coincidence rates. It has to be highlighted that a potential Monte Carlo simulation causes fluctuations in the activity predictions and subsequently uncertainties of the ΔPCC. By using the ΔPCC threshold (ΔPCC=1) as explained above such fluctuations are already considered [Kuess2012]. Thus, the presented results include this potential distortion and are also applicable for clinical patient data. Using a median filter improved the automated detection of range deviations with the PCC algorithm which is in accordance with a previous study [Kuess2012]. The influence of the median filter increases with increasing difference and seems to be more pronounced for positive differences. Regarding the detection of cavity filings (type B phantom) the median filter did not reveal an improvement. The benefit regarding small range differences was negligible also for phantom A. As the improvement for detecting a 4 mm range difference by use of a median filter is less than 2% a pre-filtering of the
image should be only optional in a dedicated software tool, as it also implies additional computational time.

In preceding studies on PT-PET images from patients it was observed that the VOI selection had a crucial impact on the detection efficiency. This could not be seen in the phantom data, most likely because of the simplified phantom composition. Therefore an easy reproducible VOI was chosen (i.e. target area).

The impact on $^{12}$C beams caused by changes of cavity fillings has never been investigated by means of phantoms so far. This study provides exact information on the cavity size and the alloy of the filling material. Furthermore, images recorded with inserted PMMA plugs showed no noticeable deviation to the ones containing a plug made of bone equivalent material. Thus, it can be concluded that the complex heterogeneous material forming cavities (soft tissue, bone, air) does not deteriorate the detection efficiency, compared to the simplified scenario of PMMA followed directly by air. The dedicated cavity filling algorithm exploits the range and the activity in volumes filled with air in the planning CT. The approach of combining different parameters from the CT, the PT-PET images, the structure set or even other imaging modalities might increase the diagnostic benefit of PT-PET considerably. A straight forward possibility is a reverse usage of the cavity detection algorithm to detect the reduction of tissue swelling. This requires a previous manual delineation of suspicious volumes in the planning CT by a physician.

The presented algorithms for an automatic detection of deviations between the actual and planned irradiation situation in PT-PET images can be combined into one software tool to enhance the automatic detection efficiency by using two different comparison approaches. This software tool decreases the clinical workload and evaluates the PT-PET in an objective way. However, the reasons for a mismatch between two PT-PET images can be manifold and hardly be assigned automatically. Therefore, the consequences with respect to the patient's treatment have always to be approved by a human supervisor. Moreover, a major aim of an automatic evaluation tool is to get objective results as visual inspection can suffer from a high inter-observer variability.

REFERENCES


