Prepared for submission to JINST

23rd International Workshop on Radiation Imaging Detectors
Riva del Garda, Italy
26 – 30 June 2022

Analysis methods for in-beam PET images in proton therapy treatment verification: a comparison based on Monte Carlo simulations

M. Moglioni[a,b] **A.C.Kraan** ✉ [a] **, A. Berti**[a,c] **, P. Carra**[a,b] **, P. Cerello**[d] **, M. Ciocca**[e] **, V. Ferrero**[d] **, E. Fiorina**[d] **, E. Mazzoni**[a] **, M. Morrocchi**[a,b] **, F. Pennazio**[d] **, A. Retico**[a] **, V. Rosso**[a,b] **, G. Sportelli**[a,b] **, V. Vitolo**[e] ***and*** **G. Bisogni**[a,b]

[a] *Istituto Nazionale di Fisica Nucleare, Sezione di Pisa, Pisa, Italy*
[b] *University of Pisa, Pisa, Italy*
[c] *Institute of Information Science and Technologies, National Research Council of Italy, Pisa, Italy*
[d] *Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Torino, Italy*
[e] *Centro Nazionale di Adroterapia Oncologica, Pavia, Italy*

E-mail: aafke@pi.infn.it

Abstract:

Background and purpose: In-beam Positron Emission Tomography (PET) is one of the modalities that can be used for in-vivo non-invasive treatment monitoring in proton therapy. PET distributions obtained during various treatment sessions can be compared in order to identify regions that have anatomical changes. The purpose of this work is to test and compare different analysis methods in the context of inter-fractional PET image comparison for proton treatment verification.

Methods: For our study we used the FLUKA Monte Carlo code and artificially generated CT scans to simulate in-beam PET distributions at different stages during proton therapy treatment. We compared the Beam-Eye-View method, the Most-Likely-Shift method, the Voxel-Based-Morphology method and the gamma evaluation method to compare PET images at the start of treatment, and after a few weeks of treatment. The results were compared to the CT scan.

Results and conclusions: Three-dimensional methods like VBM and gamma are preferred above two-dimensional methods like MLS and BEV if much statistics is available, since the these methods allow to identify the regions with anomalous activity. The VBM approach has as disadvantage that a larger number of MC simulations is needed. The gamma analysis has the disadvantage that no clinical indication exist on tolerance criteria. In terms of calculation time, the BEV and MLS method are preferred. We recommend to use the four methods together, in order to best identify the location and cause of the activity changes.

Keywords: Instrumentation for hadron therapy, PET, data analysis methods

[1]Corresponding author.

Contents

1 Introduction

Proton therapy is a radiation therapy in which proton beams are used to destroy tumor cells. With respect to conventional radiotherapy with X-rays and electrons, proton therapy allows for more conformal dose distributions, thanks to the characteristic depth dose profile of charged particles. However, proton therapy is more sensitive to uncertainties, that can possibly lead to dose distortions. Among the sources of uncertainties are anatomical changes for instance from organ motion, tumour regression, or weight loss/gain, which can occur during the course of treatment [1, 2].

Patients treated with radiotherapy for head and neck cancer frequently experience anatomical changes. For these patients a control CT is generally acquired after a few weeks of treatment (see for instance [3, 4]). In-vivo range monitoring can be a useful tool to support clinical personnel in the decision on when to schedule this control CT, which is normally scheduled according to clinical experience of the radiation oncologist. Positron Emission Tomography (PET) is an in-vivo monitoring techniques that can be used for this scope [5–9]. Nuclear interactions of the particle beams with the patient tissue can result in the production of β^+-isotopes, like ^{15}O, ^{11}C, ^{10}C. These isotopes decay emitting a positron, that annihilates into two 511 keV photons, that can be detected with a PET system. The PET activity is indirectly correlated with the delivered dose. By comparing the acquired PET images along the treatment course, it is possible to estimate whether modifications in delivered dose occur.

There are different PET data acquisition modalities, differing in acquisition time. In-beam PET are acquired inside the treatment room, allowing to obtain real-time information and to exploit short-lived isotopes [5, 9, 10]. At the National Center of Oncological Hadrontherapy (CNAO), in Pavia, Italy, a bimodal imaging system, featuring in-beam PET and charged particle detection, is installed, called INSIDE (INnovative Solution for In-beam Dosimetry in hadronthErapy) [11, 12].

The goal of the present study is to assess the effectiveness and compare various methods for the analysis of in-beam PET images. This will be done with the help of simulated PET images, that

are based on the INSIDE in-beam PET geometry and artificially modified CT scans. Specifically, four different methods will be compared, including the Beam-Eye-View (BEV) method [13], the Most-Likely-Shift method [14], the Voxel-Based-Morphometry approach [15] and the gamma-index analysis [16].

2 Methods and Materials

2.1 FLUKA MC simulations

For our study we used a CT scan of a 70-year-old patient treated in 2018 for Squamous Cell Carcinoma (SCC) with proton therapy at CNAO. This patient was used in previous studies [13, 15] and only the most relevant information is reported here. A planning CT scan was available with corresponding structure set. In Figure 1(a) a slice of the planning CT scan is given, where the Clinical Target Volume (green zone) is shown. Figure 1(b) shows the same slice, with the sinonasal cavity region highlighted with a yellow circle. The tumor region received 2 Gy per treatment session (66 Gy in 33 fractions). An intensity modulated proton therapy (IMPT) treatment plan was available for this CT scan. We introduced an anatomical change in the patient CT, simulating a cavity emptying [15], by modifying the Hounsfield units, as can be seen in Fig. 1(c). The total volume that we changed was 13 ml. Such a volume change can realistically occur over weeks of treatment.

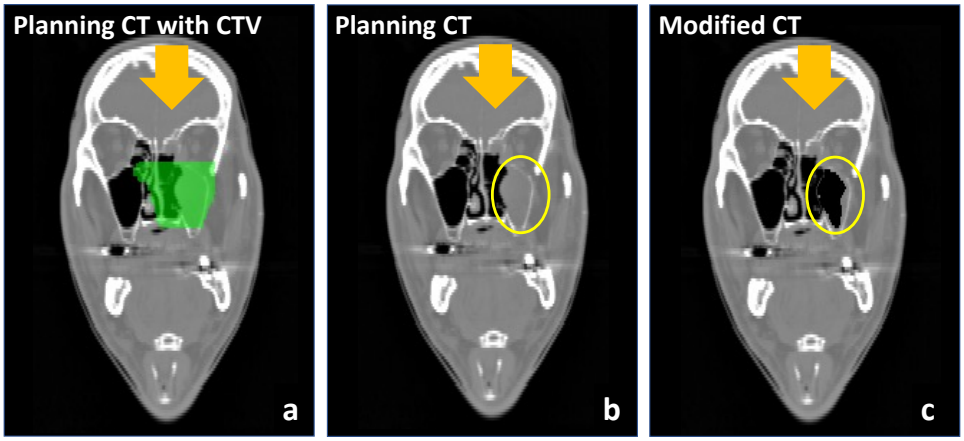

Figure 1. (a) Slice of the planning CT of the SCC patient, with the CTV indicated in green. (b) The same, but with region of interest indicated as yellow circle, containing the sinonasal cavity. (c) The same slice, but now including an artificially introduced anatomical change. The emptying of the sinonasal cavity is clearly visible.

For simulating the PET activity, a tool previously developed for INSIDE [19, 20] was used, based on FLUKA Monte Carlo simulations [22]. Shortly summarizing, it included the simulation of the CNAO beamline, the pencil beam transverse dimension and shape, the time structure of the delivered protons, and the patient geometry, obtained by importing the CT scan in a voxel FLUKA geometry. Radiation transport and interactions of protons in the patient were simulated, as well as

the generation and decay of the β^+ emitting isotopes. The geometry of the INSIDE PET detector was included, with the pixelated LSF crystals and the energy depositions of the photons. We selected coincidences, that occurred between the start of treatment, in between the beam spills, up to about 6 minutes. The lines-of-responses (LORs) were created and subsequent image reconstruction was done with an iterative maximum likelihood estimation maximization (MLEM) procedure. In the end, a 3-D PET image was obtained, with a field-of-view (FOV) of $22.4 \times 11.2 \times 26.4$ cm^3, with $1.6 \times 1.6 \times 1.6$ mm^3 voxels. The obtained images suffered from reconstruction artefacts in the direction perpendicular to the two PET detector planes, a well-known problem for planar PET scanners. We applied a median filter of 1.6 mm (1 voxel) in all directions in order to decrease statistical fluctuations. The PET activity was simulated for the original and the modified CT scan, yielding a PET activity distribution without (Figure 2 a) and one with (Figure 2 b) anatomical changes. These are hereafter referred to as PET_{ref} and PET_{mod}, respectively.

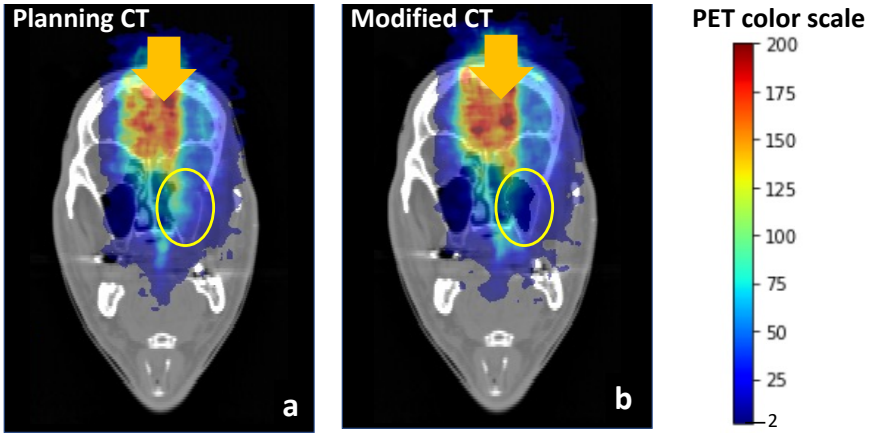

Figure 2. (a) Slice of the planning CT of the SCC patient together with the simulated PET distribution. (b) Same but now including an artificially introduced anatomical change.

2.2 Analysis methods

To compare the two PET images of Fig. 2, various analysis techniques have been developed in the past. In this work we compare four different methods:

- The Beam-Eye-View method [13]. This method focuses on the distal part of the activity (see Figure 3 a). This method is based on a multi-threshold approach to extract range differences. For each pair of coordinates x, y in the transverse plane, an activity profile along z of the reference image was constructed, and its integrated activity was determined. The maximum value of all activity profiles was evaluated, and all profiles with integrated activity less than 30% of this value were excluded. For the remaining (x, y) coordinates, we determined the z coordinate where the activity value was above a threshold t, R^t_{ref}. The value of R^t_{ref} was determined for thresholds from 2% up to 8% of the maximum of the entire PET image, with 0.5% steps. The same was done for the corresponding profiles of the modified image. For a given x, y pair, the range shift between the two profiles, $\delta R_{\mathrm{BEV}}(x, y)$, was defined as:

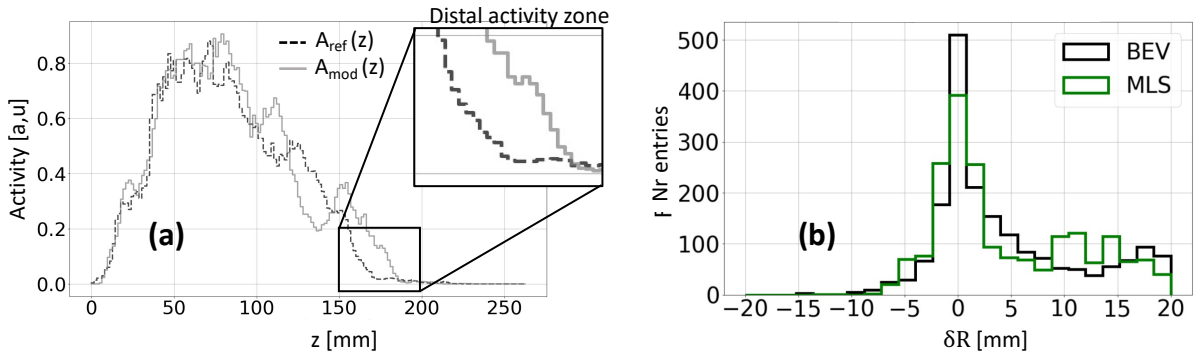

Figure 3. (a) Example of a reference PET image profile ($A_{ref}(z)$, black dashed) and the corresponding modified PET profile ($A_{mod}(z)$, grey solid) in point $(x, y) = (35, 70)$. The BEV and MLS methods focus on the very last part of an activity profile. (b) The distribution of range differences found between PET_{ref} and PET_{mod} for all (x, y) points for the BEV (black) and MLS (green) method.

$$\delta R_{\mathrm{BEV}}(x,y) = \frac{1}{N}\sum_{t=1}^{N}\delta R^{t}_{\mathrm{BEV}}(x,y) = \frac{1}{N}\sum_{t=1}^{N}\left(R^{t}_{mod}(x,y) - R^{t}_{ref}(x,y)\right) \tag{2.1}$$

where N is the total number of thresholds considered, which was 13, R^{t}_{mod} and R^{t}_{ref} are the range values of the modified and the reference profile, respectively, and $\delta R^{t}_{BEV}(x,y)$ is the difference between them.

Then a map $O_{BEV}(x,y,z)$ was created by filling each grid point (x,y,z) as:

$$\begin{cases} O_{\mathrm{BEV}}(x,y,z) = \delta R_{\mathrm{BEV}}(x,y) & \text{if } z \le R^{mod}_{8\%}(x.y) \text{ and } |\delta R_{BEV}(x.y)| \ge 5 \text{ mm} \\ O_{\mathrm{BEV}}(x,y,z) = 0 & \text{otherwise} \end{cases} \tag{2.2}$$

This map was superimposed to the CT images. The highlighted zones represent the zones where activity is present (larger than 8% of the maximum value) and where a range difference larger than 5 mm was found. Positive and negative values for δR indicate an overshoot (displayed with red) and undershoot (displayed with blue), respectively, with respect to the reference situation.

- Most-Likely-Shift method. This is also a two-dimensional range analysis, evaluating range differences in the (x, y) plane, originally proposed by Frey et al [14]. Considering two PET images that must be compared, we developed an algorithm, that provided for each x, y pair an optimal shift distance along z, called δ_{MLS}, between the PET images. Exactly the same (x, y) profiles were included as in the BEV analysis. The range difference δR_{MLS} is the δ_{MLS} value that minimizes the absolute differences in the distal part of the two activity depth profiles [14]:

$$\delta R_{MLS}(x,y) = \arg\min_{\delta_{MLS}}\left(\sum_{z_{MLS}}^{z_{max}} |A_{mod}(x,y,z-\delta_{MLS}) - A_{ref}(x,y,z)|\right) \tag{2.3}$$

with A_{ref} and A_{mod} corresponding to the reference activity value and that of the modified activity value, respectively. Then a three-dimensional map $O_{MLS}(x, y, z)$ was defined as:

$$\begin{cases} O_{\text{MLS}}(x, y, z) = \delta R_{\text{MLS}}(x, y) & \text{if } z \leq z_{max} \text{ and } |\delta R_{MLS}(x.y)| \geq 5 \text{ mm} \\ O_{\text{MLS}}(x, y, z) = 0 & \text{otherwise} \end{cases} \tag{2.4}$$

These maps were re-oriented on the patient's CT reference frame. The interpretation of the maps is just like for the BEV method.

- The Voxel-Based-Morphometry method [15]. This analysis is a three-dimensional voxel-by-voxel analysis. For each voxel of the PET_{mod} image, a statistical analysis is done, testing the compatibility of the voxel value with the null hypothesis (situation without morphological changes). Rejection of the null hypothesis implies that the voxel intensity has significantly changed. For this purpose, not one distribution PET_{ref} must be obtained, but a large number of reference distributions must be obtained. This was done by simulating 120 times the reference PET image, each time with a new random seed, as documented in an earlier work of our group [15]. Let us define v_{ref} as the intensity value in a certain voxel in the reference PET image, $\bar{v}_{ref}$ as the average intensity value over all N replicates of the reference, and v_{mod} as the intensity value in the modified PET image in the same voxel. The empirical p-values in that voxel is:

$$p = \frac{r + 1}{N + 1}, \tag{2.5}$$

where N was the number of replicates (120) and r was the number of replicates that produced a test statistic of either $v_{ref} \geq v_{mod}$ if $v_{mod} \geq \bar{v}_{ref}$, or $v_{ref} \leq v_{mod}$ if $v_{mod} < \bar{v}_{ref}$. The resulting p-values evaluated for each voxel of the modified PET image can be reported as 3-D voxelized maps. As significance threshold we chose $p \leq 0.025$. Significantly more and less activity in a certain voxel is indicated by purple and cyan, respectively. These colored three-dimensional maps can be overlaid onto the original CT image.

- The γ-index analysis. The gamma test [16] is a metric that is mostly used for dose comparisons. When an evaluated dose distribution D_e is compared to a reference dose distribution D_r, given at grid points $\vec{r}_e$ and $\vec{r}_r$, respectively, the gamma index $\gamma(\vec{r}_e)$ is given by:

$$\gamma\left(\vec{r}_e\right) = \min\left[\Gamma\left(\vec{r}_e, \vec{r}_r\right)\right] \quad \forall\, \vec{r}_r \tag{2.6}$$

where

$$\Gamma\left(\vec{r}_e, \vec{r}_r\right) = \sqrt{\frac{\Delta r^2\left(\vec{r}_e, \vec{r}_r\right)}{\text{DTA}^2} + \frac{\Delta D^2\left(\vec{r}_e, \vec{r}_e\right)^2}{\text{DD}^2}} \tag{2.7}$$

Here $\Delta r(\vec{r}_e, \vec{r}_e) = |\vec{r}_e - \vec{r}_r|$ is the distance between $\vec{r}_e$ and $\vec{r}_r$, $\Delta D(\vec{r}_e, \vec{r}_r) = D(\vec{r}_e) - D(\vec{r}_r)$ is the value of the dose difference, and the parameters DTA and DD are the acceptance criteria for distance-to-agreement and dose-to-agreement respectively. If $\gamma \leq 1$, the point is accepted (compatible distributions in that grid point), and if $\gamma > 1$ the point is rejected (incompatible distributions in that grid point). The dose tolerance is usually a percentage of prescribed dose (global gamma index). In dose comparisons, widely used parameters for DTA and

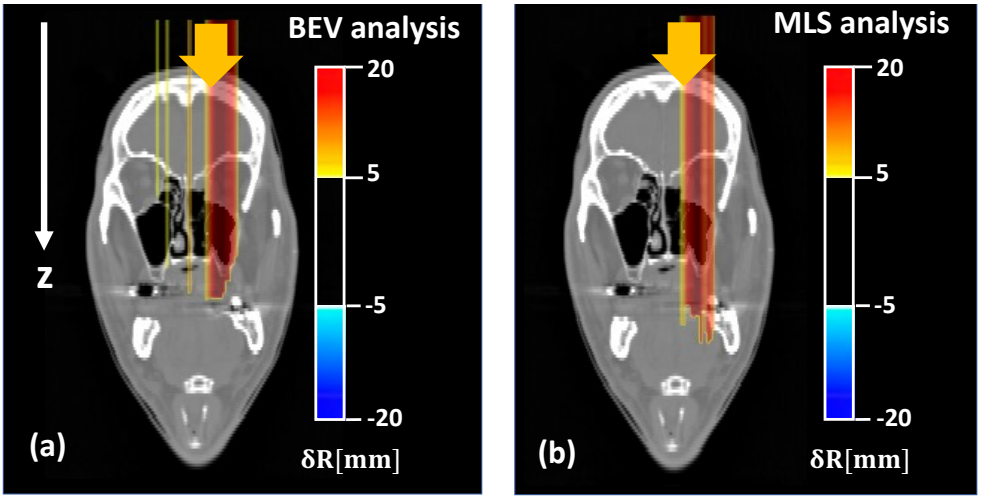

Figure 4. The maps O_{BEV} (a) and $O_(MLS)$ representing the estimated zones where the pencil beams pass that are affected by a range difference between PET_{ref} and PET_{mod}. The red color indicates that there is a beam overshoot in the modified PET image with respect to the reference. The beam direction is indicated with a yellow arrow.

DD acceptance criteria are 3 mm and 3% of the maximum dose, respectively [16, 18]. In this work, we applied the gamma index analysis directly on the activity distributions, rather than on the dose. This allows to compare, as in the other proposed methods, inter-fractional IB-PET images (see discussion). We made use of the three-dimensional γ-index function implemented in python3 [23].

3 Results

In figure 3 b we show the distribution of δR values that was obtained when considering all (x, y) points. For both methods, an asymmetric distribution is seen towards positive values, indicating that the range value of the modified PET image is larger than that in the reference situation. In other words, we expect that there is somewhere a range overshoot in the modified PET image with respect to the reference image. This was expected: the emptying of the cavity, were liquid material is replaced by air. results in a range increase (overshoot) of the pencil beams that cross the cavity.

In Figure 4 we show the O_{BEV} (a) and O_{MLS} (b) distribution resulting from comparing PET_{ref} with PET_{mod}, overlaid on the modified CT image. The red zones represent the location of the activity profiles along z, that have resulted in positive range differences, i.e., a beam overshoot, as a result of the emptied cavity.

In Figure 5 a we show the result of the VBM analysis. Here the purple and cyan zone represent zones, where significantly more and significantly less activity was found in the modified PET image with respect to the reference image, respectively. We observe that the cyan region corresponds well to the air cavity. The purple region is the zone where significantly more activity was found with respect to what was expected. This a result of the beam overshoot.

Regarding the γ-index analysis, we show in Figure 5 b the result of the analysis with tolerance criteria 3 mm/3%. Only gamma-index values above 1 are displayed, so that only the regions

where PET_{ref} and PET_{mod} are incompatible are highlighted. First of all, we observe that the emptied volume is clearly highlighted with these tolerance criteria. Second, we see some very small highlighted regions in zones where no morphological changes occur. Third, the region of the beam-overshoot, is seen to be somewhat smaller than the region identified with the VBM analysis (purple zone). This has to do with the parameters in the γ-index analysis (see Section 4).

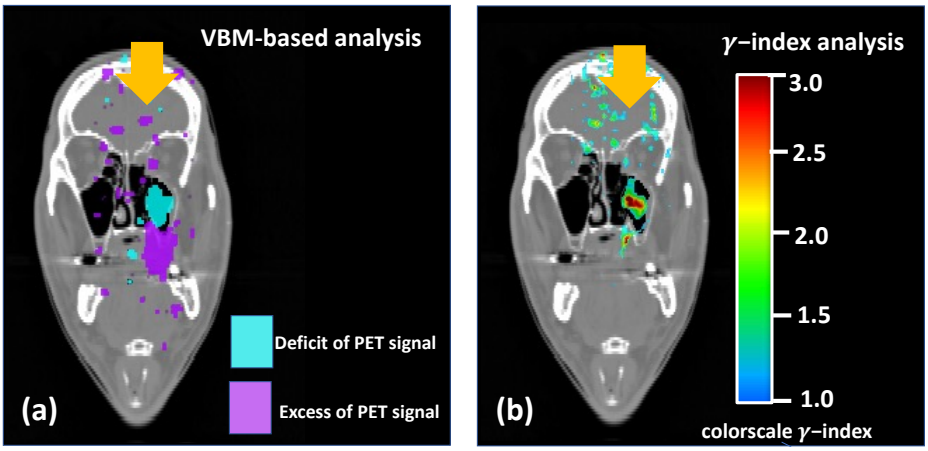

Figure 5. (a) Example of a CT slice with superimposed the VBM maps, with the zones exhibiting a deficit (cyan) and excess (purple) or activity with respect to expectations. (b) Example of a CT slice with superimposed the γ-index map (3 mm/3%.) The threshold here was 5% of the maximum activity value. In the colored regions, discrepancies were found between the reference and the modified image. The beam direction is indicated with a yellow arrow.

4 Discussion

We investigated four different methods on the same Monte Carlo simulated patient treatment. Advantages and disadvantages of the methods are the following.

In the BEV method the range differences between two images are calculated, based on the distal part of the activity along the beam axis (z). No information in the rest of the profiles is used. The maps allow to reveal a beam overshoot or undershoot. The information is 2- dimensional: δR in the (x, y) plane. Advantages of this method are that it is conceptually simple and the information is direct. The validity of this method was tested on data [13].

The MLS method gives, just like the BEV method, 2-dimensional information. With respect to the BEV method, the MLS method is much more complex, while giving more or less the same information. On the other hand, it is be somewhat less sensitive to fluctuations in the values in distal activity, since it does not look at a threshold value, but tries to find the best matching δR value for the end-of-range profile. This method is being tested on data, as reported in [21].

The VBM method is a 3-dimensional methods, and can possibly provide more spatial information. It displays precisely where the excess or lack of activity is located in appropriate color codes. This is interesting from a clinical point of view. The drawbacks of the method are the large number

of simulations that are needed to be able to calculate voxel-by-voxel the the p-value. Also, it is a MC driven method as present, and it was only validated on MC simulations.

The γ-index analysis is also a 3-dimensional method, indicating where the activity of a given distribution is anomalous with respect to another distribution, but does not indicate whether there is a lack or excess of activity. Moreover, the excess of activity is not clearly revealed. We believe that an optimization of the free parameters in the analysis can change this, and this is part of future research. Eventually, this method could help in transforming the activity distributions into a dose distribution [24, 25]. A disadvantage of the γ-index analysis is the long calculation time, and the choice of the parameters, for which no recommendations exist for PET image comparisons.

5 Conclusion

The four methods presented above are all valuable methods for comparing in-beam PET images taken during subsequent fractions in proton therapy. On one hand, the VBM and γ-index analysis methods can provide 3-dimensional information about discrepancies, which cannot be done with the BEV and MLs methods. On the other hand, they suffer more from falsely identified voxels, and their calculation time is much longer. It would be beneficial, to combine the various methods, so that regions that seem to have changed in one method, are confirmed with the other methods.

References

[1] H Paganetti, et al., *Adaptive proton therapy*, Phys. Med. Biol. 66 (2021)

[2] F. Albertini et al., *Online daily adaptive proton therapy*, Brit. J. Rad. 93 (2020) 20190594.

[3] H. Morgan and D. Sher, *Adaptive radiotherapy for head and neck cancer*, Canc. of the Head and Neck 5 (2020) 1.

[4] A.C. Kraan et. al., *Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors,* Int. J Radiat Oncol Biol Phys. 87(5):888-96 (2013).

[5] K. Parodi and J. Polf, *In vivo range verification in particle therapy.* Med. Phys. 45 (2018) e1036–e1050.

[6] W. Enghardt et al., *Charged hadron tumour therapy monitoring by means of pet.* Nuc. Instr. Meth. Phys. Res. A, 525 (2004) 284 – 288.

[7] T. Nishio et al., *The development and clinical use of a beam on-line pet system mounted on a rotating gantry port in proton therapy*, Int. J. Radiat. Oncol. Biol. Phys. 76 (2010) 277 – 286.

[8] A. Knopf and A. Lomax, *In vivo proton range verification: a review.* Phys. Med. Biol. 58 (2013) R131–160.

[9] A.C. Kraan, *Range verification methods in particle therapy: Underlying physics and Monte Carlo modeling* Front. in Oncol. 5 (2015).

[10] P. Dendooven et al. *Short-lived positron emitters in beam-on pet imaging during proton therapy*. Phys. Med. Biol. 60 (2015) 8923–8947.

[11] G. Bisogni et al., *INSIDE in-beam positron emission tomography system for particle range monitoring in hadrontherapy.* J. Med. Im. 4 (2016) 011005.

[12] Traini G, et al., *Review and performance of the dose profiler, a particle therapy treatments online monitor*. Phys. Med. 65 (2019) 84–93.

[13] E. Fiorina, et al., *Detection of interfractional morphological changes in proton therapy: A simulation and in vivo study with the INSIDE In-Beam PET.* Front. Phys. 8 (2021) 578388.

[14] K. Frey, et al., *Automation and uncertainty analysis of a method for in-vivo range verification in particle therapy.*. Phys. Med. Biol. 59 (2014) 5903–5919.

[15] A.C. Kraan et al,, *Localization of anatomical changes in patients during proton therapy with in-beam PET monitoring: A voxel-based morphometry approach exploiting Monte Carlo simulations,* . Med Phys 2022 49(1):23-40.

[16] D.A. Low, *A technique for the quantitative evaluation of dose distributions Comparative Study*, Med Phys (1998) 25(5):656-61.

[17] A-C. Knopf. et al., *Accuracy of Proton Beam Range Verification Using Post-Treatment Positron Emission Tomography/Computed Tomography as Function of Treatment Site*, Int. J. Radiat, Oncol.Biol.Phys. 79, 1, (2011) 297–304

[18] M. Miften et al., *Task Group Report, Tolerance limits and methodologies for IMRT measurement based verification QA: Recommendations of AAPM Task Group No. 218*, Med.Phys. 45, 4 (2018), e53-e83

[19] F. Pennazio et al. *Carbon ions beam therapy monitoring with the INSIDE in-beam PET* Phys. Med. Biol. 63 (2018).

[20] E. Fiorina et al. *Monte Carlo simulation tool for online treatment monitoring in hadron-therapy with in-beam PET: A patient study*, Phys. Med. 51 (2018), 71–80.

[21] M. Moglioni et al. *In-vivo range verification analysis with in-beam PET data for patients treated with proton therapy at CNAO*, Front. Oncol. 12:929949 (2022).

[22] G. Battistoni et al., *The FLUKA Code: an Accurate Simulation Tool for Particle Therapy*, Front. Oncol. 6.116 (2016).

[23] M. Wendling, et al., *Fast 3D gamma evaluation, gamma function*, Med. Phys.34(5):1647-54 (2007)

[24] S. Remmele, et al., *A deconvolution approach for PET-based dose reconstruction in proton radiotherapy*, Phys. Med. Biol.7;56(23):7601-19. (2011)

[25] M. Aiello, et al., *A Dose Determination Procedure by PET Monitoring in Proton Therapy: Monte Carlo Validation*, IEEE Trans. Nuc. Sc.60; 5, 3298-3304 (2013)