

A Set Theory- And ROC-Based Patient-Specific QA Measure To Compare Planned And Delivered Isodose Distributions In Photon Radiotherapy

Mateusz Baran

AGH University of Science and Technology: Akademia Gorniczo-Hutnicza imienia Stanislaw Staszica w Krakowie

Zbislaw Tabor (✉ tabor.zbislaw@gmail.com)

AGH University of Science and Technology: Akademia Gorniczo-Hutnicza imienia Stanislaw Staszica w Krakowie <https://orcid.org/0000-0002-9688-9718>

Damian Kabat

Narodowy Instytut Onkologii im Marii Sklodowskiej-Curie Panstwowy Instytut Badawczy Oddzial w Krakowie

Monika Tulik

Narodowy Instytut Onkologii im Marii Sklodowskiej-Curie Panstwowy Instytut Badawczy Oddzial w Krakowie

Kinga Jeleń

Instytut Fizyki Jadrowej im Henryka Niewodniczanskigo Polskiej Akademii Nauk

Michael PR Waligórski

Tadeusz Kosciuszko Cracow University of Technology: Politechnika Krakowska im Tadeusza Kosciuszki

Methodology

Keywords: quality assurance, dose distribution, set theory, confusion matrix

Posted Date: July 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-662679/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**A set theory- and ROC-based patient-specific QA measure to compare planned and delivered isodose distributions in
photon radiotherapy**

Mateusz Baran^{1,3}, Zbysław Tabor^{1,3}, Damian Kabat², Monika Tulik², Kinga Jeleń⁴, and Michael P. R. Waligórski^{3,4}

¹ AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Krakow, Poland

² Maria Skłodowska-Curie National Research Institute of Oncology Krakow Branch, Garncarska 11, 31-115 Krakow, Poland

³ Cracow University of Technology, Faculty of Materials Science and Physics, Podchorążych 1, 30-084 Krakow, Poland

⁴ Institute of Nuclear Physics Polish Academy of Sciences, Radzikowskiego 152, 31-342 Kraków, Poland,

Author to whom correspondence should be addressed: Zbysław Tabor, tabor.zbislav@gmail.com

Abstract

Background:

The DVH-based and Gamma Index patient-specific QA measures commonly applied in radiotherapy planning are unable to simultaneously deliver detailed locations and magnitudes of discrepancy between isodoses of planned and delivered dose distributions. By exploiting the receiver operating characteristic (ROC) statistical classification tool, compliance between a planned and delivered isodose may be locally evaluated, both for OAR and PTV, at any given isodose level. Thus, a patient-specific QA tool may be developed to supplement those presently available in clinical radiotherapy.

Materials and Methods:

A method to locally establish and report dose delivery errors in 3-D isodoses of planned (reference) and delivered (evaluated) dose distributions as a function of both the dose level and spatial location was developed. At any given isodose level, the total volume of dose delivery containing the reference and the evaluated isodoses is locally decomposed into four subregions: true positive – subregions within both reference and evaluated isodoses, true negative – outside of both of these isodoses, false positive – inside the evaluated isodose occurs but not the reference one, and false negatives – inside the reference isodose but not the evaluated one, as established over the total volume of dose delivery. From this decomposition a confusion matrix is derived and various indices calculated to quantify the discrepancies between the given planned and delivered isodose distributions.

Results:

Examples of clinical photon radiotherapy plans underwent analysis using the method developed. At some isodose levels, at anatomically significant locations, dose delivery errors were found which would not be highlighted either by dose volume histogram (DVH)-based QA metrics or by gamma analysis.

Conclusions:

The proposed method which generalizes the DVH-based QA method approach and is able to spatially locate delivery errors at selected isodose levels may supplement the presently applied gamma analysis and DVH-based QA measures in patient-specific radiotherapy planning.

Key words: quality assurance, dose distribution, set theory, confusion matrix

I. BACKGROUND

Advances in radiotherapy are reflected in increasingly complex tests of commissioning, therapy planning, on-board imaging, beam geometry, or beam delivery systems [1-5], and of dosimetry tools [6]. Among such tests, patient-specific procedures [1,2,7,8] remain the core component of clinical quality assurance (QA). While several computational methods to support patient-specific QA procedures have been developed [9-14], whether they can fully replace physical measurements in this area is still debatable.

Measurement-based patient-specific QA procedures have recently been reviewed by AAPM Task Group No. 218 [15]. Patient-specific QA, including point-wise dose difference between planned and delivered spatial dose distributions, distance to agreement analysis (DTA) [16], the composite test [17], and gamma analysis [18] are used in clinical practice and often integrated within commercial radiotherapy systems. However, limitations of these methods are also recognized. For example, gamma analysis, even with stringent tolerance criteria, may not preclude the occurrence of clinically significant dose delivery errors [19, 20]. In particular, apart from the passing rate, gamma index analysis returns the spatial locations of regions classified as containing dose delivery errors. However, in indicating such regions of dose delivery errors the global gamma index operates only marginally better than a random classifier [21].

Such drawbacks stimulate the search for alternative approaches to verify the distributions of dose delivered (as measured by the dosimetry system) against distributions of dose planned by the treatment planning system. Following the introduction of measurement-guided dose reconstruction [22], application of dose volume histograms (DVHs) to compare measured and planned dose distributions became a subject of debate [23-25]. Currently, the proposed DVH-based procedure for patient-specific QA is performed individually for volumes of interest (VOI), e.g., for the planning target (PTV) or organ at risk (OAR) volumes, outlined in computed tomography images. For a given VOI, one DVH is constructed for the planned dose distribution, and another for the dose distribution reconstructed from measurements. Various quantities are then derived as quality indicators from these DVHs, such as D_{90} - doses delivered to 90% of the volumes of the compared PTVs. If differences between such quantities exceed some pre-determined thresholds, the presence of a clinically significant dose delivery error is reported. Notably, in its current form, such DVH-based verification of dose delivery is inconclusive because even if, for a given VOI, the compared DVHs closely coincide (e.g., for PTV, their D_{90} - dose values are similar), the spatial distributions of planned and delivered doses may still significantly differ, both quantitatively and qualitatively. Moreover, correspondence between errors established from DVH analysis and errors established from gamma analysis has always been observed to be mild [23, 24], this raising some concern [25].

Clearly, there is still room for further development of methods supporting patient-specific QA. Here, we demonstrate an application of measures derived from set theory and receiver operating characteristics (ROC) to quantitatively compare a reference (planned) against a measured (evaluated) dose distribution. The proposed approach is supplementary to existing QA methods like DVH or gamma analysis because, as demonstrated in this paper, errors which do not manifest themselves in DVH or gamma can be captured and evaluated with the proposed approach.

II. MATERIALS AND METHODS

II.1. Materials

II.1.1. Illustrative 2D cases

The performance of the proposed measures is firstly demonstrated for several simplified examples relevant from the point of view of practical radiotherapy QA. The examples described in the current section were neither generated by a TPS nor they are measured dose distributions. They were generated with image processing tools with the intention to demonstrate the background of the proposed methods and to show how these methods highlight some aspects of the dose comparison task. Moreover, to demonstrate that the proposed methods are complementary to the current standard QA tools, the examples were selected so that the difference between the reference and evaluated dose distributions are reflected in neither corresponding DVHs nor gamma passing rates.

As a first example we consider a case of structure (which can be e.g. PTV) shown in Fig. 1a, a reference (planned) dose distribution shown in Fig. 1b and an evaluated (delivered) dose distribution shown in Fig. 1c. To obtain the distribution shown in Fig. 1b, the structure shown in Fig. 1a was blurred with an isotropic Gaussian filter. The size of this filter was equal to 5% of the width of the structure shown in Fig. 1a. In contrast, there is a clear anisotropy of blurring in the evaluated distribution shown in Fig. 1c: the horizontal size of the blurring kernel is only 3% and vertical size of the blurring kernel is 7% of the size of structure in Fig. 1a. The different amount of blurring in dose distributions may depend in real settings on, for example, the speed of MLC leaves moving from some initial to some terminating position or from amount of radiation leaking through the leaves.

In the next example of application of quantitative measures for isodose comparison it is assumed that dose must be uniformly delivered to a structure shown in Fig. 2a. The reference (planned) spatial dose distribution, shown in Fig. 2b is a blurred version of Fig. 2a (the size of the blurring kernel is equal to 5% of the size of the bright square in Fig. 2a). The distribution shown in Fig. 2b can be for example a plan generated by a TPS based on a prescription defined by the spatial distribution shown in Fig. 2a. The spatial dose distribution shown in Fig. 2c is larger by 4% and shifted horizontally and vertically by 4% of the size of the reference spatial distribution with respect to the spatial distribution

shown in **Fig. 2b**. Also, the penumbra region in **Fig. 2c** is 40% more wide than in **Fig. 2b**. The distribution shown in **Fig. 2c** can be for example an erroneously delivered dose based on a plan shown in **Fig. 2b**. In clinical settings the difference between the sizes of the reference (**Fig. 2b**) and evaluated (**Fig. 2c**) dose distributions may result from MLC leafs mispositioning. The difference between the penumbra sizes of the reference (**Fig. 2b**) and evaluated (**Fig. 2c**) dose distributions may result from differences between planned and executed speeds of MLC leaves during dose delivery.

In real settings, differences between reference and evaluated dose distribution can also result from other factors like differences in the performance of the real accelerator compared to the planning system (TPS), differences due to the accuracy of measurements and due to accuracy of measurement device settings, differences due to the accuracy of the TPS calculations, and differences in the geometry of a real multi-leaf collimator (MLC) compared to the TPS definition of MLC.

11.1.2. Clinical data

To illustrate the application of the proposed measures for comparison of isodoses, two radiotherapy plans: of prostate and brain tumours, were analysed. Both plans were prepared for a 6 MV photon beam of a Clinac 2300C/D linac. The analysed data sets consisted of CT scans, outlines of OARs and PTV, a radiotherapy plan (prepared using the Varian Eclipse 11.00 TPS and AAA algorithm), and a TPS-calculated dose distribution (planned doses).

In the prostate tumour plan only the prostate gland was irradiated, without lymph nodes. The total prescribed dose to the PTV was 63 Gy (split into 21 fractions). A five-field 6 MV 2300 C/D photon beam IMRT plan was prepared. The OARs were the rectum, bladder, and femoral heads. For the brain tumour plan the 6 MV 2300 C/D photon beam VMAT plan consisted of two partial-arcs. The dose to the PTV was 81 Gy (40.5 Gy per each arc) in 15 fractions. The OARs were the optic nerves, optic chiasm, and brain stem, all contained within the target volume. The planning constraints were to reduce the dose within the OARs to the necessary tolerance values while assuring that the therapeutic goal would be met.

The respective distributions of the delivered doses were calculated by Monte Carlo (MC) simulations, using the PRIMO software (www.primoproject.net) [26]. This software is based on the general-purpose PENELOPE 2011 MC engine [27]. The initial settings of the simulator were selected to reproduce real dose profiles measured in a water phantom. The simulator dose units were converted to Gy units to match planned and computed average doses to the PTV.

11.2. Methods

II.2.1. Pairwise comparison of isodoses

Pairwise comparison of isodoses is based on calculating the so-called confusion matrix which is widely used to summarize results of classifier testing in predictive analysis and in machine learning. Numbers of false positives, false negatives, true positives, and true negatives are reported in cells of the confusion matrix - which can be further used to compute other quantities such as sensitivity or specificity, along principles of receiver operating characteristic (ROC) statistics.

The proposed quantitative method compares regions enclosed by isodoses of two spatial dose distributions: the reference dose distribution, D_R , and the evaluated dose distribution, D_E . Both these spatial dose distributions are defined over some volume of interest V which is formed by a grid of spatial locations $x_{i,j,k}$ such that:

$$V = \{x_{i,j,k} = (i\Delta_x, j\Delta_y, k\Delta_z) : i \in \{0,1,2,\dots,I\}, j \in \{0,1,2,\dots,J\}, k \in \{0,1,2,\dots,K\}\} \quad (1)$$

where Δ_x , Δ_y , and Δ_z are grid spacings along x -, y -, and z -axes, respectively. We denote by $D_R(i,j,k)$ the reference dose at $x_{i,j,k}$. Analogously, $D_E(i,j,k)$ denotes the evaluated dose at $x_{i,j,k}$.

To evaluate conformance between D_R and D_E with respect to some dose value Th , we define the reference region $D_{R,Th}$ and the evaluated region $D_{E,Th}$:

$$D_{R,Th} = \{x_{i,j,k} \in V : D_R(i,j,k) \geq Th\}, \quad (2)$$

$$D_{E,Th} = \{x_{i,j,k} \in V : D_E(i,j,k) \geq Th\}, \quad (3)$$

i.e. $D_{R,Th}$ (or $D_{E,Th}$) are sub-volumes of D_R (or D_E) in which all doses are larger than or equal to Th . Neglecting discretization errors, the boundaries of $D_{R,Th}$ and $D_{E,Th}$ correspond to the Th isodose in D_R and D_E , respectively. Given the sets of voxels $D_{R,Th}$, and $D_{E,Th}$, the numbers of true positives TP_{Th} , true negatives TN_{Th} , false positives FP_{Th} and false negatives FN_{Th} are defined as follows (**Fig. 3**):

$$\begin{aligned} TP_{Th} &= |D_{R,Th} \cap D_{E,Th}| \\ TN_{Th} &= |(D_{R,Th} \cup D_{E,Th})^c|, \\ FP_{Th} &= |D_{E,Th} \setminus D_{R,Th}| \\ FN_{Th} &= |D_{R,Th} \setminus D_{E,Th}| \end{aligned} \quad (4)$$

where $|S|$ denotes the number of elements in the set S , and S^c denotes the complement of set S . The numbers TP_{Th} , TN_{Th} , FP_{Th} and FN_{Th} together form a 2x2 confusion matrix, M_{Th} :

$$M_{Th} = \begin{pmatrix} TP_{Th} & FN_{Th} \\ FP_{Th} & TN_{Th} \end{pmatrix}. \quad (5)$$

Various overlap measures can be derived from M_{Th} . In the present study we focus on sensitivity, SEN_{Th} , specificity, SPE_{Th} , and Dice coefficient, DC_{Th} , but other measures may also prove to be useful in clinical applications. We define our overlap measures for the isodose Th as follows [28]:

$$SEN_{Th} = 1 - FNR_{Th} = \frac{TP_{Th}}{TP_{Th} + FN_{Th}}, \quad (6)$$

$$SPE_{Th} = 1 - FPR_{Th} = \frac{TN_{Th}}{TN_{Th} + FP_{Th}}, \quad (7)$$

$$DC_{Th} = \frac{2TP_{Th}}{2TP_{Th} + FP_{Th} + FN_{Th}}, \quad (8)$$

where FPR_{Th} , and FNR_{Th} are, respectively, the false positive rate and the false negative rate for dose Th . Note that $FNR_{Th} > 0$ indicates the presence of cold regions, i.e. regions where the planned doses are higher than Th and delivered doses are lower than Th . For example, FNR_{Th} of 2.5% for a certain isodose level Th means that over 2.5% of the volume within that isodose, the delivered dose is lower than that Th value— it is a cold region within the isodose Th .

Similarly, $FPR_{Th} > 0$ indicates the presence of hot regions, i.e. regions where planned doses are lower than Th , while delivered doses are higher than Th . An FPR_{Th} of 0.1% for the isodose Th means that over 0.1% of the volume exterior to that isodose Th , the dose delivered is greater than Th – it is then a hot region exterior to isodose Th .

Finally, Dice's coefficient for a certain isodose Th , DC_{Th} , is the ratio of the intersection volume of the regions covered by the planned and realized isodose Th to the arithmetic mean of the volumes of two regions: one covered by the planned isodose Th , and the other one covered by the delivered isodose Th .

Note that the DVH for some region can be considered to be equivalent to SEN (Eq. 6) if an artificial reference dose distribution is appropriately constructed from the CT-based outlines of this region (i.e., while constructing this reference distribution it is sufficient to assume dose to be zero outside the outlines and equal to the maximum value of delivered dose inside these outlines – then all isodoses of this reference distribution coincide with the region outlines).

Calculation of overlap measures as a function of dose Th concludes our evaluation of pair-wise overlap between isodoses of the reference dose distribution D_R and the evaluated dose distribution D_E .

II.2.2. Evaluating the spatial shift between two isodoses

Overlap measures alone are not sufficient to assess differences between two isodoses. **Fig. 4a** shows two cases of dose distributions which are characterized by exactly the same dose volume histograms but are otherwise very different. **Fig. 4b** shows two cases of reference and evaluated isodoses which are characterized by exactly the same overlap measures but are also otherwise very different especially in terms of their relative shift. To account for such cases, this relative shift of reference and evaluated isodoses should also be assessed. For this purpose, the Hausdorff distance [29] may be computed - the greatest of all the distances from a point in one subregion to the closest point in the other subregion.

The generalized Hausdorff distance $HD(D_{R,Th}, D_{E,Th}, q_{min}, q_{max})$ between a reference region $D_{R,Th}$ and an evaluated region $D_{E,Th}$ corresponding to an isodose Th is defined as follows [29]:

$$HD(D_{R,Th}, D_{E,Th}, \underline{q}, \bar{q}) = \max\{h(D_{R,Th}, D_{E,Th}, \underline{q}, \bar{q}), h(D_{E,Th}, D_{R,Th}, \underline{q}, \bar{q})\} \quad (9)$$

$$h(A, B, \underline{q}, \bar{q}) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (10)$$

where $h(A, B, \underline{q}, \bar{q})$ is the generalized directed Hausdorff distance, $\max_{a \in A} f(a)$ over a in A is the \bar{q} th quantile of the values $f(a)$ of the function f over elements a of a set A , and $\|a - b\|$ is the Euclidean distance between points a and b . It can be derived from distance-to-agreement (DTA) analysis that Eq. (10) represents the maximum DTA between the isodose Th in D_R and in D_E .

It follows from the definition of generalized Hausdorff distance that replacing the two quantiles in Eq. (10) with, respectively, maximum and minimum operators, one computes the largest distance between a point in one set and the nearest point in the other set, i.e. the ordinary Hausdorff distance. While being sensitive to outliers, the ordinary Hausdorff distance offers the highest QA safety margin.

III. Results

III.1. 2D illustrative cases

In **Fig. 5a** we show reference and evaluated DVHs, both calculated for, respectively, reference and evaluated dose distributions shown in **Fig. 1b-c** with respect to the structure shown in **Fig. 1a**. Clearly, the substantial difference between

the two dose distribution is not reflected in the difference between the two DVHs. It is also not reflected in gamma (3%/3mm) passing rates, which are equal to 99.7% for the whole image and 97.4% for the structure shown in **Fig. 1a**. In contrast, the proposed indices for isodose comparison, plotted in **Fig. 5b**, demonstrate that there is discrepancy between the two distributions. The plots of FPR and FNR vs. dose level show that there are both false positive (hot) regions and false negative (cold) regions for both low and high reference (planned) dose levels. Isodoses at level of 100 units are close by while both lower and higher isodoses deviate from one another as follows from the plots of Dice's coefficient and Hausdorff distance. The Dice's coefficient for isodose of 100 units is approximately equal to 1 which means that the reference and evaluated isodoses of level 100 are close by which suggest that the discrepancy between reference and evaluated distributions may be related to some distortion with isodose of 100 unit being unchanged by the distortion. **Fig. 5c** demonstrates the difference between the two distributions shown in **Fig. 1b** and **Fig. 1c**. In **Fig. 5d** the results of gamma (3%/3mm) analysis are shown. To calculate gamma index, the image shown in **Fig. 1b** was used as the reference distribution and the image shown in **Fig. 1c** was used as the evaluated distribution. White regions of this image correspond to gamma index values larger than 1. Note that the presence of hot (red) and cold (blue) regions is clear in the difference image (**Fig. 5c**). The proposed indices help in determining how these cold and hot regions are associated with dose levels which association does not follow directly from either the difference image or gamma index analysis. For example, as the plot of FNR vs. dose level demonstrates presence of cold regions for high planned doses e.g. dose equal to 90 units, these particular cold regions can be directly associated with spatial locations corresponding to this dose level and their clinical significance can be assessed. This association of dose difference and dose level is absent in the standard approaches used in clinical practice.

In **Fig. 6a** we show reference and evaluated DVHs, both calculated for, respectively, reference and evaluated dose distributions shown in **Fig. 2b-c** with respect to the structure shown in **Fig. 2a**. The difference between the two distribution shown in **Fig. 2b** and **Fig2c** is even more prominent than for the case depicted in **Fig. 1** but again this difference is reflected in neither the difference between the two DVHs nor in gamma (3%/3mm) passing rates (equal to 99.1% for the whole image and 99.4% for the structure shown in **Fig. 2a**). As for the previous case, the proposed indices for isodose comparison demonstrate that there is discrepancy between the two distributions. The plots of 1-SPE and 1-SEN vs. dose level show that there are false positive (hot) regions for low reference dose levels and false negative (cold) regions for high reference dose levels. Note that from the viewpoint of radiotherapy goals the presence of hot regions for low doses and cold regions for high doses is especially dangerous as it may possibly correspond to delivering too high doses to organs at risk or to delivering too low doses to PTV. The Dice's coefficient is not larger than 0.96 and the Hausdorff distance is relatively big which means that, besides some distortion, there is a relative shift between the two

distributions. **Fig. 6c** demonstrates the difference between the two distributions shown in **Fig. 2b** and **Fig. 2c**. In **Fig. 6d** the results of gamma (3%/3mm) analysis are shown. As in the previous case the presence of hot (red) and cold (blue) regions is clear in the difference image (**Fig. 6c**). However, looking at **Fig. 6b** it is clear that cold regions are associated with high doses while hot regions are associated with low doses which is not evident from either **Fig. 6c** or **Fig. 6d**.

In the next section we demonstrate the application of the proposed methods to clinical cases. In contrast to the cases discussed above the interpretation of both dose difference and gamma index images is, due to noise, no longer straightforward while the proposed methods provide direct insight into the nature of the difference between a reference and an evaluated dose distributions.

III.2. Clinical data

To illustrate the clinical application of the proposed method we first consider the PTV region of the IMRT prostate plan. As seen in **Fig. 7a**, DVHs calculated for the planned (TPS-calculated) and the delivered (MC-calculated) doses almost coincide. Doses to 90% of the PTV are 62 Gy and 61.8 Gy for the planned and evaluated dose distributions, respectively, so they differ only by some 0.3%. The gamma passing rate for the PTV is 99.97%, meaning that both planned and delivered dose distributions coincide in terms of common QA measures. In **Fig. 7b** we show how our proposed measures of discrepancy between isodoses highlight the differences between these two distributions, and in particular, how do such differences depend on the level of the selected isodoses, for dose values present within the PTV region. Note that at all values of selected isodose levels both rates - false positive (FPR) and false negative (FNR) are greater than zero, indicating the presence of both hot and cold regions within the PTV, especially at isodose levels above and including 63 Gy (which is the total dose to the PTV; this dose level is marked by vertical dashed lines in **Fig. 7b**). The Hausdorff distance between planned and delivered isodoses can be as large as 2 cm, at the 63 Gy isodose. The volume of the cold region for the 63 Gy isodose is about 11% of the PTV volume. The volume of the hot region for isodose 63 Gy is also about 11% of the PTV volume. In **Fig. 8** we show samples of CT slices where the PTV is marked as green, cold regions marked as blue, and hot regions marked as red (the same colour convention is applied throughout). The hot and cold regions were always determined for the 63 Gy isodose - for which we observe a substantial discrepancy between the planned and delivered dose distributions, as demonstrated in **Fig. 7b**. To assess the significance of the observed cold and hot regions, we calculated histograms of differences between planned and delivered dose distributions over these two regions for the 63 Gy isodose, shown in **Fig. 9**. The mean dose difference, standard deviation, and maximum dose difference within the cold region of the 63 Gy isodose are, respectively, 0.88 Gy, 0.54 Gy, and 4.0 Gy. The respective values within the hot region of the 63 Gy isodose are 0.56 Gy, 0.28 Gy, and 2.2 Gy. Note that the observed differences within both cold and

hot regions do not exceed 5% of the total planned dose to the PTV (63 Gy) - which could indicate that these hot and cold regions do not pose a clinically significant problem.

In our second clinical example we consider the PTV region of the VMAT brain plan. Doses to 90% of the PTV are 81.4 Gy and 82.3 Gy for the planned and evaluated dose distributions, respectively, indicating their relative difference of 1.1%. The gamma passing rate for the PTV is 94.5%. The DVH's for reference and evaluated distributions closely coincide (**Fig. 10a**). In **Fig. 10b** we show our measures of discrepancy between isodoses versus isodose level for dose values observed within the PTV volume. Noting large discrepancies between isodoses above and including 81 Gy (which is the total planned dose to the PTV; this dose level is marked by vertical dashed lines in **Fig. 10b**), with FPR increasing to over 0.6 and HD of the order of 10mm, the hot and cold regions were determined for the 81 Gy isodose. The volume of the cold region for this isodose is about 2.7% of the PTV volume. The volume of the hot regions for the 81 Gy isodose is about 5.4% of the PTV volume. In **Fig. 11** we show samples of CT slices, with the PTV, cold and hot regions with respect to the 81 Gy isodose, marked appropriately. Histograms of the dose differences for cold and hot regions for the 81 Gy isodose are shown in **Fig. 12**. The mean dose difference, standard deviation, and maximum dose difference within the cold region of the 81 Gy isodose are 5.3 Gy, 2.5 Gy, and 16 Gy, respectively. The respective values within the hot region of the 81 Gy isodose are 4.3 Gy, 2.0 Gy, and 11.8 Gy. Note that the observed dose differences within the cold regions of the 81 Gy isodose exceed 5% of the total planned dose to the PTV within 69% of the volume of cold regions, which consists about 1.9% of the total PTV volume. The observed dose differences within hot regions of the 81 Gy isodose exceed 7% of the total planned dose to the PTV within 24% of the volume of hot regions, which is about 1.3% of the total PTV volume. These differences, while reflected neither by DVH nor by gamma analysis, may have potentially important clinical implications.

In our third clinical example, we consider one of the OAR regions (the brain) of the VMAT brain plan. Doses to 5% of the OAR are 58.3 Gy and 58.6 Gy for the planned and evaluated dose distributions, respectively. The gamma passing rate for this OAR is 99.5%. DVHs for reference and evaluated dose distributions are also very similar (**Fig. 13a**). In **Fig. 13b** we show our measures of discrepancy between planned and measured isodoses. The hot and cold regions were determined for the 40 Gy isodose (denoted by vertical dashed lines in **Fig. 13b**). The relative volumes of the cold and hot regions for this 40 Gy isodose are about 3% and 1.7% of the OAR volume. In **Fig. 14** we show samples of CT slices where the OAR, cold and hot regions with respect to the 40 Gy isodose are appropriately marked. Histograms of the dose differences for cold and hot regions for the 40 Gy isodose are shown in **Fig. 15**. While the hot region of the 40 Gy isodose is relatively small (1.7% of OAR volume), substantial parts of this hot region receive doses higher than those planned, by more than 5% of the total planned PTV dose.

IV. Discussion

We introduce a novel quantitative method of comparing two dose distributions, encoded as collections of regions delimited by isodoses specified within the range of doses encountered in the compared dose distributions. We have demonstrated our method by analysing illustrative 2D cases as well as examples of clinical plans: a prostate IMRT, and a brain VMAT plan. While, for clinical cases, the Monte Carlo-calculated dose distribution was compared here against one calculated by the TPS to represent a true composite dose measurement, the proposed method can be applied without modification to any other dose measurement protocols, such as perpendicular composite or perpendicular field-by-field.

Notably, *DVH* analysis is a special case of our proposed framework, corresponding to sensitivity, calculated when comparing the evaluated dose distribution (either planned or measured) with an artificial distribution constructed from the CT outlines of the analysed irradiated region. The framework we propose extends this analysis by several other options, such as the Dice coefficient, Hausdorff distance etc. Its main advantage over *DVH* analysis is in presenting dose delivery errors as a function of the investigated dose level. In contrast to *DVH*, the proposed approach analyses dose delivery errors from a perspective considerably broader than only via true positive rate. The proposed features Another important advantage of our approach over *DVH* is that it enables measured and delivered dose distributions to be compared directly and comprehensively -unlike in common *DVH*-based analysis where comparisons between CT-outlined structures of interest are made separately. This particular advantage over common *DVH* analysis is especially important. We have demonstrated clinically relevant cases where no differences between measured and planned dose distribution being stated by *DVH* analysis, substantial discrepancies could nevertheless be recognised using our approach.

Analysis of the dependence of our proposed indices on the isodose level delivers information about discrepancies between planned and delivered isodoses. Because each isodose set is associated with a collection of spatial locations, one may readily identify spatial locations where discrepancies between the reference and evaluated dose distributions are observed. This decomposition simultaneously uncovers regions of excessively (false positive – hot) or insufficiently (false negative – cold) delivered doses. This spatial analysis option is analogous to the detection of spatial locations of dose delivery errors via gamma index analysis. After determining either cold or hot regions, the dose differences between planned and delivered doses can be further analysed to demonstrate dose delivery errors of particular clinical significance, e.g., exceeding certain pre-determined acceptability thresholds. In a manner similar to gamma index analysis, the proposed quantitative isodose comparison may also be summarized in the form of single indices. For instance, the area under the plots of the index values, considered as functions of the dose level and

normalized to the range of available dose values, may be readily interpreted: if the normalized area under the plot of DC as a function of dose level is equal to 1, then agreement between the two dose distributions is perfect.

Our clinically relevant therapy plans underwent isodose comparison methods based on measures of intersection between two regions enclosed by isodoses, and based on the distance between such regions. These two approaches deliver complementary information. There is close analogy between simultaneous analysis of intersection volume-based indices and Hausdorff distance, and the common composite analysis which considers both dose differences and DTA when comparing two dose distributions.

Notably, the proposed method of comparison of planned and delivered dose distributions highlights local discrepancies which in the case of other methods, such as gamma analysis, may not be apparent. As such, it can be a supplementary method to the existing QA approaches. As seen in **Fig. 16**, the differences between isodoses are not merely limited to isolated voxels - attributable to noise - but may be structured over larger volumes - due to other factors, perhaps multileaf sliding window (**Fig. 16**, top) or arc misalignment (**Fig. 16**, bottom). The close vicinity of tissues of widely differing densities may also present as hot and cold regions in **Fig. 16**, as over such regions the TPS algorithms may have limited accuracy against MC simulations. Also, volume rendering of CT slices, or technical limitations in dynamic plan implementation by the TPS could contribute. For whatever reason, such discrepancies - even over quite large volumes - would not be discernible by gamma analysis if their DTA remained below the usually applied passing rates. Thus, with the proposed method, a truly comprehensive analysis of the prepared treatment plan is possible with simultaneous attention to all volumes of interest, especially to those of the target and organs at risk – which up to now have always been considered separately. Finally, particular attention may be focused on anatomically relevant areas where accuracy of dose delivery is critical to the overall success of the radiotherapy procedure.

V. Conclusions

Quantitative methods for comparing two dose distributions (a reference one and an evaluated one) based on the confusion matrix and set theory have been introduced. Two clinical plans (prostate IMRT and brain VMAT) have been analyzed and the application of the proposed methods for these cases has been demonstrated. The proposed dose delivery verification method reports dose delivery errors as a function of both dose levels and spatial locations. It is shown that DVH analysis is a special case of the proposed approach. Similarly to gamma index analysis the results of the proposed approach indicates spatial regions of erroneous dose delivery and can be summarized in a form of single indices.

Abbreviations

MC: Monte Carlo, DVH: dose volume histogram, DTA: distance to agreement, TP: true positive, TN: true negative, FP: false positive, FN: false negative, SEN: sensitivity, SPE: specificity, FPR: false positive rate, FNR: false negative rate, DC: Dice coefficient, HD: Hausdorff distance.

Author's contributions

MB: conceived the work, conducted the simulations, wrote the manuscript and revised it. ZT conceived the work, performed the subsequent analyses, wrote the manuscript and revised it. DK conducted the measurements, wrote the manuscript and revised it. MT conducted the measurements, wrote the manuscript and revised it. KJ conducted the measurements, wrote the manuscript and revised it. MW wrote the manuscript and revised it.

Funding

K. Jeleń acknowledges the support of InterDokMed project No. POWR.03.02.00-00-I013/16

M. Baran, Z. Tabor, D. Kabat, K. Jeleń and M. P. R. Waligórski acknowledge the support of TEAM-NET project POIR.04.04.00-00-0015E5/18 from the Foundation for Polish Science (FNP)

Availability of data and materials

Not applicable

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors consent to the publication.

Competing interests

The author declares that she has no competing interests.

References

- [1] Ezzell GA, Galvin JM, Low D, Palta JR, Rosen I, Sharpe MB, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT subcommittee of the AAPM radiation therapy committee. *Med Phys.* 2003;30:2089–2115.
- [2] Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D et al. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. *Med Phys.* 2009;36:5359–5373.
- [3] Tabor Z, Kabat D, Tomaszuk M, Kycia R, Latała Z. A generic multi-modular phantom for testing geometry of a linac c-arm as a part of quality control in radiotherapy, *Med Phys.* 2017;44:4989–5000.

- [4] Tulik M, Kabat D, Kycia R, Tabor Z, Woszczyzna A, Latała Z. A framework for calibration of on-board imagers of medical linear accelerators, *Phys Med*. 2018;47:80-85.
- [5] Baran M, Rzecki K, Kabat D, Tulik M, Wydra A, Derda Z et al. A simulation-based method for evaluating geometric tests of a linac c-arm in quality control in radiotherapy. *J App Clin Med Phys*. 2019;20:133-142.
- [6] Low DA, Moran JM, Dempsey JF, Dong L, Oldham M. Dosimetry tools and techniques for IMRT. *Med Phys*. 2011;38:1313–1338.
- [7] Moran JM, Dempsey M, Eisbruch A, Fraass BA, Galvin JM, Ibbott GS et al. Safety considerations for IMRT: executive summary. *Med Phys*. 2011;38:5067.
- [8] Hartford AC, Galvin JM, Beyer DC, Eichler TJ, Ibbott GS, Kavanagh B et al. American College of Radiology (ACR) and American Society for Radiation Oncology (ASTRO) practice guideline for intensity-modulated radiation therapy (IMRT). *Am J Clin Oncol*. 2012;35:612–617.
- [9] Pawlicki T, Yoo S, Court LE, McMillan SK, Rice RK, Russell JD et al. Moving from IMRT QA measurements toward independent computer calculations using control charts. *Radiother Oncol*. 2008;89:330–337.
- [10] Fan J, Li J, Chen L, Luo W, Du Plessis F, Xiong W et al. A practical Monte Carlo MU verification tool for IMRT quality assurance. *Phys Med Biol*. 2006;51:2503–2514.
- [11] Leal A, Sanchez-Doblado F, Arrans R, Rosello J, Pavon EC, Lagares JI. Routine IMRT verification by means of an automated Monte Carlo simulation system. *Int J Radiat Oncol Biol Phys*. 2003;56:58–68.
- [12] Agnew A, Agnew CE, Grattan MWD, Hounsell AR, McGarry CK. Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries. *Phys Med Biol*. 2014;59:N49–N63.
- [13] Rangaraj D, Zhu M, Yang D, Palaniswaamy G, Yaddanapudi S, Wooten OH et al. Catching errors with patient-specific pretreatment machine log file analysis. *Pract Radiat Oncol*. 2013;3:80–90.
- [14] Stell AM, Li JG, Zeidan OA, Dempsey JF. An extensive log-file analysis of step-and-shoot intensity modulated radiation therapy segment delivery errors. *Med Phys*. 2004;31:1593–1602.
- [15] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys*. 2018;45:e53-e83.
- [16] Van Dyk J, Barnett RB, Cygler JE, Shragge PC. Commissioning and quality assurance of treatment planning computers. *Int J Radiat Oncol Biol Phys*. 1993;26:261–273.
- [17] Harms WB Sr, Low DA, Wong JW, Purdy JA. A software tool for the quantitative evaluation of 3D dose calculation algorithms. *Med Phys*. 1998;25:1830–1836.

- [18] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys.* 1998;25:656–661.
- [19] Blanck O, Masi L, Damme MC, Hildebrandt G, Dunst J, Siebert FA et al. Film-based delivery quality assurance for robotic radiosurgery: Commissioning and validation. *Phys Med.* 2015;31:476-83.
- [20] Heilemann G, Poppe B, Laub W. On the sensitivity of common gamma-index evaluation methods to MLC misalignments in Rapidarc quality assurance. *Med Phys.* 2013;40:031702
- [21] Tulik M, Kabat D, Baran M, Kycia R, Tabor Z. Use of statistical approaches to improve the quality control of the dose delivery in radiotherapy. *Phys Med Biol* 2019;64:145018.
- [22] Olch AJ. Evaluation of the accuracy of 3DVH software estimates of dose to virtual ion chamber and film in composite IMRT QA. *Med Phys.* 2012;39:81–86.
- [23] Nelms BE, Opp D, Robinson J, Wolf TK, Zhang G, Moros E et al. VMAT QA: measurement-guided 4D dose reconstruction on a patient. *Med. Phys.* 2012;39:4228–4238.
- [24] Stasi M, Bresciani S, Miranti A, Maggio A, Sapino V, Gabriele P. Pretreatment patient-specific IMRT quality assurance: a correlation study between gamma index and patient clinical dose volume histogram. *Med Phys.* 2012;39:7626–7634.
- [25] Visser R, Wauben DJ, de Groot M, Steenbakkens RJHM, Bijl HP, Godart J et al. Evaluation of DVH-based treatment plan verification in addition to gamma passing rates for head and neck IMRT. *Radiother Oncol.* 2014;112:389–95.
- [26] Rodriguez M, Sempau J, Brualla L. PRIMO: a graphical environment for the Monte Carlo simulation of Varian and Elekta linacs. *Strahlentherapie und Onkologie* 2013;189: 881–886.
- [27] Salvat F, Fernández-Varea JM, Sempau J. PENELOPE 2011—a code system for Monte Carlo simulation of electron and photon transport, Issy-les-Moulineaux. France: OECD Nuclear Energy Agency (2011).
- [28] Sokolova M, Lapalme G. Systematic analysis of performance measures for classification tasks. *Information Processing and Management* 2009; 45: 427–437
- [29] Crum WR, Camara O, Hill DLG. Generalized Overlap Measures for Evaluation and Validation in Medical Image Analysis. *IEEE Trans Med Imag* 2006;25:1451 - 1461

Figure titles and legends

Fig. 1 The first illustrative case used to demonstrate application of the proposed methods. (a) A structure of interest, (b) a reference isotropic (planned) dose distribution, and (c) an anisotropic evaluated (delivered) dose distributions.

Fig. 2 The second illustrative case used to demonstrate application of the proposed methods. (a) A structure of interest, (b) a reference (planned) dose distribution, and (c) an evaluated (erroneously delivered) dose distribution.

Fig. 3 Split of the irradiated volume into subvolumes. Within the irradiated volume V there is a subvolume DR,Th (symbolized by the bottom shape) enclosed by an isodose $IsoR(Th)$ (solid line) of reference distribution DR and a subvolume DE,Th (symbolized by the top shape) enclosed by an isodose $IsoE(Th)$ of evaluated distribution (dashed line) DE .

Fig. 4 Illustrative cases demonstrating importance of analysing shift between isodoses. (a) Dose volume histograms are not sensitive to shift of isodoses provided that the volume enclosed by the isodoses remains unchanged – DVHs for both left and right dose distribution within the same PTV region are exactly the same. (b) Bottom (brighter) squares within volumes V are regions enclosed by a reference isodose $IsoR(Th)$. Top (darker) squares are regions enclosed by an evaluated isodose $IsoE(Th)$. $SENTh$, $SPETH$, and $DCTH$ calculated for $IsoR(Th)$ and $IsoE(Th)$ are the same for the right and left cases. The spatial shift between $IsoR(Th)$ and $IsoE(Th)$ is however very different.

Fig. 5 The results of the proposed analysis of the first illustrative 2D case. (a) DVHs calculated for distributions shown in Fig. 1b and Fig. 1c with respect to the structure shown in Fig. 1a do not highlight the clear difference between these distributions. (b) The proposed indices for pair-wise comparison highlight differences between the two distributions shown in Fig. 1b and Fig. 1c, in particular presence of cold and hot regions. (c) The difference between the two distributions shown in Fig. 1b and Fig. 1c. (d) The results of gamma analysis - in white regions gamma index, computed for the two distributions shown in Fig. 1b and Fig. 1c is larger than 1.

Fig. 6 The results of the proposed analysis of the second illustrative 2D case. (a) DVHs calculated for distributions shown in Fig. 2b and Fig. 2c with respect to structure shown in Fig. 2a. (b) The proposed indices for pair-wise comparison of the distributions shown in Fig. 2b and Fig. 2c. (c) The difference between the two distributions shown in Fig. 2b and Fig. 2c. (d) The results of gamma analysis - in white regions gamma index, computed for the two distributions shown in Fig. 2b and Fig. 2c is larger than 1.

Fig. 7 The results of the proposed analysis for PTV of the IMRT plan. (a) Reference (planned) and evaluated (delivered) DVHs for PTV of the IMRT plan. (b) Measures of discrepancy between isodoses in planned and delivered dose distributions for PTV of the IMRT plan. Vertical dashed line denotes isodose level for which cold and hot regions are subsequently analysed.

Fig. 8 Visualisations of QA analysis for PTV of the IMRT plan. Left column: Sample CT slices demonstrating presence of cold (blue) and hot (red) regions within PTV (green). The hot and cold regions are determined for isodose 63 Gy. Middle column: dose differences. Right column: gamma index values.

Fig. 9 Histogram of differences between planned and delivered doses for cold and hot regions of isodose 63 Gy within PTV of the IMRT plan.

Fig. 10 The results of the proposed analysis for PTV of the VMAT plan. (a) Reference (planned) and evaluated (delivered) DVHs for PTV of the VMAT plan. (b) Measures of discrepancy between isodoses in planned and delivered dose distributions for PTV of the VMAT plan. Vertical dashed line denotes isodose level for which cold and hot regions are subsequently analysed.

Fig. 11 Visualisations of QA analysis for PTV of the VMAT plan. Sample CT slices demonstrating presence of cold (blue) and hot (red) regions within PTV (green). The hot and cold regions are determined for isodose 81 Gy. Middle column: dose differences. Right column: gamma index values.

Fig. 12 Histogram of differences between planned and delivered doses for cold and hot regions of isodose 81 Gy within PTV of the VMAT plan.

Fig. 13 The results of the proposed analysis for OAR of the VMAT plan. (a) Reference (planned) and evaluated (delivered) DVHs for OAR (brain) of the VMAT plan. (b) Measures of discrepancy between isodoses in planned and delivered dose distributions for OAR (brain) of the VMAT plan. Vertical dashed line denotes isodose level for which cold and hot regions are subsequently analysed.

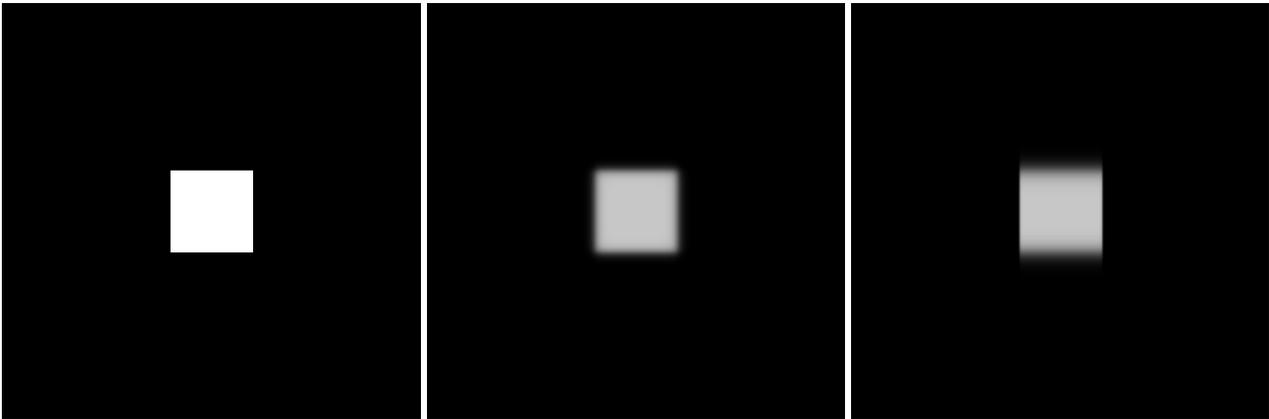
Fig. 14 Visualisations of QA analysis for OAR of the IMRT plan. Sample CT slices demonstrating presence of cold (blue) and hot (red) regions within OAR (green). The hot and cold regions are determined for isodose 40 Gy of OAR (brain) of the VMAT plan. Middle column: dose differences. Right column: gamma index values.

Fig. 15 Histogram of differences between planned and delivered doses for cold and hot regions of isodose 40 Gy within OAR (brain) of the VMAT plan.

Fig. 16 3-D renderings of the PTV (green), hot (red), and cold (blue) regions for clinical cases. Top - the IMRT prostate plan: hot and cold regions determined for the 63 Gy isodose. Bottom - for the VMAT brain plan: hot and cold regions determined for the 81 Gy isodose.

Figures:

Fig. 1

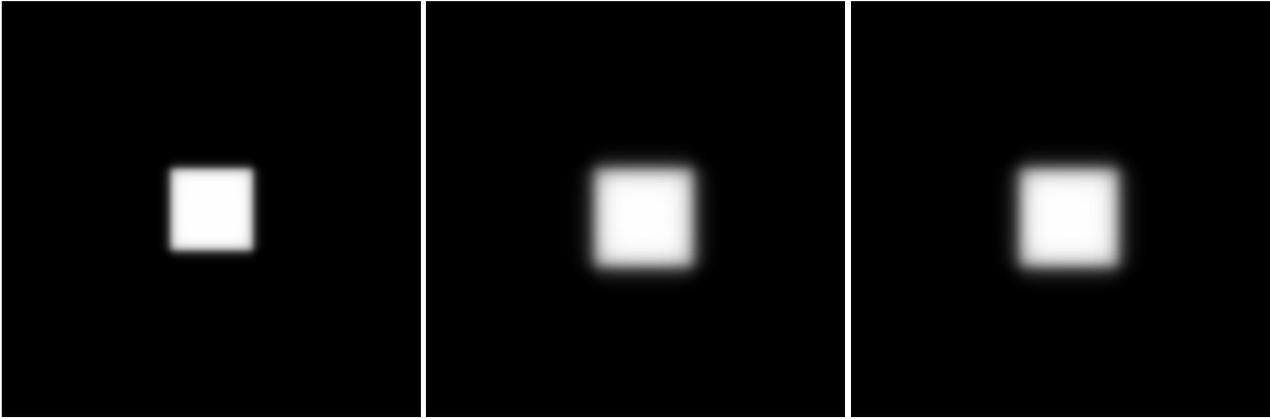


(a)

(b)

(c)

Fig. 2



(a)

(b)

(c)

Fig. 3

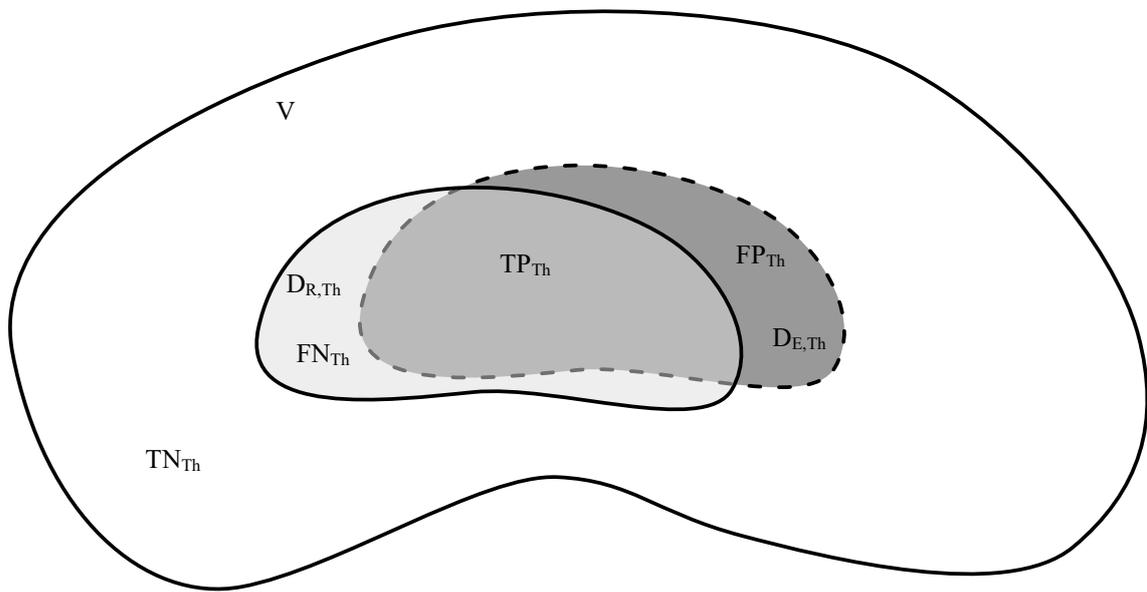
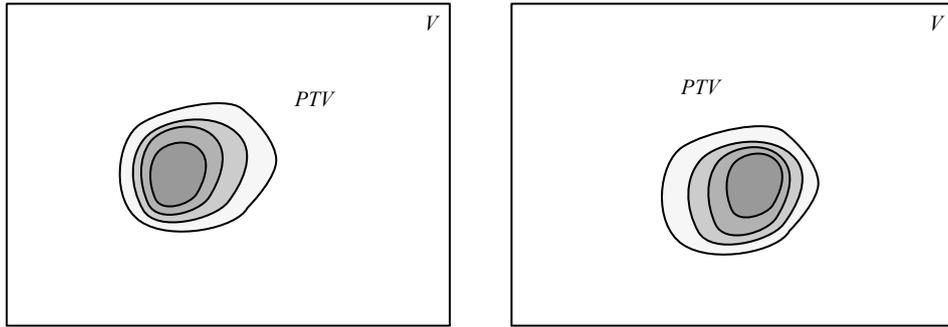


Fig. 4

(a)



(b)

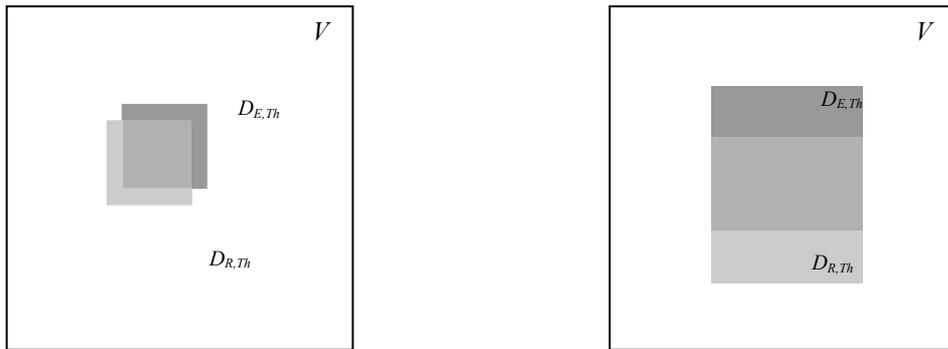
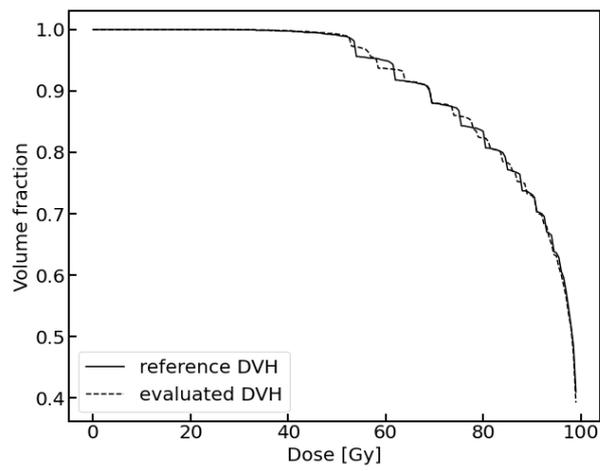
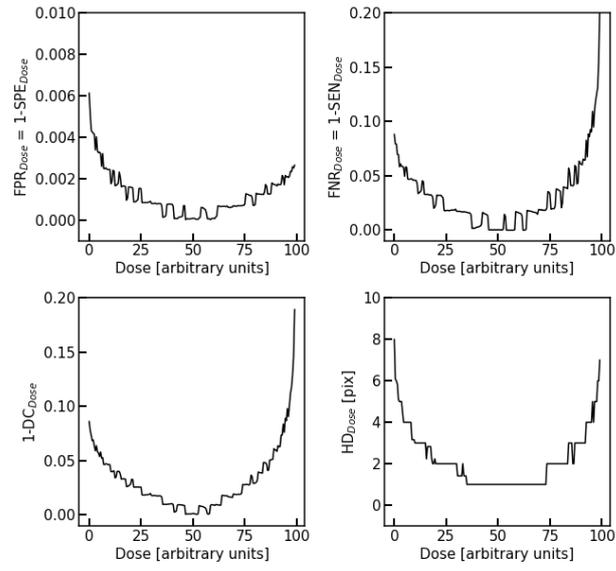


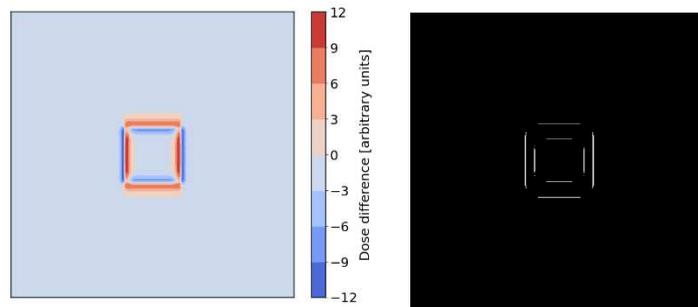
Fig. 5



(a)



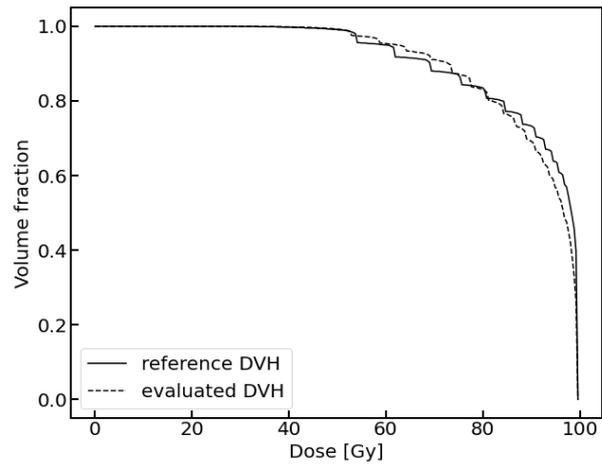
(b)



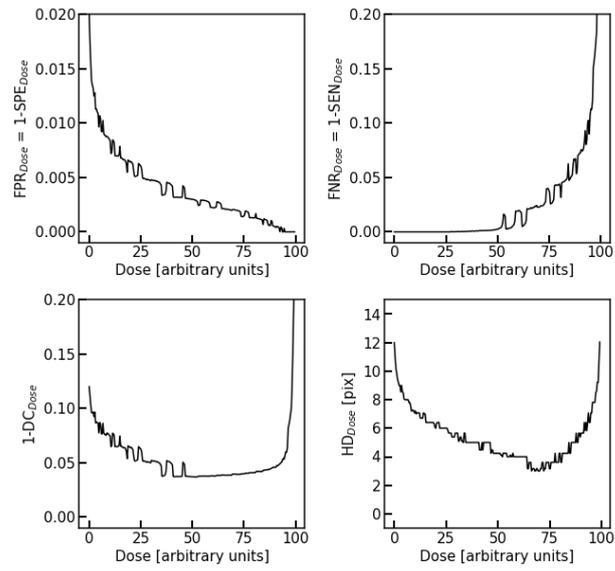
(c)

(d)

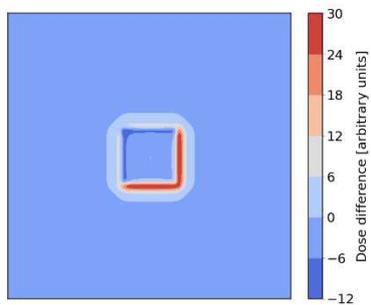
Fig. 6



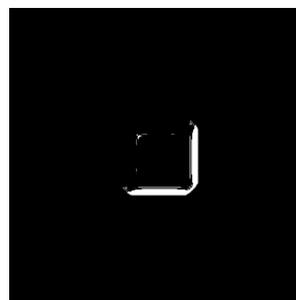
(a)



(b)

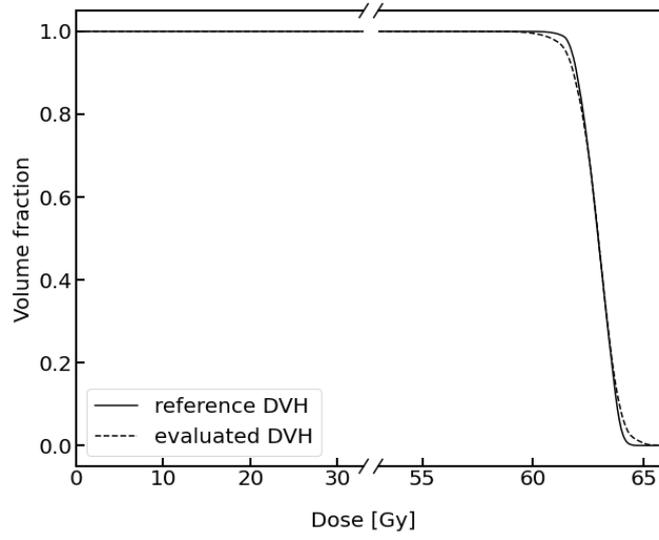


(c)

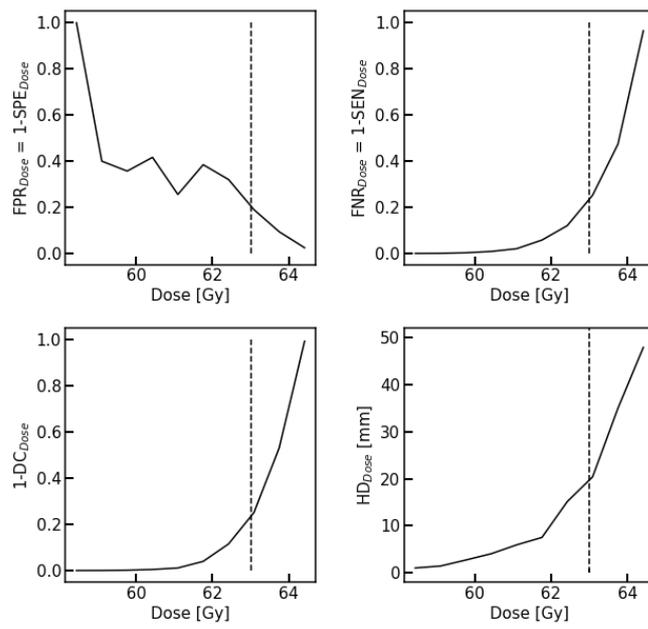


(d)

Fig. 7



(a)



(b)

Fig. 8

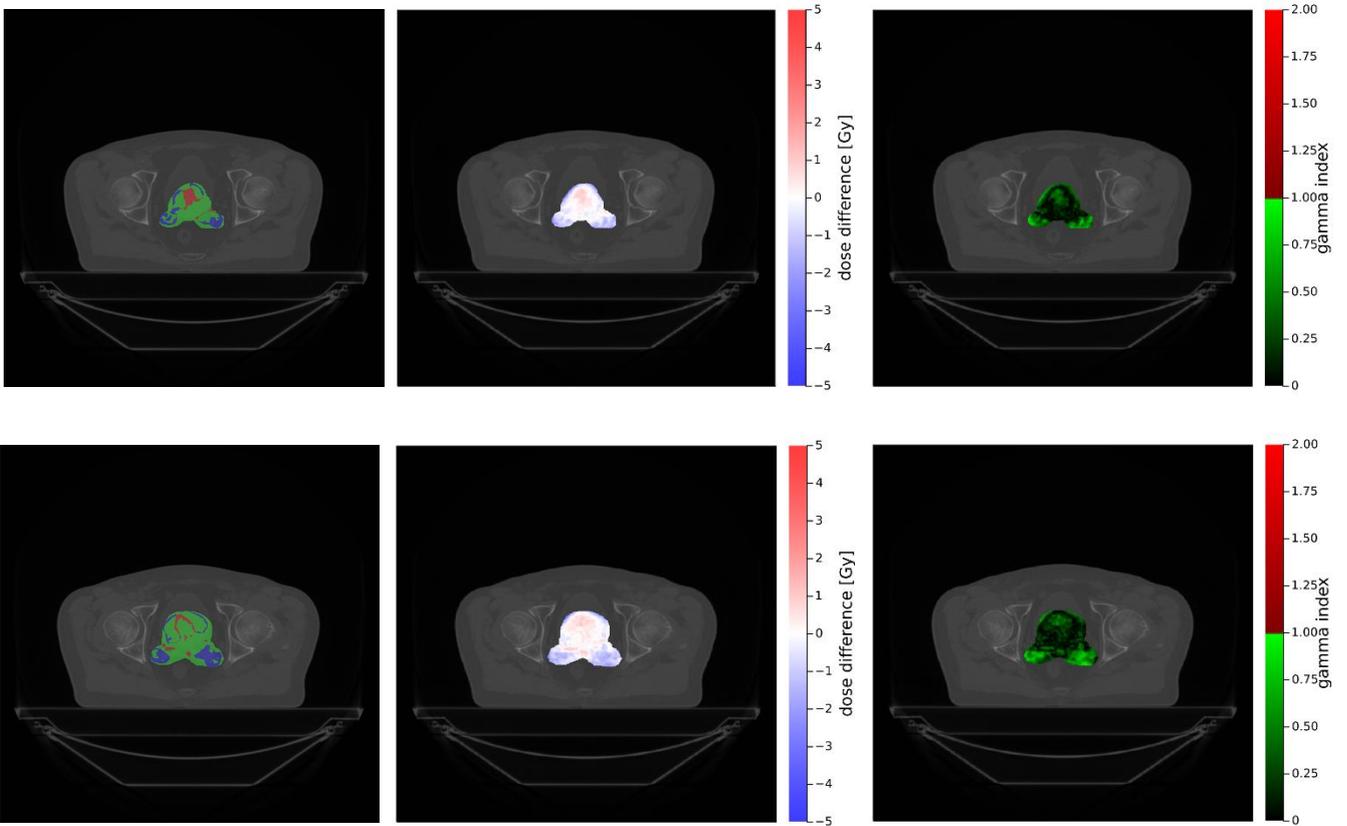


Fig. 9

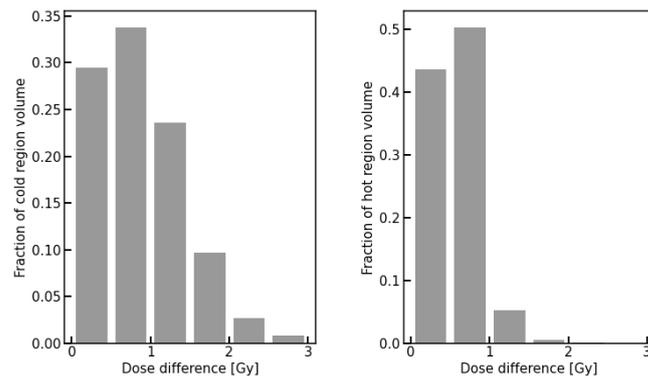
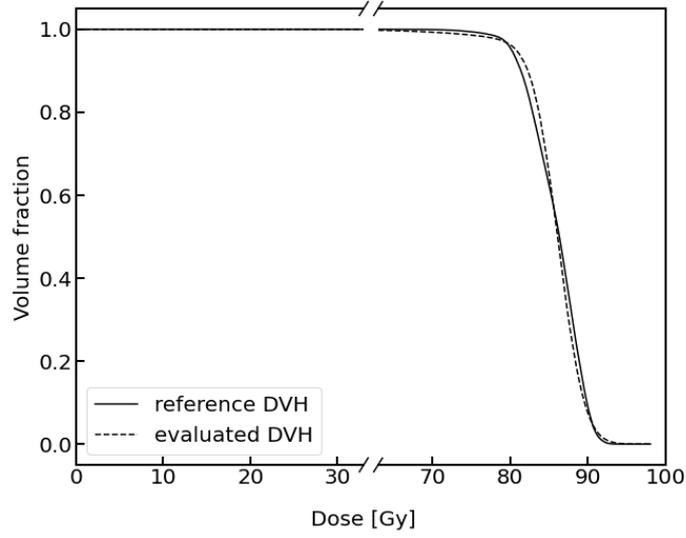
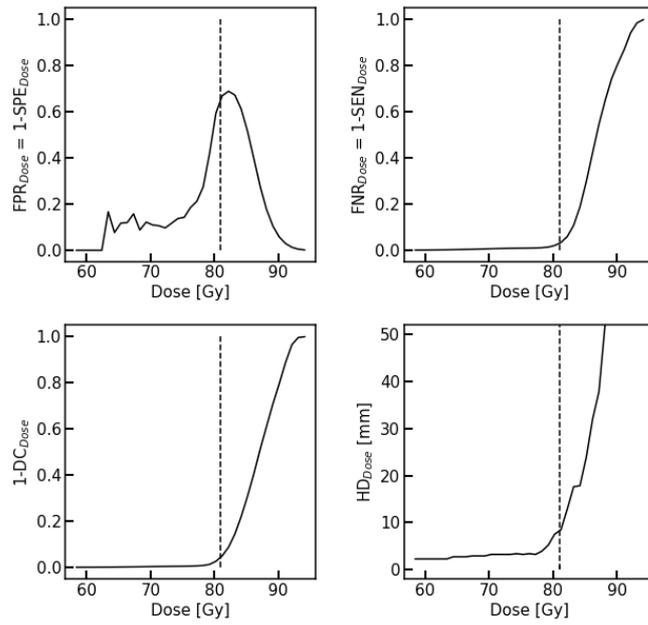


Fig. 10



(a)



(b)

Fig. 11

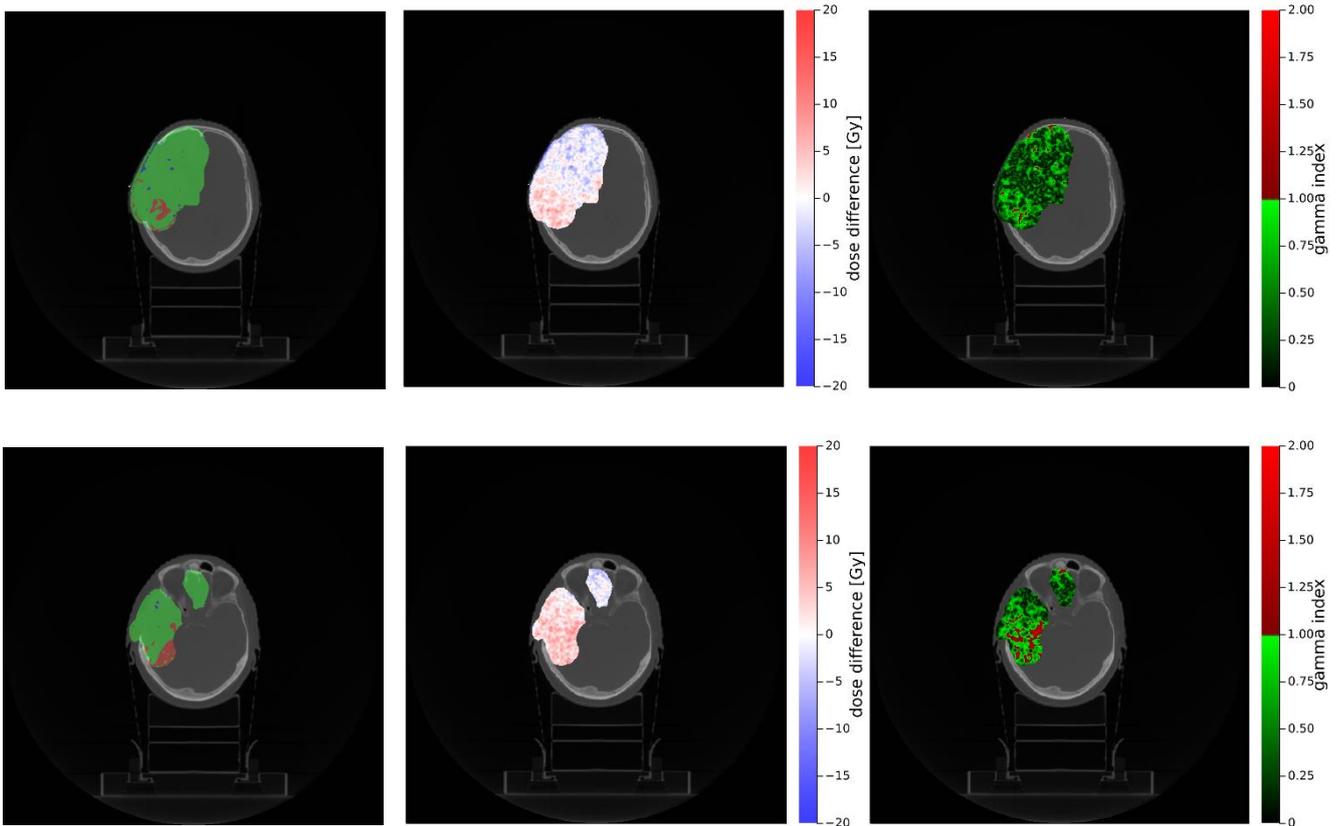


Fig. 12

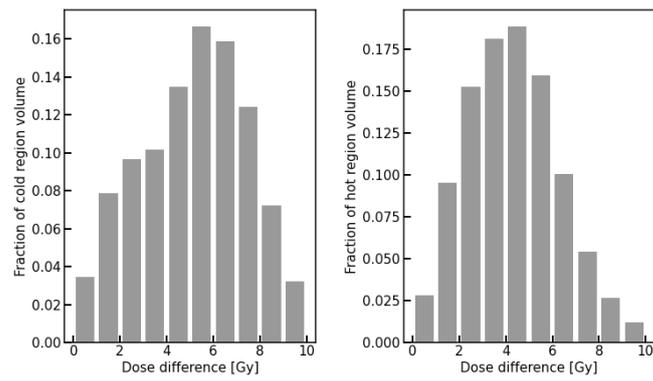
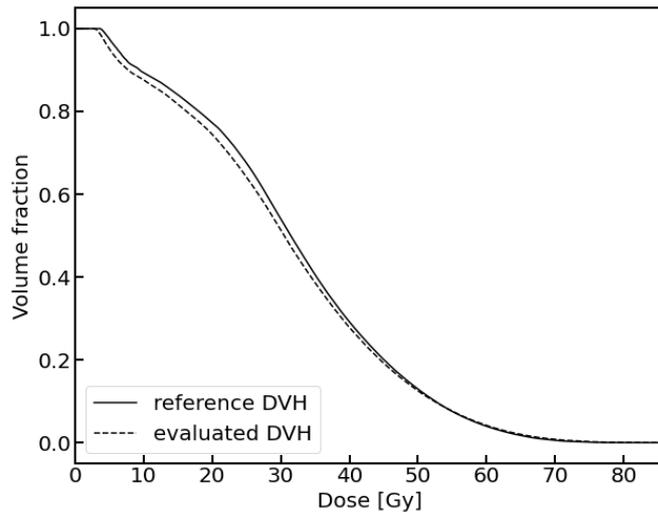
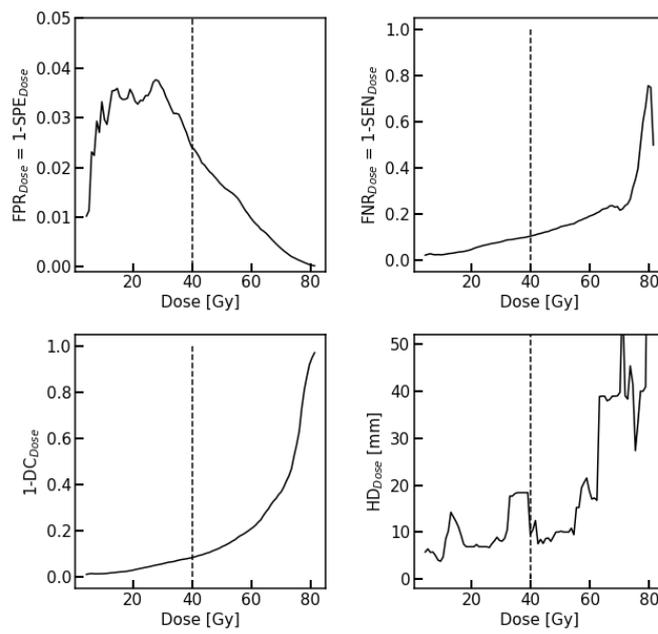


Fig. 13



(a)



(b)

Fig. 14

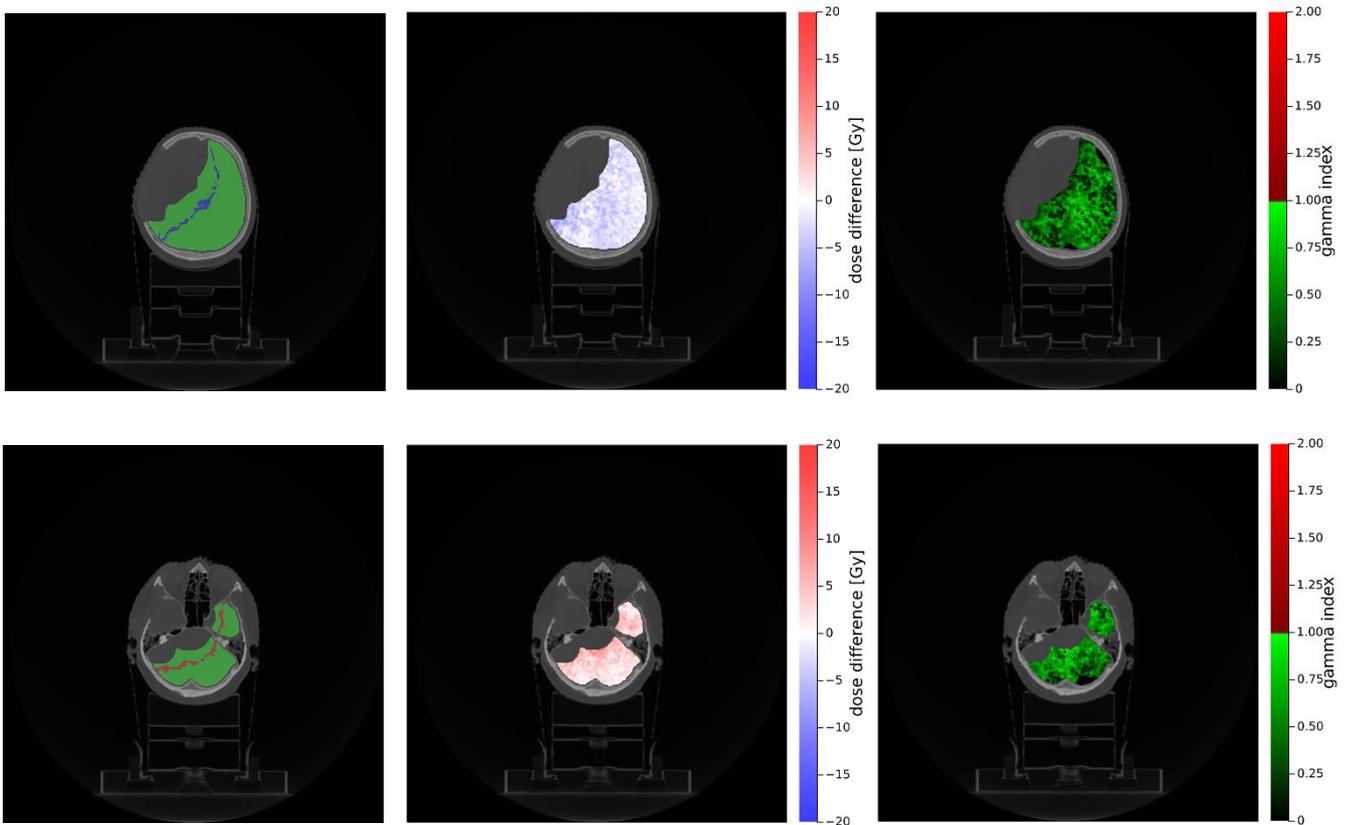


Fig. 15

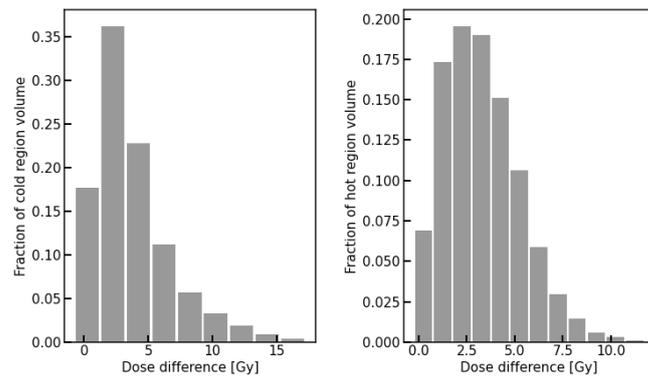


Fig. 16

