An approach for estimating dosimetric uncertainties in deformable dose accumulation in pencil beam scanning proton therapy for lung cancer

Florian Amstutz[1,2], Lena Nenoff[1,2], Francesca Albertini[1], Cássia O Ribeiro[3], Antje C Knopf [3,4], Jan Unkelbach[5], Damien C Weber[1,5,6], Antony J Lomax[1,2], Ye Zhang[1]

[1] Paul Scherrer Institute, Center for Proton Therapy, Switzerland
[2] Department of Physics, ETH Zurich, Switzerland
[3] Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, the Netherlands
[4] Division for Medical Radiation Physics, Carl von Ossietzky University Oldenburg, Germany
[5] Department of Radiation Oncology, University Hospital Zurich, Switzerland
[6] Department of Radiation Oncology, University Hospital Bern, Switzerland

Email: ye.zhang@psi.ch

Abstract

Deformable image registration (DIR) is an important component for dose accumulation and associated clinical outcome evaluation in radiotherapy. However, the resulting deformation vector field (DVF) is subject to unavoidable discrepancies when different algorithms are applied, leading to dosimetric uncertainties of the accumulated dose. We propose here an approach for proton therapy to estimate dosimetric uncertainties as a consequence of modeled or estimated DVF uncertainties.

A patient-specific DVF uncertainty model was built on the first treatment fraction, by correlating the magnitude differences of five DIR results at each voxel to the magnitude of any single reference DIR. In the following fractions, only the reference DIR needs to be applied, and DVF geometric uncertainties were estimated by this model. The associated dosimetric uncertainties were then derived by considering the estimated geometric DVF uncertainty, the dose gradient of fractional recalculated dose distribution and the direction factor from the applied reference DIR of this fraction. This estimated dose uncertainty was respectively compared to the reference dose uncertainty when different DIRs were applied individually for each dose warping. This approach was validated on seven NSCLC patients, each with nine repeated CTs.

The proposed model-based method is able to achieve dose uncertainty distribution on a conservative voxel-to-voxel comparison within ±5% of the prescribed dose to the 'reference' dosimetric uncertainty, for 77% of the voxels in the body and 66-98% of voxels in investigated structures.

We propose a method to estimate DIR induced uncertainties in dose accumulation for proton therapy of lung tumor treatments.

Keywords: deformable image registration, uncertainties, proton therapy, dose accumulation, lung cancer

This document is the accepted manuscript version of the following article:
Amstutz, F., Nenoff, L., Albertini, F., Ribeiro, C. O., Knopf, A. C., Unkelbach, J., … Zhang, Y. (2021). An approach for estimating dosimetric uncertainties in deformable dose accumulation in pencil beam scanning proton therapy for lung cancer. Physics in Medicine and Biology, 66(10), 105007 (12 pp.). https://doi.org/10.1088/1361-6560/abf8f5

1. Introduction

Deformable image registration (DIR) is an important tool for many radiotherapy applications such as dose accumulation, motion extraction and mitigation, as well as contour propagation (Jingu 2014, Rigaud *et al* 2019, Sarrut *et al* 2017). There are many different DIR algorithms proposed in the literature and in use for medical applications, particularly in radiotherapy (Maintz and Viergever 1998, Sotiras *et al* 2013, Oh and Kim 2017). However, in DIR, every voxel can have a different three-dimensional displacement vector associated with it, leading to a large number of degrees of freedom and making it mathematically an ill-posed problem (meaning no unique solution). In addition, the ground-truth displacement is unknown, particularly for indications in presence of inter-fraction motion (e.g. tumor shrinkage) (Oh and Kim 2017). Nevertheless, quantification and evaluation of the associated geometric and dosimetric uncertainties are strongly recommended in the AAPM TG132 report and by Paganelli et al (Brock *et al* 2017, Paganelli *et al* 2018), before employing DIR results for radiotherapy treatment planning.

Contrary to the recommendation, most state-of-the-art DIR algorithms in clinical radiotherapy applications do not provide any patient-specific estimation for geometric nor dosimetric uncertainty, partially due to the missing definition of an accurate and efficient validation measure for radiotherapy applications (Brock *et al* 2017, Paganelli *et al* 2018). As a research topic, there are a few large-scale international competitions on DIR geometric accuracy comparison (Brock 2010, Murphy *et al* 2011, Kadoya *et al* 2014), but only a few investigations highlight the importance of incorporating its induced geometric uncertainty in radiotherapy applications (e.g. dose accumulation) (Kashani *et al* 2007, Saleh-Sayah *et al* 2011). Furthermore, different quantitative metrics to evaluate image registration are discussed in research (Brock *et al* 2017, Paganelli *et al* 2018, Varadhan *et al* 2016, Torsten Rohlfing 2014, Kashani *et al* 2007, Saleh-Sayah *et al* 2011, Salguero *et al* 2011, Murphy *et al* 2012, Hub and Karger 2013, Saleh *et al* 2014). The target registration error (TRE) measures the accuracy between implanted or anatomical landmarks on image pairs, which can be detected manually or automatically (e.g. with a scale-invariant feature transform) (Brock *et al* 2017, Paganelli *et al* 2013). Mean distance to agreement (MDA) gives the mean distance between two contours on registered images (Brock *et al* 2017). Alternatively, it is also possible to quantify the volumetric overlap of contours on registered images with the dice similarity coefficient (DSC) (Brock *et al* 2017). Limitations of these measures are their dependency on feature (TRE) or contour regions (MDA, DSC). It was shown that contrast rich features may not represent the true DIR performance in the low contrast regions (Varadhan *et al* 2016). Additionally, also tissue overlap and image similarity provided no valid evidence for a completely accurate registration (Torsten Rohlfing 2014). Indeed, DIR algorithms are never uniformly accurate and tend to be worse in regions with significant anatomy changes (i.e. interface between tumor and lung for lung cancer) and less featured region (e.g. abdomen region in CT image), leading to the necessity to estimate the error for each individual registration algorithm and each investigated patient (Kashani *et al* 2007). For individual algorithms, different attempts have been taken to tackle the patient-specific uncertainties, such as finding the point-wise variance of the deformation vector field (DVF) after iterative DIR applications, followed by blurring of the dose map (Salguero *et al* 2011). Alternatively, one approach was proposed to model the DIR error distribution by conducting a principal component analysis of the errors in a training set with different regions of interest and sampling the error maps afterwards (Murphy *et al* 2012). Furthermore, taking the local reproducibility of the DIR has been suggested as an uncertainty measure (Hub and Karger 2013). The distance discordance metric was suggested as a quantitative

metric by measuring the distance between registered points from at least four co-registered images, but this approach is limited by the clinical feasibility, as multiple samples and registrations of the images are required for every application (Saleh *et al* 2014). Moreover, some studies, mostly for the dosimetric uncertainty quantification, discussed the possibility of using inter-algorithm uncertainties as a measurement of DIR accuracy (Samavati *et al* 2016, Zhang *et al* 2012, Ribeiro *et al* 2018, Nenoff *et al* 2020). Until now these uncertainty measures have not found the way into the clinics.

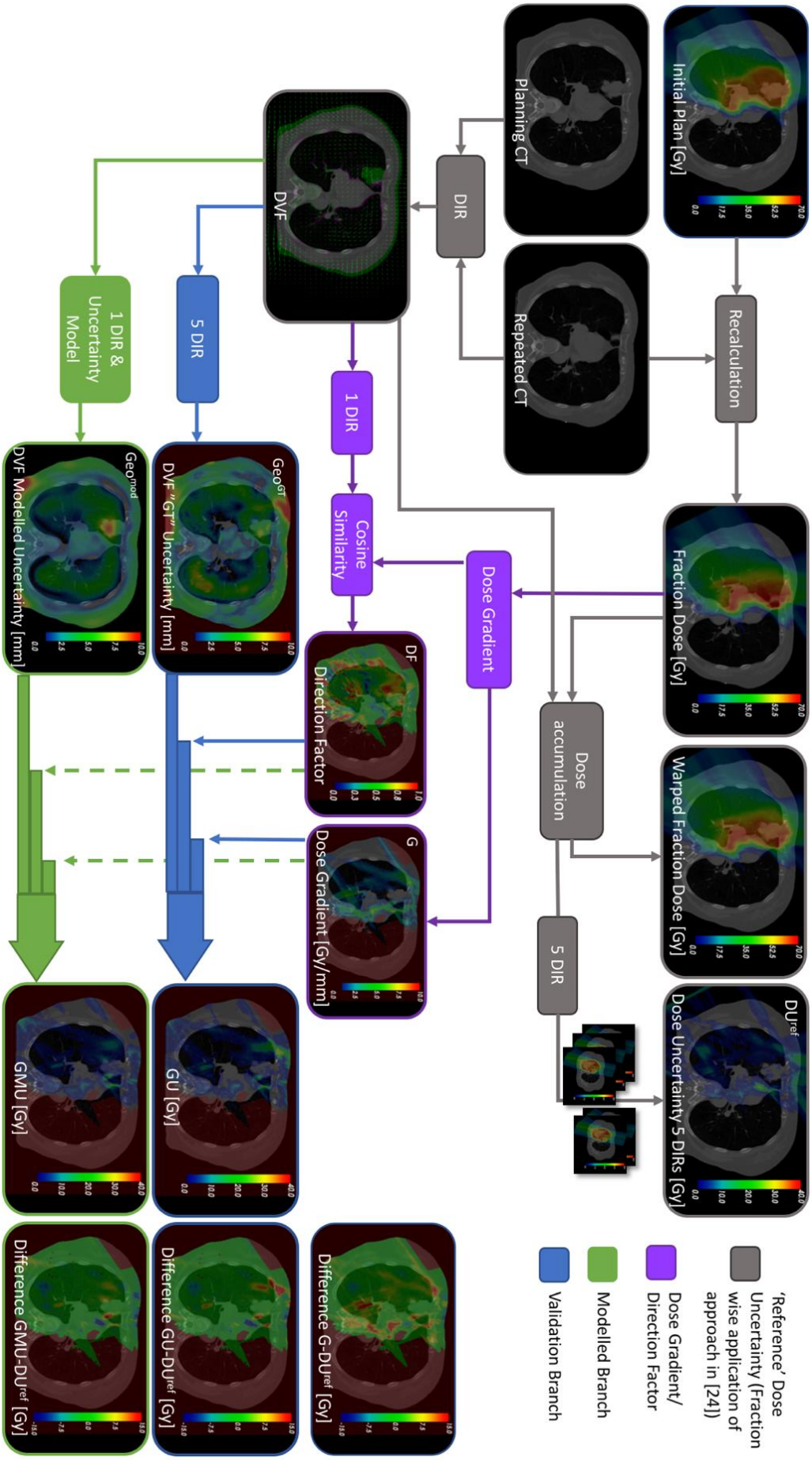
Proton therapy allows for high target dose coverage while reducing the integral dose to healthy tissue compared to photon therapy (Palm and Johansson 2007, Hill-Kayser *et al* 2011). However, intra- and inter-fractional anatomical changes can have a substantial impact on the dose distribution, due to the finite range of protons and the resulting steep dose gradients (Szeto *et al* 2016, Stuschke *et al* 2012). Furthermore, especially for pencil beam scanned protons, the impact can be even more pronounced due to the interplay effects (Grassberger *et al* 2013, Dowdell *et al* 2013, Kardar *et al* 2014). DIR therefore plays even a greater role for analyzing the impact of anatomical changes on proton therapy, as well as for assessing the effects of intra-fractional motion using 4D dose calculations (Rigaud *et al* 2019, Sarrut *et al* 2017, Zhang *et al* 2012, Ribeiro *et al* 2018). In addition, for inter-fractional anatomical changes in adaptive treatments, DIR is needed to accumulate the dose distributions of recalculated or adapted fraction plans to the reference geometry of the initial plan, in order to better evaluate treatment outcomes (Rigaud *et al* 2019, Sarrut *et al* 2017, Nenoff *et al* 2020, Chetty and Rosu-Bubulac 2019). Indeed, studies have shown significant dosimetric discrepancies when different DIR algorithms were used (Zhang *et al* 2012, Ribeiro *et al* 2018, Nenoff *et al* 2020) and DIR showed clear dose uncertainties from the accumulated 4D dose distribution for liver tumors (Zhang *et al* 2012, Ribeiro *et al* 2018). Furthermore, the influence of inter-algorithm DIR uncertainties on dose accumulation for adaptive proton therapy was discussed for non-small cell lung cancer (NSCLC) patients in a previous study (Nenoff *et al* 2020). The dose degradation due to anatomical changes and the uncertainties from applying different DIRs for dose accumulation were investigated by warping nine fraction doses with different DIRs and summing up the warped doses of each DIR to a single accumulated dose distribution. This showed potential dosimetric uncertainties if different DIRs were applied, and suggested estimating DIR uncertainties for clinically relevant DVH parameters by using multiple DIRs (Nenoff *et al* 2020). The difficulty for this approach is the transfer into clinical practice, for which multiple DIRs would have to be applied for each fraction, and subsequently the dose warped by each resulting DVF. Therefore, the authors recommended developing a fast, automated quality assurance on both the image and dosimetric level.

Based on these previous experiences (Nenoff *et al* 2020), in this paper, we aim to develop a fast dose uncertainty prediction approach for improving the DIR based dose accumulation workflow. The hypothesis is that dosimetric uncertainty is a function of local dose gradients (magnitude and direction) and corresponding geometric uncertainties in the DVF magnitude and direction. The ideal utilization of this approach would be reached if clinically used DIRs can provide a geometric uncertainty estimation along each registration output. However, here, we firstly focus on testing this hypothesis, and developing a first patient-specific model for predicting geometric DVF uncertainties that does not require, on a fraction-by-fraction basis, the use of multiple DIRs. We have validated the effectiveness of this approach in the presence of inter-factional anatomical changes for NSCLC, a particularly challenging anatomical site for proton therapy and dose accumulation (Han 2019).

2. Materials and Methods

The overview of the proposed method together with its reference and validation branches are illustrated in Figure 1. Additionally, a summary of the most important abbreviations can be found later in Table 1. Generally, this framework makes use of a patient-specific model for the geometric DVF uncertainty, in combination with the fraction dose gradient and the direction weighting factor, in order to efficiently derive the associated dosimetric uncertainty (GMU), in the following called the 'modelled branch'. With a validation branch, the concept of combining geometric DVF uncertainty and dose gradients to derive a dosimetric uncertainty map (GU) in the absence of geometric uncertainty modelling is investigated. In a third approach, the dose gradient is upscaled to derive another dose uncertainty estimation (G). The different dose uncertainty estimations are compared to the 'reference' uncertainty (DU^{ref}) which is derived by warping each fraction dose with five DIRs.

Figure 1: Scheme of the proposed methodology. The top branch in grey is a fraction-wise version of the method described in our earlier study (Nenoff et al 2020), for which each fraction dose is warped with multiple DIR algorithms back to the planning CT and by taking the maximum-minimum value a dosimetric uncertainty estimation is generated. In our study, this estimation is considered as the reference dosimetric uncertainty. The validation branch (blue) combines the 'ground-truth' geometric uncertainty calculated from all five DIRs together with the dose gradient (G) and the direction factor (DF) to get a dosimetric uncertainty estimation (GU). For the modelled branch (green) the geometric uncertainty map (Geo^{mod}) is modelled by a linear model built in the first fraction. The dosimetric uncertainty estimation (GMU) is again a combination of the dose gradient and the direction factor.


Planning CT
Initial Plan [Gy]
DIR
DVF
Repeated CT
Recalculation
1 DIR & Uncertainty Model
5 DIR
1 DIR
Cosine Similarity
Dose Gradient
Fraction Dose [Gy]
Dose accumulation
Warped Fraction Dose [Gy]
5 DIR
DUref
Dose Uncertainty 5 DIRs [Gy]
DF
Direction Factor
G
Dose Gradient [Gy/mm]
GeoGT
DVF "GT" Uncertainty [mm]
Geomod
DVF Modelled Uncertainty [mm]
GMU [Gy]
GU [Gy]
Difference GMU-DUref [Gy]
Difference GU-DUref [Gy]
Difference G-DUref [Gy]
'Reference' Dose Uncertainty (Fraction wise application of approach in [24])
Dose Gradient/ Direction Factor
Modelled Branch
Validation Branch

2.1 Patient cohort and deformable image registration algorithms

The data used for this study has previously been described (Nenoff *et al* 2020, Josipovic *et al* 2016) and is therefore only briefly summarized here. The patient cohort includes seven NSCLC patients, each with a planning (reference) CT and nine repeated CTs. The repeated CTs were acquired for fraction 2, 16 and 31 during radiation therapy, with three repeat CT acquisitions for each day. All CTs were acquired in visually guided, voluntary deep inspiration breath-hold. The treatment plans were individually computed 3-field IMPT plans with prescribed doses of 60 Gy-RBE. Each repeated CT was registered to the corresponding planning CT as a reference and was sequentially registered using five DIR algorithms, namely Plastimatch Demons & B-spline, Velocity (Varian Medical Systems, Palo Alto, United States), Mirada (Mirada Medical, Oxford, UK), and Raystation Anaconda (RaySearch, Stockholm, Sweden). The resulting five deformation vector fields for each registered image pair were used for this study.

2.2 'Reference' scenario for dose uncertainty

The top branch (in grey) of Figure 1 shows the dosimetric uncertainty estimation as a consequence of DVF uncertainties, using the method described in our earlier study. However, instead of summing the warped dose over all repeated CTs, each fraction dose distribution was investigated separately (Nenoff *et al* 2020). For this purpose, the initial plan is recalculated on each repeated CT first. Then the resulting dose distribution is individually warped back to the planning CT, using each of the five DVFs from the different DIRs.

The voxel-wise maximum-minimum dosimetric uncertainty between the five warped doses can be calculated for each CT as follows

$$DU_i^{ref} = \max_m d_i^m - \min_m d_i^m \quad (1)$$

where $\vec{r_i}$ denote the 3-dimensional position of a specific voxel *i* in the reference CT, $\overrightarrow{DVF_i^m}$ is the 3-dimensional deformation field vector that maps voxel *i* to the repeat CT with correlation derived by DIR algorithm *m*, and d_i^m is the dose evaluated on the repeated CT at position $\vec{r_i} + \overrightarrow{DVF_i^m}$. This is considered in this study as the 'reference' dose uncertainty distribution due to the inter-algorithm wise DIR ambiguity.

2.3 Fast dosimetric uncertainty estimation and its validation

In order to derive the uncertainty in dose distribution more efficiently than performing dose warping using each individual DVFs as described above, we first propose a fast estimation method based on the multiplication of three distributions.

A geometric uncertainty map of the DVF (deduced as described in the following sections), the local dose gradient as derived from the fraction specific recalculated dose distribution and the direction of this gradient (Figure 1):

$$DU_i^{est} = Geo_i \cdot G_i \cdot DF_i \quad (2)$$

With DU_i^{est} being the resulting dose uncertainty, Geo_i the geometric DVF uncertainty, G_i the dose gradient, and DF_i the direction factor at the *i*-th voxel.

The *dose gradient* G_i is used to convert the geometric DVF uncertainties into the dose uncertainty estimation, similarly as proposed by Hub et al. (Hub *et al* 2012). By multiplying, for each voxel, the magnitude of the DVF uncertainty (mm) with the dose gradient (Gy-RBE/mm), an estimation of the dosimetric uncertainty is obtained. To calculate the dose gradient, a Sobel filter is applied on the fraction dose distribution. However, this estimation is valid only if the direction of the dose gradient is the same as the one from the estimated DVF uncertainty. Therefore, an additional direction weighting factor is further incorporated.

The *direction weighting factor* DF_i is constructed by calculating the absolute value of the voxel-wise cosine similarity between the dose gradient and the DVF direction of one reference DIR algorithm:

$$DF_{\mathrm{i}} = \| \cos(\theta_i) \| = \| \left\{ \frac{\overrightarrow{DVF_i^{ref}} \cdot \overrightarrow{G_i}}{\left\| \overrightarrow{DVF_i^{ref}} \right\| \cdot \left\| \overrightarrow{G_i} \right\|} \right\} \| \tag{3}$$

with θ_i being the angle between DVF vector and the gradient direction at the i-th voxel; $\overrightarrow{DVF_i^{ref}}$ the vector of the reference displacement vector field at the *i*-th voxel and $\overrightarrow{G_i}$ dose gradient at the *i*-th voxel. If the DVF uncertainty is pointing in the same direction as the dose gradient, Equ. 3 results in a value of 1, the highest weighting factor. For the case of a perpendicular DVF to the gradient, the uncertainty does not have an impact and the value is 0, the lowest weighting factor. Then, as shown in Equ. 2, by multiplying the DVF uncertainty map with the dose gradient and the cosine similarity direction factor, we can get an estimation of the dosimetric uncertainty (green and blue branch in Figure 1). As a comparison, we have also investigated whether the dose gradient on its own is a predictor of dose uncertainty, by uniformly up-scaling dose gradients such that the highest gradient takes the same value as the highest dose uncertainty in the reference

$$G_i' = \frac{\max_i d_i}{\max_i G_i} \cdot G_i. \tag{4}$$

2.4 Testing the dose uncertainty estimation method with 'ground-truth' DVF uncertainties

To test the fast dose uncertainty estimation concept described in 2.3, a 'ground truth' representation of DVF uncertainty (Geo_i) has first been used. This DVF uncertainty representation is calculated as the maximum-minimum magnitude of the five DVFs for each voxel i, as,

$$Geo_i^{GT} = \max_m \left\| \overrightarrow{DVF_i^m} \right\| - \min_m \left\| \overrightarrow{DVF_i^m} \right\| \tag{5}$$

This 'ground-truth' DVF uncertainty was calculated for all fractions. These DVF uncertainty maps were then multiplied with the dose gradient and the direction factor, resulting in a 'best-case' estimation of the dosimetric uncertainty for each fraction that can be predicted by the above described approach (blue branch in Figure 1)

$$GU_i = Geo_i^{GT} \cdot G_i \cdot DF_i \tag{6}$$

With this validation branch, we aim to investigate to what extent the proposed dose uncertainty estimation concept works if the geometric DVF uncertainty is definitive. However, this approach would still require multiple DIRs applied for each fraction. This branch was therefore performed as a reference, in order to validate whether dose uncertainty can be estimated from the combination of dose gradient, direction and DVF uncertainty. In case of a future DIR algorithm which outputs a geometric uncertainty estimation together with the DVF, this estimation can directly be incorporated into the method.

2.5 Generating geometric DVF uncertainty model

In a next step, a first model for estimating geometric DVF uncertainty on a fraction-by-fraction basis has been tested, without applying multiple DIR algorithms each day. To implement this concept, on the first treatment fraction, a patient-specific linear uncertainty model has been constructed. This model was built by correlating the DVF vector magnitude of any reference DIR (of the five in this study) with the Geo^{GT} from the first fraction (Equ. 7a).

$$\hat{a} = corr(Geo_{CT1}^{GT}, \left\|\overrightarrow{DVF_{CT1}^{ref}}\right\|) \tag{7a}$$

$$Geo_{CTn}^{mod} = \hat{a} \cdot \left\|\overrightarrow{DVF_{CTn}^{ref}}\right\| \tag{7b}$$

With linear regression coefficient $\hat{a}$, Geo_{CT1}^{GT} refers to the 'ground-truth' geometric uncertainty representation of CT1 and $\left\|\overrightarrow{DVF_{CT1}^{ref}}\right\|$ the DVF magnitudes of one reference DIR at CT1. For the consecutive fractions, only the reference DIR is applied, with the DVF uncertainty being estimated from this model (Equ.7b).

The dosimetric uncertainty is then calculated by using the modelled geometric uncertainty as input to our proposed method

$$GMU_i = Geo_i^{mod} \cdot G_i \cdot DF_i \tag{8}$$

The scheme of the DVF uncertainty modelling can be seen in Figure 2.

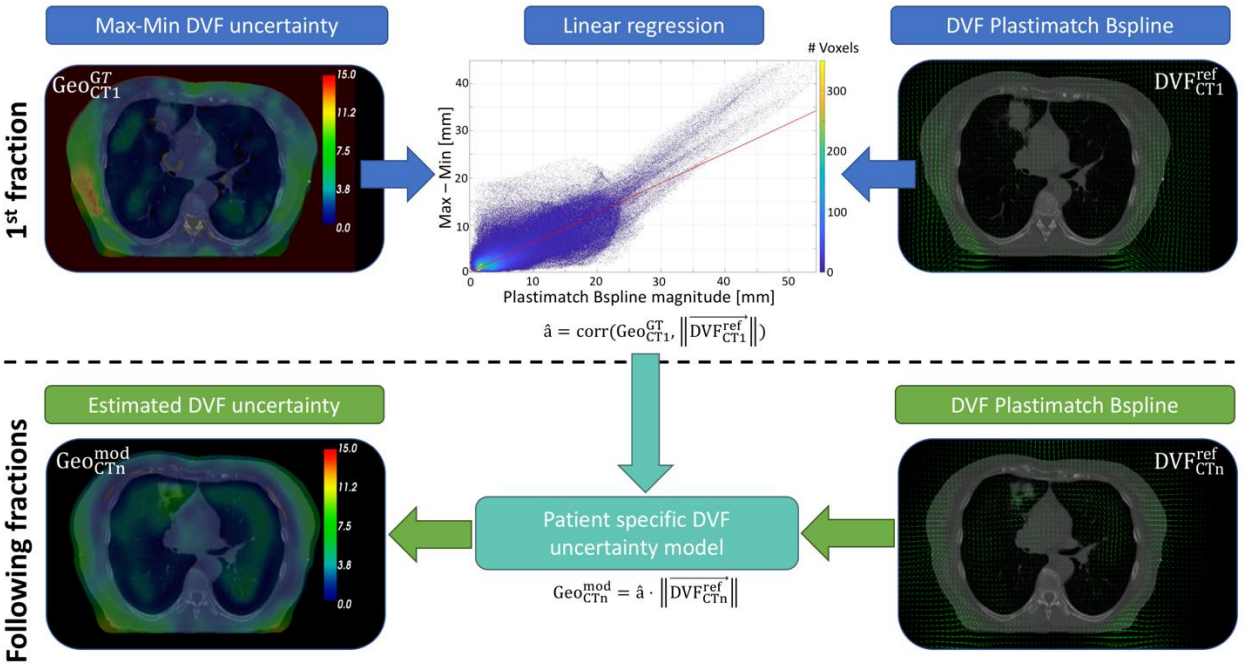


Figure 2: Scheme of the patient-specific DVF uncertainty modelling. In a first fraction the maximum difference of the five deformation vector magnitudes at each voxel are calculated (Geo_{CT1}^{GT}) and then correlated to the DVF magnitude of one reference DIR ($\overrightarrow{DVF_{CT1}^{ref}}$; e.g. Plastimatch B-spline). This gives a linear model for the DVF uncertainty. For the following fractions only the reference DIR is applied and the DVF uncertainty is modelled with the built model from the first fraction by multiplying for each voxel the magnitude of the deformation vector with the linear regression coefficient â.

Table 1: Summary of Abbreviations

Abbreviation	Description / Formula
$\vec{r_i}$	Position of voxel ***i*** in the reference CT
$\overrightarrow{DVF_i^m}$	Deformation field vector mapping voxel ***i*** to repeated CT under consideration using DIR algorithm ***m***
d_i^m	Dose evaluated on repeated CT at position $\vec{r_i} + \overrightarrow{DVF_i^m}$
$\mathbf{DU_i^{ref}}$	Reference dose uncertainty: $DU_i^{ref} = \max_m d_i^m - \min_m d_i^m$
G	Dose gradient
θ_i	Angle between DVF vector and dose gradient direction at voxel ***i***
$\mathbf{Geo^{GT}}$	"Ground-truth" geometric uncertainty:

	$Geo_i^{GT} = \max_m \left\|\overrightarrow{DVF_i^m}\right\| - \min_m \left\|\overrightarrow{DVF_i^m}\right\|$
â	Linear regression coefficient: $\hat{a} = corr(Geo_{CT1}^{GT}, \left\|\overrightarrow{DVF_{CT1}^{ref}}\right\|)$
Geo^{mod}	Modelled geometric uncertainty: $Geo_{CTn}^{mod} = \hat{a} \cdot \left\|\overrightarrow{DVF_{CTn}^{ref}}\right\|$
DF	Direction factor; $DF_i = \| \cos(\theta_i) \|$
$\mathbf{DU_i^{est}}$	$DU_i^{est} = Geo_i \cdot G_i \cdot DF_i$
GU	$GU_i = Geo_i^{GT} \cdot G_i \cdot DF_i$
GMU	$GMU_i = Geo_i^{mod} \cdot G_i \cdot DF_i$

2.6 Evaluation of DVF uncertainty model

For evaluation of this uncertainty estimate concept, we compared three different estimations of dosimetric uncertainty (gradient-only, gradient & 'ground-truth' geometric DVF uncertainty, and gradient & estimated geometric DVF uncertainty), with respect to the reference dosimetric uncertainty map (grey branch). As quantification, we first subtracted the reference dosimetric uncertainty (DU^{ref}) from *A)* the estimated dosimetric uncertainty including the modelled DVF uncertainty map (modelled branch: GMU-DU^{ref}); *B)* the DVF uncertainty from five DIRs (validation branch: GU-DU^{ref}) and *C)* the up-scaled gradient uncertainty map (gradient branch: G-DU^{ref}). Afterwards, these differences in dose uncertainty between the estimations and the reference were evaluated for various structures (e.g. body, heart, clinical target volume [CTV], planning target volume [PTV], medulla, ipsilateral lung). Only the voxels with a non-zero dose in the initial plan were considered in the analysis. Note that the *dose uncertainties* (DU^{ref}, G, GU and GMU) always take positive values, as they are defined as the difference between the maximum and minimum dose at each voxel. Conversely, the *differences* in dose uncertainty (G-DU^{ref}, GU-DU^{ref} and GMU-DU^{ref}) can be either positive or negative, depending on whether the method over- or underestimates the dose uncertainty. The differences in the dose uncertainty are plotted in boxplots containing the 25^{th}-75^{th} percentile in the box, the 10^{th}-90^{th} percentile in the whiskers, the location of the 5^{th} and 95^{th} percentile as squares, and the 1^{st} and 99^{th} percentile as triangles. In the evaluation, we focus on the number of voxels within +/-5% and within +/-10% to the reference dosimetric uncertainty.

The method was validated on all seven NSCLC patients, each with nine repeated CTs. The DVF uncertainty model was built using the first repeated CT (acquired on treatment day 2), and afterwards the workflow was validated for the remaining eight repeated CTs.

2.7 Dependence on reference DIR

In order to evaluate the impact of the reference DIR selection on the direction factor and modelling of the geometric DVF uncertainties, we applied the above dosimetric uncertainty workflow using each of the five DIRs as input.

3. Results

3.1 Visual inspection

The 'reference' dose uncertainty (DU^{ref}) together with the estimated dosimetric uncertainties from the three estimation approaches (G, GU and GMU) for one example patient on one fraction CT are shown in Figure 3 (a1-a4). Additionally the differences of the three estimation approaches to the 'reference' are included (Fig.3 b1-b3). For all patients the 'reference' dose uncertainty together with the difference maps are included in the supplement (Figure A.1). Although there are similarities of the three approaches to the reference dose uncertainty, there are clear differences. The uncertainty maps based on DVF uncertainties (GU-DU^{ref}, GMU-DU^{ref}), are visibly closer to the reference scenario than the simple gradient maps (G-DU^{ref}). For most cases (6 out of 7 patients), the derived dose uncertainty maps using geometrical uncertainties directly from the five DIRs (GU) are closer to the reference than the ones based on the estimated DVF uncertainties (GMU) (for the example in figure 3 and example slices in A.1). However, differences are marginal.

3.2 Quantitative analysis

More quantitatively, voxel-wise differences are compared to the reference dose uncertainties as boxplots for the three methods (G, GU and GMU-DU^{ref}) respectively for the same patient in Figure 3c (and for another patient in the supplement [Figure A.2]). Depending on the volume of the evaluated structure, the whiskers include 10^2 to 10^8 voxels. Of note, all whiskers are within +/-10% of the reference, with the majority of voxels (G: 77%/ GU: 89%/ GMU: 88%) within +/-5%. In addition, over all CTs of this patient, the trends among the three scenarios were very similar. This trend has also been observed for the other structures. In the following therefore, we will only compare the averaged result (over all eight CTs) of each patient.

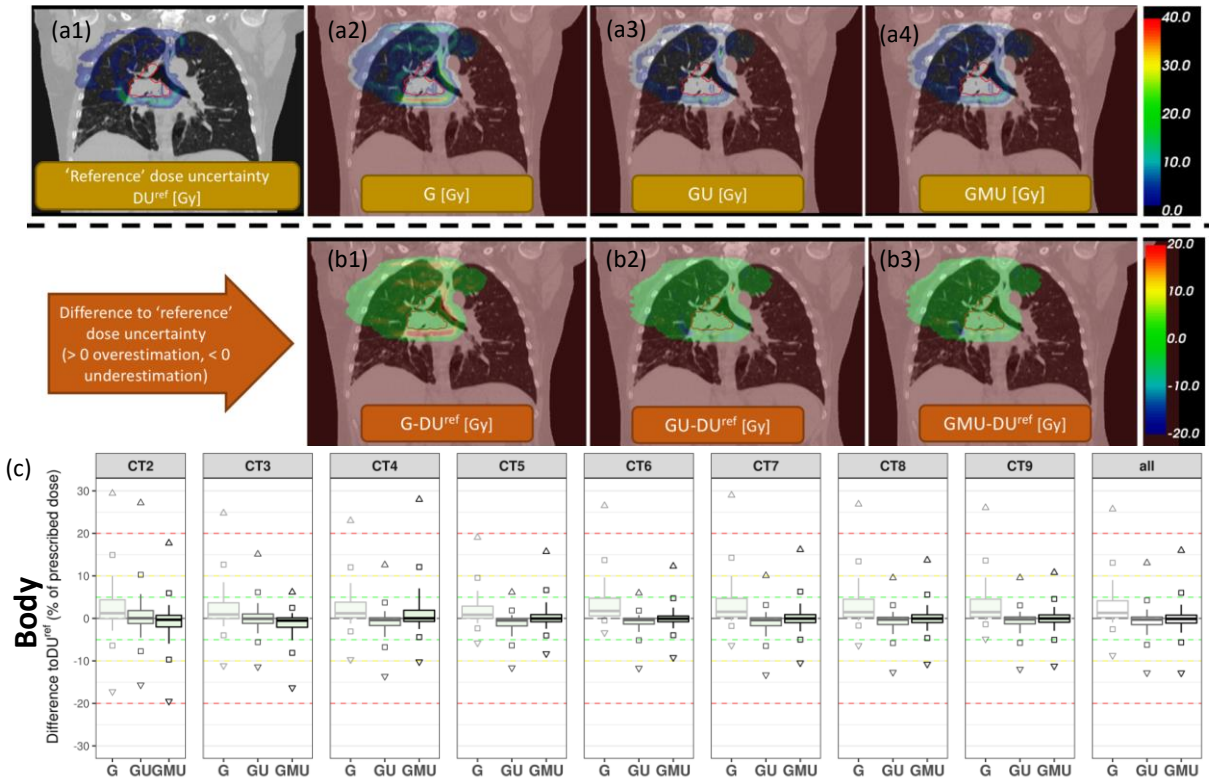


Figure 3: (a1) Example CT slice from patient 3 showing the reference dose uncertainty (DU^{ref}) with the CTV in red; (a2) the estimated dose uncertainty from the upscaled gradient (gradient branch: G); (a3) the estimation based on the geometric DVF uncertainty of all five DIRs (validation branch: GU); (a4) the approach including the modelled DVF uncertainty map (modelled branch: GMU). (b1-3) Distributions of the voxel wise dose differences between reference dose uncertainty and (b1) gradient branch ($G\text{-}DU^{ref}$), (b2) validation branch ($GU\text{-}DU^{ref}$); (b3) modelled branch ($GMU\text{-}DU^{ref}$). (c) Boxplots for the differences of the three estimated to the reference dose uncertainty for the non-zero dose voxels in the whole body structure. The boxplot contains the 25th-75th percentiles of voxels in the box, the 10th-90th percentiles (80%) in the whiskers, the 5th and 95th percentiles as circles and the 1st and 99th percentiles as triangles.

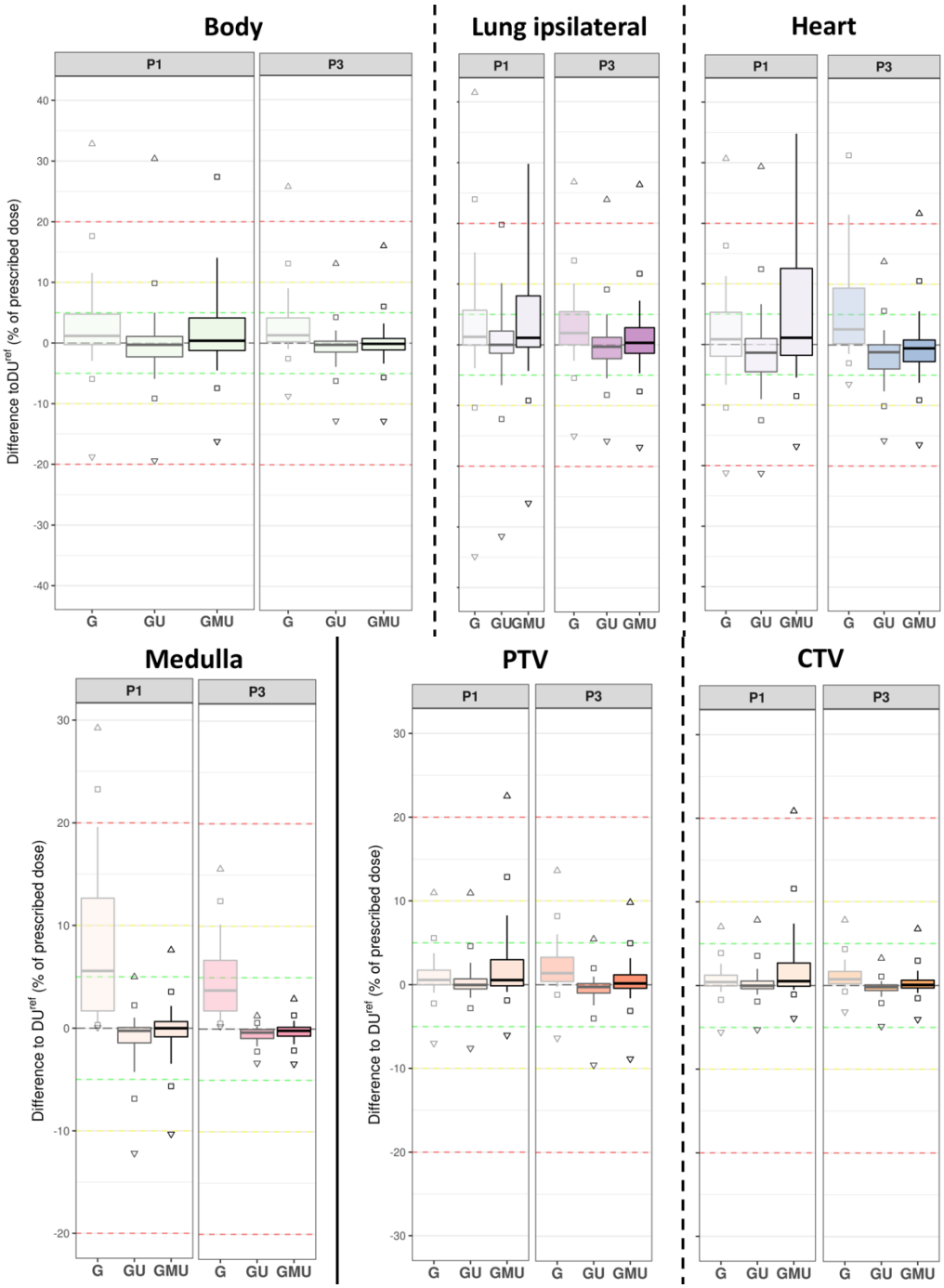


Figure 4: Difference to reference dose uncertainty voxel values for the body, ipsilateral lung, heart, medulla, PTV and CTV structures. For two example patients P1 and P3 the difference from the gradient-only branch (G), the branch including "ground-truth" geometric uncertainties (GU) and the branch with modelled geometric uncertainties (GMU) are plotted.

In Figure 4, the boxplots for the body, ipsilateral lung, heart, medulla, CTV and PTV for two example patients P1 and P3 are shown, each of which include voxels inside the respective structure from the eight repeated CTs. Patient 3 represents one of the cases in-line with the trend over the patients and patient 1 is the patient showing the worst results for the modelled branch. Fig. A.3 in the appendix includes the results for all patients. Over all patients, for the body structure including all non-zero dose voxels inside the body, most voxels (70%/80%/77%) for the three methods (G/GU/GMU) have differences to the reference within ±5% of the prescribed dose (Table A.1 in the supplement). For the

PTV (84%/91%/89%), CTV (90%/94%/92%) and the medulla (69%/98%/98%), more voxels are found to be within ±5% of the prescribed dose, while for the ipsilateral lung (68%/78%/75%) and the heart (59%/68%/66%) comparable respectively less voxels are in this range.

Furthermore, except for patient 1, it can be seen that the 'ground-truth' or estimated DVF uncertainties (GU-DU^{ref} or GMU-DU^{ref}) are generally closer to the reference dosimetric uncertainty than the upscaled gradient (G) only. Looking at individual structures of interest, this trend becomes more pronounced, such as results for the target structures CTV and PTV (again except for patient 1). In particular, in the case of the medulla, taking the upscaled gradient method can lead to substantial differences to the reference.

The 'ground-truth' DVF uncertainty (GU-DU) shows to be the closest to the reference dosimetric uncertainty with 92% of the voxels in the body, averaged over all patients, within ±10% difference, compared to 90% for the estimated DVF uncertainty (GMU-DU). The same is observed for other structures, for the ipsilateral lung 90% vs. 88%, for the heart 84% vs. 81% and for the PTV 97% vs. 95% (Table A.1 in the supplement).

Looking at the results for patients 3,4,6 and 7, it can be seen that for all structures, at least 86% of the voxels are within ±10% difference for the GU and GMU approaches, with 72% and more of the voxels having a difference of less than ±5% (Table A.1 in the supplement). However, for patient 2 and 5, a substantial number of voxels in the heart have differences larger than ±10% of the prescribed dose to the reference, whereas for other structures, the difference to the reference remain predominantly below ±10% for these patients. Worse results are found for patient 1, where for all structures, the estimated DVF uncertainties (GMU) differ substantially from those of the reference. On the other hand, dose uncertainties estimated using the GU approach give similar results to those observed for the other patients.

3.3 Impact of reference DIR

Finally, the impact of the reference DIR algorithm selection on direction factor calculation and the geometric uncertainty modelling were also investigated. For this purpose, the GUs from the validation branch were repeatedly calculated by using the direction factor derived from each of the five DIR algorithms, while the GMUs were estimated individually by using each of the DIRs as the reference for the geometric uncertainty model. In Figure 5, the results for two example CTs of patient 3 are shown, with CT2 being one of the worst-cases and CT9 one of the best cases. The results for all CTs are included in the appendix (Fig. A.4). The selection of the DIR for the direction factor has only a marginal impact on the uncertainty map GU but can influence GMU estimation slightly more. Nevertheless, the range of the differences is independent on the chosen reference DIR, except for certain CTs/structures (Body: CT 2, 4 & 5; Medulla: CT6; CTV: CT2 & 4).

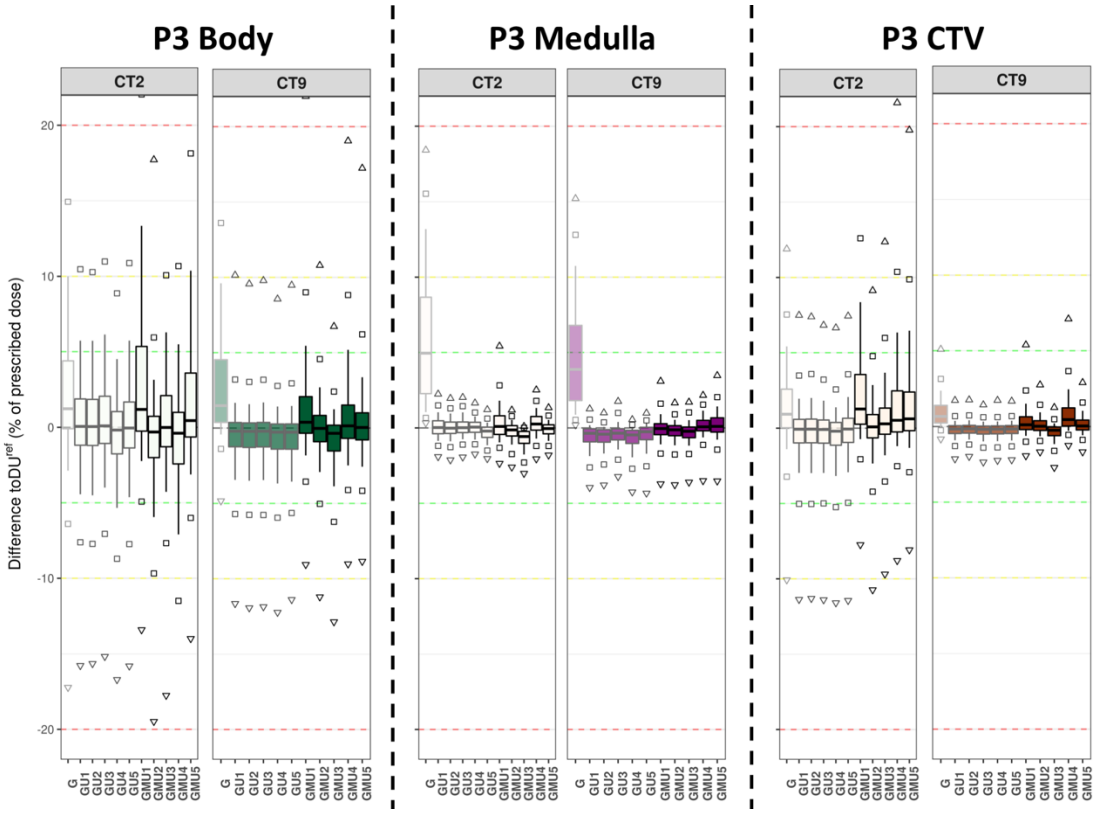


Figure 5: Impact of the reference DIR selection on direction factor calculation and the geometric uncertainty modelling for the body, medulla and CTV structure (G: gradient-only branch, GU: validation branch, GMU: the modelled branch; 1 Mirada, 2 Plastimatch B-spline, 3 Plastimatch Demons, 4 Raystation Anaconda, 5 Velocity) for CT2 and CT9 of patient 3.

4. Discussion

In this paper, we have proposed and evaluated a new method for estimating dosimetric uncertainties due to inter-algorithm DIR geometric uncertainty for seven NSCLC patients with multiple repeated CTs. The method is able to provide accurate dosimetric uncertainty estimation, with differences within ±5% to the reference scenarios for 77% of all non-zero dose voxels over all seven patients. The median difference of all voxels in the investigated structures is within ±2% for all patients, except for the heart in patient 5 (Table A.1 in the supplement). Note however, that our voxel to voxel comparison is a very conservative measure, which compares approximately 10^8 voxels across seven patients with nine CTs.

The closest estimations to the reference dosimetric uncertainty resulted from the validation branch (GU), indicating the potential of determining dose uncertainties by combining DVF variability, local dose gradients and directions. The model-based approach (GMU) can be considered as a special scenario of the concept, for which a first attempt to model the DVF variability is made. In addition, we proved that using only the upscaled dose gradient as a DIR uncertainty indication is not sufficient for proton therapy dose accumulation. Overall, the worst results were seen for patients with large tumor volumes and substantial visible changes in tumor size, especially for the modelled uncertainties (Patient 1, 2 & 5). For these patients, the used DIRs are expected to be less accurate in general, as they cannot properly handle tumor tissue loss. Furthermore, for these three patients, substantial parts of the PTV were separated from the mediastinal structures, while for patients 3,4,6 & 7, the PTV was located relatively close to this structure.

From the estimation branches, the model-based approach naturally showed larger differences to the reference than the results from the validation branch, where the geometric uncertainty is obtained by applying each DIR individually for all fractions. However, comparing the validation branch (GU) with the modelled branch (GMU), it was observed that for some cases (e.g. P2), the modelled branch had smaller differences to the reference (Fig. A.2). That would be due to the use of a simple linear model to predict DVF uncertainty. This could cause uncertainties at some points, resulting in even smaller differences than the validation path. Considering optimized models, for example by using machine learning, is planned as follow-up work.

Although the proposed model for the DVF uncertainties only initially considered the spread of vector magnitudes between the 5 DIRs, and does not include the spread of their directions, by comparing results in Figure 5 when different DIRs were used as reference for the direction factor calculation, magnitude differences were demonstrated to be the main contributing factor for the potential dosimetric uncertainty. On the other hand, for the modelling part, the choice of the DIR has a slightly higher impact on the estimation accuracy when selecting different DIRs as reference.

A limitation of this study is the fact that the ground-truth of the DIRs is not available. In the worst case, the ground-truth may neither be represented by any of the five DIR results, nor within the range of them. In addition, we need to point out that the presented results were based on the assumption that the five DIRs give a high enough variability to capture the possible dosimetric uncertainty span. Nevertheless, the method itself is not restricted to the exact number of DIRs, providing the possibility to include as many DIRs as available in the clinic. Moreover, besides the investigated inter-algorithm uncertainty, the impact of intra-algorithm uncertainty is worthy of further study, when different parameters are used for the same DIR algorithm to get a geometric uncertainty map of the DVF uncertainties. In general, the proposed method for the dosimetric uncertainty is not restricted to the geometric uncertainty map discussed in this paper but could be another geometric uncertainty measure defined on a dense grid. Additionally, the approach discussed here for the management of lung cancer treatments with pencil beam scanned proton therapy has the potential for broader applicability. The described method includes no factors limiting the approach to a specific tumor indication or treatment modality. However, in-depth validations are needed to verify the transfer of the method to other indications (e.g. head and neck cancer) or to photon treatments.

Overall, we present a promising method to estimate dosimetric uncertainties due to geometric discrepancies resulting from application of different DIR algorithms. The knowledge of the possible dosimetric uncertainties could give rise to robustness-like considerations for DIR. Moreover, if DIR algorithms come in the future together with a spatial uncertainty estimation itself, the proposed methodology could be directly plugged in to estimate the dosimetric uncertainties in an even faster manner without modelling. However, the actual approach could be refined further, for example the modeling of the geometric uncertainties could be improved in the following ways:

- Including image and/or image gradient information
- A possible inclusion of the direction into the modelling
- Studying a possible extension of the geometric uncertainty model from patient-specific to a population based model (application of the model still on patient-level);

- Trying other possible advanced modelling techniques (e.g. machine learning or deep learning) to improve the estimation accuracy.

5. Conclusion

We have proposed and evaluated a framework for estimating DIR induced dosimetric uncertainties in dose accumulation for proton therapy, by building a patient-specific model of the geometric uncertainties from multiple DIRs in a first fraction and applying the model in consecutive fractions. The estimation of the dosimetric uncertainty is a combination of the geometric uncertainties, the dose gradient and a direction factor, relating the dose gradient direction and the uncertainty direction. The proposed method was evaluated on multiple lung cancer patients, each with multiple CTs, and showed promise in obtaining dosimetric DIR uncertainties in dose accumulation with additional knowledge of the geometric uncertainties. Therefore, we were able to provide useful uncertainty information in DIR related applications for pencil beam scanning proton therapy.

Acknowledgments

This work was funded by Krebsliga Schweiz (KFS-4528-08-2018). We kindly acknowledge Dr. Mirjana Josipovic and Dr. Gitte F Persson from Copenhagen University Hospital, for sharing their valuable repeated BH lung CT dataset. Lena Nenoff was funded by SNSF (Project: 165961).

References

Brock K K 2010 Results of a Multi-Institution Deformable Registration Accuracy Study (MIDRAS) *Int. J. Radiat. Oncol. Biol. Phys.* **76** 583–96 Online: https://www.sciencedirect.com/science/article/pii/S0360301609009432?via%3Dihub

Brock K K, Mutic S, McNutt T R, Li H and Kessler M L 2017 Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132: Report *Med. Phys.* **44** e43–76

Chetty I J and Rosu-Bubulac M 2019 Deformable Registration for Dose Accumulation *Semin. Radiat. Oncol.* **29** 198–208 Online: https://www.sciencedirect.com/science/article/pii/S1053429619300116?via%3Dihub

Dowdell S, Grassberger C, Sharp G C and Paganetti H 2013 Interplay effects in proton scanning for lung: A 4D Monte Carlo study assessing the impact of tumor and beam delivery parameters *Phys. Med. Biol.* **58** 4137–56

Grassberger C, Dowdell S, Lomax A, Sharp G, Shackleford J, Choi N, Willers H and Paganetti H 2013 Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **86** 380–6 Online: http://dx.doi.org/10.1016/j.ijrobp.2013.01.024

Han Y 2019 Current status of proton therapy techniques for lung cancer *Radiat. Oncol. J.* **37** 232–48

Hill-Kayser C E, Both S and Tochner Z 2011 Proton therapy: Ever shifting sands and the opportunities and obligations within *Front. Oncol.* **1** 1–9

Hub M and Karger C P 2013 Estimation of the uncertainty of elastic image registration with the demons algorithm *Phys. Med. Biol.* **58** 3023–36 Online: http://stacks.iop.org/0031-9155/58/i=9/a=3023?key=crossref.305b6f895330665c0f623ce34d088fd3

Hub M, Thieke C, Kessler M L and Karger C P 2012 A stochastic approach to estimate the uncertainty of dose mapping caused by uncertainties in b-spline registration *Med. Phys.* **39** 2186–92 Online: http://doi.wiley.com/10.1118/1.3697524

Jingu K 2014 Use of Deformable Image Registration for Radiotherapy Applications *J Radiol Radiat Ther* **2** 1042–9

Josipovic M, Persson G F, Dueck J, Bangsgaard J P, Westman G, Specht L and Aznar M C 2016 Geometric uncertainties in voluntary deep inspiration breath hold radiotherapy for locally advanced lung cancer *Radiother. Oncol.* **118** 510–4 Online: http://dx.doi.org/10.1016/j.radonc.2015.11.004

Kadoya N, Fujita Y, Katsuta Y, Dobashi S, Takeda K, Kishi K, Kubozono M, Umezawa R, Sugawara T, Matsushita H and Jingu K 2014 Evaluation of various deformable image registration algorithms for thoracic images *J. Radiat. Res.* **55** 175–82 Online: https://academic.oup.com/jrr/article-lookup/doi/10.1093/jrr/rrt093

Kardar L, Li Y, Li X, Li H, Cao W, Chang J Y, Liao L, Zhu R X, Sahoo N, Gillin M, Liao Z, Komaki R, Cox J D, Lim G and Zhang X 2014 Evaluation and mitigation of the interplay effects of intensity modulated proton therapy for lung cancer in a clinical setting *Pract. Radiat. Oncol.* **4** E259–68 Online: http://dx.doi.org/10.1016/j.prro.2014.06.010

Kashani R, Balter J, Kessler M, Hub M, Dong L, Zhang L, Xing L, Xie Y, Hawkes D, Schnabel J, Mcclelland J and Joshi S 2007 TU C M100J 03: Objective Assessment of Deformable Image Registration in Radiotherapy; a Multi Institution Study *Med. Phys.* **34** 2545

Maintz J B A and Viergever M A 1998 A survey of medical image registration *Med. Image Anal.* **2** 1–36

Murphy K, Van Ginneken B, Reinhardt J M, Kabus S, Ding K, Deng X, Cao K, Du K, Christensen G E,

Garcia V, Vercauteren T, Ayache N, Commowick O, Malandain G, Glocker B, Paragios N, Navab N, Gorbunova V, Sporring J, De Bruijne M, Han X, Heinrich M P, Schnabel J A, Jenkinson M, Lorenz C, Modat M, McClelland J R, Ourselin S, Muenzing S E A, Viergever M A, De Nigris D, Collins D L, Arbel T, Peroni M, Li R, Sharp G C, Schmidt-Richberg A, Ehrhardt J, Werner R, Smeets D, Loeckx D, Song G, Tustison N, Avants B, Gee J C, Staring M, Klein S, Stoel B C, Urschler M, Werlberger M, Vandemeulebroucke J, Rit S, Sarrut D and Pluim J P W 2011 Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge *IEEE Trans. Med. Imaging* **30** 1901–20

Murphy M J, Salguero F J, Siebers J V., Staub D and Vaman C 2012 A method to estimate the effect of deformable image registration uncertainties on daily dose mapping *Med. Phys.* **39** 573–80 Online: http://doi.wiley.com/10.1118/1.3673772

Nenoff L, Ribeiro C O, Matter M, Hafner L, Josipovic M, Langendijk J A, Persson G F, Walser M, Weber D C, Lomax A J, Knopf A-C, Albertini F and Zhang Y 2020 Deformable image registration uncertainty for inter-fractional dose accumulation of lung cancer proton therapy. *Radiother. Oncol.* **0** Online: http://www.ncbi.nlm.nih.gov/pubmed/32380117

Oh S and Kim S 2017 Deformable image registration in radiation therapy *Radiat. Oncol. J.* **35** 101–11 Online: http://e-roj.org/journal/view.php?doi=10.3857/roj.2017.00325

Paganelli C, Meschini G, Molinelli S, Riboldi M and Baroni G 2018 “Patient-specific validation of deformable image registration in radiation therapy: Overview and caveats” *Med. Phys.* **45** e908–22 Online: http://doi.wiley.com/10.1002/mp.13162

Paganelli C, Peroni M, Riboldi M, Sharp G C, Ciardo D, Alterio D, Orecchia R and Baroni G 2013 Scale invariant feature transform in adaptive radiation therapy: A tool for deformable image registration assessment and re-planning indication *Phys. Med. Biol.* **58** 287–99

Palm Å and Johansson K A 2007 A review of the impact of photon and proton external beam radiotherapy treatment modalities on the dose distribution in field and out-of-field; implications for the long-term morbidity of cancer survivors *Acta Oncol. (Madr).* **46** 462–73

Ribeiro C O, Knopf A, Langendijk J A, Weber D C, Lomax A J and Zhang Y 2018 Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours *Radiother. Oncol.* **128** 174–81 Online: https://www.sciencedirect.com/science/article/pii/S0167814018301336?via%3Dihub

Rigaud B, Simon A, Castelli J, Lafond C, Acosta O, Haigron P, Cazoulat G and de Crevoisier R 2019 Deformable image registration for radiation therapy: principle, methods, applications and evaluation *Acta Oncol. (Madr).* **58** 1225–37 Online: https://doi.org/10.1080/0284186X.2019.1620331

Saleh-Sayah N K, Weiss E, Salguero F J and Siebers J V. 2011 A distance to dose difference tool for estimating the required spatial accuracy of a displacement vector field *Med. Phys.* **38** 2318–23

Saleh Z H, Apte A P, Sharp G C, Shusharina N P, Wang Y, Veeraraghavan H, Thor M, Muren L P, Rao S S, Lee N Y and Deasy J O 2014 The distance discordance metric - A novel approach to quantifying spatial uncertainties in intra- and inter-patient deformable image registration *Phys. Med. Biol.* **59** 733–46

Salguero F J, Saleh-Sayah N K, Yan C and Siebers J V. 2011 Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping *Med. Phys.* **38** 343–53 Online: http://doi.wiley.com/10.1118/1.3528201

Samavati N, Velec M and Brock K K 2016 Effect of deformable registration uncertainty on lung SBRT dose accumulation *Med. Phys.* **43** 233–40

Sarrut D, Baudier T, Ayadi M, Tanguy R and Rit S 2017 Deformable image registration applied to lung SBRT: Usefulness and limitations *Phys. Medica* **44** 108–12 Online: https://doi.org/10.1016/j.ejmp.2017.09.121

Sotiras A, Davatzikos C and Paragios N 2013 Deformable medical image registration: A survey *IEEE Trans. Med. Imaging* **32** 1153–90 Online: http://www.ncbi.nlm.nih.gov/pubmed/23739795

Stuschke M, Kaiser A, Pöttgen C, Lübcke W and Farr J 2012 Potentials of robust intensity modulated scanning proton plans for locally advanced lung cancer in comparison to intensity modulated photon plans *Radiother. Oncol.* **104** 45–51 Online: http://dx.doi.org/10.1016/j.radonc.2012.03.017

Szeto Y Z, Witte M G, van Kranen S R, Sonke J J, Belderbos J and van Herk M 2016 Effects of anatomical changes on pencil beam scanning proton plans in locally advanced NSCLC patients *Radiother. Oncol.* **120** 286–92 Online: http://dx.doi.org/10.1016/j.radonc.2016.04.002

Torsten Rohlfing 2014 Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable *Bone* **23** 1–7 Online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf

Varadhan R, Magome T and Hui S 2016 Characterization of deformation and physical force in uniform low contrast anatomy and its impact on accuracy of deformable image registration *Med. Phys.* **43** 52–61

Zhang Y, Boye D, Tanner C, Lomax A J and Knopf A 2012 Respiratory liver motion estimation and its effect on scanned proton beam therapy *Phys. Med. Biol.* **57** 1779–95

Supplementary material

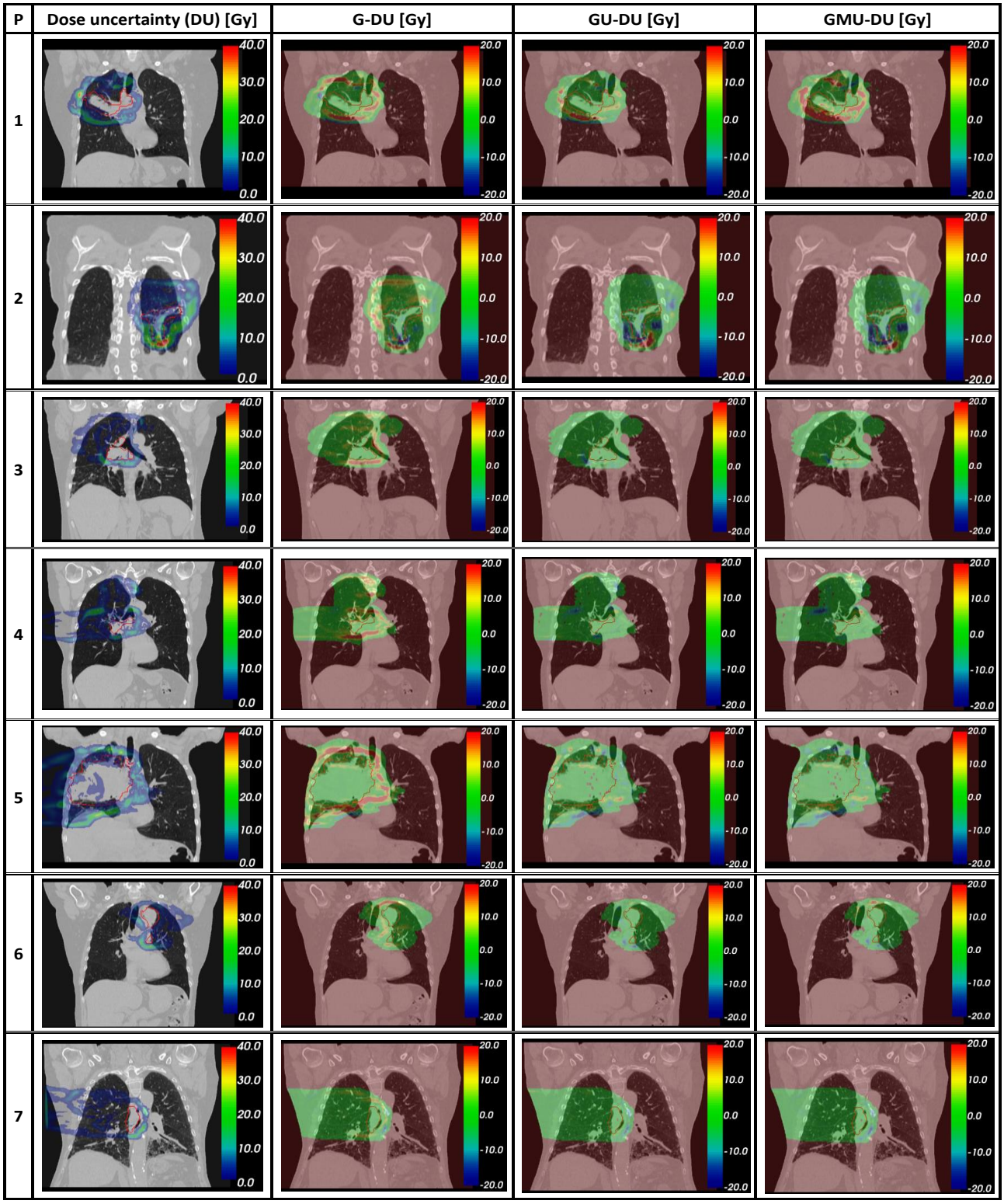


Fig. A.1: Example slices of the three difference maps together with the reference dose uncertainty for all patients (1-7). The CTV structure is shown in red.

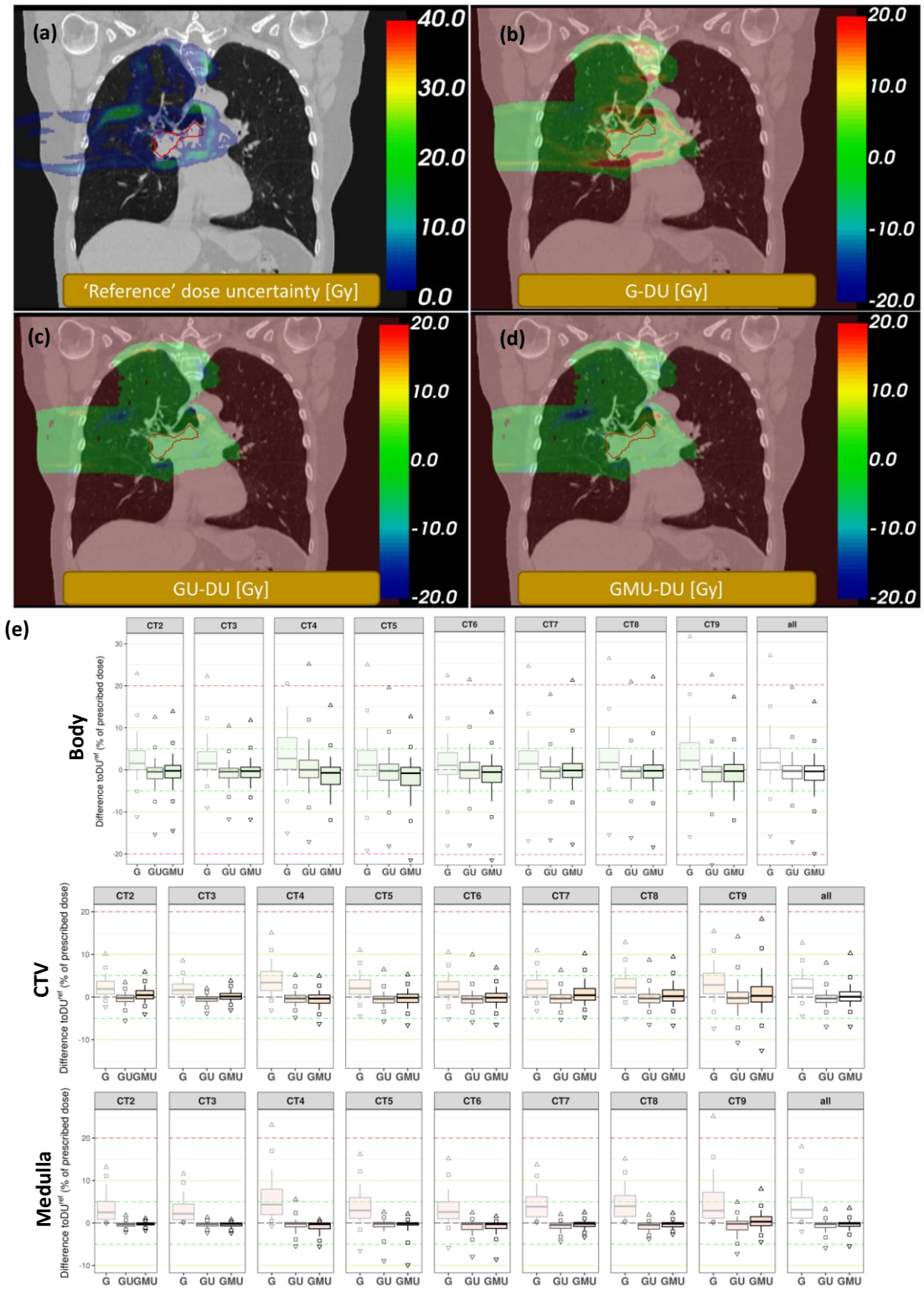


Fig. A.2: (a) Example CT slice from patient 4 showing the reference dose uncertainty (DU) and the three resulting differences to the reference dose uncertainty (CTV in red). Distribution of the voxel wise dose differences between reference dose uncertainty and (b) the upscaled gradient uncertainty map (gradient branch: G-DU), (c) the DVF uncertainty from five DIRs (validation branch: GU-DU); (d) the estimated dosimetric uncertainty including the modelled DVF uncertainty map (modelled branch: GMU-DU). (e) Comparison of the 10th-90th percentile of the dose difference in body, medulla and CTV.

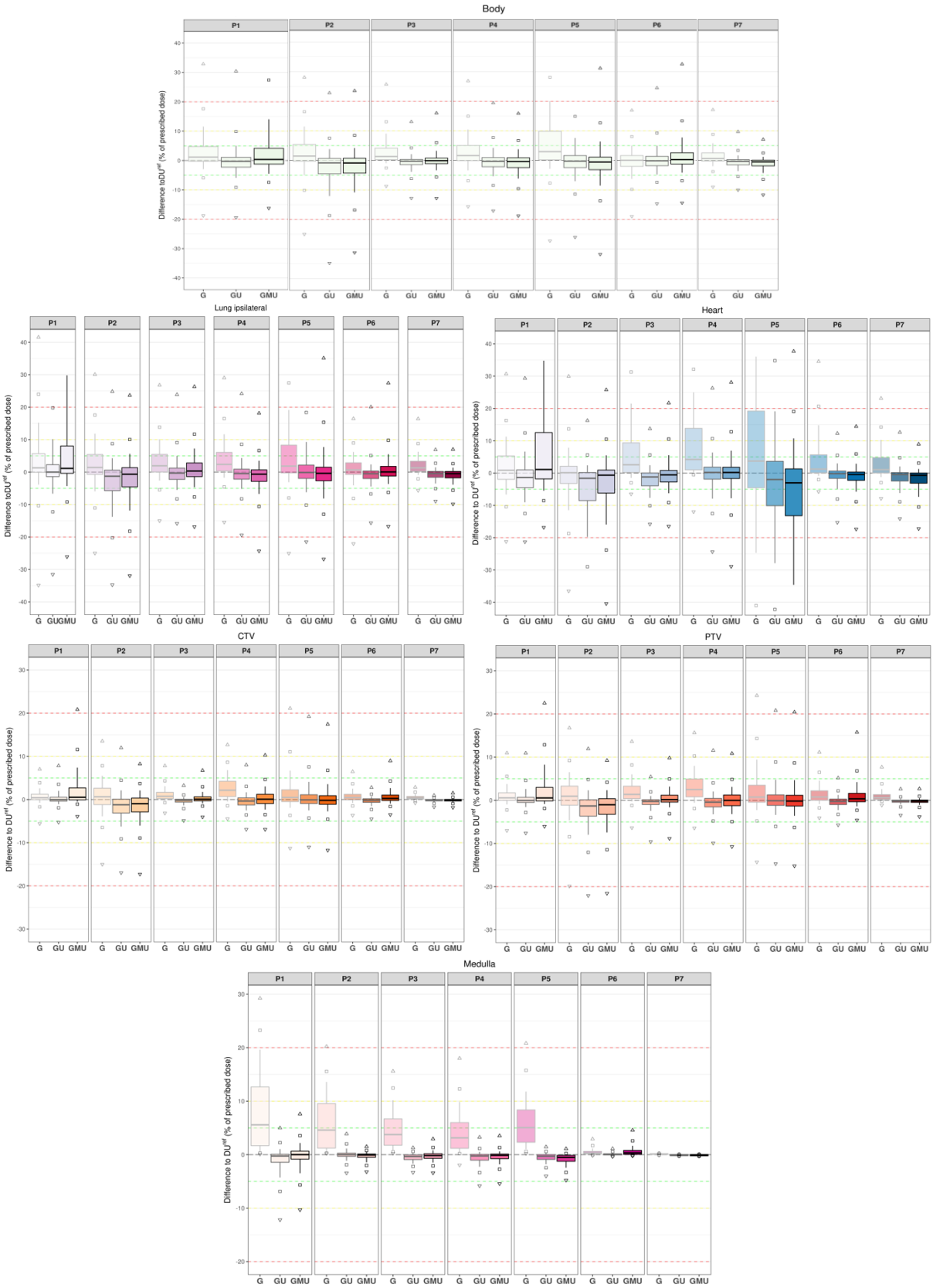


Fig. A.3: Difference to reference dose uncertainty voxel values for the body, ipsilateral lung, heart, medulla, CTV and PTV structures. For each patient (P1-P7) the difference from the gradient-only branch (G), the branch including “ground-truth” geometric uncertainties (GU) and the branch with modelled geometric uncertainties (GMU) are plotted.

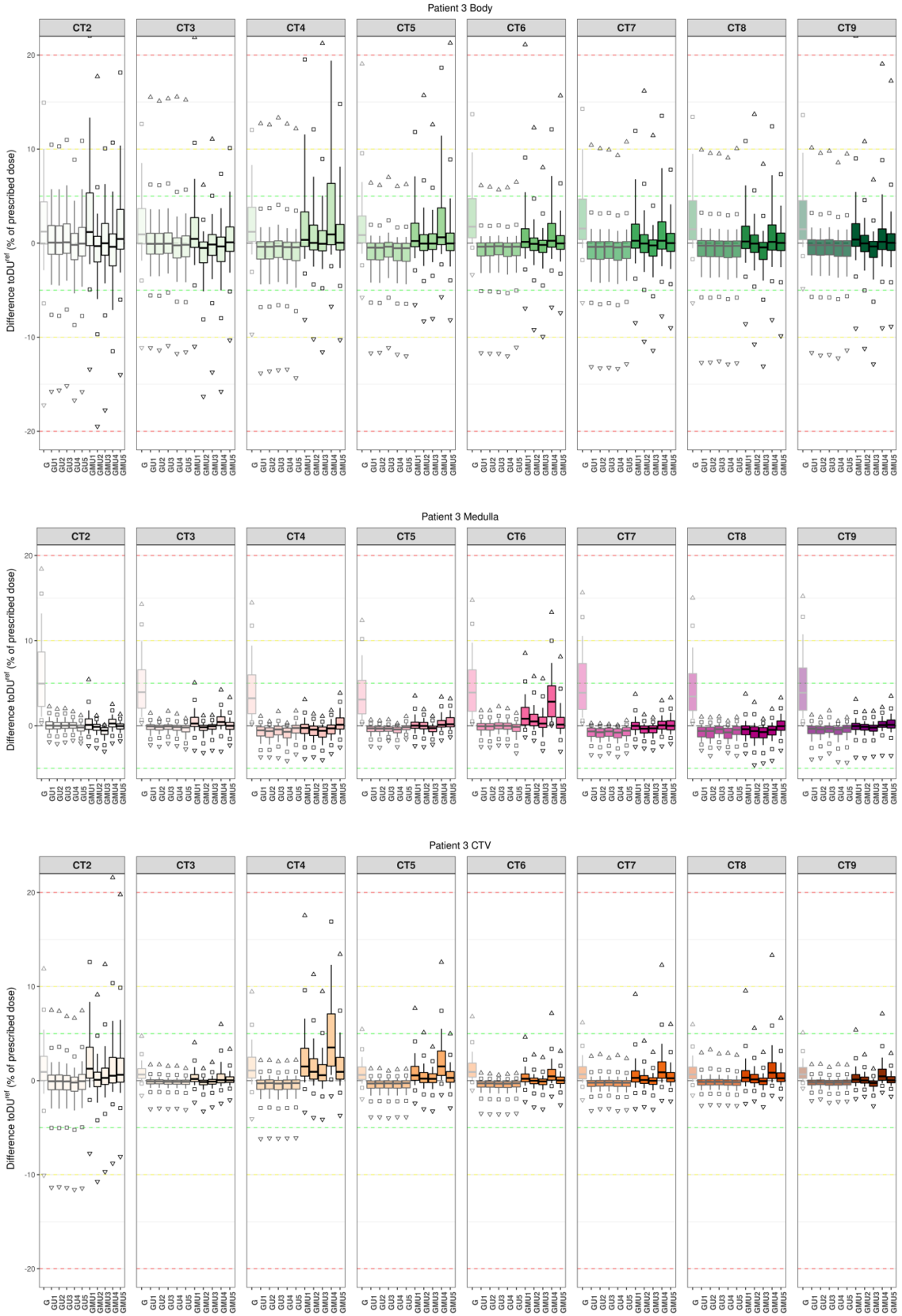


Fig. A.4: Impact of the reference DIR selection on direction factor calculation and the geometric uncertainty modelling for the body, medulla and CTV structure (G: gradient-only branch, GU: validation branch, GMU: the modelled branch; 1 Mirada, 2 Plastimatch B-spline, 3 Plastimatch Demons, 4 Raystation Anaconda, 5 Velocity) for patient 3.

		Patient 1			Patient 2			Patient 3			Patient 4		
		G	GU	GMU	G	GU	GMU	G	GU	GMU	G	GU	GMU
Body	Voxels	15'382'145	15'382'145	15'382'145	9'573'821	9'573'821	9'573'821	14'924'913	14'924'913	14'924'913	21'650'393	21'650'393	21'650'393
	Median [%]	1.17	-0.29	0.37	1.46	-0.85	-0.89	1.30	-0.28	-0.10	1.61	-0.34	-0.39
	25^{th}-75^{th} percentile [%]	[-0.26 , 4.76]	[-2.28,1.08]	[-1.23,4.13]	[-0.32,5.38]	[-4.56,0.66]	[-4.33,0.71]	[0.24,4.14]	[-1.46,0.33]	[-1.11,0.77]	[-0.10,5.06]	[-2.15,0.95]	[-2.49,0.93]
	< ±5 % difference	70%	78%	69%	63%	69%	69%	77%	89%	88%	68%	81%	79%
	< ±10 % difference	85%	91%	83%	83%	84%	85%	91%	97%	96%	86%	93%	93%
ipsilateral Lung	Voxels	4'518'423	4'518'423	4'518'423	4'482'011	4'482'011	4'482'011	3'862'375	3'862'375	3'862'375	7'900'865	7'900'865	7'900'865
	Median [%]	1.31	0.00	1.14	1.43	-1.27	-0.67	1.94	-0.31	0.33	2.37	-0.45	-0.66
	25^{th}-75^{th} percentile [%]	[-0.07,5.68]	[-1.39,2.27]	[-0.38,8.04]	[-0.72,5.42]	[-5.73,0.60]	[-4.56,1.37]	[0.04,5.46]	[-2.27,1.24]	[-1.39,2.83]	[0.35,6.07]	[-2.19,0.89]	[-2.86,0.68]
	< ±5 % difference	64%	71%	59%	62%	64%	65%	67%	79%	75%	65%	81%	78%
	< ±10 % difference	79%	83%	74%	82%	81%	83%	88%	92%	91%	85%	92%	92%
Heart	Voxels	211'009	211'009	211'009	882'012	882'012	882'012	247'515	247'515	247'515	244'525	244'525	244'525
	Median [%]	0.86	-1.36	1.10	0.12	-1.61	-0.65	2.57	-1.20	-0.56	4.20	0.20	0.19
	25^{th}-75^{th} percentile [%]	[-1.94,5.31]	[-4.51,0.97]	[-1.79,12.54]	[-3.20,2.11]	[-8.52,0.19]	[-6.19,0.92]	[0.21,9.37]	[-3.96,0.05]	[-2.76,0.82]	[0.98,13.84]	[-1.89,1.86]	[-1.69,1.81]
	< ±5 % difference	60%	65%	50%	65%	60%	61%	61%	75%	75%	51%	72%	72%
	< ±10 % difference	82%	85%	68%	81%	75%	77%	76%	93%	91%	66%	87%	86%
Medulla	Voxels	41'280	41'280	41'280	22'840	22'840	22'840	46'244	46'244	46'244	62'751	62'751	62'751
	Median [%]	5.57	-0.26	0.00	4.58	-0.01	-0.07	3.76	-0.35	-0.17	3.14	-0.24	-0.13
	25^{th}-75^{th} percentile [%]	[1.69,12.66]	[-1.43,0.07]	[-0.84,0.65]	[1.22,9.55]	[-0.33,0.29]	[-0.51,0.11]	[1.76,6.69]	[-0.92,-0.04]	[-0.69,0.18]	[1.18,6.00]	[-1.04,0.05]	[-0.75,0.11]
	< ±5 % difference	47%	91%	91%	54%	100%	100%	63%	100%	100%	67%	98%	98%
	< ±10 % difference	66%	98%	99%	77%	100%	100%	90%	100%	100%	90%	100%	100%
CTV	Voxels	1'552'903	1'552'903	1'552'903	598'789	598'789	598'789	342'979	342'979	342'979	627'732	627'732	627'732
	Median [%]	0.41	-0.02	0.54	0.63	-1.21	-0.97	0.74	-0.19	0.07	2.15	-0.36	0.07
	25^{th}-75^{th} percentile [%]	[-0.05,1.24]	[-0.38,0.54]	[-0.08,2.75]	[-1.20,2.62]	[-3.16,0.09]	[-2.90,0.33]	[0.21,1.67]	[-0.60,0.071]	[-0.29,0.63]	[0.69,4.23]	[-1.26,0.39]	[-0.93,1.22]
	< ±5 % difference	96%	96%	84%	81%	81%	84%	96%	99%	97%	80%	95%	93%
	< ±10 % difference	99%	99%	93%	95%	94%	95%	99%	100%	100%	97%	99%	99%
PTV	Voxels	2'393'186	2'393'186	2'393'186	1'069'145	1'069'145	1'069'145	706'252	706'252	706'252	1'187'142	1'187'142	1'187'142
	Median [%]	0.56	-0.03	0.57	0.92	-1.33	-1.05	1.37	-0.26	0.16	2.50	-0.43	-0.01
	25^{th}-75^{th} percentile [%]	[-0.03,1.73]	[-0.50,0.69]	[-0.13,3.1]	[-1.20,3.28]	[-3.77,0.12]	[-3.35,0.35]	[0.38,3.27]	[-0.99,0.14]	[-0.43,1.18]	[0.80,4.94]	[-1.53,0.44]	[-1.25,1.23]
	< ±5 % difference	92%	93%	81%	75%	76%	79%	85%	95%	92%	74%	91%	90%
	< ±10 % difference	98%	98%	91%	91%	91%	92%	97%	99%	98%	94%	98%	97%

Patient 5			Patient 6			Patient 7			Average over Patients		
G	GU	GMU	G	GU	GMU	G	GU	GMU	G	GU	GMU
21'867'320	21'867'320	21'867'320	13'505'638	13'505'638	13'505'638	16'005'648	16'005'648	16'005'648			
2.96	-0.26	-0.56	0.04	-0.15	0.26	0.63	-0.36	-0.51	1.31	-0.36	-0.26
[0.19,9.85]	[-2.45,1.62]	[-3.15,1.11]	[-2.05,1.63]	[-1.82,1.36]	[-1.24,2.59]	[-0.27,2.51]	[-1.53,0.24]	[-1.86,0.10]	[-0.37,4.76]	[-2.3,0.89]	[-2.2,1.48]
53%	72%	70%	77%	81%	77%	84%	91%	90%	70%	80%	77%
71%	86%	85%	92%	93%	90%	95%	98%	98%	86%	92%	90%
8'645'430	8'645'430	8'645'430	4'000'671	4'000'671	4'000'671	5'382'746	5'382'746	5'382'746			
1.85	-0.13	-0.38	0.67	-0.50	0.04	1.31	-0.45	-0.48	1.55	-0.45	-0.10
[-0.27,8.31]	[-1.96,2.16]	[-2.69,1.49]	[-0.73,2.87]	[-2.00,0.36]	[-1.22,1.77]	[0.21,3.36]	[-1.62,0.22]	[-1.78,0.20]	[-0.17,5.31]	[-2.4,1.11]	[-2.13,2.34]
58%	71%	70%	78%	87%	82%	84%	93%	92%	68%	78%	75%
75%	85%	84%	92%	95%	93%	96%	99%	99%	85%	90%	88%
712'460	712'460	712'460	137'087	137'087	137'087	655'606	655'606	655'606			
3.71	-2.00	-3.03	1.24	-0.18	-0.43	1.06	-0.46	-0.69	1.97	-0.94	-0.58
[-4.55,19.20]	[-10.10,3.64]	[-13.20,1.30]	[0.09,5.66]	[-1.66,0.64]	[-2.25,0.32]	[0.03,4.79]	[-2.48,0.17]	[-3.15,0.03]	[-1.20,8.61]	[-4.73,1.07]	[-4.43,2.53]
29%	40%	41%	72%	85%	82%	73%	82%	81%	59%	68%	66%
44%	59%	60%	84%	95%	94%	89%	95%	94%	75%	84%	81%
32'865	32'865	32'865	3'323	3'323	3'323	10	10	10			
5.05	-0.29	-0.52	0.32	0.04	0.31	0.10	-0.06	-0.06	3.22	-0.17	-0.09
[2.31,8.36]	[-0.90,0.01]	[-1.25,-0.13]	[0.15,0.60]	[-0.06,0.16]	[0.04,0.79]	[0.08,0.19]	[-0.26,0.04]	[-0.28,0.03]	[1.20,7.24]	[-0.71,0.08]	[-0.61,0.25]
50%	100%	99%	100%	100%	99%	100%	100%	100%	69%	98%	98%
86%	100%	100%	100%	100%	100%	100%	100%	100%	87%	100%	100%
3'929'624	3'929'624	3'929'624	274'667	274'667	274'667	418'991	418'991	418'991			
0.49	-0.05	-0.17	0.45	-0.18	0.25	0.35	-0.12	-0.13	0.75	-0.30	-0.05
[-0.48,2.34]	[-0.91,1.24]	[-1.10,0.93]	[-0.08,1.22]	[-0.66,0.17]	[-0.23,1.04]	[0.12,0.71]	[-0.32,0.02]	[-0.34,0.03]	[-0.11,2.00]	[-1.04,0.36]	[-0.84,0.99]
83%	86%	88%	97%	99%	96%	100%	100%	100%	90%	94%	92%
92%	94%	95%	100%	100%	99%	100%	100%	100%	98%	98%	97%
5'397'126	5'397'126	5'397'126	524'756	524'756	524'756	779'504	779'504	779'504			
0.78	-0.06	-0.16	0.77	-0.23	0.34	0.59	-0.15	-0.15	1.07	-0.36	-0.04
[-0.43,3.70]	[-1.14,1.49]	[-1.32,1.27]	[0.01,2.21]	[-1.00,0.29]	[-0.34,1.64]	[0.21,1.33]	[-0.45,0.05]	[-0.48,0.07]	[-0.04,2.92]	[-1.34,0.46]	[-1.04,1.18]
75%	82%	83%	91%	97%	91%	97%	99%	99%	84%	91%	88%
88%	91%	92%	99%	100%	97%	100%	100%	100%	95%	97%	95%

Table A.1: Summary over all investigated patients and structures showing included voxels, median difference to the reference dosimetric uncertainty, the 25th and 75th percentile of the differences, and the percentage of voxels having a difference <±5% of the prescribed dose, respectively <±10%.