Towards an optimal clinical protocol for the treatment of moving targets with pencil beam scanned proton therapy

Ribeiro, Cássia O.

DOI:
10.33612/diss.126443635


E

CHAPTER 3

ASSESSMENT OF DOSIMETRIC ERRORS INDUCED BY DEFORMABLE IMAGE REGISTRATION METHODS IN 4D PENCIL BEAM SCANNED PROTON TREATMENT PLANNING FOR LIVER TUMOURS

PUBLISHED IN:
Radiotherapy and Oncology 2018 July; 128(1): 174-181

• Cássia O. Ribeiro / 1
• Antje Knopf / 1
• Johannes A. Langendijk / 1
• Damien C. Weber / 2 / 3
• Antony J. Lomax / 2 / 4
• Ye Zhang / 2

1 / Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, **The Netherlands**
2 / Center for Proton Therapy, Paul Scherrer Institut, Villigen-PSI, **Switzerland**
3 / Department of Radiation Oncology, University Hospital Zurich, **Switzerland**
4 / Department of Physics, ETH Zurich, **Switzerland**

ABSTRACT

Purpose: Respiratory impacts in pencil beam scanned proton therapy (PBS-PT) are accounted by extensive 4D dose calculations, where deformable image registration (DIR) is necessary for estimating deformation vector fields (DVFs). We aim here to evaluate the dosimetric errors induced by different DIR algorithms in their resulting 4D dose calculations by using ground truth(GT)-DVFs from 4DMRI.

Materials and methods: Six DIR methods: ANACONDA, Morfeus, B-splines, Demons, CT Deformable, and Total Variation, were respectively applied to nine 4DCT-MRI liver data sets. The derived DVFs were then used as input for 4D dose calculation. The DIR induced dosimetric error was assessed by individually comparing the resultant 4D dose distributions to those obtained with GT-DVFs. Both single-/three-field plans and single/rescanned strategies were investigated.

Results: Differences in 4D dose distributions among different DIR algorithms, and compared to the results using GT-DVFs, were pronounced. Up to 40 % of clinically relevant dose calculation points showed dose differences of 10 % or more between the GT. Differences in V_{95}(CTV) reached up to 11.34 ± 12.57 %. The dosimetric errors became in general less substantial when applying multiple-field plans or using rescanning.

Conclusion: Intrinsic geometric errors by DIR can influence the clinical evaluation of liver 4D PBS-PT plans. We recommend the use of an error bar for correctly interpreting individual 4D dose distributions.

INTRODUCTION

When treating moving targets in the thorax or abdomen with pencil beam scanned proton therapy (PBS-PT), due to the presence of breathing motion, a time-resolved 3D (4D) image is necessary for quantifying the motion characteristics and performing a 4D dose calculation. For treating this type of tumour with a highly precise technique such as PBS-PT, a 4D dose calculation is crucial in order to take into account the deterioration of the dose distribution due to the relative motion between the target and the delivered pencil beams (interplay effects) [1–3].

To calculate motion induced geometric differences between two image phases, deformable image registration (DIR) is the standard approach for building up a point-to-point correlation between corresponding features. To perform DIR, a fixed and a moving image are pre-defined to estimate the patient's deformable motion between these two images [4]. The result of DIR is a deformation vector field (DVF), which contains vectors for each voxel pointing from the fixed image towards the moving image.

For any form of radiotherapy, DIR is one of the irreplaceable components for both inter- and intra- fractional dosimetric evaluation. It is especially important for PBS-PT, due to its high sensitivity to geometric accuracy. However, it is well known that DIR is an ill-posed problem intrinsically [5]. When applying different DIR methods to the same image pair, the resulting motion estimations can be inaccurate and differ significantly from each other [6]. Some of these errors are quantifiable, and can be calculated by comparing the DIR estimated motion of well-defined landmarks to their actual positions in both images (the so-called ground truth (GT) data). This is the classic approach of evaluating any DIR algorithm performance, as used by many previous publications [7,8]. Despite compromising the efficiency for the error quantification, the more landmarks that are defined, the more reliable the results will be. In contrast, there are also unquantifiable errors in featureless regions of

the images, where the deformable problem is intrinsically ill-defined. Motion vectors in these regions will directly depend on the transformation model and regularization of the selected algorithm, and it is therefore unavoidable that ambiguity between different algorithms will exist.

Any form of registration uncertainty can directly lead to distinguishable differences in dose distributions, which consequently influence any further dosimetric analysis and clinical decision-making [9]. In the literature, a number of studies have investigated the dosimetric uncertainties induced by a particular DIR method [10,11]. However, their conclusions were restricted to their selected DIR method, and a consensus on the clinical impact of DIR uncertainty is still difficult to achieve. Yeo *et al.* [12] compared calculated doses based on results from several available DIR algorithms with a measured dose using a deformable 3D dosimeter. However, DIR errors for real patient geometries may perform differently in contrast to the rather simple experimental setup used in that work.

Zhang *et al.* [6] showed that the ambiguity of two DIR approaches can lead to significant differences in the estimated motion maps, and subsequent 4D dose distributions, among liver cancer patients for PBS-PT, even if landmark registration errors were similar. Due to the lack of a comprehensive GT-DVF however, it is often impossible to validate the accuracy of DIR in the whole region of interest.

In this work, we would like to improve the above studies in two aspects. First, to investigate the systematic errors induced by DIRs in 4D dose calculations, using the unique advantage of comprehensive GT-DVFs extracted from synthetic 4DCT-MRI [13]. Second, we include multiple DIR methods to reveal the extent of potential variation induced by different algorithms. As such, six DIR methods (five commercially available and one research version) have been applied to nine 4DCT-MRI data sets to estimate deformable motion within the abdomen region. Compared to previous works, we also consider comprehensive GT-DVFs as reference to quantify the absolute accuracy for deformable motion estimation.

Consequently, the resulting 4D dose distributions generated using different DIR algorithms can be directly compared under conditions of varying plan configurations, rescanning scenarios, patient geometries, and motion scenarios.

MATERIALS AND METHODS

Synthetic 4DCT-MRI and ground truth deformation vector fields (GT-DVFs)

4DCT-MRI data sets consist of end-of-exhalation 3DCTs (reference phases) modulated by consecutive and extended breathing motion extracted from 4DMRI data through a validated image processing method [13–15] (Fig. 1, upper left). Through this process, synthetic 4DCT-MRI data sets within the liver are obtained by warping the reference phase with DVFs extracted from 4DMRI using a combination of multiresolution affine registration and B-spline non-rigid registration [15].

Nine such 4DCT-MRI data sets, generated from motion artefact-free 3DCTs of three liver cancer patients (denoted as PI, PII, and PIII respectively), were included in this study. The reference phases of the three patients were modulated by three different 4DMRI motion scenarios indicated as motions A, B, and C [16]. Clinical target volumes (CTVs) at the reference phase were 122, 264, and 403 cm^3 for patients I, II, and III respectively. Only 4DCT-MRI data sets corresponding to the first breathing cycle were analysed, and no consideration of motion irregularity has been included in the study. For the nine data sets, the amplitude for the first breathing cycle (given by the mean of the amplitude of all different points within the whole liver region) of motion scenarios A, B, and C were 7.82 (SD = 2.01), 20.61 (SD = 3.39), and 16.88 (SD = 2.78) mm respectively. Additionally, motion periods (extracted using Fourier analysis) for this first cycle equalled 3.66, 4.62, and 7.22 s for A, B, and C respectively. The corresponding DVFs extracted from 4DMRI to generate these nine 4DCT-MRI data sets were then defined as the

GT-DVFs. Subsequently, new DVFs were extracted from these nine 4DCT-MRI data sets using the different DIR methods being investigated (see Fig. 1). These GT-DVFs and DIR estimated DVFs were used for the 4D dose calculation analysis.

Deformable image registration (DIR) methods and derived deformation vector fields (DVFs)

Six DIR methods have been included in this study. DIR1 and DIR2 are available in the RayStation (RaySearch Laboratories, Stockholm, Sweden) treatment planning system used in the UMCG, whereas DIR3 and DIR4 [6] are algorithms provided in open source software (Plastimatch; www.plastimatch.com) and used at PSI. DIR5 and DIR6 were developed in turn by the commercial medical imaging software company Mirada Medical (Oxford, UK) and by the Computer Vision Laboratory in ETH Zurich (Zurich, Switzerland) respectively. The different DIR methods are based on the ANACONDA [17], Morfeus [18,19], B-splines, Demons, CT Deformable [20,21], and Total Variation [22] algorithm respectively (Suppl. 1). For each data set, all six approaches were applied to the reference phase as the fixed image. The remaining phases were defined as successive moving images (see Table S.1).

4D dose calculation

The DVFs resulting from the application of the six DIR methods were used as input to the in-house 4D dose calculation engine at PSI, which is an extension of the 3D dose calculation algorithm. The gantry (beam) coordinate system is defined as (s,t,u), in which s is the pencil beam central axis direction and (t_0, u_0) its position orthogonal to the field direction (Fig. 2(a)). The clinically used dose grid size in this coordinate system is $4 \times 4 \times 2.5\,mm^3$.

To extend the 3D dose calculation to a 4D dose calculation, time-dependent displacements of dose grid points for motions in the t and u directions are taken into account using displacement and density-variation maps derived from each phase of the relevant 4DCT-MRI data. The 4D dose calculation algorithm first estimates the time stamp

of each delivered pencil beam [13]. The DIR extracted DVFs are then geometrically translated and rotated into the gantry (*s*,*t*,u) coordinate system, and sampled by the dose grid size to provide displacement maps for each dose grid point [6]. Density-variation maps are derived from the different 4DCT-MRI phases using Siddon's algorithm [23]. With these displacement and density-variation maps, the offsets of the dose grid points from their nominal positions are calculated and a 4D dose distribution obtained.

4D planning configurations

Static, single-field uniform dose (SFUD) plans [24] were calculated on each of the reference phases of the three patients. Both single- and three-field plans were investigated. Field arrangements were anterior-posterior (F1), right lateral (F2), and anterior-inferior oblique (F3), with the three-field plan being a combination of all fields. 4D dose distributions for these were then subsequently obtained by using either the GT-DVFs or the DVFs resulting from the six DIR methods in the 4D dose calculation algorithm (Fig. 1). Single scan or five times layered rescanning [16] were simulated with the scanning parameters of Gantry 2 at PSI [25–27]. Plan delivery started at the reference phase of the corresponding 4DCT-MRI. All analysed plan configurations and respective notations are given in Suppl. 2.

Assessing DIR performance

The impact of using the different DVFs derived from the applied DIR methods in the 4D dose calculation has been assessed by (1) quantifying the geometric error in the beam coordinate system and (2) analysing dosimetric errors in the 4D dose distribution.

The DIR induced geometric error was quantified by the absolute differences of the derived DVFs from each DIR method with respect to the GT-DVFs (Suppl. 3). Since this study is focused on the impact of DIR on 4D dose distributions, the assessment of the geometric error was performed in the field direction, by analysing the *t* and u coordinates

within a defined field-specific volume of interest (fVOI) (Fig. 2(b)). To more precisely correlate the geometric error to the dosimetric error, we defined the fVOI as the dose region where the dose calculation grid covers patient geometry (to exclude dose regions outside the patient). As such, the geometric evaluation was performed only in regions that directly contribute to the 4D dose calculation.

The 4D dose distributions obtained from the DVFs derived from the six DIRs were individually calculated and compared to those resulting when using the GT-DVFs. Both dose-volume histograms (DVHs) with error bands, as well as V_{95}(CTV) values were analysed to quantify the impact of DIR in clinical practice. Difference DVHs (DDVHs), which correspond to the histograms of the absolute dose difference between the individual DIR estimated and GT 4D plans in an extension of the CTV volume (CTV + 1 cm), were computed. Finally, for each investigated DIR method, percentages of the extended CTV volume with absolute dose differences higher than 10 % were extracted from the DDVH for analysis.

RESULTS

DIR induced geometric error

Geometric registration errors with respect to motion scenarios and DIR methods for all single-field arrangements are given in Table 1. The lowest and highest mean errors in the field direction for the smallest motion cases (motion scenario A) were achieved by DIR6 and DIR1 (0.72 ± 0.16 mm and 1.05 ± 0.29 mm respectively). Despite motion B having the highest amplitude within the liver (Suppl. 4.1), motion C actually showed larger movement in the different single-field directions within the selected fVOI. For motion C, all the errors increased up to 3.21 ± 0.54 mm for DIR3, 2.84 ± 0.52 mm for DIR1, and 2.22 ± 0.42 mm for DIR6. Overall, the best performance was achieved by DIR6 for all three motion scenarios. Most importantly, it was also observed that all DIR methods underestimated the GT motion amplitudes (Table S.4.1).

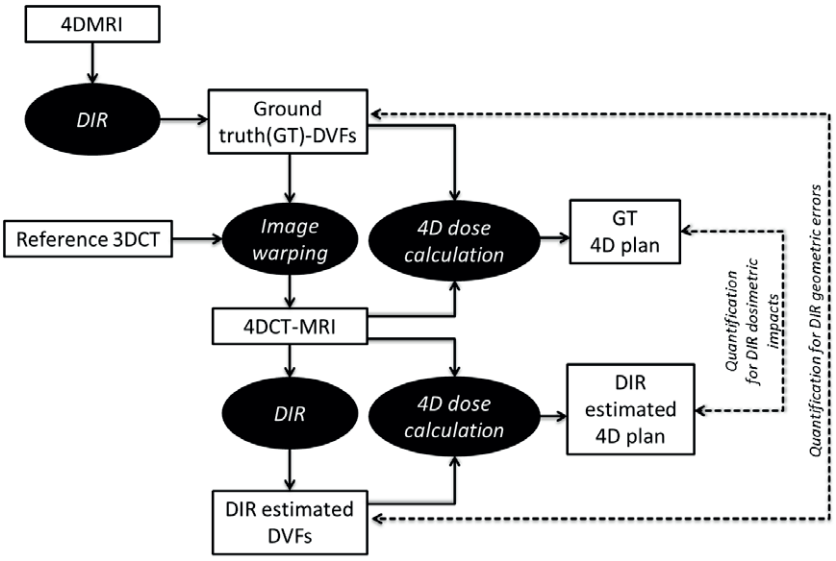


Fig. 1. Schematic representation of the workflow adopted in this paper.

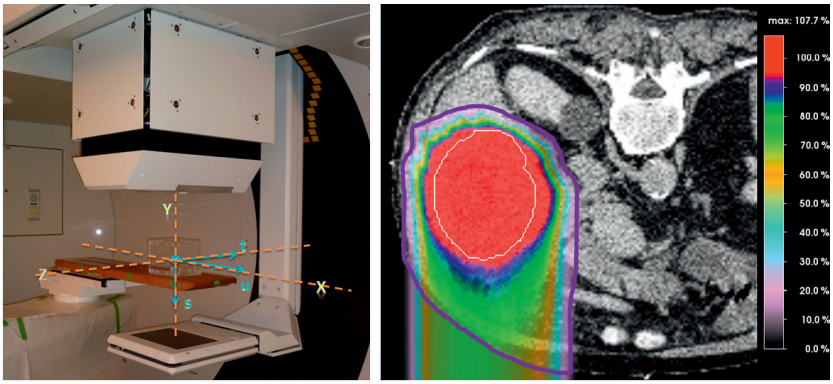


Fig. 2. (a) Gantry (or beam) coordinate system defined by (s,t,u) represented in blue. **(b)** Anterior-posterior (F1) fVOI of patient PIII (in purple) for the corresponding geometric field-specific registration error quantification.

Table 1 Field-specific geometric errors before and after DIR for the analysed configurations: three different single-field directions and three liver cancer patients, each modulated by motion scenarios A, B, or C.

	Before DIR	**After DIR**						
	(mean ± SD) [mm]	(mean ± SD) [mm]						
Motion	GT	DIR1	DIR2	DIR3	DIR4	DIR5	DIR6	Averaged
A	1.47 ± 0.26	1.05 ± 0.29	1.01 ± 0.15	1.04 ± 0.23	0.76 ± 0.15	0.74 ± 0.17	0.72 ± 0.16	0.89 ± 0.06
B	3.91 ± 0.57	2.44 ± 0.88	2.34 ± 0.33	2.76 ± 0.56	1.97 ± 0.46	1.93 ± 0.52	1.85 ± 0.55	2.22 ± 0.18
C	4.47 ± 0.57	2.84 ± 0.52	2.66 ± 0.35	3.21 ± 0.54	2.38 ± 0.35	2.34 ± 0.38	2.22 ± 0.42	2.61 ± 0.08

DIR induced dosimetric error

Without rescanning, pronounced differences in 4D dose distributions were observed among the different DIR scenarios, and in comparison to the GT generated 4D dose distributions (Fig. 3(a)). Indeed, the impact of interplay was clearly different when using different DVFs as input to the 4D dose calculation. For the GT, DIR2, DIR4, DIR5, and DIR6 the effects of interplay were more pronounced than for the other two methods (DIR1 and DIR3), in which the dose distributions look more homogeneous. In fact, for this particular 4D plan configuration, DVHs of the CTV and DDVHs of the extended CTV volume, obtained from the six DIR 4D plans, clearly differed from the GT DVH (Fig. 3(b.i) and (b.ii)), with all DIR plans underestimating the dose in-homogeneity due to the interplay effects in comparison to the GT plan. Additionally, substantial absolute V_{95}(CTV) differences (between the GT and all investigated DIR methods) were observed (Fig. 4). For motion A, and single fields with single scans, V_{95}(CTV) differences of 7.91 ± 3.46 % were observed for DIR3, and 2.02 ± 1.28 % for DIR2. Moderate motions in the fVOI (motion scenario B) and single fields without rescanning showed the greatest differences, with DIR3 and DIR2 having differences of 10.58 ± 14.08 % and 1.43 ± 1.37 % respectively. Of the six tested DIR methods, DIR2 achieved the lowest errors in V_{95}(CTV) for most situations with the best prediction of interplay effects in comparison to the GT.

Rescanning and/or multiple-field plans smoothed out the V_{95}(CTV) differences for all three motion scenarios. For instance, for motion scenario B, three-field plans with rescanning resulted in absolute V_{95}(CTV) differences to the GT of 3.46 ± 1.40 % and 0.23 ± 0.19 % for DIR3 and DIR2 respectively. For the smallest motion, negligible dosimetric errors were obtained for all DIR methods (0.37 ± 0.38 % for DIR3, 0.30 ± 0.17 % for DIR2, and 0.24 ± 0.18 % for DIR4). V_{95}(CTV) values for all 4D plan configurations can be found in Table S.4.2, showing that overestimation of target coverage is consistent for the single-field single/rescanned, or multiple-field single scan DIR generated plans.

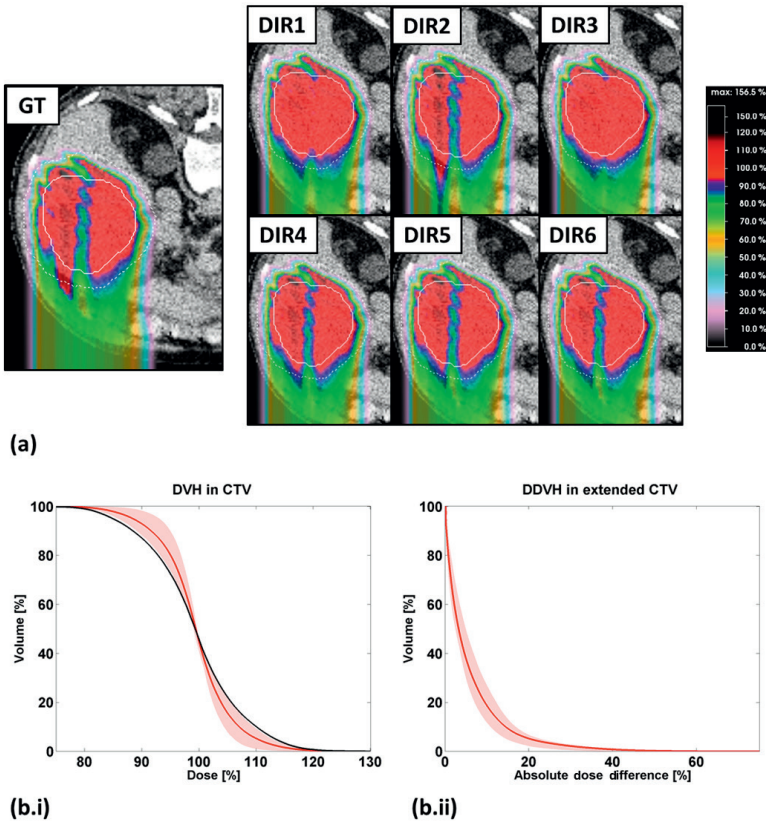
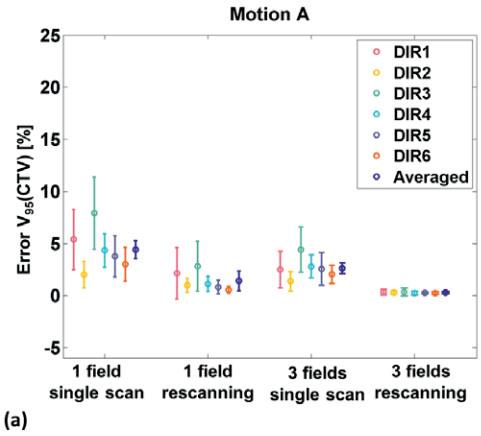
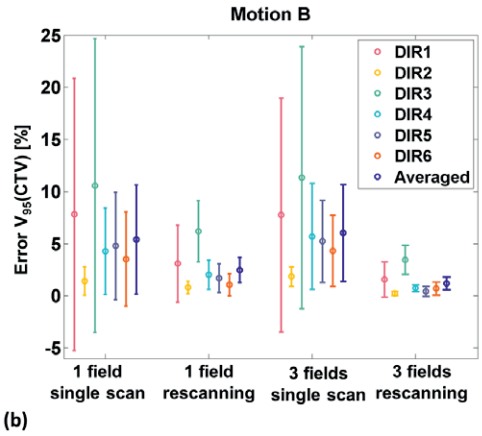


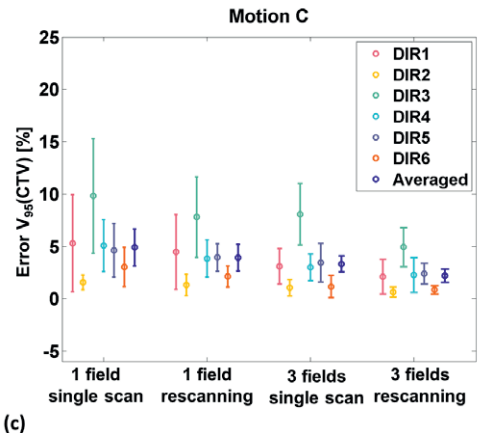
Fig. 3. 4D dose calculation results for the example 4D plan configuration of patient geometry PIII modulated by motion scenario C, and treated with the single anterior-posterior field F1 applied without any rescanning. This corresponds to the patient case with the largest tumour volume and moderate motion amplitude within the liver, but largest movement in the field direction within the selected fVOI. **(a)** 4D dose distributions using GT, DIR1, DIR2, DIR3, DIR4, DIR5, and DIR6 DVFs. The white normal and white dashed lines represent the CTV and CTV + 1cm (extended CTV) delineations respectively. **(b.i)** CTV DVH and **(b.ii)** extended CTV DDVH curves obtained with the six DIR methods. The black solid line gives the DVH curve calculated with the GT. The red shadow represents the band obtained by using different motion estimation methods, and the solid red line is showing the mean value from the six DIRs.



(a)



(b)



(c)

Fig. 4. Absolute differences between the V_{95}(CTV) of a particular DIR method and the GT V_{95}(CTV) for all analysed 4D plan configurations. These overall V_{95}(CTV) errors are calculated by the mean ± D of the individual errors given by all three patient geometries combined, with respect to modulated motion scenario **(a)** A, **(b)** B, or **(c)** C, single- or three-field plans, and single scan or five times layered rescanning deliveries.

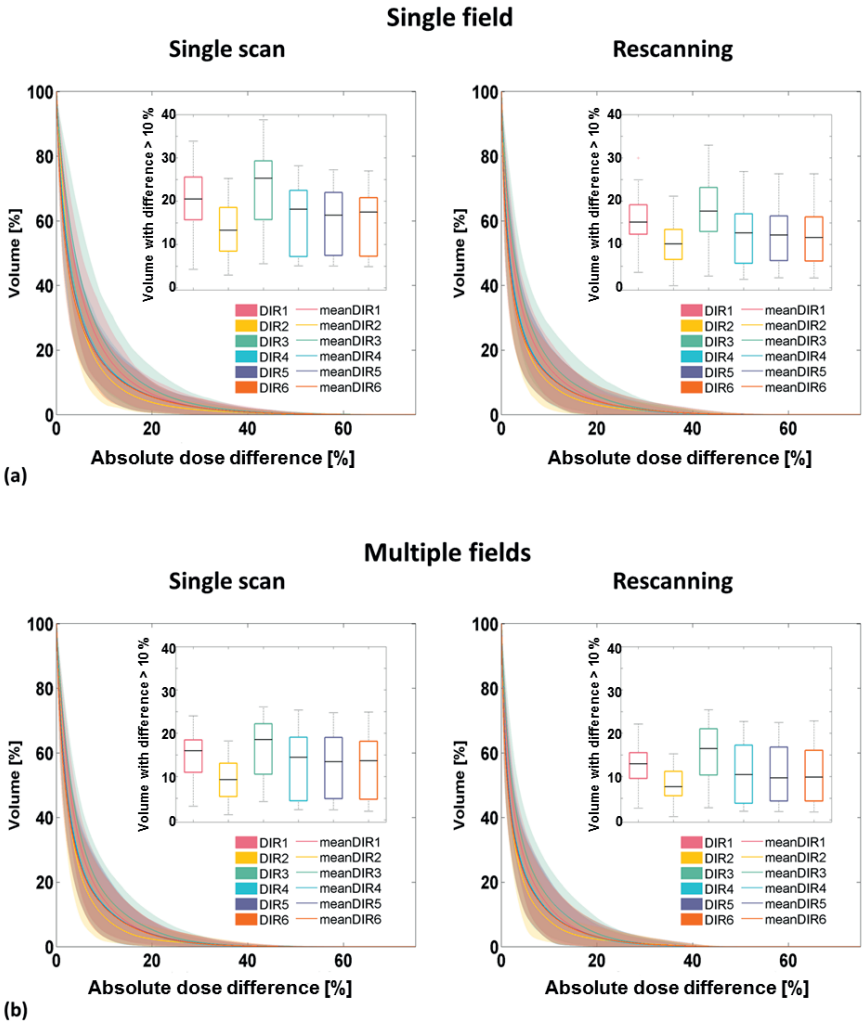


Fig. 5. DDVH band between investigated DIR and GT in the extended CTV (CTV + 1cm) for all **(a)** single- vs. **(b)** three-field 4D plan configurations. Each individual plot gives the information for all motion scenarios (A, B, and C). The boxplots give the statistics of the volume of extended CTV with absolute dose differences higher than 10 %, with respect to DIR method used as input for the 4D dose calculation.

DDVH bands in the extended CTV for all 4D plan configurations (as a function of DIR for each motion scenario (Fig. S.4.2.1 and Fig. S.4.2.2)) show that the accuracy of a particular DIR method does not depend just on the motion characteristics, or that any single DIR algorithm performed better for one particular motion scenario. Generally however, dose differences between DIR2 and the GT were the smallest for most of the 4D plan configurations and for all analysed motion scenarios. For the others, up to 40 % of the volume of extended CTV had absolute dose differences in comparison to the GT of more than 10 % for single-field plans delivered with a single scan (Fig. 5(a)). The largest dosimetric differences occurred for single-field plans delivered with a single scan for the largest motions (B and C). Dose differences using all DIRs could however be smoothed out when applying rescanning or adding fields to the treatment plan (Fig. 5(b)).

DISCUSSION

We have shown in this study that the application of different DIR methods to extract motion from 4D data sets can result in pronounced differences. Geometric differences of up to 1.05 ± 0.29 mm for the smallest motion amplitude (motion A) and 3.21 ± 0.54 mm for larger motions (motion C) have been observed. Without motion mitigation, associated dosimetric differences in target coverage (assessed as differences in V_{95}(CTV)) were as high as 7.91 ± 3.46 % for motion A and 11.34 ± 12.57 % for motion B. Moreover, up to 40 % of the extended CTV volumes have been predicted to have absolute dose differences in comparison to the GT dose of more than 10 %. However, differences in 4D dose distributions among different DIR scenarios, and compared to the GT, were significantly smoothed out in most cases when using multiple-field treatment plans and/or rescanning [26]. In addition, it was confirmed that larger motion amplitudes and deformations contribute to larger geometric registration errors and consequently larger dosimetric errors [6].

However, target coverage (quantified by the V_{95}(CTV)) was highly dependent on other factors, such as magnitude of the interplay effects, and no linear relationship between target coverage and motion amplitude could be found.

The 4D dose calculation algorithm selected for this project is the one developed at PSI. It obtains a 4D dose distribution by deforming the dose grid as a function of time, instead of performing multiple calculations of the dose on different 3DCT phases, which need to be subsequently warped back to a reference phase for dose accumulation [28–30]. Thus, it has the potential advantage of allowing for a high number of recalculations in an acceptable time-frame. Additionally, unlike the more standard 4D algorithm, the 4D dose calculation of PSI employs linear interpolation for the present motion between the phases of the 4D imaging and this approach has been shown to provide more accurate results [14,31,32].

Instead of focusing solely on the assessment of registration ambiguity as in the work performed by Zhang *et al.* [6], this study systematically investigates the performance of a variety of DIR algorithms, with respect to GT-DVFs, as part of a 4D dose calculation. Naturally, such GT-DVFs have also been extracted from 4DMRI using DIR and one could argue therefore, that DIR induced error is already present in these GT-DVFs. As such, these do not represent 'real' GT anatomical motion, but rather an estimation that will depend on the characteristics of the extraction method used. Indeed, we cannot deny that DIR induced error is present in the GTs. However, these are treated as GTs for the 4DCT-MRI data set itself (where the different DIRs are applied), and not for the 4DMRI, and so these errors turn out to have minimal impact on our conclusions. Additionally, due to the much higher contrast in abdomen MRI images, we believe the DIR errors for the GTs are rather limited. Therefore, for DIR error assessment applications, the GT-DVFs used here as reference provide dense image features for comparison, being a great advantage over the common approach that relies on the sparse distribution of identifiable landmarks [7,8]. Moreover, the recently published American Association

of Physicists in Medicine task group 132 report provided recommendations for clinical DIR quality assurance, such as a digital phantom [32].

Although several 4D treatment plan configurations using SFUD have been analysed, further research still needs to be performed analysing 4D intensity-modulated proton therapy (IMPT) plan configurations. In addition, statistics could be improved by setting different starting phases for the treatment delivery, and only one breathing cycle has been considered, which in the 4D dose calculation was repeated, cycle-to-cycle, over the duration of the full delivery time per field. Results may vary even more if true, variable breathing patterns were taken into account. As such, the influence of irregular breathing patterns, as provided by the 4DCT-MRI approach, will be exploited in future work.

Interestingly, the geometric and dosimetric accuracies provided by all tested DIR methods are not directly correlated. For most situations, Total Variation (DIR6) provided the lowest field-specific geometric errors and Morfeus (DIR2) the lowest dosimetric errors. However, it is important to remember the complex nature of 4D dose calculations, and so multiple other variables could have influenced this lack of correlation. Furthermore, DIR2 is the only algorithm that requires a contour of the liver to be delineated on all the registered image pairs, indicating that the addition of such anatomical information to the algorithm is a safe approach to improve accuracy. However, defining such regions of interest requires a considerable amount of manual work. Therefore, there should be careful consideration about which registration method to choose and whether the additional work is justified. As such, the choice of a slightly less accurate, but more time efficient, DIR method can have advantages as long as the uncertainties resulting from these approaches are understood. Indeed, we would recommend that, given the lack of accuracy demonstrated in this work between the different DIRs, it would make sense to provide error bars on dose calculations that depend on such algorithms. Such error bars could be generated by systematic analysis of dose

calculations using different DIR algorithms for different treatment sites, as the one performed on this paper, which is specific for PBS-PT 4D dose calculations from liver cases.

In summary, to understand the influence of a particular registration algorithm on the 4D dose calculation, the accuracy of different DIR methods to estimate the GT results has been analysed. Regarding field-specific geometric registration error, it has been shown that all DIR methods underestimate motion amplitude in the fVOI for all applied fields. This consequently resulted in an overestimation of the calculated plan index of V_{95}(CTV) for all DIR algorithms in comparison to the GT for most of the 4D plan configurations analysed. The performance of this study is of great importance for the proton therapy community in general, and particularly for PBS-PT, due to its sensitivity to respiratory-induced impacts. As well for passively scattered proton therapy, or even conventional radiotherapy, where DIR is widely used for dose distribution warping and accumulation, we believe that our outcomes, especially on the DIR induced geometric error assessment, still remain noteworthy [33,34].

CONCLUSION

The demonstrated dosimetric errors induced by different DIR methods indicate the necessity to interpret individual 4D dose distributions for PBS-PT plans for liver cases with caution, and ideally with an error bar. However, by adding fields to the treatment plan and/or using motion mitigation techniques such as rescanning, the impact of DIR motion estimation uncertainties on the 4D dose distributions could be reduced.

REFERENCES

[1] Bert C, Grözinger SO, Rietzel E. Quantification of interplay effects of scanned particle beams and moving targets. Phys Med Biol 2008;53:2253–65.

[2] Grassberger C, Dowdell S, Lomax A, Sharp G, Shackleford J, Choi N, et al. Motion Interplay as a Function of Patient Parameters and Spot Size in Spot Scanning Proton Therapy for Lung Cancer. Int J Radiat Oncol 2013;86:380–6.

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

[4] Noe KØ. Deformable Image Registration for Use in Radiotherapy. 2009.

[5] Jingu K. Use of deformable image registration for radiotherapy applications. J Radiol Radiat Ther 2014;2:1–7.

[6] Zhang Y, Boye D, Tanner C, Lomax AJ, Knopf A. Respiratory liver motion estimation and its effect on scanned proton beam therapy. Phys Med Biol 2012;57:1779–95.

[7] Kadoya N, Nakajima Y, Saito M, Miyabe Y, Kurooka M, Kito S, et al. Multi-institutional Validation Study of Commercially Available Deformable Image Registration Software for Thoracic Images. Int J Radiat Oncol Biol Phys 2016;96:422–31.

[8] Brock KK. Results of a Multi-Institution Deformable Registration Accuracy Study (MIDRAS). Int J Radiat Oncol Biol Phys 2010;76:583–96.

[9] Bender ET, Hardcastle N, Tomé WA. On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation. Med Phys 2011;39:272–80.

[10] Salguero FJ, Saleh-Sayah NK, Yan C, Siebers JV. Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping. Med Phys 2010;38:343–53.

[11] Murphy MJ, Salguero FJ, Siebers JV, Staub D, Vaman C. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. Med Phys 2012;39:573–80.

[12] Yeo UJ, Taylor ML, Supple JR, Smith RL, Dunn L, Kron T, et al. Is it sensible to "deform" dose? 3D experimental validation of dose-warping. Med Phys 2012;39:5065–72.

[13] Boye D, Lomax T, Knopf A. Mapping motion from 4D-MRI to 3D-CT for use in 4D dose calculations: A technical feasibility study. Med Phys 2013;40:61702.

[14] Boye D. Applications of 4D-MRI in proton therapy. 2016.

[15] Bernatowicz K, Peroni M, Perrin R, Weber DC, Lomax AJ. 4D

dose reconstruction for scanned proton therapy using liver 4DCT-MRI. Int J Radiat Oncol 2016;95:216–23.

[16] Zhang Y, Huth I, Wegner M, Weber DC, Lomax AJ. An evaluation of rescanning technique for liver tumour treatments using a commercial PBS proton therapy system. Radiother Oncol 2016;121:281–7.

[17] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. Med Phys 2015;42:40–53.

[18] Press release. RaySearch licenses groundbreaking technology from Princess Margaret Hospital 2011:7–8.

[19] Samavati N, Velec M, Brock K. A hybrid biomechanical intensity based deformable image registration of lung 4DCT. Phys Med Biol 2015;60:3359–73.

[20] Mirada Medical. Why choose Mirada Registration ? 2002. http://www.mirada-medical.com/.

[21] Lucas BD. Generalized Image Matching by the Method of Differences. 1984.

[22] Vishnevskiy V, Gass T, Szekely G, Tanner C, Goksel O. Isotropic Total Variation Regularization of Displacements in Parametric Image Registration. IEEE Trans Med Imaging 2017;36:385–95.

[23] Siddon RL. Fast Calculation of the Exact Radiological Path for a Three-Dimensional CT Array. Med Phys 1985;12:252–5.

[24] Lomax A. SFUD, IMPT, and Plan Robustness. Part. Radiother. Emerg. Technol. Treat. Cancer, 2016, p. 169–94.

[25] Zenklusen SM, Pedroni E, Meer D. A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning at the new Gantry 2 at PSI. Phys Med Biol 2010;55:5103–21.

[26] Knopf A-C, Hong TS, Lomax A. Scanned proton radiotherapy for mobile targets—the effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. Phys Med Biol 2011;56:7257–71.

[27] Bernatowicz K, Lomax AJ, Knopf A. Comparative study of layered and volumetric rescanning for different scanning speeds of proton beam in liver patients. Phys Med Biol 2013;58:7905–20.

[28] Rietzel E, Chen GTY, Choi NC, Willet CG. Four-dimensional image-based treatment planning: Target volume segmentation and dose calculation in the presence of respiratory motion. Int J Radiat Oncol Biol Phys 2005;61:1535–50.

[29] Lin L, Kang M, Huang S, Mayer R, Thomas A, Solberg TD, et al. Beam specific planning target volumes incorporating 4DCT for pencil beam scanning proton therapy of thoracic tumors. J Appl Clin Med Phys 2015;16:281–92.

[30] Inoue T, Widder J, van Dijk LV, Takegawa H, Koizumi M,

Takashina M, et al. Limited Impact of Setup and Range Uncertainties, Breathing Motion, and Interplay Effects in Robustly Optimized Intensity Modulated Proton Therapy for Stage III Non-small Cell Lung Cancer. Int J Radiat Oncol 2016;96:661–9.

[31] Zhang Y, Mueller F, Weber DC, Lomax AJ. TH-AB-605-4 Dosimetric Influence of Temporal Resolution Used by the PBS Proton 4D Dose Calculations. AAPM 59th Annu. Meet. Exhib., Denver, CO, USA: 2017.

[32] Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. Med Phys 2017;44:e43–76.

[33] Flampouri S, Jiang SB, Sharp GC, Wolfgang J, Patel AA, Choi NC. Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4D-CT data and Monte Carlo simulations. Phys Med Biol 2006;51:2763–79.

[34] Abe Y, Kadoya N, Arai K, Takayama Y, Kato T, Kimura K, et al. Effect of DIR uncertainty on prostate passive-scattering proton therapy dose accumulation. Phys Medica 2017;39:113–20.

SUPPLEMENTARY DATA

Suppl. 1. Deformable image registration (DIR) methods

The ANACONDA (Anatomically Constrained Deformation Algorithm) algorithm (DIR1) is a hybrid registration solution, in which the objective function includes both the intensity and geometric information of the images. The additional anatomical information needs to be provided by introducing contoured data sets. In RayStation, a region encompassing the liver was focused in order to achieve a higher accuracy in that targeted area. Morfeus (DIR2) is a set of algorithms for DIR, based on the biomechanical modelling of anatomical structures. The biomechanical behaviour of soft tissue deformation is included in this method in order to improve accuracy and extensibility of the motion estimation. Despite not being an intensity-based approach, for lung applications, DIR2 already proved that it is able to efficiently model the biomechanics of respiration through the breathing motion, the sliding interface between the lungs and the chest cavity, and the nonlinearity of heterogeneous lung soft tissues. For the application of DIR1 and DIR2, the liver of the three patient cases were manually delineated in the reference 3DCT using Mirada, and subsequently checked by an experienced physician. In DIR1, the manually countered liver was extended by 1 cm (excluding the rib cage), and then used as a focus region of interest (ROI) for the registration. In DIR2, a triangular surface mesh of liver for each of the nine 4DCT-MRI data sets was used as a controlling ROI for the registration. It is created by using DIR1 for warping the respective manually contoured livers to all the remaining phases of the first breathing cycle and latter correcting them.

In DIR3 (B-splines algorithm), multi-resolution optimization is performed, so that a gradual refinement of the image resolution along with a gradual increase in the number of free transformation parameters (control points) can be achieved. DIR4 was implemented as symmetric log domain diffeomorphic Demons, initially by an affine transformation to obtain a global alignment. Moreover, the deformation

field is smoothened by a Gaussian filter in order to avoid unrealistic deformation. Both DIR3 and DIR4 use CT image intensity itself as a similarity measurement. The mean square differences of intensity and the sum of squared differences, as an image similarity measure, were used respectively. The registration parameter settings for these two DIR algorithms have been extensively investigated in previous works for liver 4DCT registration.

Mirada CT Deformable (DIR5) is a multi-resolution highly optimised derivative approach of the method of differences for handling large deformations and CT image artefacts in high speed. The method of differences iteratively uses the image intensity gradient, together with the intensity differences between the images, to improve an initial estimate. DIR5 was applied by choosing a global approach (without any focus or controlling ROIs), and the setting super-fine was selected in order to obtain an optimal result.

ETH DIR (DIR6) uses an isotropic Total Variation regularization that is able to estimate displacement fields that occur on both sides of sliding interfaces in the thorax and abdomen in 4D parametric image registration. Since organ sliding motion (non-smooth displacement fields) are handled by this DIR method, no organ masks as input were used, and so a global approach was also selected for its application.

Table S.1 The six different DIR methods investigated in this paper, and the information regarding the addition of contoured structures in the images to the algorithm when applied.

DIR method	Algorithm	Additional geometry
DIR1	ANACONDA	Focus ROI
DIR2	Morfeus	Controlling ROIs
DIR3	B-splines	None
DIR4	Demons	None
DIR5	CT Deformable	None
DIR6	Total Variation	None

Suppl. 2. 4D planning configurations

Table S.2 Configurations and respective notations of the static treatment plans.

Static plans		
	Number of fields	
Patient name	1 field	3 fields
PI	PI_1F1 PI_1F2 PI_1F3	PI_3F
PII	PII_1F1 PII_1F2 PII_1F3	PII_3F
PIII	PIII_1F1 PIII_1F2 PIII_1F3	PIII_3F

An individual plan is denoted as P_F_S_M_E, where P represents the patient case (P ∈ {PI, PII, PIII}) and F stands for the number of fields (F ∈ {1FX, 3F}), where X represents the single-field orientation applied (X ∈ {1, 2, 3}). The remaining part of the plan denotation concern the 4D plans, where the scanning mode is given by S (S ∈ {single scan, five times layered rescanning}), M is the 4DMRI motion scenario (M ∈ {A, B, C}), and E stands for the used DVFs in the 4D dose calculation (E ∈ {GT, DIR1, DIR2, DIR3, DIR4, DIR5, DIR6}). The absolute dose difference between the plan obtained with the GT-DVFs (the GT 4D plan) and each one of the plans using the derived DVFs from the applied DIR methods (the DIR 4D plan) is denoted as P_F_S_M_DIR_diff, in which DIR ∈ {DIR1, DIR2, DIR3, DIR4, DIR5, DIR6}.

Suppl. 3. Field-specific geometric error

The field-specific registration error was defined as the Euclidean distance ($e_{h,p,f}$)for each time phase h, patient geometry p and single field f, between the investigated DIR estimated position ($tDIR_{h,p,f}$, $uDIR_{h,p,f}$) and the GT position ($tGT_{h,p,f}$, $uGT_{h,p,f}$), calculated by **(1)**. The geometric error of each of the six different DIR methods (error after DIR) was then quantified by *e*DIR. Conversely, the geometric error before the registration ($eGT_{h,p,f}$) was the magnitude of the GT-DVFs in the machine coordinate system, given by **(2)**.

$$eDIR_{h,p,f} = \sqrt{(tDIR_{h,p,f}^2 - tGT_{h,p,f}^2) + (uDIR_{h,p,f}^2 - uGT_{h,p,f}^2)},$$

$$eDIR_{p,f} = \{eDIR_{h,p,f}|_{h\in H}\}, \quad (1)$$

$$eGT_{h,p,f} = \sqrt{tGT_{h,p,f}^2 + uGT_{h,p,f}^2},$$

$$eGT_{p,f} = \{eGT_{h,p,f}|_{h\in H}\} \quad (2)$$

and *H* belongs to the set of the image phases corresponding to the first breathing cycle of the 4DCT-MRI.

Suppl. 4. Results

Suppl. 4.1. DIR induced geometric error

Motions in the gantry coordinate system were evaluated for one exemplary case (PIII, 1F1, motion scenario B). This choice for example configuration was made since larger motion amplitudes and deformations proved to be more difficult to predict with DIR, and so they tend to challenge the DIR performance more significantly. The overall statistics of the errors before and after the registration over all time phases ($eGT_{p,f}$ and $eDIR_{p,f}$ respectively) are shown in Fig. S.4.1 in form of boxplots (N = 822250). Several outliers (represented in black for GT and in colour for the six DIR methods) are present in the plots. The outliers of DIR1 proved the largest errors, whilst DIR2 the lowest among all

methods. The magnitude of estimated motion errors induced by DIR including outliers (with respect to the GT) were found to be as high as 32.59 mm (for DIR1), whilst no more than 19.67 mm for DIR2. The maximum and minimum median values (obtained with DIR2 and DIR6) were 1.41 mm and 0.58 mm respectively. DIR3 and DIR6 showed the largest and lowest mean values (3.71 mm and 2.40 mm respectively). The 25 %–75 % quartile for DIR (coloured boxes range) reached up to 2.82 mm for DIR6 and 6.01 mm for DIR3.

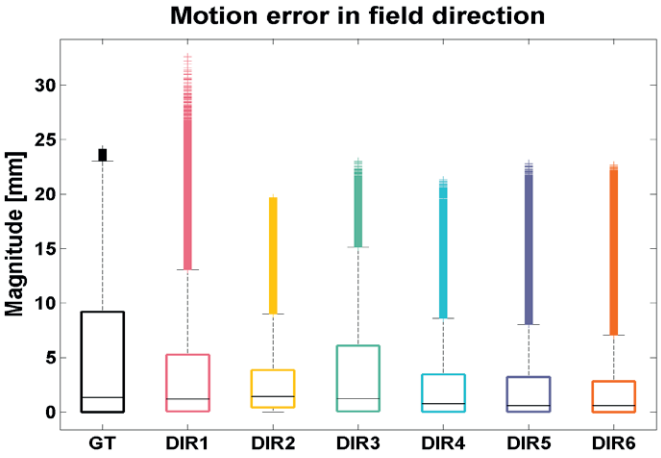


Fig. S.4.1. Motion error information for the GT and respective estimation from DIR in the gantry coordinate system over all phases of the first breathing cycle ($\boldsymbol{eGT}_{p,f}$ and $\boldsymbol{eDIR}_{p,f}$) for the exemplary configuration (PIII, 1F1, motion scenario B).

Table S.4.1 Field-specific motion amplitudes before and after DIR for the analysed configurations: three different single-field directions and three liver cancer patients, each modulated by motion scenarios A, B, or C.

	Before DIR (mean ± SD) [mm]	**After DIR** (mean ± SD) [mm]						
Motion	GT	DIR1	DIR2	DIR3	DIR4	DIR5	DIR6	Averaged
A	1.47 ± 0.26	0.56 ± 0.27	1.25 ± 0.27	0.53 ± 0.27	0.84 ± 0.29	0.89 ± 0.35	0.87 ± 0.37	0.82 ± 0.05
B	3.91 ± 0.57	2.06 ± 0.90	3.42 ± 0.58	1.47 ± 0.66	2.18 ± 0.65	2.22 ± 0.72	2.26 ± 0.78	2.27 ± 0.11
C	4.47 ± 0.57	2.23 ± 0.51	3.80 ± 0.53	1.60 ± 0.64	2.34 ± 0.54	2.37 ± 0.56	2.43 ± 0.65	2.46 ± 0.06

Suppl. 4.2. DIR induced dosimetric error

Table S.4.2 Obtained V_{95}(CTV) (by using the GT-DVFs or the derived DVFs from the six investigated DIRs in the 4D dose calculation) for all 4D plan configurations: three patient geometries modulated by three motion scenarios (A, B, or C), single- and three-field plans, and single scan and five times layered rescanning deliveries.

			V95(CTV) (mean ± SD) [%]							
Motion	**Number of fields**	**Scanning**	GT	DIR1	DIR2	DIR3	DIR4	DIR5	DIR6	Averaged
A	1 field	Single scan	81.11 ± 7.54	86.50 ± 7.82	83.13 ± 6.74	89.02 ± 5.48	85.45 ± 6.52	84.89 ± 6.09	84.10 ± 6.65	85.51 ± 0.77
		Rescanning	92.36 ± 3.99	94.51 ± 3.15	93.21 ± 3.37	95.18 ± 2.66	93.42 ± 3.55	93.13 ± 3.50	92.70 ± 3.74	93.69 ± 0.38
	3 fields	Single scan	90.30 ± 5.57	92.80 ± 6.54	91.67 ± 4.63	94.70 ± 3.83	93.10 ± 4.53	92.86 ± 4.06	92.35 ± 4.86	92.91 ± 0.96
		Rescanning	97.24 ± 1.53	97.31 ± 1.55	97.23 ± 1.13	97.44 ± 1.50	97.28 ± 1.51	97.09 ± 1.58	96.98 ± 1.51	97.22 ± 0.17
B	1 field	Single scan	67.02 ± 20.30	74.31 ± 10.89	67.07 ± 19.34	77.31 ± 10.40	70.63 ± 16.40	70.47 ± 15.35	69.49 ± 16.17	71.55 ± 3.46
		Rescanning	80.97 ± 6.31	84.06 ± 7.14	81.35 ± 6.15	87.17 ± 6.69	82.98 ± 6.31	82.66 ± 6.45	81.61 ± 6.61	83.31 ± 0.35
	3 fields	Single scan	77.50 ± 9.88	85.17 ± 6.70	77.61 ± 8.22	88.85 ± 7.00	81.97 ± 4.07	81.24 ± 4.27	80.19 ± 5.01	82.51 ± 1.67
		Rescanning	91.27 ± 5.24	92.84 ± 6.26	91.33 ± 4.92	94.73 ± 4.45	92.01 ± 5.27	91.69 ± 5.47	91.04 ± 6.25	92.27 ± 0.72
C	1 field	Single scan	66.35 ± 7.49	70.97 ± 8.32	67.63 ± 7.67	76.18 ± 10.55	71.42 ± 7.30	70.98 ± 8.02	69.09 ± 8.05	71.04 ± 1.15
		Rescanning	74.86 ± 7.43	79.34 ± 9.96	75.76 ± 7.48	82.67 ± 10.92	78.62 ± 7.51	78.82 ± 8.30	77.00 ± 7.90	78.70 ± 1.43
	3 fields	Single scan	75.13 ± 6.73	77.46 ± 9.46	75.92 ± 7.33	83.21 ± 9.60	78.15 ± 7.85	78.59 ± 8.56	76.24 ± 7.89	78.26 ± 0.92
		Rescanning	84.22 ± 8.61	86.33 ± 9.48	84.07 ± 7.83	89.17 ± 9.34	86.20 ± 7.23	86.62 ± 7.75	84.73 ± 8.02	86.19 ± 0.92

Abbreviations: V_{95}(CTV) = volume of CTV receiving at least 95 % of the prescribed dose.

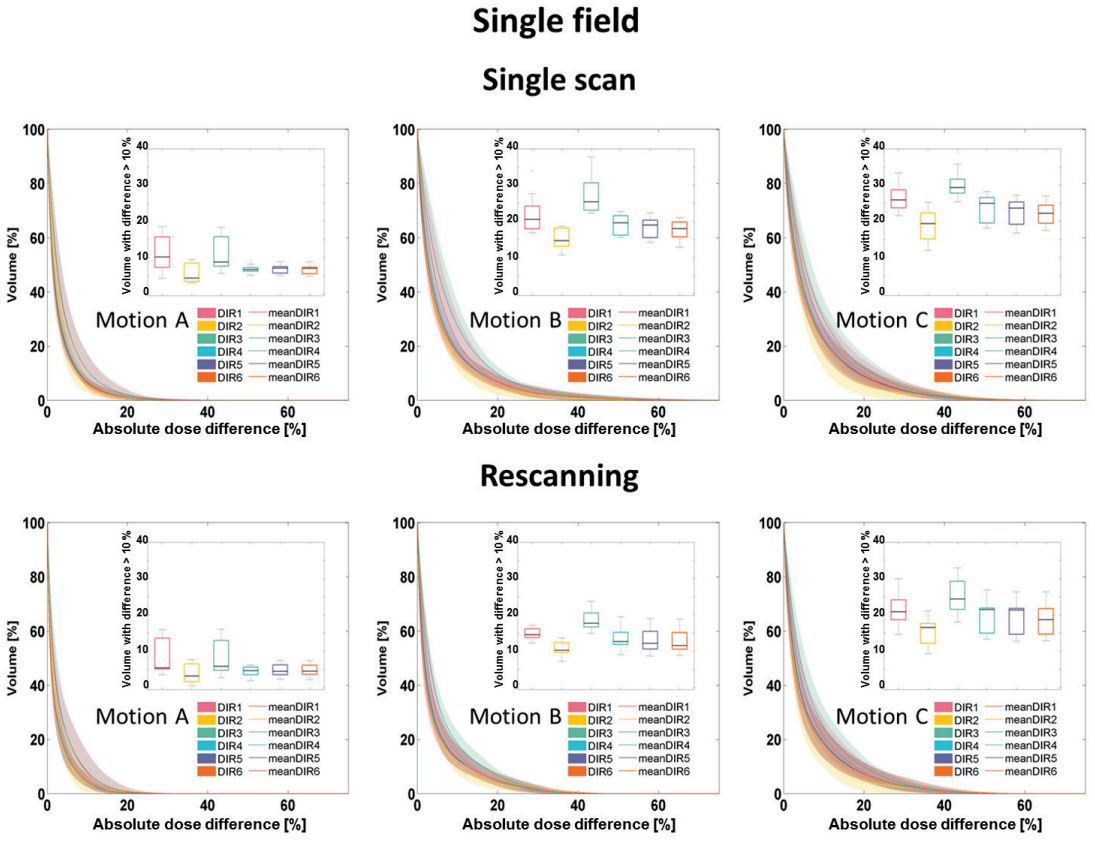


Fig. S.4.2.1. DDVH band between investigated DIR and GT in the extended CTV (CTV + 1cm) for all single-field 4D plan configurations. Each individual plot gives the information for one motion scenario (A, B, or C). The boxplots give the statistics of the volume of extended CTV with absolute dose differences higher than 10 %, with respect to DIR method used as input for the 4D dose calculation.

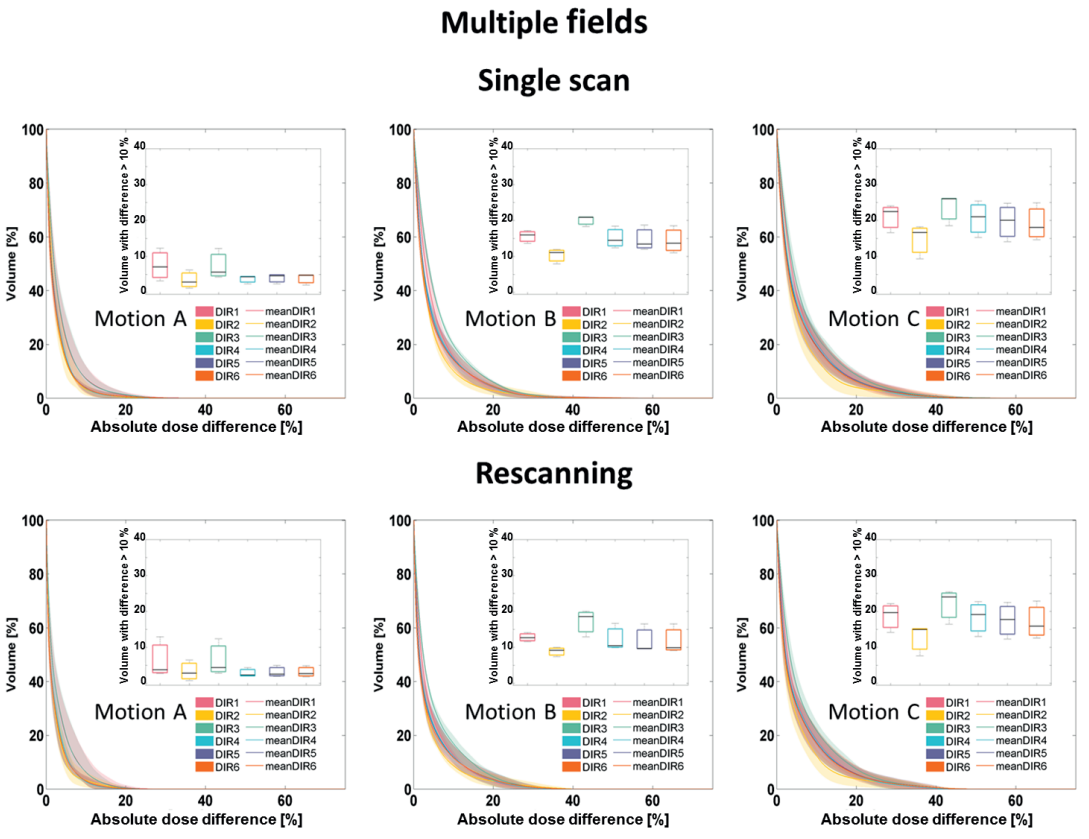


Fig. S.4.2.2. DDVH band between investigated DIR and GT in the extended CTV (CTV + 1cm) for all three-field 4D plan configurations. Each individual plot gives the information for one motion scenario (A, B, or C). The boxplots give the statistics of the volume of extended CTV with absolute dose differences higher than 10 %, with respect to DIR method used as input for the 4D dose calculation.