



HAL
open science

Quantitative investigation of dose accumulation errors from intra-fraction motion in MRgRT for prostate cancer

Lando Bosma, Cornel Zachiu, Mario Ries, Baudouin Denis de Senneville, Bas
Raaymakers

► **To cite this version:**

Lando Bosma, Cornel Zachiu, Mario Ries, Baudouin Denis de Senneville, Bas Raaymakers. Quantitative investigation of dose accumulation errors from intra-fraction motion in MRgRT for prostate cancer. *Physics in Medicine and Biology*, 2021, 66 (6), pp.065002. 10.1088/1361-6560/abe02a . hal-03453543

HAL Id: hal-03453543

<https://hal.science/hal-03453543>

Submitted on 28 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantitative investigation of dose accumulation errors from intra-fraction motion in MRgRT for prostate cancer

Lando Bosma¹, Cornel Zachiu¹, Mario Ries², Baudouin Denis de Senneville^{1,3}, Bas Raaymakers¹

¹ Department of Radiotherapy, UMC Utrecht, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands

² Imaging Division, UMC Utrecht, Heidelberglaan 100, 3508 GA Utrecht, The Netherlands

³ Institut de Mathématiques de Bordeaux (IMB), UMR 5251 CNRS/University of Bordeaux, F-33400 Talence, France

E-mail: lbosma@umcutrecht.nl

October 2020

Abstract. Accurate spatial dose delivery in radiotherapy is frequently complicated due to changes in the patient's internal anatomy during and in-between therapy segments. The recent introduction of hybrid MRI radiotherapy systems allows unequaled soft-tissue visualization during radiation delivery and can be used for dose reconstruction to quantify the impact of motion. To this end, knowledge of anatomical deformations obtained from continuous monitoring during treatment has to be combined with information on the spatio-temporal dose delivery to perform motion-compensated dose accumulation (MCDA).

Here, the influence of the choice of deformable image registration algorithm, dose warping strategy, and MR image resolution and SNR on the resulting MCDA is investigated. For a quantitative investigation, four 4D MRI-datasets representing typical patient observed motion patterns are generated using finite element modeling and serve as a gold standard. Energy delivery is simulated intra-fractionally in the deformed image space and, subsequently, MCDA-processed. Finally, the results are substantiated by comparing MCDA strategies on clinically acquired patient data.

It is shown that MCDA is needed for correct quantitative dose reconstruction. For prostate treatments, using the energy per mass transfer dose warping strategy has the largest influence on decreasing dose estimation errors.

Keywords: Motion-compensated dose accumulation, intra-fraction plan adaption, MR-guided radiotherapy, dose reconstruction

Submitted to: *Phys. Med. Biol.*

1. Introduction

Radiotherapy for cancer treatment aims to deliver a high dose to the tumor while keeping the dose to the surrounding healthy tissue as low as possible. Biological motion (*e.g.* respiration, bladder and rectal filling, peristalsis) during treatment might result in geometric uncertainties and, thus, dose delivery errors to the tumor. Margins around the tumor are typically used to optimize dose coverage (De Luca et al. 2010). However, such margins generally lead to over-exposing healthy tissue or organs-at-risk (OAR) situated within the margins and are often dose-limiting.

Image-guided radiotherapy offers the possibility to reduce these uncertainties, allowing a reduction of these margins and alleviating the dose limitations, potentially improving treatment outcomes (Jaffray 2012, Mundt & Roeske 2010). For example, on-board imaging such as cone-beam computed tomography (CBCT) can be used to compare the daily treatment anatomy to the anatomy at the time of planning, allowing the correction of translations and rotations of the anatomy on an inter-fraction basis (Guckenberger 2011, Thilmann et al. 2006). CBCT-guidance can thus be used to decrease geometric uncertainties to a particular extent; however, the poor soft tissue contrast renders the estimation of more complex deformations of individual anatomical areas challenging (McBain et al. 2006). Additionally, CBCT-guidance also leads to the delivery of imaging-related radiation, which is undesirable. Recently, magnetic resonance (MR) imaging has also been integrated with radiotherapy delivery systems (Lagendijk et al. 2014, Mutic & Dempsey 2014), allowing MR guided radiotherapy (MRgRT). MRgRT offers high soft-tissue contrast suitable for tumor tracking at high frame-rates without additional radiation exposure to the patient. This enables direct tumor visualization and OAR localization from the treatment table, not only prior to the daily treatment but also during radiation delivery. An example of dose reconstruction for prostate treatments at the 1.5T MRI radiotherapy system is presented by Kontaxis et al. (2020), where the information on the anatomical deformations is coupled with knowledge of the delivered dose under the assumption of correct image registration of the different time points. Thus, having knowledge of the delivered dose can aid more precise reporting for treatment response assessment. Additionally, the time-resolved dose reconstruction can also be used as input to adapt the treatment, *i.e.* for intra-fraction plan adaption.

In particular, for radiotherapy for prostate cancer, several therapy innovations have been suggested that rely on an accurate and precise motion-compensated dose accumulation (MCDA), or motion-compensated dose reconstruction, as an input. Hypofractionation has recently gained traction as it combines positive therapy results with cost benefits and decreased patient burden (den Hartogh et al. 2019, Widmark et al. 2019). However, when increasing the dose-per-fraction while decreasing the number of fractions, there is less to none averaging out of the dose error over fractions, making precise dose delivery and repeated adaptation even more of an imperative. Furthermore, since fractions in a hypofractionated plan are generally of

longer duration, there is a higher chance for (more extensive) biological motion to occur (de Muinck Keizer, Pathmanathan, Andreychenko, Kerkmeijer, van Zyp, Tree, van den Berg & de Boer 2019, de Muinck Keizer, Kerkmeijer, Maspero, Andreychenko, van Zyp, Van den Berg, Raaymakers, Lagendijk & de Boer 2019). **Note that the possibility of large prostate motion is somewhat dependent on the treatment method used as for modern non-MR-hybrid linacs, the duration of a standard fraction can be as short as 2 minutes, decreasing the possibility of large intra-fraction prostate motion.**

Moreover, due to the relatively slow and episodic **component of the** changes in the anatomy due to bladder filling and peristalsis, intra-fraction dose reconstruction and plan adaptation are conceptually feasible for prostate MRgRT, **as the temporal MR-resolution is about one image per 10 to 20 seconds.** For example, if the reconstructed dose deviates from the planned dose distribution, the treatment might be interrupted and continued with an adapted treatment plan, which accounts for the dose delivered so far. This ensures convergence of the delivered dose to the planned dose, irrespective of motion (Kontaxis et al. 2015b, Kontaxis et al. 2015a, Kontaxis et al. 2017), thus increasing the quality of the treatment.

Since MCDA can potentially serve as an input for such motion correction strategies, its reliability, precision, and accuracy become paramount, as any errors may end-up accumulating over time. Moreover, such errors could even amplify due to feed-back into the (re-)planning algorithm, up to a point where the (erroneous) corrections become detrimental for therapy. The quality of MCDA (*i.e.* accuracy and precision) depends on multiple factors such as (a) the deformable image registration (DIR) algorithm, (b) the image resolution, (c) the signal-to-noise-ratio (SNR), and (d) the employed dose warping strategy.

Here we investigated the impact of these factors on the accumulated dose of MCDA for intra-fraction MRgRT for prostate cancer. To this end, a quantitative comparison of the different MCDA strategies is performed on finite-element simulated motion patterns representing ‘typical’ motion patterns observed in patients. **A quantitative comparison is then possible as the ground truth deformations are known, and the different MCDA strategies can be compared to a gold standard.** Subsequently, all MCDA strategies are qualitatively evaluated on a single patient clinical dataset. The aim of this study is to give an overview of the impact of the most relevant methodological choices on the quality of the resulting MCDA and to provide guidance for making well-considered choices for clinical implementations.

2. Materials and methods

In the following section, the experimental setup used for analyzing the impact of the choice of registration algorithm, image resolution and SNR, and dose warping method on the dose accumulation process will be described. Four finite element models (FEM) of different typical prostate motion patterns are simulated. From these, a set of cine-MR images depicting the motion trace is generated for a quantitative assessment of the

DIR algorithms. A single radiotherapy fraction is simulated, using a typical clinical treatment plan and incorporating the simulated motion patterns.

2.1. The data

As a basis for the FEM-generated images used here for the MRgRT simulation, we have used a clinical 3D MR scan acquired on a prostate cancer patient treated on the 1.5T MR-Linac Unity system (Elekta Unity AB, Stockholm, Sweden) installed at the UMC Utrecht, The Netherlands. **The current study was conducted in agreement with the required standards and regulatory requirements. Ethical approval was provided by the Ethics Board of the University Medical Center Utrecht.** The original spatial dimensions of the image are $448 \times 448 \times 63$, with a reconstructed voxel spacing of $0.93 \times 0.93 \times 2 \text{ mm}^3$. The image was acquired using 3D bTFE with fat suppression (relaxation time $T_R = 4.7 \text{ ms}$, echo time $T_E = 2.3 \text{ ms}$, flip angle = 50° , field strength $B_0 = 1.5 \text{ T}$). The delineations of the bladder, prostate, and rectum performed on the image by an experienced radiation oncologist were then imported into the FEM software FEBio (Maas et al. 2012). The prostate was modeled as an isotropic elastic material with a density $\rho = 1000 \text{ g/L}$, Young's modulus $E = 21 \text{ kPa}$ (Barr et al. 2012) and Poisson's ratio $\nu = 0.4$ (Bharatha et al. 2001). The rectum and bladder were modeled as biphasic materials, allowing them to expand. Volumetric changes and movements of the bladder and rectum were subsequently induced and used as actuators within the FEM simulation, which in turn led to movements and deformations of the surrounding anatomy.

We modeled four different anatomical deformations, mimicking typical patient observed motion patterns, which we have outlined below (from largest to smallest average prostate motion amplitude).

- **Rectal filling motion.** The rectum expands and moves in the superior direction, modeling the passage of a gas-bubble. Under this effect, the prostate moves in the anterior-superior direction, following an individual patient motion pattern observed and reported within a previous study (de Muinck Keizer, Kerkmeijer, Maspero, Andreychenko, van Zyp, Van den Berg, Raaymakers, Lagendijk & de Boer 2019).
- **Bladder filling motion.** The bladder expands, pushing the prostate in the inferior-posterior direction. This motion is based on individual patient-motion, as observed during 6 minutes in (de Muinck Keizer et al. 2020).
- **Average prostate motion.** The bladder expands, pushing the prostate in the inferior-posterior direction. This motion is modeled after the population-average motion of patients during 10 minutes, as reported in (de Muinck Keizer, Pathmanathan, Andreychenko, Kerkmeijer, van Zyp, Tree, van den Berg & de Boer 2019) and (de Muinck Keizer, Kerkmeijer, Maspero, Andreychenko, van Zyp, Van den Berg, Raaymakers, Lagendijk & de Boer 2019).
- **Residual motion only.** The most common anatomical motion in the abdomen during treatment is the absence of large scale deformations or displacements, with

only residual motion due to the digestive tract’s peristalsis. In this simulation, the bladder expands only slightly, pushing the prostate marginally in the inferior direction. This dataset is added mostly as a control in order to investigate the effect of MCDA on dose accumulation errors when minimal motion is present.

The resulting average displacements of the prostate for all motions are shown in figure 1 and the resulting anatomical deformations for the sixth time point of the simulation for the **rectal filling** motion are shown in figure 2. The simulated motion was extracted in

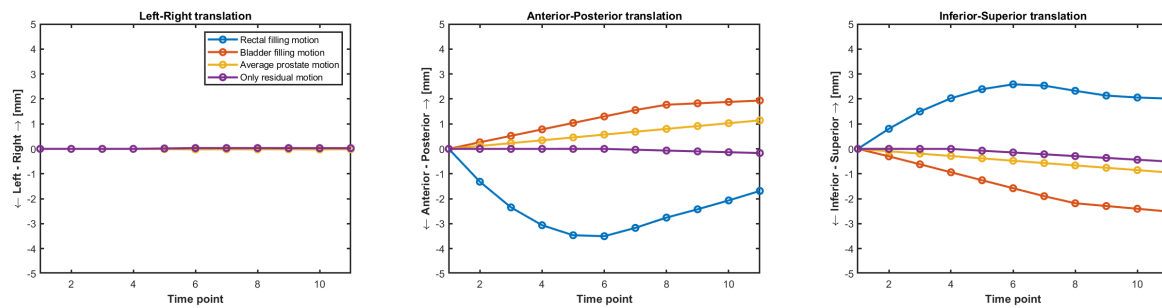


Figure 1: Average displacements of the prostate in millimeters for the four different motion patterns exported from the finite element models. **The motion-patterns are modelled after motions observed in patients during 6 to 10 minutes.**

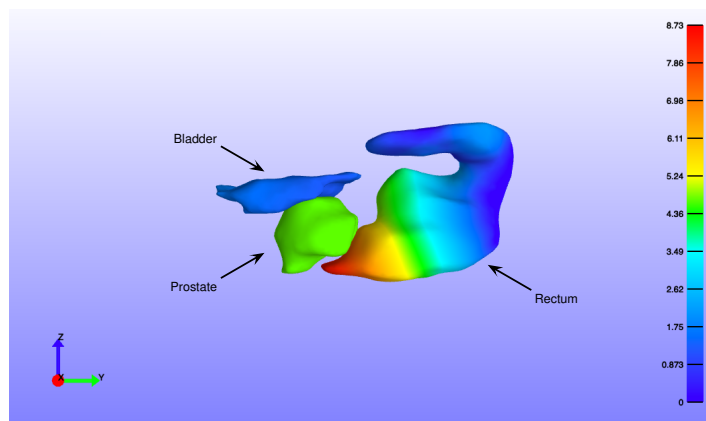


Figure 2: The modeled anatomical deformation for the sixth time point of the **rectal filling** motion simulation. The total displacement is color-coded, with the scale matching 0.93 mm/voxel and running up to 8.73 voxels.

ten equally distanced time-steps, resulting in ten 3D deformation vector fields (DVF), which relate each voxel’s position in 3D space to their original position. This set of vector fields serves as the gold standard for all DIR algorithms investigated in the scope of this study. **The 11 time-steps correspond to motions observed during 6 to 10 minutes.**

The FEM-generated deformations were then used to warp the aforementioned base 3D MR scan, resulting in a series of eleven MR images (the original scan was also

included), which depict the simulated anatomical changes under the influence of the simulated motion. Due to hardware limitations, the MR images and gold standard vector fields were re-sampled onto an image matrix of size $256 \times 256 \times 128$ (corresponding to a resolution of $1.63 \times 1.63 \times 0.98 \text{ mm}^3$).

In addition to the synthetic MR sequence, a clinical cine-MR series acquired on the MR-Linac consisting of sixteen images with a temporal resolution of 16.9 seconds acquired during beam-on time (patient in the supine position, same acquisition parameters as above) was also submitted to the analysis.

2.2. Simulating a radiation therapy fraction

For the systematic dose error analysis, we simulated the delivery of a single fraction for a prostate tumor. We established a treatment plan using MatRad (Wieser et al. 2017) with clinical constraints that are frequently used at the Radiotherapy department of the UMC Utrecht for prostate cancer therapy (8 Gy to the prostate, maximally 4 Gy to the rectum and bladder, **using step-and-shoot IMRT as is clinically performed on the MR-Linac with** 7 photon beams at 0° , 50° , 100° , 150° , 210° , 260° and 310° angles). The simulated treatment time is 10 minutes and 14 seconds. The electron density and Hounsfield units were extracted from the clinical planning CT-image of the patient, pre-registered to the first MR-image in the cine series. Subsequently, both electron density and Hounsfield units were warped using the FEM-generated DVFs, which in turn allowed the calculation of the absorbed radiation dose for each of the deformed anatomical states. To simulate the aggregate dose absorption under the influence of motion, MatRad was modified to allow the simulation of sequential dose absorption for all different gantry and MLC positions with an updated anatomical situation and to store the corresponding dose fragments separately. In summary, this strategy provided for each of the simulated time points a complete set of (I) anatomical MRIs, (II) electron density and Hounsfield units, and (III) the resulting dose absorption maps for each delivery configuration of the Linac in the known frame-of-reference of the deformed anatomy. Finally, both the gold standard DVFs and those estimated by the DIR algorithms were used to accumulate the dose **onto the original anatomy**, as detailed in section 2.3.3.

2.3. Considered factors influencing motion-compensated dose accumulation

2.3.1. Deformable image registration algorithms

In the scope of this study, we have selected five different variational DIR algorithms previously proposed in the context of MRgRT (Zachiu, Papadakis, Ries, Moonen & Denis de Senneville 2015, Zachiu, Denis de Senneville, Willigenburg, de Boer, Raaymakers, Ries et al. 2020). A variational registration algorithm typically implies the minimization of a two-part cost function: $\mathcal{E} = \mathcal{D} + \alpha\mathcal{R}$. The data fidelity part \mathcal{D} is a function quantifying the dissimilarity between the images. Since the minimization of the data fidelity term alone is usually an under-determined problem, a regularization term \mathcal{R} is added to ensure well-posedness.

This regularization represents an assumption on the pattern of estimated deformations, which essentially reduces the dimensionality of the solution space. The regularization parameter α determines the relative weight of the regularization. One of the most frequently used regularization terms for DIR algorithms constrains the estimated deformations to be spatially smooth, and is defined as

$$\mathcal{R}_s = \|\vec{\nabla} \mathbf{u}(\vec{r})\|_2^2, \quad (1)$$

where $\mathbf{u} = (u_1, u_2, u_3)$ is the 3D DVF and $\|\cdot\|_2$ is the Euclidean norm. Aside from smoothness, previous studies have also employed incompressibility as a constraint on the estimated deformations (Yang et al. 2000, Haber & Modersitzki 2004, Haber & Modersitzki 2007, Zachiu et al. 2018). This can be applied when, due to their high water content, biological soft tissues (*e.g.* liver, kidney, prostate etc.) are not expected to undergo volumetric changes at the time scale of a radiotherapy fraction. The corresponding mathematical formulation of this incompressibility regularization is

$$\mathcal{R}_i = \|J(\vec{r} + \mathbf{u}(\vec{r})) - 1\|_2^2, \quad (2)$$

where we have used the Jacobian determinant of the deformation:

$$J(\vec{r} + \mathbf{u}(\vec{r})) = \begin{vmatrix} 1 + \frac{\partial u_1(\vec{r})}{\partial x} & \frac{\partial u_1(\vec{r})}{\partial y} & \frac{\partial u_1(\vec{r})}{\partial z} \\ \frac{\partial u_2(\vec{r})}{\partial x} & 1 + \frac{\partial u_2(\vec{r})}{\partial y} & \frac{\partial u_2(\vec{r})}{\partial z} \\ \frac{\partial u_3(\vec{r})}{\partial x} & \frac{\partial u_3(\vec{r})}{\partial y} & 1 + \frac{\partial u_3(\vec{r})}{\partial z} \end{vmatrix}. \quad (3)$$

The five DIR algorithms used in this project are listed below. Their specifics are detailed in the Appendix A.

- The original optical flow (OF22), introduced in the seminal paper by Horn & Schunck (1981). Assumes that moving voxels conserve their brightness, combined with a smooth regularisation. The implementation was adapted from (Zachiu, Denis de Senneville, Moonen & Ries 2015a)‡.
- Improved optical flow (OF21), introduced in (Zachiu, Papadakis, Ries, Moonen & Denis de Senneville 2015). Substitutes the L_2 -norm of the original optical flow by an L_1 -norm, making the algorithms less sensitive to local grey-level intensity variations.
- EVOLution (EVO), introduced by Denis de Senneville et al. (2016). This algorithm aims to align similar contrast patterns by matching image gradients. Also uses a smooth regularisation.
- Incompressible EVOLution (EVI), introduced in (Zachiu et al. 2018). Using the same data-fidelity term as EVO, combined with an incompressibility-regularisation.
- Spatially adaptive EVOLution (AEVO), introduced by Zachiu, Denis de Senneville, Willigenburg, de Boer, Raaymakers, Ries et al. (2020). Uses a spatially-adaptive

‡ See <http://bsenneville.free.fr/RealTITracker/>

combination of both regularisations, allowing incompressibility in certain regions (e.g. bones and the prostate) and smoothness in others, combined with the same data-fidelity term as EVO.

In this study, an exhaustive search in the parameter space of the regularization parameter(s) was employed, minimizing the endpoint error (further discussed in section 2.4) between the estimated and gold standard deformations. The results are shown in section 3.1.

2.3.2. Resolution & signal-to-noise-ratio of the MR-images MRI acquisitions typically imply a compromise between spatial resolution, temporal resolution and SNR. Since both accuracy and precision of image registration algorithms depend on both the spatial resolution and SNR of the underlying images, we evaluate the impact of both factors on the MCDA results.

To assess the influence of the spatial resolution, the synthetic cine-MR data were downsampled with an isotropic factor of 2, 3, and 4. A sagittal slice for the four resolutions is shown in figure 3. After performing the registration for each downsampling

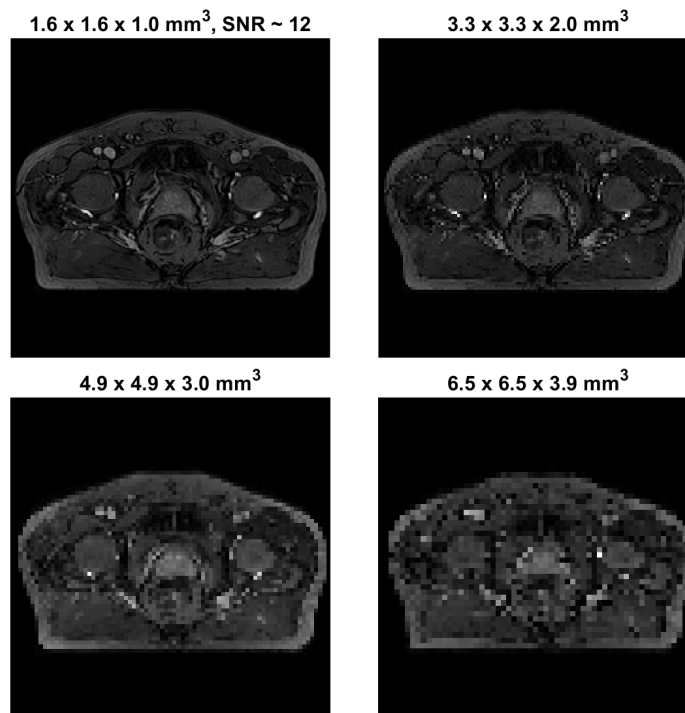


Figure 3: A sagittal slice of the reference MR-image for the four evaluated resolutions. The top-left image displays the original resolution.

factor, the estimated DVF is then re-sampled to the original resolution ($256 \times 256 \times 128$) and compared to the gold standard.

To vary the SNR of the images, increasing levels of Rician noise were added to the images (Henkelman 1985, Gudbjartsson & Patz 1995, Rice 1944). The resulting images

have SNR-values of approximately (12, 9, 6, 4, 3) or PSNR-values of (30, 26, 22, 19, 17). A **sagittal** slice for the resulting images is shown in figure 4.

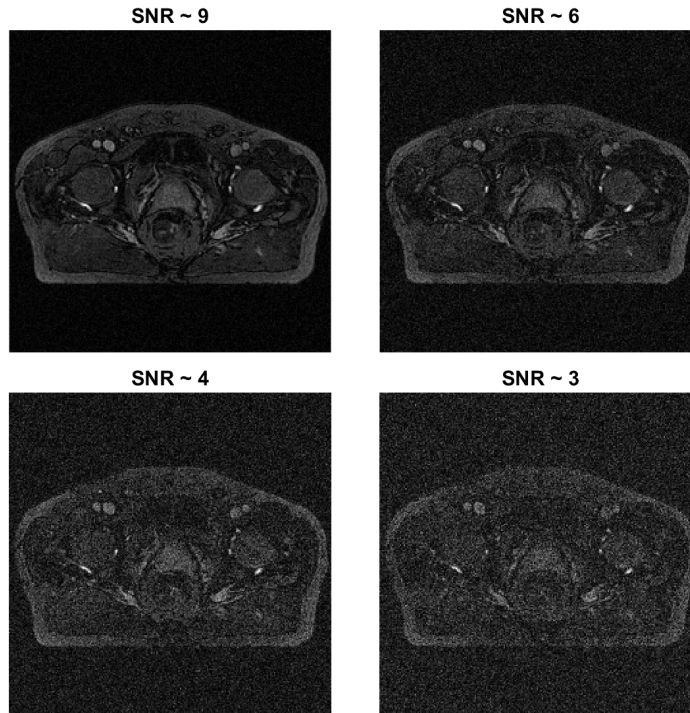


Figure 4: A **sagittal** slice of the reference MR-image for the different levels of SNR (excluding the original SNR already shown in figure 3).

2.3.3. Dose warping strategies For a motion-compensated dose accumulation, the delivered dose for all moving images is warped to the reference image using their respective DVF. In this study, two different strategies are used for this warping: direct dose mapping (DDM) and energy per mass transfer (EMT), as detailed in (Li et al. 2013). Due to the deformable character of the motion, tissues from multiple voxels in the moving anatomy might end up, for example, in a single reference voxel. Using DDM in this situation, the reference voxel dose is found by averaging the doses from the moving voxels. EMT on the other hand, is motivated by the physical definition of dose as the energy absorbed by a mass. This method warps both the energy and mass from the moving voxels and then divides the energy by the mass to get the reference dose. The EMT method thus requires a mass-map for the anatomy, which we created from the CT-image, differentiating between air (a density of $\rho = 1.225$ g/L was assumed) and soft tissue (density $\rho = 1000$ g/L).

2.4. Error measures

To assess the quality of an estimated DVF, it is compared with the gold standard DVF in terms of absolute, component-wise differences, or the endpoint error (Baker et al. 2011):

$$\Delta DVF(\mathbf{x}, t) = \|\mathbf{DVF}_{gs}(\mathbf{x}, t) - \mathbf{DVF}_{est}(\mathbf{x}, t)\|_2, \quad (4)$$

i.e. the Euclidean distance between the 3D deformation vectors of the gold standard and the estimated vector field. This results in a scalar for each voxel, providing a distribution of errors. In a clinical setting, however, the gold standard DVF is not available. Therefore, an alternative surrogate quality assurance criterion is required for evaluations of the clinical data included in this study. Evaluated on our simulated datasets, we found the structural similarity index (SSI) (Wang et al. 2004) and image intensity differences (IID) of the MR-data inside a contour or set of contours to have a Pearson correlation coefficient with the mean endpoint error of 0.90 and 0.95, respectively. This value is much higher than found for the Dice similarity coefficient (Dice 1945, Sorensen 1948) (0.79) or Hausdorff distance (Hausdorff 1914) (0.57). Therefore, the former two criteria will be used to evaluate the registration performances on the clinical dataset.

Additionally, we use the Jacobian determinant in incompressible regions to investigate the physiological plausibility of a DVF, as suggested in (Zachiu et al. 2018, Zachiu, Denis de Senneville, Raaymakers & Ries 2020). For the intra-fraction investigation, only minor volume variations of the prostate are expected (and therefore modeled), such that large deviations of the Jacobian determinant from one are an indication of implausible deformations.

To assess the performance of MCDA, the accumulated dose using an estimated DVF and the dose warping method of choice (*i.e.* DDM or EMT) is compared to the accumulated dose resulting from the gold standard DVF in combination with the EMT dose warping method. This choice is motivated by the physical definition of dose (*i.e.* amount of absorbed radiation energy per unit mass). The relative difference in grays is what we will refer to as the dose (accumulation) error. **As for the endpoint error, this gives a value for each voxel, providing a distribution of errors.**

3. Results

3.1. Deformable image registration algorithm configuration

The approach used to investigate and optimize the regularization parameters for the five DIR algorithms is described in section 2.3.1. The optimal values for the regularization parameters used in this study are reported in table 1. Of the four single-parameter algorithms, EVO and EVI are most robust with respect to a choice of α , while OF22 is the most sensitive. Furthermore, OF22 has its optimum with respect to the endpoint error at the lowest analyzed regularization level. For EVO and EVI, the error is rather stable for $0.3 \leq \alpha \leq 1$, but EVI does not converge in all voxels for $\alpha \leq 0.2$. For the interested reader, the complete results of the endpoint error averaged over the bladder,

Table 1: Optimal values for the regularization parameters.

DIR algorithm	α	β
OF22	0.05	
OF21	0.35	
EVO	0.30	
EVI	0.35	
AEVO	0.10	0.30

prostate, and rectal wall as a function of α for the four single-parameter algorithms for the four different simulated motion patterns is shown in figure A1 in the appendix.

AEVO has two free parameters: α for the smooth regularization and β for the strength of the incompressible regularization. Their influence on the endpoint error can be seen in the appendix in figure A2. To reduce computation time, only the last time point for the **rectal filling** motion is evaluated. The regularization parameter space displays a global minimum for the endpoint error at $0.1 \leq \alpha \leq 0.2$, $0.2 \leq \beta \leq 0.8$.

3.2. Influence of the deformable image registration algorithm & dose warping method

In figure 5, a boxplot of the distribution of the absolute dose accumulation errors on the prostate in [Gy] (grays) for the four different motions are shown. The central red line in the boxplots indicates the median while the bottom and top edges of the box mark the 25th and 75th percentiles, respectively. The indicators connected by the dotted lines (the ‘whiskers’) give the most extreme data points not considered outliers, while the outliers are shown as black dots (for a normally distributed variable they consist of 0.7% of the data). The results for performing no registration and the five DIR algorithms are presented, using both DDM and EMT to accumulate the dose. From the figure, it is clear that the choice of dose mapping method generally has a much larger effect on the error than choosing the optimal DIR algorithm. When performing no MCDA, the median dose error on the prostate is 0.42 Gy for the **rectal filling** motion. Using DDM, this is reduced to 0.24-0.28 Gy for the different algorithms. When using EMT, however, the median dose error decreases by a factor 6-8 to 0.06-0.07 Gy. Furthermore, using EMT outperforms DDM in every case.

A noteworthy observation is that for the residual motion pattern (bottom right graph in figure 5), OF22 and OF21 coupled with a DDM dose warping strategy lead to larger dose errors compared to the case where no motion compensation is performed. On the other hand, when using EMT or one of the EVolution-based algorithms for MCDA, the resulting dose error is slightly decreasing.

It should be noted that all y-axes are scaled to the maximum dose error of performing no registration. The (out-of-range) maxima are OF22 (7.1 Gy) for the **rectal filling** motion, OF22 (3.3 Gy) for the bladder filling motion, OF22 (6.0 Gy) and OF21 (3.0 Gy) for the average prostate motion, and OF22 (4.3 Gy) and OF21 (2.0 Gy)

for only residual motion, all associated with a DDM dose warping strategy.

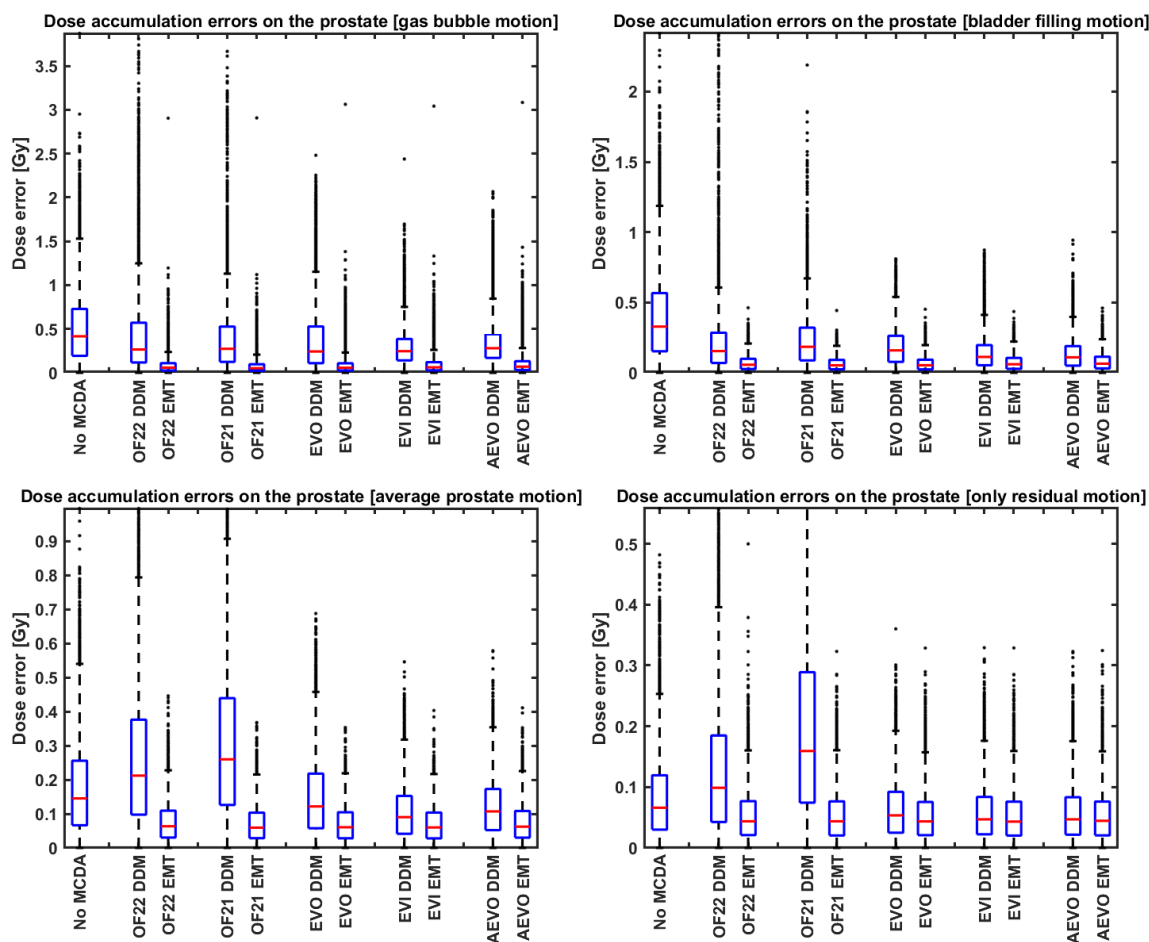


Figure 5: Absolute dose accumulation errors for the four motion patterns on the prostate in grays. From left to right in each graph are shown the results for performing no registration, the original optical flow, the improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution, both using direct dose mapping (DDM) and energy per mass transfer (EMT). Noticeable is that EMT (greatly) outperforms DDM in every case and that MCDA using EMT can decrease dose errors by a factor of 2.3 to 8.

In figure 6, the dose accumulation errors on the rectal wall are shown for the **rectal filling** motion. Here, the *relative* (*i.e.* not using the absolute value) errors are shown. This offers little additional information for the prostate region as all errors are fairly symmetric around zero for this anatomy. For the rectal wall, however, the non-compensated dose error is not centered around zero. For this particular deformation pattern, where the rectum would move into the high dose region, performing no MCDA *underestimates* the dose received by the rectal wall, by a median of almost 2 Gy, ranging up to 6 Gy, on a plan with a dose constraint of 4 Gy to the rectum. Performing *any* MCDA centers the dose error around zero.

The median absolute dose accumulation error on the rectal wall for no MCDA is 1.9 Gy. This error is reduced to 0.2-0.4 Gy when using DDM. When using EMT, the error is reduced by a factor of 12-16 to 0.12-0.16 Gy. For the other three motion patterns, the results are qualitatively similar but of a lower magnitude. It is also worth noting that MCDA always decreases the maximal and median dose error, except when using an optical flow algorithm combined with DDM. Moreover, similar to the results obtained on the prostate (see figure 5), EMT outperforms DDM in every case.

For the interested reader, we have also included figure A3 in the appendix, showcasing an example of the spatial distribution of the dose accumulation errors for one of the motion patterns (*i.e.* **rectal filling** motion). The resulting observations are well in line with the ones made above. For the reader interested in comparing the quality of the registration only, the distribution of endpoint errors within the prostate, bladder and rectal wall for all time point of the four different simulated motion patterns are shown in figure A5 in the appendix. The lowest maximum and median endpoint errors are achieved by either OF21 or EVO, while the highest outliers arise from OF22.

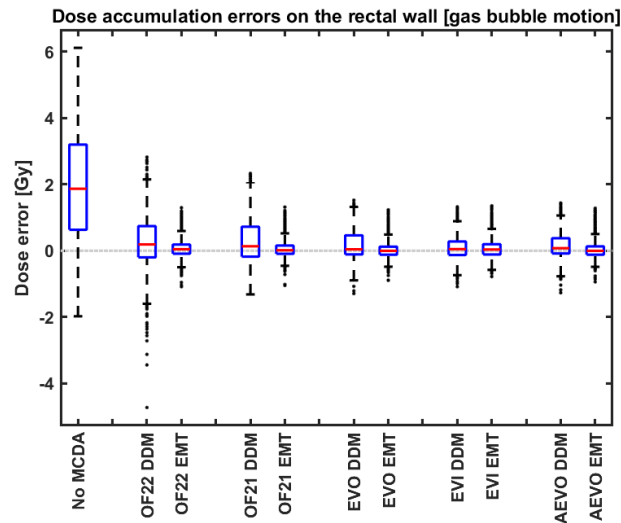


Figure 6: Relative dose accumulation errors for the **rectal filling** motion on the rectal wall in grays. Shown from left to right are the results for performing no registration and for the five DIR algorithms using direct dose mapping (DDM) and energy per mass transfer (EMT). Without MCDA, there is an underestimation of the dose delivered to the rectal wall of 2 Gy. This asymmetry vanishes for all algorithms, and the median absolute dose error decreases by a factor of 12-16 when using EMT.

Figure 7 shows the Jacobian determinants within the prostate of the simulated and estimated DVFs for the bladder filling motion. The results for the other three motion patterns (not shown) are qualitatively similar. The deformations from all the simulated time points were included in the analysis. OF21, OF22, and EVO have a wider range of Jacobian determinants (0.93 - 1.09, 0.91 - 1.10, and 0.93 - 1.07, excluding outliers) compared to EVI and AEO (0.94 - 1.04 and 0.96 - 1.04), which are closer to the gold

standard values (1.00 - 1.00). For OF22, the outliers even range from -0.25 to 3.0, indicating tissue folding and extreme expansion.

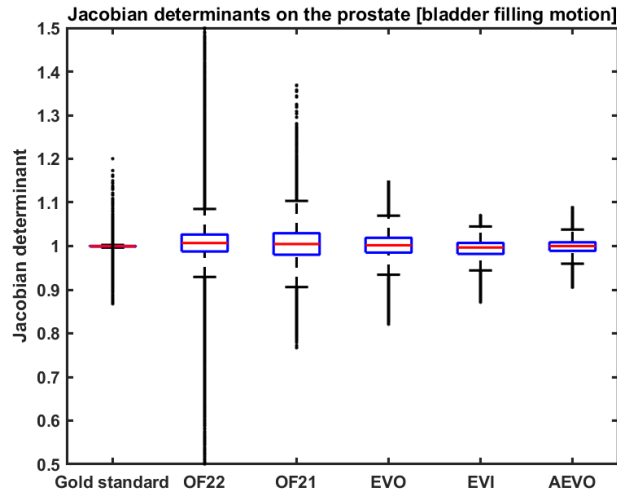


Figure 7: Jacobian determinants on the prostate for the bladder filling motion pattern for the gold standard, the original optical flow, the improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution (from left to right). Note that the gold standard shows only slight deviations from 1. EVI and AEVO produce a narrow range around 1, while OF21, OF22, and EVO show a wider spread.

For completion, the endpoint error distribution on the prostate, bladder and rectal wall is shown in figure A5 in the appendix, showing that performing image registration considerably reduces the endpoint error with errors after registration mostly under half a voxel.

3.3. The influence of image resolution and signal-to-noise-ratio

Figure 8 shows the influence of resolution on the dose accumulation errors within the prostate and rectal wall for the **rectal filling** and bladder filling motion patterns. The dose accumulation errors for the other two motion patterns (not shown) are qualitatively in-between those shown, with slight relative differences between the optical flow and the EVOLution-based algorithms. Given the observations made in section 3.2, we only report the results obtained by using EMT, for better graph readability. Note that the six leftmost boxes correspond to registrations performed at the original image resolution and are the same as in figure 5.

Except for OF22 on the lowest resolution on the prostate for the bladder filling motion, the maximum and median absolute dose accumulation error is decreased by performing MCDA with any DIR algorithm on any resolution, for both anatomies and shown motion patterns. For the bladder filling motion, the median absolute dose accumulation error on the prostate decreases with a factor 5-6 on the highest resolution,

4-5 on the second-highest, 3-4 on the second-lowest, and 1.3-2.6 on the lowest resolution. On the rectal wall, this becomes a factor 10-17 on the highest, 2-9 on the second-highest, 3-9 on the second-lowest, and 1.4-3.2 on the lowest resolution. EVO and AEVO are the most robust algorithms for lower resolutions, but the differences between algorithms are relatively small.

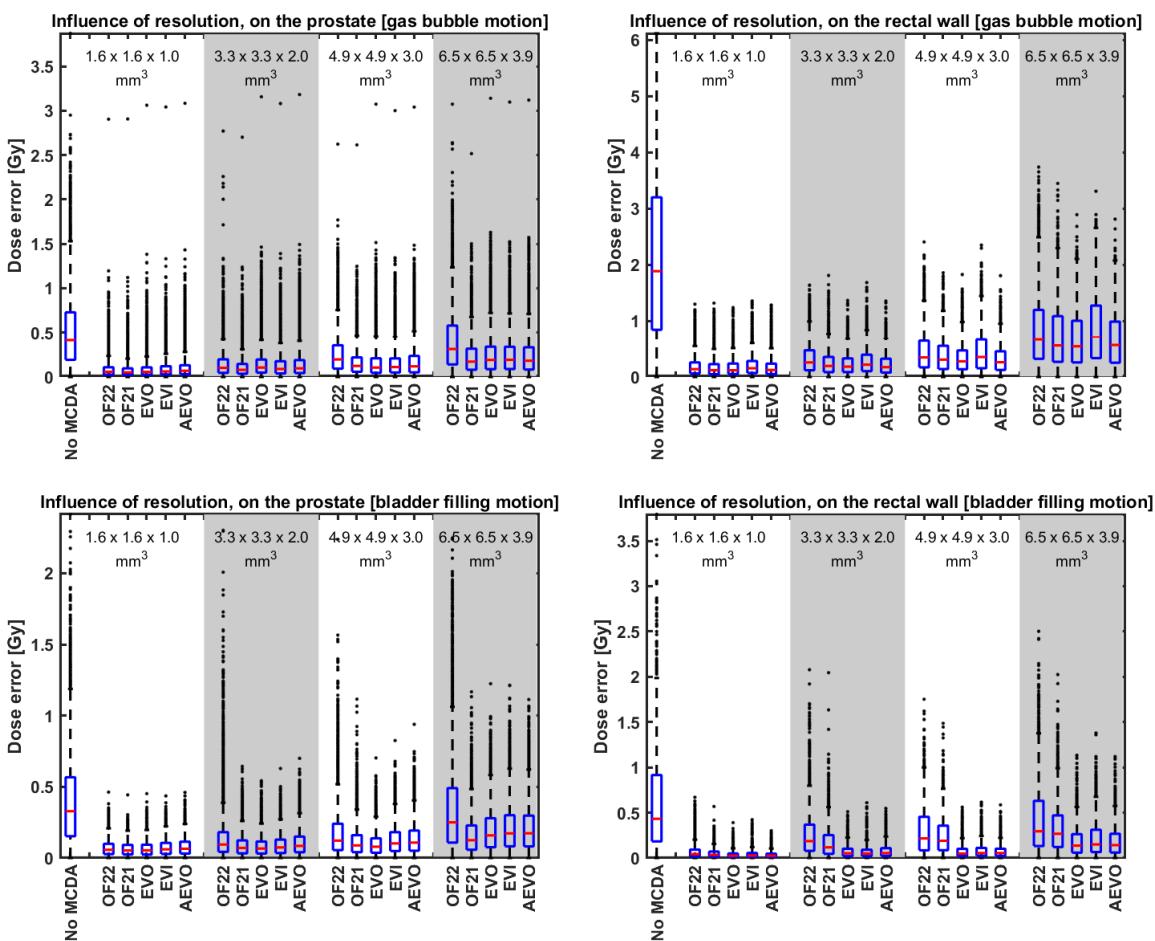


Figure 8: Absolute dose accumulation error on the prostate and rectal wall for different algorithms at no, two-fold, three-fold and four-fold down-sampled images for the **rectal filling** and bladder filling motions, respectively. The resolutions in millimeters are provided in the graphs. The DIR algorithms shown from left to right are no registration, the original optical flow, the improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution, all combined with energy per mass transfer. The effect of lowering the resolution is multiple factors higher than the differences between algorithms. Even on the lowest resolution performing MCDA decreases the dose error notably, while a resolution below 2 mm³ is optimal.

Figure 9 shows the influence of SNR on the absolute dose accumulation errors within the prostate and rectal wall, again for the **rectal filling** and bladder filling motions. The results for the other two motion patterns (not shown) are qualitatively similar to those

of the bladder filling motion. It can be observed that all DIR algorithms decrease the median as well as the maximum absolute dose accumulation error on both anatomies for all levels of the SNR. For the bladder filling motion, the median absolute dose accumulation errors on the prostate decrease by a factor of 2.0 to 4.3, even for the lowest SNR. On the rectal wall, this becomes a factor of 3.4 to 6.8.

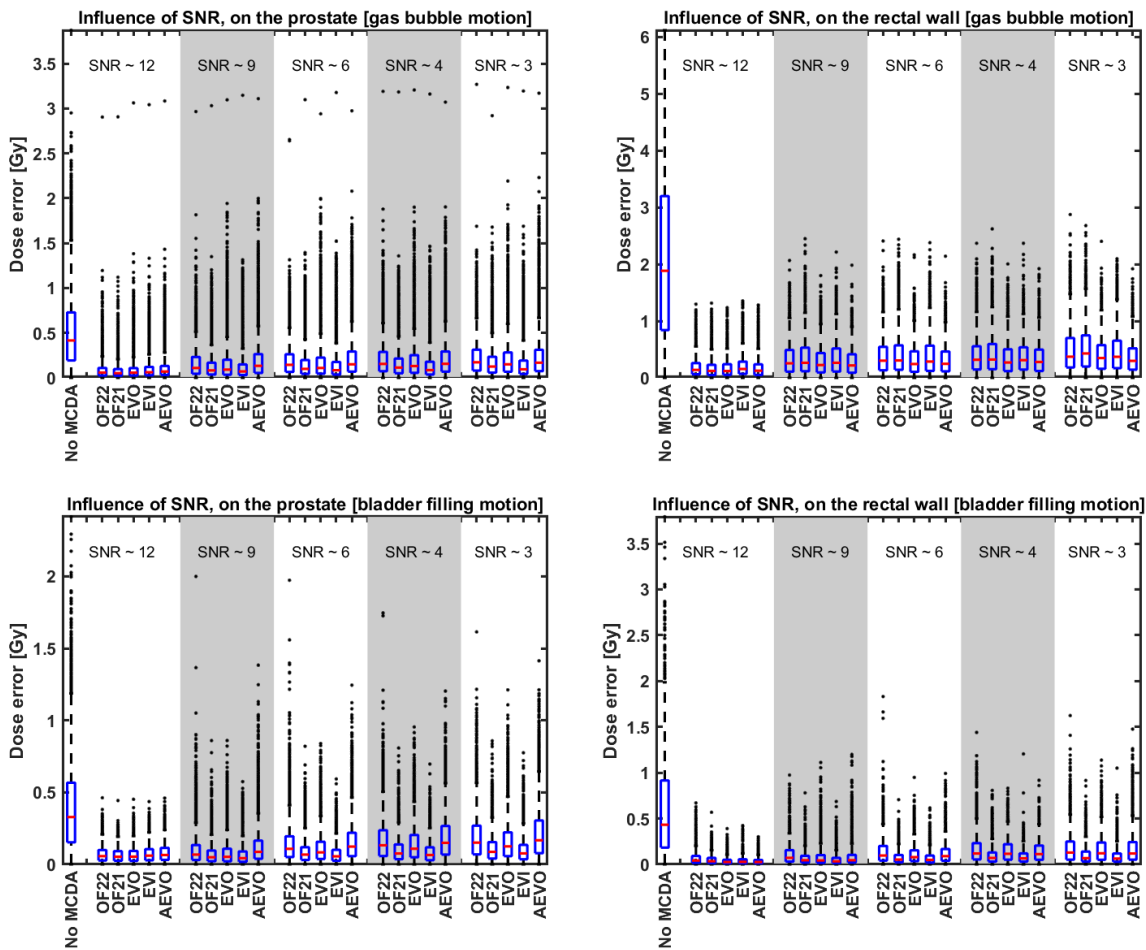


Figure 9: Absolute dose accumulation errors on the prostate and rectal wall for different values of the SNR and different algorithms. The **rectal filling** and bladder filling motion patterns are shown. The SNR is provided within the graph. No registration, original optical flow, improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution are shown from left to right, all using energy per mass transfer. The difference in dose errors between an SNR of 12 and 9 is about as big as the difference between an SNR of 9 and 3.

3.4. Evaluation on the clinical data

For the clinical dataset, the prostate experiences a sudden displacement between the 8th and 10th time point of about 4-5 millimeters in the anterior direction and about 3

millimeters in the superior direction. The prostate then only slowly drifts towards its original position for about 0.5-1.5 and 1 millimeters, respectively, during the final 6-7 time-steps. For the interested reader, an assessment of the motion as estimated by the five used DIR algorithms is shown in figure A4 in the appendix.

As discussed in section 2.4, for the clinical cine-MR series, we use the SSI and IID to evaluate the registration performances since no gold standard DVF is available. Note that we evaluate 1-SSI (as the SSI gives a value between 0 and 1) such that it increases for larger registration errors, as does the endpoint error. Figure 10 showcases the SSI (in (a)) and the IID (in (b)) quantified jointly inside the prostate, bladder, and rectum wall. Note that the same regularization parameters as for the simulated data were used. On a qualitative level, the results are similar in nature to those obtained for the simulated data illustrated in figure A5: the differences between algorithms are relatively small, and the registration error decreases notably for all algorithms compared to performing no registration. In terms of both SSI and IID, it can be observed that EVI performs worst of all algorithms. Excluding EVI, the median registration error decreases by a factor of 1.4 to 3.0, depending on the DIR algorithm and quality assurance criterion used.

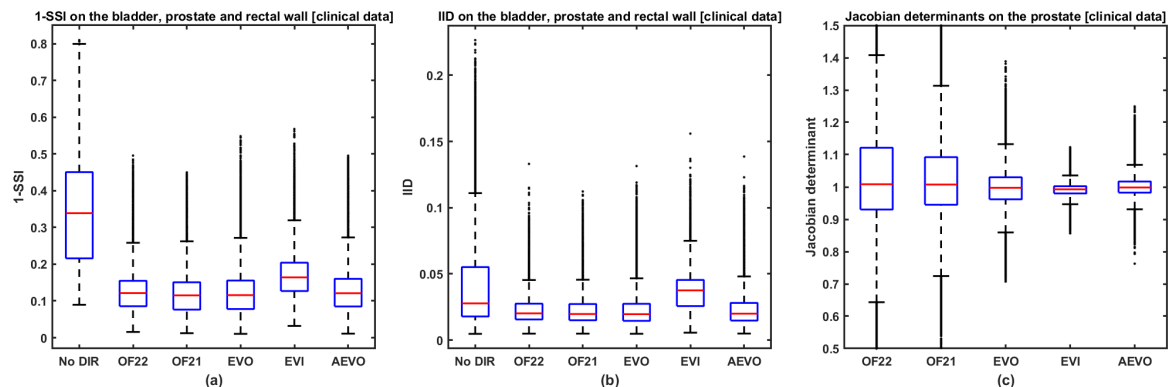


Figure 10: (a) the time-averaged structural similarity index (shown as 1-SSI) and (b) image intensity differences (IID) within the bladder, prostate, and rectal wall for the clinical data. The results for performing no registration, the original optical flow, the improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution are shown, from left to right. For both error measures, the differences between algorithms are relatively small, and the error decreases considerably compared to no registration. (c) the Jacobian determinants for the original optical flow, the improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution on the prostate for the clinical data. OF22, OF21, and EVO show a wider range of values than EVI and AEVO.

Figure 10 (c) displays the Jacobian determinants on the prostate for all time points of the clinical dataset. As for the simulated data, OF22, OF21, and EVO show a wider range of Jacobian determinants for voxels inside the prostate (0.64 - 1.41, 0.72 - 1.31,

and 0.86 - 1.13, excluding outliers), compared to EVI and AEVO (0.95 - 1.04 and 0.93 - 1.07, respectively). For OF22, the outliers even range from -0.5 to 6.2, indicating tissue folding and extreme expansion.

Due to its competitive registration errors (figure 10 (a) and (b)) and better anatomical plausibility of the estimated deformations in terms of Jacobian determinant values (figure 10 (c)), the accumulated dose provided by the AEVO algorithm was chosen as a point of comparison to analyze dose accumulation errors on the clinical data (the results on the simulated data also supported this choice). Note that for dose warping, the EMT strategy was employed as before.

The resulting dose accumulation errors are shown in figure 11, showing great similarities to the results for the simulated data. Without registration, the dose accumulation errors on the prostate range towards 2.5 Gy, with outliers towards 5.6 Gy. As for the simulated data, the dose warping method has a larger impact than the choice of the DIR algorithm. Morevoer, as for the simulated **rectal filling** motion, there is a median *underestimation* of the dose delivered to the rectal wall when not performing any MCDA of almost 2 Gy, ranging up to 4.5 Gy while the rectum had a dose constraint of 4 Gy in the therapy plan. This asymmetry is again addressed by all algorithms and warping methods.

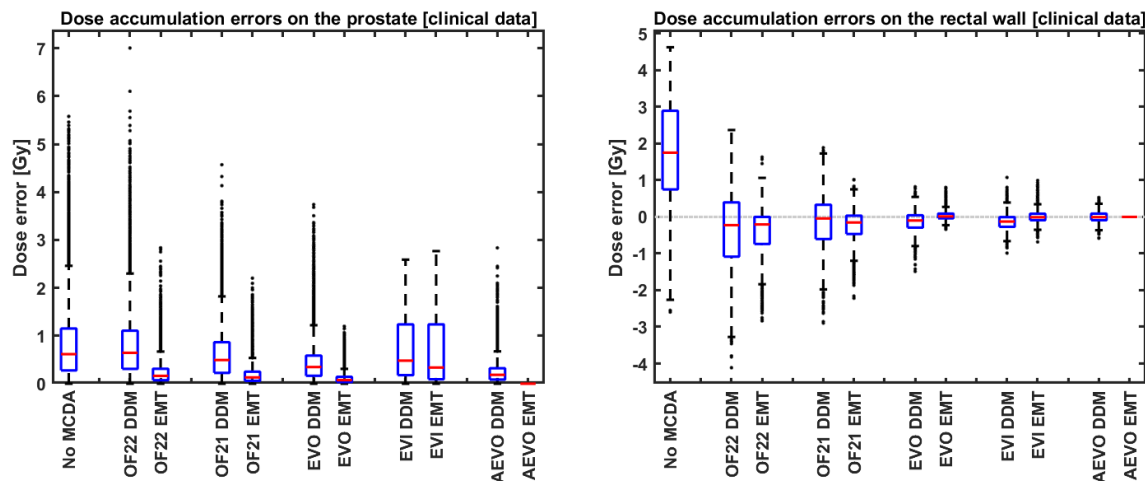


Figure 11: Dose accumulation errors for the clinical data (AEVO with EMT was used as a point of comparison). The absolute error on the prostate (left) and the relative error on the rectal wall (right) are shown for the five different DIR algorithms, as well as no registration, for energy per mass transfer (EMT) and direct dose mapping (DDM). The difference between the dose warping methods is much larger than between DIR algorithms, with EMT outperforming DDM substantially. All methods resolve the underestimation of the dose received by the rectal wall when not performing MCDA.

4. Discussion

All presented dose accumulation results show that performing MCDA using variational algorithms for motion mitigation can considerably reduce errors in the estimated delivered dose for prostate cancer therapy. Figure 5 shows that using an EVOLUTION-based algorithm, or EMT, *always* results in a decrease in the dose error. When using EMT, the median dose error on the prostate is reduced by a factor of 1.5 to 7, depending on the motion pattern and DIR algorithm. For the rectal wall, the median dose error decreases with a factor of 2.5 to 16 for the non-residual motion patterns. Additionally, MCDA resolves any asymmetry or bias in the uncorrected dose error. Of similar importance is the observation that this type of MCDA in the absence of large-scale biological motion does not degrade dosimetry compared to the uncorrected dose estimate.

4.1. Dose warping strategy

The choice of dose warping strategy has the most substantial impact on the dose accumulation error of all factors considered in this project, with EMT always outperforming DDM. This reconfirms the findings of Siebers & Zhong (2008) on a deformable phantom (1.1% difference between EMT and DDM along the beam and up to 25% in the penumbra), Zhong & Siebers (2009) on a lung plan (7.3% difference) and Li et al. (2013) on a lung plan (11.3% difference in the planning target volume and 4.4% in the internal target volume minimum doses). We observe median dose differences of approximately 2.5% on the prostate and 10% on the rectal wall.

These earlier works evidenced that the typical situations where DDM becomes increasingly problematic are deformations of boundary interfaces displaying high tissue density variations and tissue areas that display internal density fluctuations. This trend can clearly be seen in figure 5, where the difference between using DDM and EMT is found to be larger for algorithms for which the Jacobian determinants are unequal to one (figure 7), *i.e.* where compression or expansion is estimated. This is due to the fact that EMT intrinsically compensates for density modifications introduced by the realignment process, while DDM is very susceptible to this effect. As a consequence, algorithms that have been optimized to minimize density variations (such as EVI and AEVO) display a considerably better performance when paired with DDM, as shown in figure 5 and 6. Nevertheless, even in this configuration, EMT resulted in a lower dose error compared to DDM. It is noteworthy that the usage of EMT compared to DDM introduces only a small level of additional complexity and thus computing time.

4.2. Different deformable image registration algorithms

First and foremost, the selection of algorithms in the scope of this paper has been motivated by the established track record of variational algorithms concerning robust registration performance even for image data of low SNR and resolution (Glitzner

et al. 2015, Zachiu, Denis de Senneville, Moonen & Ries 2015b, Zachiu, Papadakis, Ries, Moonen & Denis de Senneville 2015, Denis de Senneville et al. 2016). Since variational algorithms also employ ‘surrounding’ information of respective voxels, accuracy and precision can, for coherent deformations, typically exceed the image resolution. We observed that the 75th percentile of the endpoint error for all motion patterns and all algorithms is around half a voxel.

When using DDM, clear differences in dose error between DIR algorithms are observed. EVI and AEVO with biophysical plausibility (in the form of an incompressibility regularization) built-in outperform the other algorithms, in particular OF22 and OF21. The biophysical plausibility is observed in the narrow range of Jacobian determinants (figure 7), as described in (Zachiu et al. 2018, Zachiu, Denis de Senneville, Raaymakers & Ries 2020). Exemplary for the implausible deformations is for example the tendency of the OF22 algorithm to estimate tissue folding in problematic regions. This phenomenon is most likely caused by violations of the brightness constancy assumption. Using all gray-level information, the optical flow algorithm is more likely to be biased by image artifacts to converge into local minima. Although the effect of these unrealistic deformations on the accumulated dose is reduced by using the EMT dose warping strategy, such an approach may not be physically applicable for warping other quantitative information such as Hounsfield units, diffusion/perfusion values, mass maps, etc.

Furthermore, an asset of the EVolution-based algorithms is that they perform well for a wide range of regularization parameters (figures A1 and A2 in the appendix). This can be of considerable clinical value in particular for online scenarios, where the tuning of regularization parameters with the patient on the treatment table is either impossible or for a smooth work-flow highly undesirable.

Based on these grounds, we consider AEVO the optimal choice for the DIR algorithm, as the clinical data shows its superiority over EVI in terms of registration errors for sudden deformations.

4.3. Resolution and SNR

Generally, a resolution of about 1 mm³ seems good, 2 mm³ seems adequate, and from 3 mm³ onward, we start to under-sample. This corresponds to approximately half the size of some relevant structures like the rectal wall (8 by 24 mm) and the dose gradient (11 to 2 Gy) between the prostate and rectal wall (7-11 mm).

The EVolution-based algorithms prove to be the most robust with respect to lower resolutions. Interestingly, for these algorithms, even at the lowest resolution, the median dose error is lower for all simulated motions and anatomies compared to the one obtained by performing no motion mitigation. For the **rectal filling** and bladder filling motion, the error on these low-resolution images is still reduced by a factor of two. It is thus overall beneficial to perform MCDA even when only compensating for bulk motion; however, the optimal resolution value resulting from the present study is above 2 mm³.

All variational methods are fairly robust and still perform well for a low SNR. The difference in median absolute dose error for the **rectal filling** motion between the lowest and highest SNR is only a factor of 1.5 to 3.5. EVI is the optimal DIR algorithm when working with images of very low SNR due to incompressibility being a strong-natured regularization. For all DIR algorithms, the step in dose errors from an SNR of 12 to an SNR of 9 is about as big as from an SNR of 9 to an SNR of 3. Using an SNR above 10 thus gives the optimal results, but even for an SNR of 3, performing MCDA (using EMT) is beneficial.

4.4. Clinical data

The dose errors found for the clinical data (figure 11) confirm the findings of the simulated data. First, performing MCDA again proves to be able to decrease errors in the estimated dose considerably. Secondly, the choice of dose warping strategy again impacts the dose errors the most. Finally, as for the simulated **rectal filling** motion, when not performing MCDA, there is a significant underestimation of the dose delivered to the rectal wall that is resolved by performing MCDA.

The dose accumulation errors for the clinical data reported in figures 11 are found using AEVO with EMT as a the point of comparison. AEVO was used as it produces particularly low registration differences (figure 10 (a) and (b)) and a narrow Jacobian determinant range within the prostate (figure 10 (c)) as well as competitive dose errors for the simulated data (figures 5 and 6).

Combined with the simulated data results, there is a strong case for performing MCDA in scenarios where the delivered dose is of interest. This will provide another step in the direction of hypofractionation and intra-fractionally adapted treatments that count on reliable MCDA for further development and implementation.

The results presented here for the prostate are expected to carry over to other tumor sites with motion patterns having a slow motion component (like peristaltic events, respiratory baseline, muscle relaxation or digestive activity). For inter-fraction dose accumulation, deformations are generally more extreme than for intra-fraction, and it is to be expected that using EMT over DDM has an even larger impact.

5. Conclusion

The presented study analyzed the influence of the choice of DIR algorithm, dose warping strategy, MR image resolution, and MR image SNR on the resulting motion-compensated accumulated dose. For a quantitative investigation, deformations based on clinical observations were simulated using finite element modeling, which provided a gold standard for both the deformations estimated by the registration algorithms and the accumulated dose. The evaluation was also extended to include clinical patient data.

The analysis conducted on both the simulated and the clinical data demonstrates that, in general, performing MCDA leads to a notable improvement of the geometric

uncertainty with respect to the radiation dose delivery. We have found that, for an optimal selection of the investigated influence factors, for large-scale biological motion such as the passage of a **rectal filling**, the dose error when performing MCDA decreases by a factor of 7 and 15 within the anatomical structures of interest. While not to the same extent, MCDA continued to provide improvements in geometric uncertainty even for the lower end of the image resolution and SNR investigated within the scope of this study. Moreover, MCDA did not introduce any relevant errors for the case in which only minimal motion is present.

Out of all the investigated factors, we conclude that the most important one impacting the accumulated dose is the dose warping strategy. The differences between the two dose warping strategies were considerably larger than the differences between the registration algorithms. Furthermore, using energy per mass transfer over direct dose mapping decreases dose errors in all cases, at no extra scan time, and only implies limited extra computing power.

In terms of the employed registration algorithm, it was found that, due to its accuracy, anatomical plausibility, and robustness to deviations of the regularization parameters, AEVO is a marginally better choice than the rest of the investigated algorithms. Finally, an isotropic image resolution of 2 mm and an image SNR above 10 were found to be optimal for an accurate MCDA.

Overall, the present work provides an extensive analysis of the dependency of MCDA on multiple contributing factors, providing a guideline for their choice in accordance with the specific technical and functional specifications of ongoing and future MRgRT workflows.

Acknowledgments

The authors wish to thank Daan de Muinck Keizer and Hans de Boer for helpful discussions. This work has been supported in part by the Topconsortia for Knowledge and Innovation - LifeSciences & Health, project no. LSHM19031 (POPCORN).

References

- Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M. J. & Szeliski, R. (2011). A database and evaluation methodology for optical flow, *International journal of computer vision* **92**(1): 1–31.
- Barr, R. G., Memo, R. & Schaub, C. R. (2012). Shear wave ultrasound elastography of the prostate: initial results, *Ultrasound quarterly* **28**(1): 13–20.
- Bharatha, A., Hirose, M., Hata, N., Warfield, S. K., Ferrant, M., Zou, K. H., Suarez-Santana, E., Ruiz-Alzola, J., D'amico, A., Cormack, R. A. et al. (2001). Evaluation of three-dimensional finite element-based deformable registration of pre-and intraoperative prostate imaging, *Medical physics* **28**(12): 2551–2560.
- De Luca, P., Jones, D., Gahbauer, R., Whitmore, G. & Wambersie, A. (2010). Prescribing, recording, and reporting intensity-modulated photon-beam therapy (imrt)(icru report 83), *Journal of the ICRU* **10**.
- de Muinck Keizer, D., Kontaxis, C., Kerkmeijer, L., van Zyp, J. v. d. V., van den Berg, C., Raaymakers,

- B., Lagendijk, J. & de Boer, J. (2020). Dosimetric impact of soft-tissue based intrafraction motion from 3d cine-mr in prostate sbrt, *Physics in Medicine & Biology* **65**(2): 025012.
- de Muinck Keizer, D. M., Kerkmeijer, L. G., Maspero, M., Andreychenko, A., van Zyp, J. v. d. V., Van den Berg, C. A., Raaymakers, B. W., Lagendijk, J. J. & de Boer, J. (2019). Soft-tissue prostate intrafraction motion tracking in 3d cine-mr for mr-guided radiotherapy, *Physics in Medicine & Biology* **64**(23): 235008.
- de Muinck Keizer, D., Pathmanathan, A., Andreychenko, A., Kerkmeijer, L., van Zyp, J. v. d. V., Tree, A., van den Berg, C. & de Boer, J. (2019). Fiducial marker based intra-fraction motion assessment on cine-mr for mr-linac treatment of prostate cancer, *Physics in Medicine & Biology* **64**(7): 07NT02.
- den Hartogh, M. D., de Boer, H. C., de Groot-van Breugel, E. N., van der Voort van Zyp, J., Hes, J., van der Heide, U. A., Pos, F., Haustermans, K., Depuydt, T., Jan Smeenk, R. et al. (2019). Planning feasibility of extremely hypofractionated prostate radiotherapy on a 1.5 t magnetic resonance imaging guided linear accelerator, *Physics and Imaging in Radiation Oncology* **11**: 16–20.
- Denis de Senneville, B., Zachiu, C., Ries, M. & Moonen, C. (2016). EVolution: an edge-based variational method for non-rigid multi-modal image registration, *Physics in Medicine and Biology* **61**(20): 7377–7396.
URL: <https://doi.org/10.1088%2F0031-9155%2F61%2F20%2F7377>
- Dice, L. R. (1945). Measures of the amount of ecologic association between species, *Ecology* **26**(3): 297–302.
- Glitzner, M., de Senneville, B. D., Lagendijk, J., Raaymakers, B. & Crijs, S. (2015). On-line 3d motion estimation using low resolution mri, *Physics in Medicine & Biology* **60**(16): N301.
- Guckenberger, M. (2011). Image-guided radiotherapy based on kilovoltage cone-beam computed tomography—a review of technology and clinical outcome, *Eur Oncol Haematol* **7**(2): 121–124.
- Gudbjartsson, H. & Patz, S. (1995). The rician distribution of noisy mri data, *Magnetic resonance in medicine* **34**(6): 910–914.
- Haber, E. & Modersitzki, J. (2004). Numerical methods for volume preserving image registration, *Inverse problems* **20**(5): 1621.
- Haber, E. & Modersitzki, J. (2007). Image registration with guaranteed displacement regularity, *International journal of computer vision* **71**(3): 361–372.
- Hausdorff, F. (1914). *Grundzüge der mengenlehre*, Vol. 7, von Veit.
- Henkelman, R. M. (1985). Measurement of signal intensities in the presence of noise in mr images, *Medical physics* **12**(2): 232–233.
- Horn, B. K. & Schunck, B. G. (1981). Determining optical flow, *Techniques and Applications of Image Understanding*, Vol. 281, International Society for Optics and Photonics, pp. 319–331.
- Jaffray, D. A. (2012). Image-guided radiotherapy: from current concept to future perspectives, *Nature Reviews Clinical Oncology* **9**(12): 688.
- Kontaxis, C., Bol, G., Lagendijk, J. & Raaymakers, B. (2015a). A new methodology for inter-and intrafraction plan adaptation for the mr-linac, *Physics in Medicine & Biology* **60**(19): 7485.
- Kontaxis, C., Bol, G., Lagendijk, J. & Raaymakers, B. (2015b). Towards adaptive imrt sequencing for the mr-linac, *Physics in Medicine & Biology* **60**(6): 2493.
- Kontaxis, C., Bol, G., Stemkens, B., Glitzner, M., Prins, F., Kerkmeijer, L., Lagendijk, J. & Raaymakers, B. (2017). Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the mr-linac, *Physics in Medicine & Biology* **62**(18): 7233.
- Kontaxis, C., de Muinck Keizer, D. M., Kerkmeijer, L. G., Willigenburg, T., den Hartogh, M. D., de Groot-van Breugel, E. N., Hes, J., Raaymakers, B. W., Lagendijk, J. J., de Boer, H. C. et al. (2020). Delivered dose quantification in prostate radiotherapy using online 3d cine imaging and treatment log files on a combined 1.5 t magnetic resonance imaging and linear accelerator system, *Physics and Imaging in Radiation Oncology* **15**: 23–29.

- Legendijk, J. J., Raaymakers, B. W. & Van Vulpen, M. (2014). The magnetic resonance imaging–linac system, *Seminars in radiation oncology*, Vol. 24, Elsevier, pp. 207–209.
- Li, H. S., Zhong, H., Kim, J., Glide-Hurst, C., Gulam, M., Nurushev, T. S. & Chetty, I. J. (2013). Direct dose mapping versus energy/mass transfer mapping for 4d dose accumulation: fundamental differences and dosimetric consequences, *Physics in Medicine & Biology* **59**(1): 173.
- Maas, S. A., Ellis, B. J., Ateshian, G. A. & Weiss, J. A. (2012). FEBio: finite elements for biomechanics, *J Biomech Eng* **134**(1): 011005.
- McBain, C. A., Henry, A. M., Sykes, J., Amer, A., Marchant, T., Moore, C. M., Davies, J., Stratford, J., McCarthy, C., Porritt, B. et al. (2006). X-ray volumetric imaging in image-guided radiotherapy: the new standard in on-treatment imaging, *International Journal of Radiation Oncology* Biology* Physics* **64**(2): 625–634.
- Mundt, A. J. & Roeske, J. C. (2010). *Image-guided radiation therapy: a clinical perspective*, PMPH-USA.
- Mutic, S. & Dempsey, J. F. (2014). The viewray system: magnetic resonance–guided and controlled radiotherapy, *Seminars in radiation oncology*, Vol. 24, Elsevier, pp. 196–199.
- Rice, S. O. (1944). Mathematical analysis of random noise, *The Bell System Technical Journal* **23**(3): 282–332.
- Siebers, J. V. & Zhong, H. (2008). An energy transfer method for 4d monte carlo dose calculation, *Medical physics* **35**(9): 4096–4105.
- Sorensen, T. A. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons, *Biol. Skar.* **5**: 1–34.
- Thilmann, C., Nill, S., Tücking, T., Höss, A., Hesse, B., Dietrich, L., Bendl, R., Rhein, B., Häring, P., Thieke, C. et al. (2006). Correction of patient positioning errors based on in-line cone beam cts: clinical implementation and first experiences, *Radiation Oncology* **1**(1): 16.
- Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity, *IEEE transactions on image processing* **13**(4): 600–612.
- Widmark, A., Gunnlaugsson, A., Beckman, L., Thellenberg-Karlsson, C., Hoyer, M., Lagerlund, M., Kindblom, J., Ginman, C., Johansson, B., Björnlinger, K. et al. (2019). Ultra-hypofractionated versus conventionally fractionated radiotherapy for prostate cancer: 5-year outcomes of the hypo-rt-pc randomised, non-inferiority, phase 3 trial, *The Lancet* **394**(10196): 385–395.
- Wieser, H.-P., Cisternas, E., Wahl, N., Ulrich, S., Stadler, A., Mescher, H., Müller, L.-R., Klinge, T., Gabrys, H., Burigo, L. et al. (2017). Development of the open-source dose calculation and optimization toolkit matrad, *Medical Physics* **44**(6): 2556–2568.
- Yang, Q., Parvin, B. & Mariano, A. (2000). Singular features in sea surface temperature data, *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Vol. 1, IEEE, pp. 516–520.
- Zachiu, C., Denis de Senneville, B., Moonen, C. & Ries, M. (2015a). A framework for the correction of slow physiological drifts during mr-guided hifu therapies: Proof of concept, *Medical physics* **42**(7): 4137–4148.
- Zachiu, C., Denis de Senneville, B., Moonen, C. & Ries, M. (2015b). A framework for the correction of slow physiological drifts during mr-guided hifu therapies: Proof of concept, *Medical physics* **42**(7): 4137–4148.
- Zachiu, C., Denis de Senneville, B., Moonen, C. T., Raaymakers, B. W. & Ries, M. (2018). Anatomically plausible models and quality assurance criteria for online mono-and multi-modal medical image registration, *Physics in Medicine & Biology* **63**(15): 155016.
- Zachiu, C., Denis de Senneville, B., Raaymakers, B. W. & Ries, M. (2020). Biomechanical quality assurance criteria for deformable image registration algorithms used in radiotherapy guidance, *Physics in Medicine & Biology* **65**(1): 015006.
- Zachiu, C., Denis de Senneville, B., Willigenburg, T., de Boer, J., Raaymakers, B. W., Ries, M. G. et al. (2020). Anatomically-adaptive multi-modal image registration for image-guided external-beam

radiotherapy, *Physics in Medicine & Biology* .

Zachiu, C., Papadakis, N., Ries, M., Moonen, C. & Denis de Senneville, B. (2015). An improved optical flow tracking technique for real-time mr-guided beam therapies in moving organs, *Physics in Medicine & Biology* **60**(23): 9003.

Zhong, H. & Siebers, J. V. (2009). Monte carlo dose mapping on deforming anatomy, *Physics in Medicine & Biology* **54**(19): 5815.

Appendix A. DIR algorithms

The specifics of the five DIR algorithms used in this project are outlined below.

- *The original optical flow (OF22)*. It relies on the assumption that voxels conserve their gray-level intensity as they move. This implies minimizing a cost function which uses the square Euclidean distance between the images as a data fidelity term, coupled with a smoothness regularization:

$$\mathcal{E}_{OF22} = \iiint_{\Omega} (I_x u_1 + I_y u_2 + I_z u_3 + I_t)^2 + \alpha \mathcal{R}_s, \quad (\text{A.1})$$

where (I_x, I_y, I_z) are the image spatial partial derivatives, I_t is the temporal derivative.

- *Improved optical flow (OF21)*. Replaces the quadratic data fidelity term from the original optical flow by a linear L^1 norm, reducing the impact of local gray-level inconsistencies between the images to be registered. In effect, this implies the minimization of the following cost function:

$$\mathcal{E}_{OF21} = \iiint_{\Omega} |I_x u_1 + I_y u_2 + I_z u_3 + I_t| + \alpha \mathcal{R}_s. \quad (\text{A.2})$$

- *EVOLution (EVO)*. This algorithm aims to align similar contrast patterns between the registered images. This is achieved by the minimization of the function

$$\mathcal{E}_{EVO} = \sum_{\vec{r} \in \Omega} \exp(-C(\mathbf{u}(\vec{r}))) + \alpha \mathcal{R}_s, \quad (\text{A.3})$$

with

$$C(\mathbf{u}(\vec{r})) = \frac{\sum_{\vec{s} \in \Gamma} \left| \vec{\nabla} I_1(\vec{s}) \cdot \vec{\nabla} I_2(\vec{s} + \mathbf{u}(\vec{s})) \right|}{\sum_{\vec{s} \in \Gamma} \|\vec{\nabla} I_1(\vec{s})\|_2 \|\vec{\nabla} I_2(\vec{s} + \mathbf{u}(\vec{s}))\|_2}, \quad (\text{A.4})$$

where Γ is a neighborhood around \vec{r} and $I_{1,2}$ are the reference and moving images, respectively.

- *Incompressible EVOLution (EVI)*. This algorithm uses the same data fidelity term as EVO, but replaces the smoothness regularization term with a global penalty on incompressibility.
- *Spatially adaptive EVOLution (AEVO)*. Employs the same data fidelity term as EVO and EVI. However, it uses the smoothness and the incompressibility regularization selectively, depending on the physical properties of the underlying

anatomy. Knowledge of the underlying anatomy can thus be used to enforce incompressibility within *e.g.* biological soft tissues, while allowing other regions to expand or compress, using the smoothness regularization. In effect, this algorithm employs two regularization parameters: one for the incompressible region and one for the smooth region.

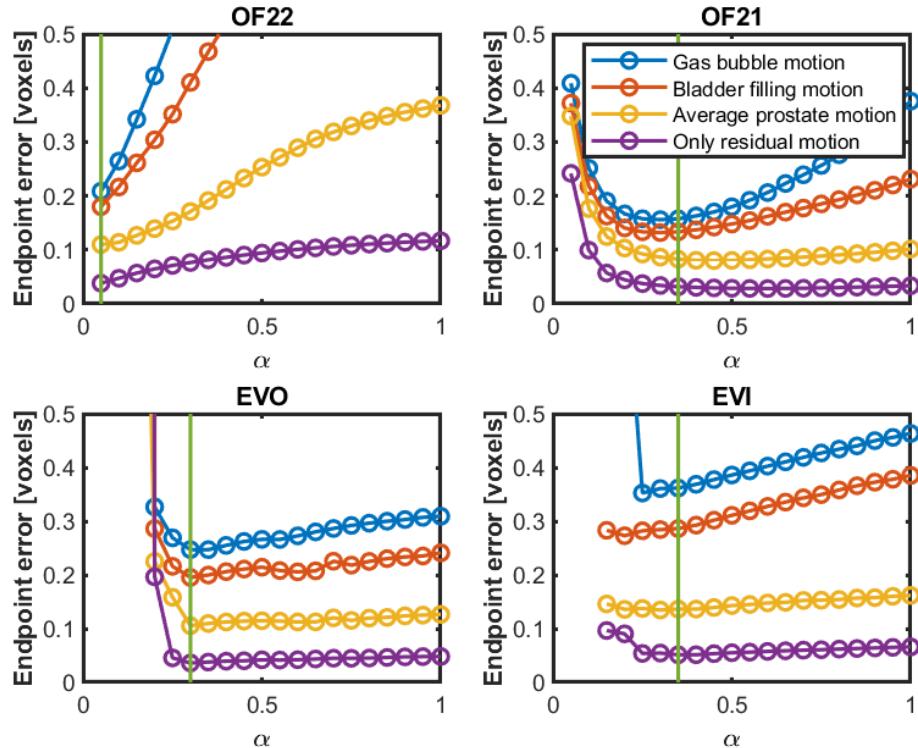


Figure A1: Endpoint error averaged over the voxels of the bladder, prostate, and rectal wall for different values of the regularization parameter α for the original optical flow (OF22), the improved optical flow (OF21), EVolution (EVO), and incompressible EVolution (EVI), respectively. The different colors correspond to the four simulated motion patterns. The vertical green line depicts the optimal choice for the regularization parameter used in this paper. OF22 in particular is very sensitive to the correct value of the regularization parameter. **TODO: change legend to rectal filling motion and adapt scale for optical flows $\alpha \rightarrow \alpha^2$!**

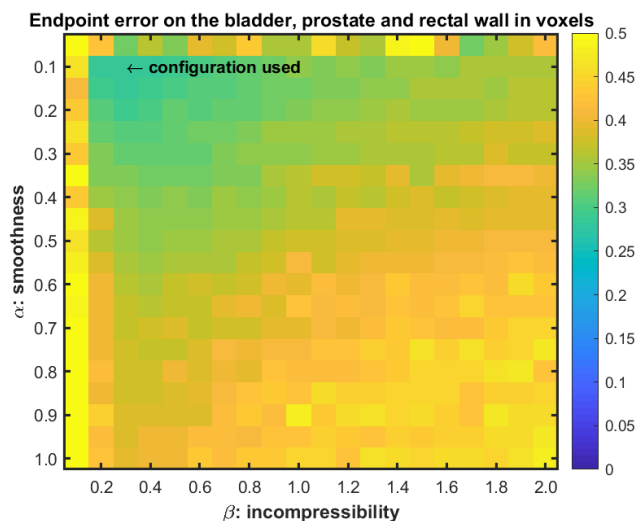


Figure A2: Average endpoint error on the bladder, prostate, and rectal wall in voxels for different values of the two regularization parameters for spatially adaptive EVolution. Results for the last time point of the **rectal filling** motion are shown. The arrow points at the optimal configuration used in this paper.

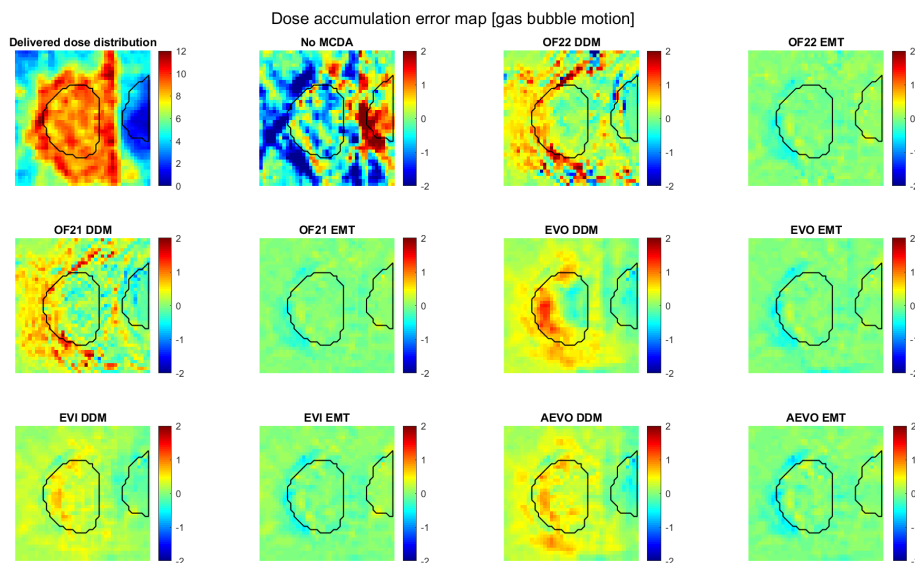


Figure A3: **Sagittal** slice of delivered dose distribution and the dose accumulation errors in grays for the **rectal filling** motion pattern. Shown are the results for no MCDA, original optical flow, improved optical flow, EVolution, incompressible EVolution, and spatially adaptive EVolution combined with direct dose mapping (DDM) or energy per mass transfer (EMT) to warp the dose. The prostate (center) and rectum (right) contours are shown in black. A large influence of the choice of dose warping method can be observed.

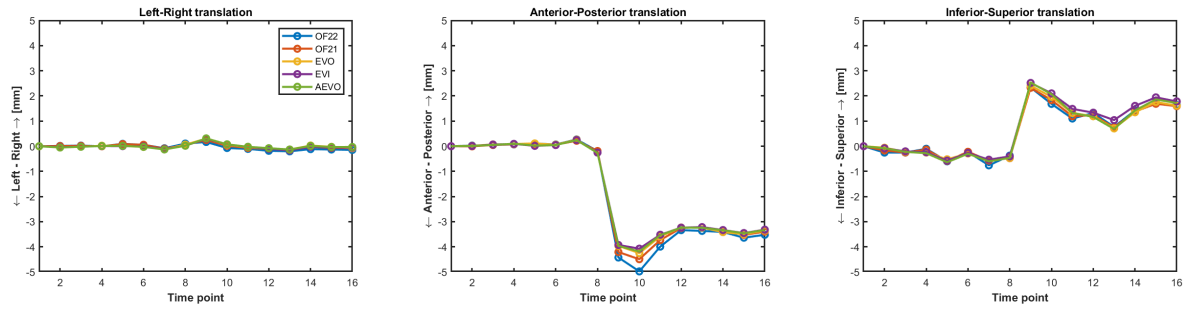


Figure A4: Average displacement of the prostate as estimated by the five DIR algorithms for the clinical dataset. There is a sudden shift in the position of the prostate in the anterior (4-5 mm) and superior (3 mm) directions between the 8th and 10th time point. Thereafter it slowly drifts back in the posterior (0.5-1.5 mm) and inferior (1 mm) directions.

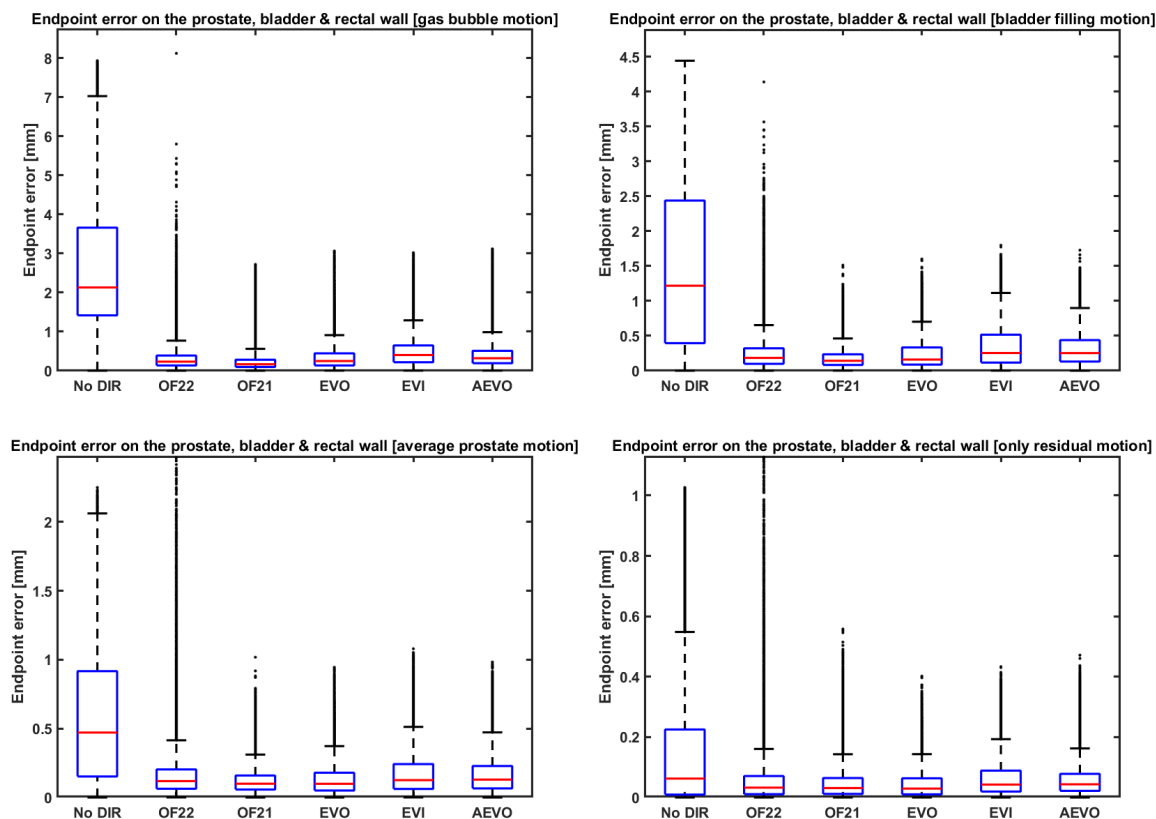


Figure A5: Image registration performance in terms of endpoint errors in millimeters on the prostate, bladder, and rectal wall for the different DIR algorithms and motion patterns. From left to right in each graph are shown the results for performing no deformable image registration, the original optical flow, the improved optical flow, EVOLution, incompressible EVOLution, and spatially adaptive EVOLution. All algorithms lower the median endpoint error considerably, from a factor of 5-13 for the **rectal filling** motion to a factor of 1.4-2.1 for residual motion. Note that the y-axis is scaled differently for the four different motion patterns.