Technical Note: Monte Carlo methods to comprehensively evaluate the robustness of 4D treatments in proton therapy

Kevin Souris[1], Ana Barragan Montero[1], Guillaume Janssens[2], Dario Di Perri[1], Edmond Sterpin[1,3], John A. Lee[1]

1. Center for Molecular Imaging and Experimental Radiotherapy, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, 1200 Brussels, Belgium.

2. Ion Beam Applications SA, 1348 Louvain-la-Neuve, Belgium.

3. Department of Oncology, Katholieke Universiteit Leuven, 3000 Leuven, Belgium.

Corresponding author: Kevin Souris <kevin.souris@uclouvain.be>

Abstract

Introduction: Proton therapy is very sensitive to treatment uncertainties. These uncertainties can induce proton range variations and may lead to severe dose distortions. However, most commercial tools only offer a limited integration of these uncertainties during treatment planning. In order to verify the robustness of a treatment plan, this study aims at developing a comprehensive Monte Carlo simulation of the treatment delivery, including the simulation of setup and range errors, variation of the breathing motion, and interplay effect.

Method: Most clinically relevant uncertainties have been modeled and implemented in the fast Monte Carlo dose engine MCsquare. Especially, variation of the breathing motion is taken into account by deforming the initial 4DCT series and generating multiple new 4DCT series with scaled motion. Systematic and random errors are randomly sampled, following a Monte Carlo approach, to generate individual erroneous treatment scenarios. The robustness of treatment plans is analyzed and reported with DVH-bands. The statistical uncertainty coming from the Monte Carlo

scenario sampling is studied.

Results: A validation demonstrated the ability of the motion model to generate new 4DCT series with scaled motion amplitude and improved image quality in comparison to the initial 4DCT. The robustness analysis is applied to a lung tumor treatment. Considering the proposed uncertainty model, the simulation of 300 treatment scenarios was necessary to reach an acceptable level of statistical uncertainty on the DVH-band.

Conclusion: A comprehensive and statistically sound method of treatment plan robustness verification is proposed. The uncertainty model presented in this paper is not specific to protons and can also be applied to photon treatments. Moreover, the generated 4DCT series, with scaled motion, can be imported in commercial TPSs.

Keywords: robustness, treatment uncertainties, Monte Carlo, interplay effect, proton therapy

1. Introduction

Proton therapy can potentially offer better dose conformity to the target and a reduced integral dose, compared with conventional radiotherapy. However, the high dose gradient in the Bragg peak also increases the sensitivity of the dose distribution to treatment uncertainties. Especially, the actual proton range in the patient may vary from the range calculated during treatment planning due to the inaccuracies along the treatment planning workflow [1]. Range uncertainties hinder harnessing the full potential of the steep distal dose fall-off to spare organs-at-risk and represent currently a major limitation for clinical indications of proton therapy.

Patient setup errors and conversion of CT units into stopping powers largely contribute to treatment uncertainties. Moreover, the dose calculation algorithm is another important source of proton range uncertainties. Indeed, the accuracy of the analytical dose calculation typically employed by most treatment planning systems (TPS) becomes inaccurate in highly heterogeneous geometries, such as lung tissues. In contrast, Monte Carlo dose calculation yields for the price of computation speed a higher dosimetric accuracy thanks to realistic simulation of particle interactions using direct random sampling of the physics laws [2-3].

Motion further complicates the situation of knowing the exact stopping location of ions [4-6]. The breathing motion causes the geometrical shift of the target with respect to the dose distribution. In proton therapy, the dose distribution might also be distorted due to the proton range variation caused by this motion. Moreover, dynamic delivery techniques, such as proton pencil beam scanning (PBS), may interfere with organ motion and distort the dose distributions. This effect is generally referred to as interplay effect.

A treatment plan is considered robust to a given level of uncertainty when the resulting dose distribution still satisfies the clinical goals, even in the presence of that uncertainty. Safety margins recipes defining a Planned Target Volume (PTV) have been historically used in photon therapy to ensure target coverage in the presence of geometrical uncertainties. The validity of these margin recipes relies on the static dose cloud approximation. However, this approximation is violated in proton therapy due to the finite range of charged particles. Adaptation of the PTV has been suggested by integrating range uncertainties along the beam direction within the margin, leading to the so-called beam-specific PTV [7-9]. However, the latter gets overall inadequate in the context of multi-field intensity modulated proton therapy (IMPT), due to possibly high in-field dose gradients. On the contrary, robust optimization can ensure plan robustness for multi-field IMPT by including uncertainties in the definition of the objective function [10-12]. However, most robust optimizers proposed in the literature and integrated in commercial software are limited to a few scenarios of systematic setup errors, a flat range uncertainty percentage, and a model of organ motion represented by a single 4DCT, neglecting the interplay effect. Moreover, the use of a single 4DCT only provides an estimation of the motion at the time it was acquired. Breathing motion may vary from one fraction to another. As this motion variation is not considered during planning, its effect on the delivered dose is unknown.

All these planning approximations and uncertainties in treatment delivery, and their limited integration in current robust optimizers, motivate the development of a tool to verify robustness in a comprehensive fashion. Such a robustness analysis should rely on realistic and comprehensive simulations of the entire space of relevant uncertainties. However, robustness

evaluation tools provided by commercial applications generally offer the same limited uncertainty model as for the treatment plan optimization. Setup and range errors,are typically simulated based on their extreme values, which does not necessarily guarantee the plan robustness for intermediate errors. Moreover, natural variations of the breathing motion should be incorporated together with a quantified assessment of the interplay effect. Effects of interplay have already been studied in the past [4, 13], but not in combination with all other treatment uncertainties. To assess the combined effect of these errors, all treatment uncertainty models should be mixed in a combinatorial and statistically consistent way. As a consequence, the number of treatment scenarios to be simulated to cover the error input space grows exponentially with the number of distinct input error sources. To be practical, such global assessment of the robustness would therefore require an adequate sampling of the space of possible scenarios, which is not trivial.

In this work, we developed a comprehensive and accurate robustness evaluation tool for proton therapy plans. Dose distributions are calculated for multiple uncertainty scenarios using MCsquare [3], an open source Monte Carlo dose engine. Most clinically relevant uncertainties are modeled and implemented directly in the Monte Carlo code. Both patient and PBS delivery motions are simulated. Variations in breathing amplitude and period are taken into account in order to provide a more comprehensive treatment simulation. A Monte Carlo approach is also adopted to select and analyze the uncertainty scenarios. The resulting dose distributions are combined and reported in DVH-bands.

2. Methods

To assess the robustness of a treatment plan, the dose distribution is computed for multiple

scenarios of treatment uncertainties. The dose calculation is performed by a fast Monte Carlo code presented in the following first sub-section. The modeling of various treatment uncertainties, including setup and range errors, as well as motion effects are then described. All error models are combined to simulate comprehensive treatment realizations. A Monte Carlo approach is then presented to randomly sample the uncertainty scenarios and analyze the robustness of the treatment plan. Lastly, the proposed solution is applied to a clinical case.

Monte Carlo dose calculation

In comparison to analytical algorithms, Monte Carlo dose calculations improve dose computation accuracy in heterogeneous geometries. However, a large number of particles must be simulated to reach an acceptable statistical precision, leading to long computation times. Recently, several fast Monte Carlo codes have been developed to address this issue [2, 3] and some commercial TPSs feature a Monte Carlo dose engine performing dose computations at practical speed.

In this study, MCsquare [3], a fast voxel-based multi-purpose Monte Carlo code was used. In order to enable accurate simulations, MCsquare was commissioned for a clinical beam line [14]. The transport algorithm was validated with the Fano test [15] and measurements in a heterogeneous phantom designed to highlight the multiple scattering process [16]. The beam model was further validated for clinical applications by comparing patient simulations calculated with MCsquare and Topas/Geant4 [17].

Density and material composition are assigned to each voxel of the CT image using the Schneider method [18]. The dose to medium resulting from the Monte Carlo simulation is converted on the fly to dose to water using Paganetti's technique [19]. Due to the stochastic nature of the Monte

Carlo process, a statistical uncertainty affects the resulting dose distributions and makes them noisy. The batch method was used to estimate the statistical precision of Monte Carlo results [2]. In this study, Monte Carlo dose calculations were performed with a maximum statistical noise of 1% inside the target.

Commonly used uncertainty models

Most robust optimization tools and commercial TPSs provide models of setup and range errors [10, 20]. These commonly used models are implemented in MCsquare. Setup errors are simulated by translating the patient image used for dose calculation. Range errors result from image conversion to stopping powers and other effects [1]. They are modeled by applying a global scaling of the mass densities extracted from the CT image.

Breathing motion is typically represented by a 4DCT, composed of multiple 3D images of the patient at various phases of the motion cycle. The impact of the moving anatomy on dose distributions can be studied with a 4D dose calculation algorithm, which computes the doses for the multiple phases of the motion cycle and accumulates them at a reference position. For this purpose, a diffeomorphic deformable registration algorithm, called Morphon [21] and implemented in OpenReggui, was used to register all 4DCT phases to a reference phase. MCsquare was then used for the 4D Monte Carlo dose calculation. Dose distributions are computed on all phases of the 4DCT and deformed to the reference phase using the pre-generated deformation fields. Lastly, deformed doses are averaged in a 3D image representing the dose absorbed by the moving tissues.

Improvements of 4DCT image quality

Dose calculation based on 4DCT is subjected to additional uncertainties due to the artifacts contained in each phase image. Especially, noisy images can lead to systematic range errors in proton therapy [22]. This sub-section describes how velocity fields obtained from the diffeomorphic deformable registration can be manipulated to generate new 4DCT series. By definition, a diffeomorphic vector field is invertible. It is also differentiable, as well as its invert. Therefore, unlike the commonly used motion fields, the velocity fields produced by the diffeomorphic algorithm, can be easily inverted, scaled, or combined, while maintaining physical consistency of deformations. In particular, rotations are preserved. After manipulation, velocity fields must then be converted to regular motion fields for image deformation purpose as explained in Janssens et al. [21].

The first step consists in registering all phases to a reference phase (phase 1 for instance, as illustrated in Figure 1.a). Multiple velocity fields can then be combined to generate intermediate or fictitious phases. For instance, the first phase (P1) can be deformed to the time-averaged mean position of the tumor (Mid-Position) using the average of all velocity fields, as illustrated in Figure 1.b. The other phases can also be deformed to the Mid-Position by subtracting their original phase field from the average field. The Mid-Position CT image (MidP-CT) [23] is then generated by computing the median HU value in each voxel over all deformed phases. This results in a CT image with less noise and fewer artifacts than the nominal 4D phases. The deformation can easily be inverted by taking the opposite of the velocity field. An improved 4DCT can therefore be reconstructed by deforming the newly created MidP-CT image back to each phase position using the inverted velocity fields as illustrated in Figure 1.c.

Model of the motion amplitude variation

Intra- and inter-fraction variations of the breathing motion have been reported in the literature [5, 6]. A new 4DCT series, with scaled motion, can be created to verify the robustness of the treatment plan against this effect. Similarly to the previous sub-section, the new 4DCT series is generated by deforming the MidP-CT, or any other reference phase. However, the inverted velocity fields are now scaled by a scalar factor in order to increase or decrease the motion amplitude as illustrated in Figure 1.d.

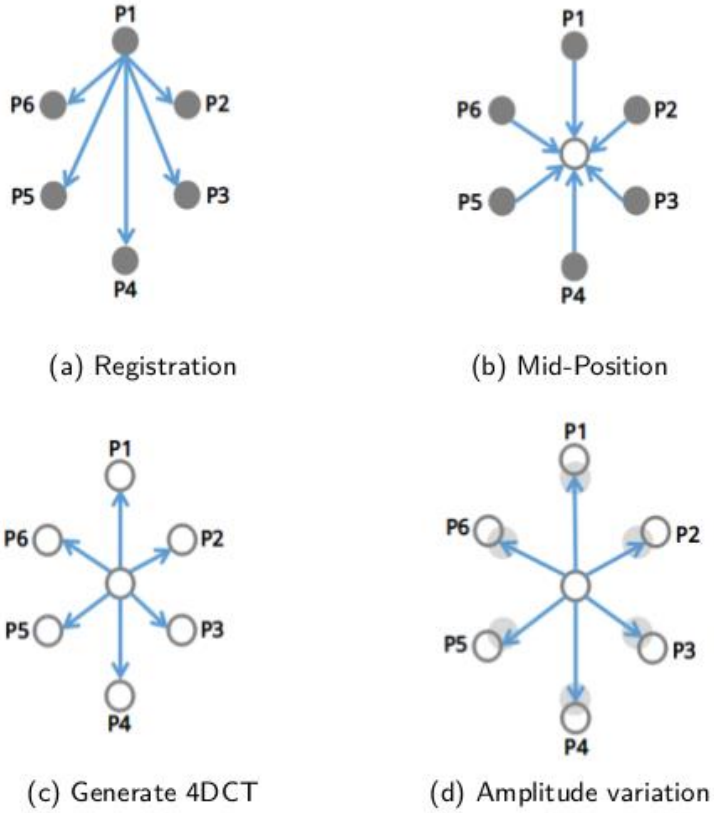

Figure 1: Illustration of the main steps of the 4DCT reconstruction with scaled motion amplitude: (a) a deformable image registration algorithm is used to generate the velocity fields from one phase to all the others; (b) the MidP-CT image is reconstructed by deforming all phases to the Mid-Position; (c) the MidP-CT is deformed back to each phase position in order to reconstruct a new 4DCT with improved image quality; (d) velocity fields are scaled to

generate a 4DCT with different motion amplitude. Tumor position at each phase (P1 to P6) is represented by the circles. Velocity fields are illustrated by the blue vectors.

The model was evaluated using two 4DCT series that were acquired on the same lung patient over a one-week interval. The breathing amplitude variation between both 4DCT series was calculated with the ratio of their maximum GTV displacement. The proposed model was then used to reconstruct the second 4DCT from the first week series. To do so, the first week velocity fields were scaled with the breathing amplitude variation factor measured between both acquisitions. The first week MidP-CT is then deformed to generate the new 4DCT, with scaled motion. The GTV displacement of the generated 4DCT is compared with the real second week 4DCT series to verify the model. For this comparison, to remove any possible setup differences and baseline shifts, the images of week 1 and week 2 were rigidly registered to align the GTV position in the Mid-Position.

Simulation of interplay effect

To assess the dose degradation due to potential interplay effects, the 4D dose calculation should take into account the dynamic of the delivery equipment. The simulation of the interplay effect has already been described in the literature [4, 24, 25], using various methods to reconstruct the delivery timing information. In our implementation, the IBA software ScanAlgo is employed to simulate the delivery process and extract the delivery time and duration for each spot of the PBS plan. Each spot can then be associated with a phase of the 4DCT in which it would be delivered. The breathing period and the starting delivery time within the motion cycle are varied for each

beam and fraction. According to the delivery timing simulation and these breathing parameters all spots are sorted and assigned to a phase of the 4DCT. A partial plan is then generated for each phase by merging all corresponding spots. Instead of computing the full plan on all 4DCT phases, like for the regular 4D dose calculation, partial plans are simulated by MCsquare on their appropriate CT phase. The resulting partial doses are then deformed to the reference phase and accumulated, providing a 3D dose map that includes both organ and delivery motion effects.

Sampling of uncertainty scenarios

A Monte Carlo approach is proposed in this study to overcome the challenges of scenario selection. Instead of mixing uncertainty models in a combinatorial way, the errors are randomly sampled for each uncertainty model. Each simulated experiment represents a full treatment simulation and is composed of multiple daily error scenarios in order to take into account inter-fraction variability. Each fraction is therefore simulated individually with MCsquare, considering systematic errors that are sampled once per full treatment experiment, and random errors, re-sampled for each fraction scenario. The resulting fraction doses are then summed together to obtain the total experiment dose that represents one of the possible full treatment realizations. This process has to be repeated many times to provide a representative set of full treatment experiments.

More specifically, in our implementation, the three spatial components of the systematic and random setup errors are randomly sampled individually from Gaussian distributions with zero mean and user defined standard deviations as illustrated in Table 1. The resulting shift is applied for all beam directions delivered at each fraction. Similarly, the global range uncertainty, the

organ motion period and amplitude are sampled from Gaussian distribution centered on their nominal value, with standard deviations defined by the user according to variations reported in the literature [1, 5, 6, 26]. The sampled global range uncertainty (in percents) is considered to be the same for all spots of the treatment plan. The initial phase involved in the simulation of the interplay effect is randomly sampled using a continuous uniform distribution from the beginning to the end of the breathing cycle.

Robustness analysis

The impact of treatment uncertainties on the dose distribution is usually reported with an uncertainty band around the nominal DVH. This band represents the envelope of all DVHs computed by taking into account multiple experiments of treatment uncertainties. However, in our Monte Carlo strategy, the sampling of treatment uncertainties can lead to very large errors due to the infinite tails of Gaussian distributions. Generally, safety margins recipes are applied in such a way that target coverage is ensured for a reasonable number of patients, typically 90% as suggested in van Herk et al [27]. To mimic this approach, uncertainty bands are generated by computing percentiles of the experiment DVHs as proposed by Gordon et al [28]. For instance, the 5^{th} and 95^{th} percentiles would define the envelope of the dose degradation expected for 90% of the treatment errors.

In this dual Monte Carlo strategy, both dose calculation and scenario sampling lead to statistical uncertainties on the final result. The number of particles simulated during the Monte Carlo dose calculation is adapted to reach a maximum statistical uncertainty of 1% after the 4D dose accumulation and fraction summation. The robustness analysis resulting from the random

scenario sampling is also subject to statistical uncertainties. If the robustness analysis is performed multiple times by randomly re-sampling the scenarios, the DVH-bands may vary from one run to another, even when the dose calculation statistical uncertainty is negligible. Specifically, the estimation of the 5^{th} and 95^{th} percentiles that determine the DVH-band is very sensitive to the rare events in the tails of the scenario distribution. This means that the edges of the DVH-band converge slowly when increasing the number of simulated scenarios. The bootstrap method [29, 30] is employed in this study to estimate the percentile values and their confidence interval based on a limited number of scenario samples.

Patient case

The robustness test presented in this paper is applied to a patient case in order to illustrate the feasibility of this concept. The MIROpt 4D robust optimizer [31], developed in house and based on MCsquare dose calculation, was employed to generate a treatment plan for a lung tumor patient. A robust plan was optimized on the CTV based on the 4DCT series composed of 10 CT phases. All robustness parameters used for the optimization are listed in Table 1 (first column). The worst-case scenario approach [11] was used for robust optimization, considering 21 scenarios to simulate combinations of systematic setup and range errors. The blurring effect of random setup errors is included in the beamlet MC dose calculation. The initial position of each particle is randomly shifted to simulate random setup errors, assuming an infinite number of fractions [20]. The variation of the breathing amplitude and period was not taken into account during the treatment plan optimization. The resulting PBS plan is composed of 2 beams (150° and 180°) with a total of 2536 spots.

The generated treatment plan was then evaluated using the robustness test method proposed in this paper. Setup errors, range errors, motion amplitude variation, interplay effect, and motion period variation were simulated. All robustness parameters used for this test are listed in Table 1 (second column). A nominal breathing period of 4 seconds was considered. Each treatment simulation is composed of 35 fractions.

Treatment uncertainties	**Planning**	**Robustness verification**
Systematic setup errors [26]	±2.5 × (0.4; 0.7; 0.6)	σ = (0.4; 0.7; 0.6) mm
Random setup errors [26]	σ = (1.7; 1.4; 1.4) mm	σ = (1.7; 1.4; 1.4) mm
Systematic range errors [1]	±2.4 %	σ = 1.6 %
Systematic motion amplitude errors [6]	*NA*	σ = 0 %
Random motion amplitude errors [5]	*NA*	σ = 36 %
Systematic motion period errors [6]	*NA*	σ = 0 %
Random motion period errors [6]	*NA*	σ = 17 %

Table 1: Summary of uncertainty parameters used for the robust optimization and the robustness evaluation of the treatment plan.

The bootstrap method is used after the computation of each experiment to estimate the DVH-bands and their convergence. The 5^{th} and 95^{th} percentiles constituting the DVH-band edges are calculated over 500 bootstrap samples randomly sampled from the pool of already computed scenarios. A 95% confidence interval for the DVH-bands is also computed with the bootstrap technique [29, 30].

Treatment uncertainties were directly sampled by MCsquare during the scenario simulations. In order to reach a maximum noise level of 1% inside the ITV, 36 million primary particles were simulated per experiment. All computations were performed on dual-processor computer (Intel Xeon E5-2660 v4) running Linux.

3. Results

The model of motion amplitude variation is first evaluated with a clinical example. Next, the robustness analysis approach presented in this paper is applied to a lung patient in order to demonstrate the feasibility of this concept.

Evaluation of motion amplitude variation model

First, a new 4DCT with improved image quality was reconstructed by deforming the reference image (MidP-CT) to each phase position. As it can be observed in Figure 2, the method can reproduce the organ and target positions of the original 4DCT while reducing noise and artifacts in the generated 4DCT. Importantly, the visible tumor in the CT phase remains inside its GTV contour after reconstruction.

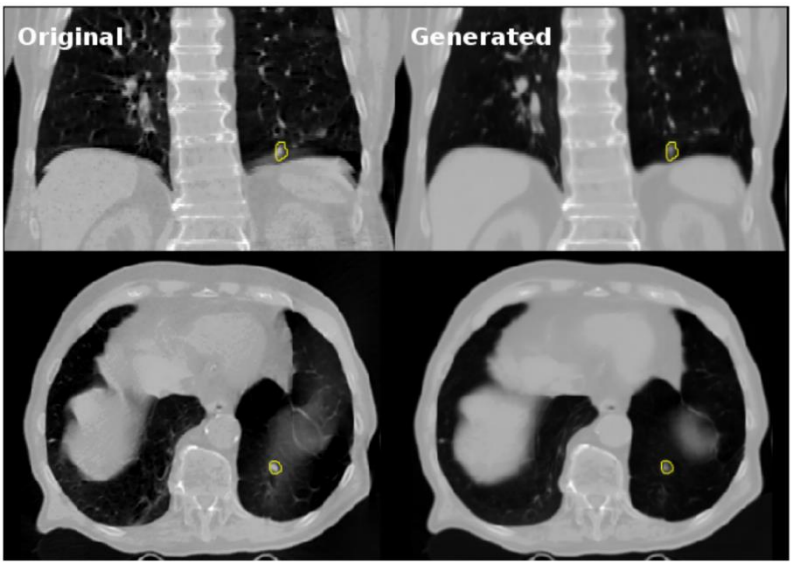

Figure 2: Comparison of the density maps calculated by MCsquare from the 8th phase of the original 4DCT (left) and generated 4DCT (right). The GTV drawn on the original CT phase is displayed in yellow on both images.

Next, the motion measured in the second week 4DCT series was reproduced by deforming the first week series using the proposed method. The motion hysteresis representing the position of the GTV center of mass at each phase is displayed in Figure 3a for both 4DCT acquisitions. From these data, maximum GTV displacements of 33.59 mm and 38.08 mm were calculated for week 1 and week 2, respectively. This represents an increase of 13.4% of the target motion amplitude. The first-week MidP-CT and its associated GTV contour were deformed using the scaled velocity fields to reproduce the motion amplitude measured at week 2. The motion hysteresis of the adjusted week 1 images is compared to week 2 in Figure 3b. The motion amplitude of the new generated 4DCT is 37.82 mm. The difference between the modeled and the expected motion amplitude is therefore 0.68%.

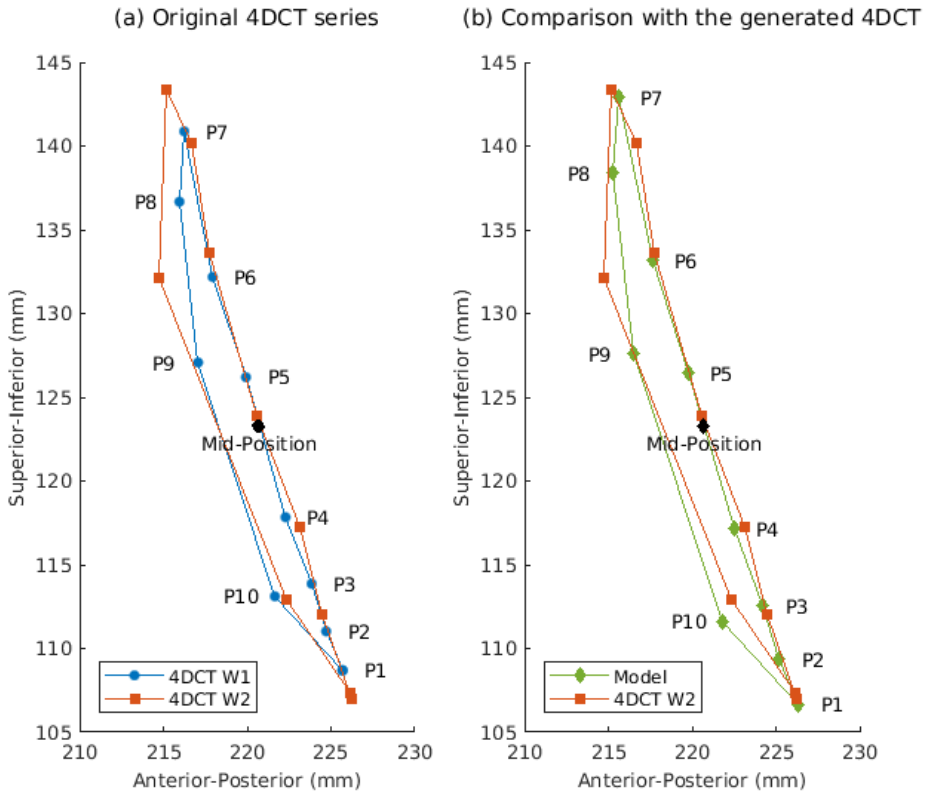

Figure 3: Motion hysteresis represented by the position of the GTV center of mass for each phase (P1 to P10). a) Comparison of the GTV motion extracted from original 4DCT series

acquired at week 1 and week 2. b) The motion amplitude measured at week 2 is reproduced by the model using week 1 images.

Treatment plan robustness evaluation

The robustness of the lung treatment plan was evaluated using our Monte Carlo approach. The DVH-bands are displayed in Figure 4. DVH metrics are reported in Table 2 for the nominal 4D dose calculation (without treatment uncertainties) and for the uncertainty band and its median curve.

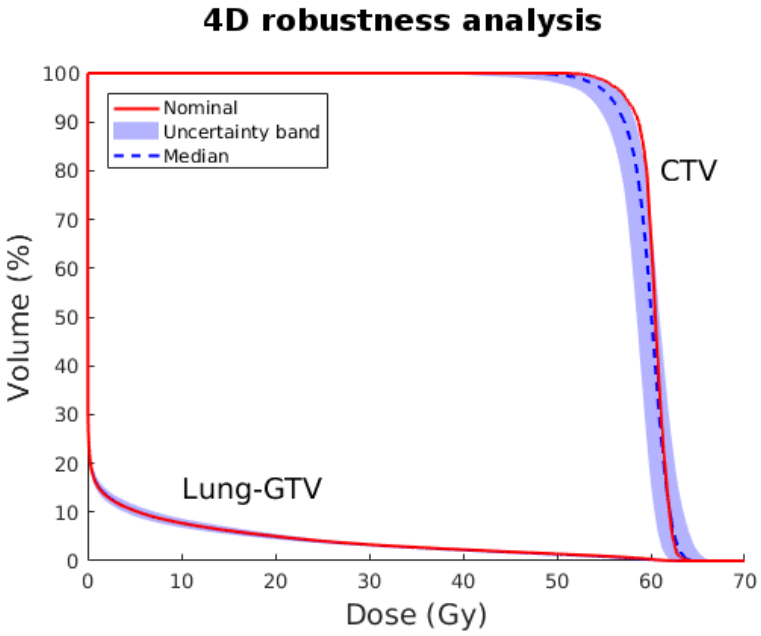

Figure 4: DVH-bands resulting from the robustness evaluation of the lung treatment plan. The light blue band contains 90% of the uncertainty scenarios. The dark blue DVH is the median of all scenarios. Uncertainty scenarios include the simulation of setup and range errors, the variation of motion amplitude, and the interplay effect with variable breathing motion period. The red curve represents the nominal plan without uncertainties.

	CTV D95	CTV D5	Lung-GTV
Nominal DVH	57.6 Gy	62.4 Gy	2.7 Gy
Uncertainty band	53.4 - 57.3 Gy	60.9 - 64.6 Gy	2.4 - 3.0 Gy
Median	55.7 Gy	62.5 Gy	2.7 Gy

Table 2: DVH metrics resulting from the robustness test.

A total of 1000 full treatment experiments, each composed of 35 fractions, were simulated to generate the DVH band results. It is, however, possible to stop the simulation earlier, when the statistical uncertainties on the DVH-band reach an acceptable level. The convergence of the uncertainty band at the D95 level is shown in Figure 5. The results look unstable below 300 scenarios, and converge slowly beyond this point. The statistical uncertainty is higher for the 5th percentile (left edge of the band). Its 95% confidence interval is 0.69 Gy wide for 300 scenarios, 0.49 Gy wide for 500 scenarios, and 0.37 Gy wide for 1000 scenarios.

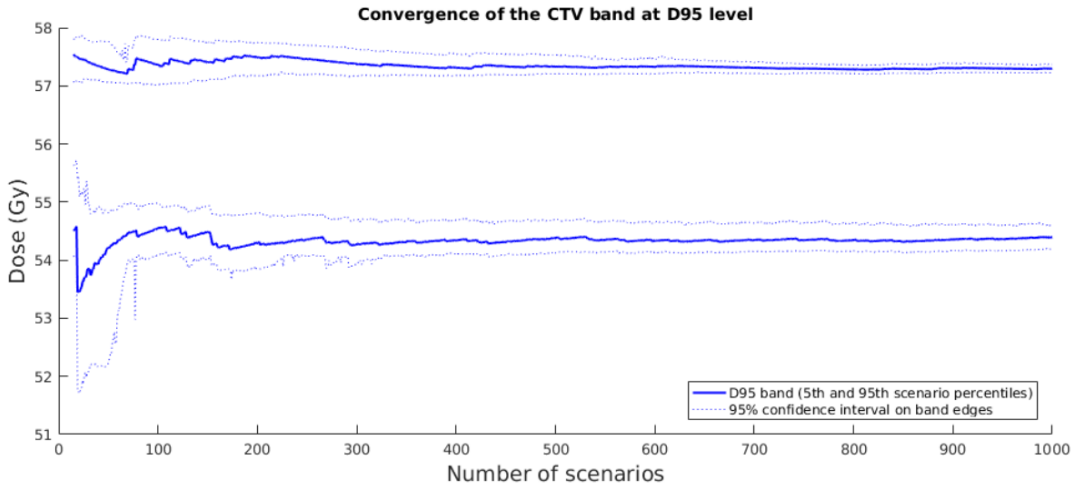

Figure 5: Variability of the uncertainty band edges and their 95 % confidence interval with respect to the number of simulated scenarios.

The computation time was measured and analyzed. The pre-processing step, consisting in registering the 4DCT phases, generating the MidP-CT, and estimating the spot delivery time, is performed automatically by a workflow in the OpenReggui application and takes 12 minutes. The

dose calculation and uncertainty models are performed with MCsquare. The computation of the nominal dose distribution takes 97 seconds for this particular lung treatment plan. An average computation time of 385 seconds is then required to simulate each uncertainty experiment. This time includes the generation of a new 4DCT (22% of computation time) with scaled motion amplitude for each fraction, other uncertainty models (computation time negligible), Monte Carlo dose calculation (49%), and accumulations of 4D and fraction doses (29%). For this example lung case, the simulation of 300 scenarios was sufficient to reach an acceptable noise level. The total computation time necessary to perform the robustness test with 300 scenarios is 32 hours on a single computer. A total of 1000 scenarios was simulated (107 hours) in this study to demonstrate the slow convergence of the DVH-band result.

4. Discussion

In this study, a comprehensive and accurate robustness verification of proton therapy treatment plans is presented. A novel method for the simulation of breathing motion variation was implemented and combined with the conventional setup and range errors, as well as the simulation of the interplay effect. Systematic and random errors are randomly sampled for each treatment scenario, following a Monte Carlo approach. The variable motion amplitude model was first validated, and the robustness analysis was illustrated for a lung tumor treatment plan.

The validation results demonstrated that the variable motion amplitude model is able to reproduce the motion measured on a second 4DCT acquisition by scaling the velocity fields and deforming the reference CT image. The difference between the motion amplitude measured on the generated 4DCT series and the expected one is 0.68%. This difference may be due to the

inaccuracies during the registration or the motion amplitude measurements. The accuracy of the registration and deformation algorithms and their impact on this particular application still need to be studied.

As the reference phase is reconstructed from all 4DCT phases, it is usually less noisy and contains fewer artifacts than the original 4DCT phases. Therefore, the generated 4DCT series, with scaled motion, also benefit from this improved image quality. For this motion reduction/amplification model, the choice of the reference phase is important. For instance, using the Mid-Position phase, as presented in this paper, will lead to a symmetric motion amplification, extending both end-inhale and end-exhale positions. On the contrary, starting from the end-exhale phase, the motion will be amplified only in the inhale direction. This asymmetric motion variation may be preferred if a better stability of the end-exhale position is observed in the patient.

The uncertainty models presented in this paper are not specific to protons and can also be applied to photon treatments, like, for instance, the variation in breathing amplitude, which is also an issue in conventional radiotherapy. Moreover, the generated 4DCT series can be imported in a commercial TPS.

Treatment plan optimization requires to compute the dose distribution for each spot individually, increasing the computation time and the memory requirement significantly. To simplify the problem, uncertainty scenarios must therefore be limited during robust optimization. The probabilistic approach described in this paper would therefore not be practical for robust optimization. In consequence, it is important to verify the actual robustness of the resulting treatment plan using more realistic and comprehensive models as presented in this technical

note.

For the studied case, the computation time was measured at 385 seconds per full treatment experiment. Moreover, this time must be multiplied by a large number of experiment to reduce the statistical uncertainty of the resulting DVH-band. For a clinical application it is crucial to further optimize the algorithms and their implementation in order to reduce the computation time. The performances can be improved by working on two aspects. First, the computation time per scenario can be reduced by further working on the algorithm implementation. For instance, the generation of 4DCT and dose accumulations represent half of the computation load and mainly consist of image deformations. These operations could be performed by a GPU while the CPU already starts the next Monte Carlo simulation. A cluster of computers could also be used to simulate multiple scenarios simultaneously. The second possible improvement is to reduce the number of experiments that are required to achieve an acceptable statistical uncertainty. As for every Monte Carlo algorithm, variance reduction techniques can be applied. For instance, the uncertainty values are currently sampled from Gaussian distributions, favoring the sampling of small error values. However, the percentiles that define the DVH-band are more sensitive to extreme events in the tails of the scenario distribution. To improve the scenario sampling efficiency, uncertainty values could be sampled from uniform distributions with a corrected weighting factor in the percentile estimation. Another improvement could also be adopted for the simulation of the random setup error. Fredriksson [32] demonstrated that for a sufficient number of fractions, the random errors can be simulated by sampling a random shift for each primary particle during the Monte Carlo transport.

5. Conclusion

This study proposes a comprehensive and statistically sound method of robustness verification for scanned proton therapy treatment plans. It enables a more accurate evaluation of the treatment plans generated with approximate robust optimization techniques implemented in commercial TPS. A Monte Carlo sampling approach was employed to calculate individual dose error scenarios. In our uncertainty model, we considered the variability of the breathing motion, the interplay effect, and the conventional setup and range errors. The proposed methodology was tested on a lung patient case. These tools are released open source within the MCsquare application [33].

The variable breathing amplitude is simulated by generating new 4DCT series with scaled motion. These 4DCT images can be easily imported in a commercial TPS for robustness optimization or verification of photon or proton treatment plans.

6. Acknowledgments

Kevin Souris is supported by a research grant from Ion Beam Application (IBA s.a., Louvain-la-Neuve, Belgium). J. A. Lee is a Research Associate with the Belgian fund of scientific research (F.R.S.-FNRS). A. Barragan Montero is funded by the Baillet-Latour Funds.

7. Bibliography

[1] Paganetti H. Range uncertainties in proton therapy and the role of Monte Carlo simulations. Phys Med Biol. 2012; 57(11):99–1172.

[2] Jia X., Schümann J., Paganetti H., Jiang S. B. GPU-based fast Monte Carlo dose calculation for proton therapy. Physics in Medicine & Biology. 2012; 57(23), 7783.

[3] Souris K., Lee J. A., Sterpin E. Fast multipurpose Monte Carlo simulation for proton therapy using multi- and many-core CPU architectures. Med Phys. 2016; 43(4):1700.

[4] Lin L., Souris K., Kang M., et al. Evaluation of motion mitigation using abdominal compression in the clinical implementation of pencil beam scanning proton therapy of liver tumors. Medical physics. 2017; 44(2), 703-712.

[5] Bissonnette J. P., Franks K. N., Purdie T. G., et al. Quantifying interfraction and intrafraction tumor motion in lung stereotactic body radiotherapy using respiration-correlated cone beam computed tomography. Int. J. Radiat. Oncol. Biol. Phys. 2009; 75(3):688–695.

[6] Bosmans G., van Baardwijk A., Dekker A., et al. Intra-patient variability of tumor volume and tumor motion during conventionally fractionated radiotherapy for locally advanced non-small-cell lung cancer: a prospective clinical study. Int. J. Radiat. Oncol. Biol. Phys. 2006; 66(3):748–753.

[7] Lin L., Kang M., Huang S., et al. Beam-specific planning target volumes incorporating 4D CT for pencil beam scanning proton therapy of thoracic tumors. J Appl Clin Med Phys. 2015; 16(6):5678.

[8] Park P. C., Zhu X. R., Lee A. K., et al. A beam-specific planning target volume (PTV) design for proton therapy to account for setup and range uncertainties. Int. J. Radiat. Oncol. Biol. Phys. 2012; 82(2):e329–336.

[9] Knopf A. C., Boye D., Lomax A., Mori S. Adequate margin definition for scanned particle therapy in the incidence of intrafractional motion. Phys Med Biol. 2013; 58(17):6079–6094.

[10] Fredriksson A., Forsgren A., Hardemark B. Minimax optimization for handling range and setup uncertainties in proton therapy. Med Phys. 2011; 38(3):1672–1684.

[11] Pflugfelder D., Wilkens J. J., Oelfke U. Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy. Phys Med Biol 2008; 53(6):1689–1700.

[12] Unkelbach J., Bortfeld T., Martin B. C., Soukup M. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. Med Phys. 2009; 36(1):149–163.

[13] Knopf A. C., Hong T. S., 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(22):7257–7271.

[14] Grevillot L., Bertrand D., Dessy F., Freud N., Sarrut D. A Monte Carlo pencil beam scanning model for proton treatment plan simulation using GATE/GEANT4. Phys Med Biol. 2011; 56(16):5203–5219.

[15] Sterpin E., Sorriaux J., Souris K., Vynckier S., Bouchard H. A Fano cavity test for Monte Carlo proton transport algorithms. Med Phys. 2014; 41(1):011706.

[16] Sorriaux J., Testa M., Paganetti H., et al. Experimental assessment of proton dose calculation accuracy in inhomogeneous media. Phys Med. 2017; 38:10–15.

458 [17] Huang S., Kang M., Souris K., et al. Validation and clinical implementation of an accurate
459 Monte Carlo code for pencil beam scanning proton therapy. J Appl Clin Med Phys. 2018; 19(5),
460 558-572.

461 [18] Schneider W., Bortfeld T., Schlegel W. Correlation between CT numbers and tissue
462 parameters needed for Monte Carlo simulations of clinical dose distributions. Phys Med Biol.
463 2000; 45(2):459–478.

464 [19] Paganetti H. Dose to water versus dose to medium in proton beam therapy. Phys Med Biol.
465 2009; 54(14):4399–4421.

466 [20] Fredriksson, A. A characterization of robust radiation therapy treatment planning
467 methods—from expected value to worst case optimization. Medical physics. 2012; 39(8):
468 5169-5181.

469 [21] Janssens G., Jacques L., Orban de Xivry J., Geets X., Macq B. Diffeomorphic registration of
470 images with variable contrast enhancement. Int J Biomed Imaging. 2011; 2011:891585.

471 [22] Brousmiche, S., Souris, K., de Xivry, J. O., Lee, J. A., Macq, B., Seco, J. (2017). Combined
472 influence of CT random noise and HU-RSP calibration curve nonlinearities on proton range
473 systematic errors. Physics in Medicine & Biology, 62(21), 8226.

474 [23] Wolthaus J. W., Sonke J. J., van Herk M., et al. Comparison of different strategies to use
475 four-dimensional computed tomography in treatment planning for lung cancer patients. Int. J.
476 Radiat. Oncol. Biol. Phys. 2008; 70(4):1229–1238.

477 [24] Dowdell S., Grassberger C., Sharp G. C., Paganetti H. Interplay effects in proton scanning for
478 lung: a 4D Monte Carlo study assessing the impact of tumor and beam delivery parameters. Phys
479 Med Biol. 2013; 58(12):4137–4156.

480 [25] Dowdell S., Grassberger C., Paganetti H.. Four-dimensional Monte Carlo simulations
481 demonstrating how the extent of intensity-modulation impacts motion effects in proton therapy
482 lung treatments. Med Phys. 2013; 40(12):121713.

483 [26] Bissonnette J. P., Purdie T. G., Higgins J. A., Li W., Bezjak A. Cone-beam computed
484 tomographic image guidance for lung cancer radiation therapy. Int. J. Radiat. Oncol. Biol. Phys.
485 2009; 73(3):927–934.

486 [27] van Herk M., Remeijer P., Rasch C., Lebesque J. V. The probability of correct target dosage:
487 dose-population histograms for deriving treatment margins in radiotherapy. Int. J. Radiat. Oncol.
488 Biol. Phys. 2000; 47(4):1121–1135.

489 [28] Gordon J. J., Sayah N., Weiss E., Siebers J. V. Coverage optimized planning: probabilistic
490 treatment planning based on dose coverage histogram criteria. Medical physics. 2010; 37(2),
491 550-563.

492 [29] Efron B., Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other
493 measures of statistical accuracy. Statistical science. 1986; pages 54–75.

494 [30] Vaisovich Vishnyakov B., Ivanovich Kibzun A. Application of the bootstrap method for
495 estimation of the quantile function. Automation and Remote Control. 2007; 68(11):1931–1944.

496 [31] Barragan Montero A. M., Souris K., Sanchez-Parcerisa D., Sterpin E., Lee J. A. Performance of
497 a hybrid Monte Carlo-Pencil Beam dose algorithm for proton therapy inverse planning. Medical
498 physics. 2018; 45(2), 846-862.

499 [32] Fredriksson A., Bokrantz R. A critical evaluation of worst case optimization methods for
500 robust intensity-modulated proton therapy planning. Medical physics. 2014; 41(8Part1), 081701.

501 [33] Souris K. *MCsquare: a fast Monte Carlo dose engine for proton PBS.*
502 http://www.openmcsquare.org (accessed February 2019)