TITLE PAGE

Title:

A comprehensive evaluation of the accuracy of CBCT and deformable registration based dose calculation in lung proton therapy

Authors:

Catarina Veiga[1], Guillaume Janssens[2], Thomas Baudier[3], Lucian Hotoiu[3], Sebastien Brousmiche[2], Jamie McClelland[4], Ching-Ling Teng[5], Lingshu Yin[5], Gary Royle[1] and Boon-Keng Kevin Teo[5]

Affiliations:

[1]Proton and Advanced RadioTherapy Group, Department of Medical Physics & Biomedical Engineering, University College London, London, WC1E 6BT, United Kingdom

[2]Ion Beam Applications SA, Louvain-la-Neuve, B1348, Belgium

[3]iMagX Project, ICTEAM Institute, Université Catholique de Louvain, Louvain-la-Neuve, B1348, Belgium

[4]Centre for Medical Image Computing, Department of Medical Physics & Biomedical Engineering, University College London, London, WC1E 6BT, United Kingdom

[5]Department of Radiation Oncology, University of Pennsylvania, Philadelphia, PA 19104, United States

Corresponding author:

Boon-Keng Kevin Teo PhD,
3400 Civic Center Boulevard, TRC 2 West
Department of Radiation Oncology
University of Pennsylvania
Philadelphia, PA, 19104

Abstract

The uncertainties in water equivalent thickness (WET) and accuracy of dose estimation using a virtual CT (vCT), generated from deforming the planning CT (pCT) onto the daily cone-beam CT (CBCT), were comprehensively evaluated in the context of lung malignancies and passive scattering proton therapy. The validation methodology utilized multiple CBCT datasets to generate the vCTs of twenty lung cancer patients. A correction step was applied to the vCTs to account for anatomical modifications that could not be modeled by deformation alone. The CBCT datasets included a regular CBCT (rCBCT) and synthetic CBCTs created from the rCBCT and rescan CT (rCT), which minimized the variation in setup between the vCT and the gold-standard image (i.e., rCT). The uncertainty in WET was defined as the voxelwise

difference in WET between vCT and rCT, and calculated in 3D (planning target volume, PTV) and 2D (distal and proximal surfaces). The uncertainty in WET based dose warping was defined as the difference between the warped dose and a forward dose recalculation on the rCT. The overall root mean square (RMS) uncertainty in WET was 3.6±1.8, 2.2±1.4 and 3.3±1.8 mm for the distal surface, proximal surface and PTV, respectively. For the warped dose, the RMS uncertainty of the voxelwise dose difference was 6±2% of the maximum dose (%mD), using a 20% cut-off. The rCBCT resulted in higher uncertainties due to setup variability with the rCT; the uncertainties reported with the two synthetic CBCTs were similar. The vCT followed by a correction step was found to be an accurate alternative to rCT.

Key words: adaptive radiotherapy, cone-beam CT, deformable image registration, lung, proton therapy

1. Introduction

The dosimetric advantage of protons over photons has been used to improve the poor outcome of lung cancer patients (DeSantis *et al* 2014, Wink *et al* 2014). Passive scattering proton therapy (PSPT) has predominantly been used as a robust strategy to deal with the intra-fractional motion of lung tumors; however, other inter-fractional changes during the course of radiotherapy may also affect the dose delivered to target and healthy tissues (Guckenberger *et al* 2011, Qin *et al* 2013). These factors include changes in tumor size and position, alterations in tissue anatomy, variations in respiratory patterns, and fluctuations in patient weight (Gomez and Chang 2011). Tumor regression was found to occur in conventional photon radiotherapy for 40% of the patients undergoing definitive treatment (van Zwienen *et al* 2008), with reductions of up 70% of their volume reported in the literature (Britton *et al* 2007). Lung changes during the course of treatment, such as increase or decrease of atelectasis (i.e., lung collapse) and pleural effusion (i.e., liquid accumulated between lungs and the ribs or diaphragm), are less frequent but dramatically modify the density locally. Additionally, movement of the tumor can evolve throughout the treatment fractions; therefore, re-evaluation of the targets may be required (Ozhasoglu and Murphy 2002). The position of the spread out Bragg peak (SOBP) is sensitive to changes in tissue density along the beam path, which may result in potential shifts of the SOBP and consequently to tumor underdosage and/or high doses delivered outside of the intended target. Inter-fractional adaptive replanning is therefore beneficial to select patients (Hui *et al* 2008). At our clinic (Roberts Proton Therapy Center, Philadelphia, PA), replanning is triggered in about 10-20% of the lung cancer patients treated with PSPT.

Gantry mounted cone beam computed tomography (CBCT) systems are now available for image-guidance in proton therapy. In comparison with repeat computed tomography (CT), CBCT has the advantage of providing three dimensional visualization of the patient's anatomy in treatment position; however, the lower image quality and poor consistency of the Hounsfield units (HU) undermine its direct usability for proton dose calculations (Stock *et al* 2009). Deformable image registration (DIR) can be used to create a virtual CT (vCT) (Peroni *et al* 2012) by deforming the CT onto the CBCT, and thus provide a solution for the issues with image quality for dose recalculation (Veiga *et al* 2014, Moteabbed *et al* 2015). vCTs have been shown to be a potential surrogate for repeat CTs for proton treatment verification in the context of head and neck malignancies (Landry *et al* 2015a, Landry *et al* 2015b, Veiga *et al* 2015, Kurz *et al* 2016); however, these retrospective studies used data from linear accelerator CBCT systems.

We previously proposed an adaptive radiotherapy (ART) workflow using on-board CBCT and DIR, where replanning is triggered after three decision-points (Veiga *et al* 2016): (1) an online evaluation of a fast range-corrected dose distribution based on a vCT, which may be followed by (2) a more accurate offline

dose recalculation on the vCT and in turn trigger (3) a rescan CT (rCT) for replanning. We critically assessed the clinical application of this ART workflow to generate indicators that prompt patient re-evaluation and replanning, and compared its performance to routine repeat CT. DIR validation is however a complex and extensive topic, which merits a separate discussion on its own. Therefore, in this companion study we provide procedures for in-house validation of CBCT and DIR based workflows in adaptive proton therapy, followed by a thorough and comprehensive analysis of the uncertainties of the workflow implemented in our clinic. Particularly, we describe and discuss different types of CBCT datasets that can be used for validation of CBCT based methods. This is the first study that validates DIR and CBCT for lung cancer malignancies in the context of proton therapy that uses CBCT acquired on a proton-gantry.

2. Materials and methods

2.1. Design of the gantry mounted CBCT system for proton therapy

The gantry mounted CBCT system (Ion Beam Applications (IBA) SA, Ottignies-Louvain-la-Neuve, BE) was first used clinically at the Roberts Proton Therapy Center in September 2014 (Fig. 1) (Seabra *et al* 2014, Brousmiche *et al* 2014). This was the first IBA CBCT clinical system and was not part of the original proton room solution, but rather retrospectively fitted by updating the existing kV imaging system. Modifications to the orthogonal kV X-ray imaging system for CBCT were made by upgrading one X-ray tube to the GS-2075 (Varian Medical Systems, Palo Alto, CA) and one flat panel imager to the Pixium RF 4343 with CsI scintillator (Thales, Neuilly-sur-Seine, France). Image acquisition can be performed with either full or half gantry rotation at up to one revolution per minute. Image reconstruction was performed using the FDK algorithm (Feldkamp *et al* 1984) implemented in the open-source Reconstruction Toolkit (RTK) (Rit *et al* 2014). The system has a source-to-axis distance of 288.4 cm, detector-to-axis distance of 58.6 cm and a maximum field of view (FoV) of 34 cm in diameter and length. In this CBCT system, the panel was not designed to be offset laterally and this limited the CBCT FoV.

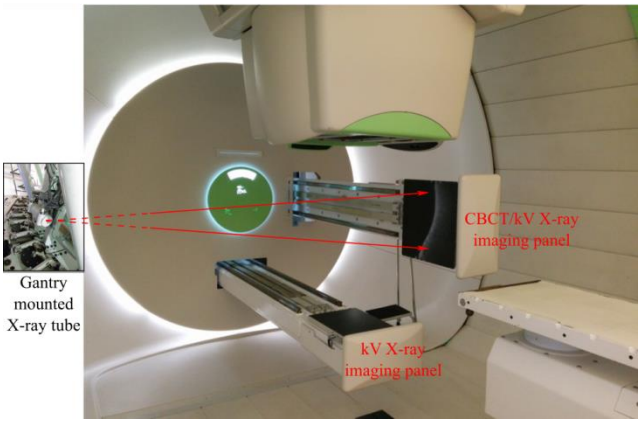

Fig. 1 - Picture of the gantry mounted CBCT system on the IBA Proteus® Plus proton therapy system. The kV X-ray source and imaging panel orthogonal to the proton beam are used for CBCT acquisition.

2.2. Patient data acquisition

Data from 20 patients treated for lung malignancies were included in this retrospective study. All patients underwent PSPT using two treatment fields, with a median daily fractionation of 1.8 Gy[Cobalt Gray Equivalent (CGE)]/fraction up to 66.6 Gy[CGE]. The patient cohort included a variety of tumor sizes, locations and anatomical changes that occurred throughout the treatment course (Veiga *et al* 2016).

The imaging protocol consisted of a 4D PET/CT (Gemini TF Big Bore PET/CT, Philips Healthcare, Andover, MA) for treatment planning, CBCT acquired online as part of the image guidance protocol, and an offline rescan 4D CT (Sensation Open, Siemens Healthcare, Malvern, PA) in treatment position for plan verification during the course of treatment. The average of the 4D CT was used for treatment planning and to generate the vCTs. One pair of CBCT and rCT at mid-treatment were selected for evaluation for each patient. The chosen rCT and CBCT pairs were acquired with up to 2 days difference. The CBCTs were acquired using a half-scan mode at 110 kVp. The image voxel sizes were $1.17\times1.17\times3.0$ mm^3, $1.33\times1.33\times2.5$ mm^3 and $0.98\times0.98\times3$ mm^3 for the planning CT (pCT), CBCT and rCTs, respectively.

2.3. CBCT and DIR based adaptive proton therapy workflow for lung malignancies

The online adaptive proton therapy workflow for lung patients implemented in our clinic can be divided in 3 main steps: (1) image registration, (2) vCT correction, and (3) calculation of WET and "dose of the day" (Veiga *et al* 2016).

2.3.1. Image registration and virtual CT correction

Rigid registration between the pCT and the CBCT was first applied prior to the DIR using six degrees of freedom by matching the bony anatomy at the target volume level; this was manually performed online. Then, the CT was registered to the CBCT using a diffeomorphic implementation of the Morphons algorithm (Knutsson and Andersson 2005). This algorithm is based on the matching of local phases. It was previously shown to be robust to changes in intensity when applied to CT images of the thorax with and without contrast enhancement (Janssens *et al* 2011) and evaluated in the context of head and neck adaptive proton therapy (Landry *et al* 2015a, Landry *et al* 2015b). Eight resolution scales were used, with ten iterations for the six coarsest scales, five and two for the second finest and finest scales, respectively. A Gaussian

regularization of 1.25 mm standard deviation was used. These settings were optimized empirically for the thoracic region. After registration, the vCT was corrected for gross registration errors caused by complex changes that occur in the thorax such as atelectasis, pleural effusion and complex tumor response to treatment (Veiga *et al* 2016).

2.3.3. Water equivalent thickness and dose warping

Water equivalent thickness (WET) represents the equivalent thickness of water that would cause a proton beam to lose the same amount of energy in a given thickness of a different medium. The CT HUs was first converted to relative stopping power using the tissue-substitute calibration method (Schneider 1996). The WET maps were obtained per beam angle by accumulating the relative stopping power on a voxel-by-voxel basis of all tissues crossed by an infinitely thin beam reaching the point from a virtual proton source. The changes between planned and daily WET can then be used to range-correct the dose distribution; hence, the "dose of the day" was estimated using a WET based dose warping method as described by Park *et al* (2012). The implementation of this algorithm was benchmarked in-house against dose recalculation on our clinically commissioned proton therapy treatment planning system (Eclipse v11.0, Varian Medical Systems, Palo Alto, CA).

2.3.4. Implementation details

The open-source REGGUI package was used to implement the vCT and "dose of the day" estimation workflow. REGGUI is a researcher friendly platform built in MATLAB (MathWorks, Natick, MA) for image processing featuring various registration methods, filtering methods, segmentation tools, CBCT simulation and other proton therapy dedicated functions, including WET computation and WET based dose warping. The software was designed in a way that facilitates the construction of clinical data workflows for research purposes. REGGUI and the clinical workflow implemented in our facility is freely available as part of the recently launched open-source initiative supported by IBA, the Proton Adaptive Therapy (PATh) project (https://openreggui.org/).

2.4. Procedures for the validation of CBCT and DIR in proton therapy

2.4.1. CBCT dataset definition

The lack of an ideal imaging dataset is one of the main challenges in validating DIR in clinical settings. In order to demonstrate that the vCT can provide similar clinical information as the rCT, the ideal imaging dataset would be a pair of anatomically matched CBCT (used to generate the vCT) and rCT (used as gold-standard for comparison) and images. Only by using such a perfect dataset the true uncertainty of using a

vCT to calculate WET and dose could be estimated. However, the CBCT and rCT cannot be acquired simultaneously and will always display residual geometrical variations from setup variation, patient and respiratory motion. Therefore, for validation purposes we propose to generate the vCTs from two synthetic CBCT datasets in addition to a regular CBCT: a simulated CBCT (created from the rCT) and a virtual CBCT (created from the rCBCT). These synthetic CBCTs were created to minimize the geometrical variations indicated above and permits a more accurate estimate of the true uncertainty in WET and dose of the vCT method arising from the DIR approach for lung patients.

1) Regular CBCT (rCBCT): rCBCTs acquired within 2-3 days from a rCT are commonly used in the literature to validate DIR (Peroni *et al* 2012, Landry *et al* 2015b). The rCBCTs have the characteristic image quality and real limitations of the cone beam geometry, namely variability in HU and reconstruction artifacts (e.g., scatter, streaks and distortions caused by the couch). The main disadvantage is setup variations between rCT and rCBCT. The rCBCTs used in this study were acquired within two days the rCT; despite the closeness in time, for some patients the apparent anatomical mismatch was considerable.

2) Simulated CBCT (sCBCT): In order to create an anatomically matched CBCT and rCT pair with no setup variations, a simulated CBCT can be generated from the rCT. We implemented a workflow using RTK (Rit *et al* 2014). The first step is to simulate raw 2D projection data from the average 4D rCT using the same geometry and acquisition parameters of the in-room CBCT system. To further mimic the rCBCT image characteristics, we added scatter and Poisson noise to the raw 2D projections. We used a scatter kernel superposition method (Ohnesorge *et al* 1999, Sun and Star-Lack 2010) with parameters estimated through Monte-Carlo simulations (GATE) (Jan *et al* 2004). The weight of these parameters was empirically fine-tuned and tested on the RANDO phantom (The Phantom Laboratory, Greenwich, NY). Finally, the 3D volume was reconstructed using the same reconstruction algorithm as in our facility (i.e., FDK implemented on RTK). The sCBCT is in principle anatomically identical to the rCT but has slightly different image artifacts from the rCBCT.

3) Virtual CBCT (vCBCT): Another way to create an anatomically matched CBCT and rCT pair is to deform the rCBCT to match the rCT. The resulting virtual CBCT contains the Hounsfield unit (HU) information of the rCBCT and the geometry of the rCT (Veiga *et al* 2014). No further correction was applied after DIR in this case since the differences between rCT and rCBCT were purely setup variation, and no large and/or non-deformable anatomical changes were verified between scans. This process introduces small geometric errors but recreates most of the artifacts present in real CBCT imaging (Veiga *et al* 2015).

The three CBCT datasets (rCBCT, sCBCT and vCBCT) were employed to generate three vCTs (vCT_{rCBCT}, vCT_{sCBCT} and vCT_{vCBCT}, respectively), which were used to report the uncertainties associated with DIR and CBCT imaging in WET and dose estimation. To distinguish between intrinsic limitations of the registration

process from the residual uncertainties due to differences in image quality and/or setup, we generated a fourth additional vCT (hereby denominated as vCT_{rCT}) by deforming the pCT directly to the rCT. Differences observed between the vCT_{rCT} and rCT are therefore attributable only to DIR errors without any influence in CBCT image quality and setup variations.

2.4.2. Validation workflow

Popular methods to validate DIR in ART applications include measuring the accuracy in landmark and/or contour propagation (Peroni *et al* 2012, Veiga *et al* 2014, Landry *et al* 2015a, Zhang *et al* 2007). However, accurate dose calculations require accurate HU information, and not necessarily accurate voxel-by-voxel mapping. In fact, we acknowledge that in certain situations DIR will fail to recover the real deformations, which lead us to include in our clinical workflow a correction step to account for registration gross errors. Therefore, we evaluate the uncertainties associated with meaningful quantities for the proposed application, i.e, WET and warped dose. The clinical usefulness of these quantities was explored in the companion study, where clinical indicators based on potential under/over-ranges (i.e., WET variations at the target location), underdosage of target volumes and/or overdose of organs at risk (OAR) were investigated (Veiga *et al* 2016). The uncertainty in WET was here defined as the voxelwise difference in WET between vCT and rCT, and was calculated in both 3D (planning target volume, PTV) and 2D (distal and proximal surfaces). The 2D WET maps are the projections of the 3D WET at the entrance and exit surfaces of the PTV (proximal and distal, respectively). The uncertainty in WET based dose warping was defined as the voxelwise difference between the warped dose and the full dose recalculation in the rCT. We report the mean, root-mean square (RMS) and 95^{th} percentile of the voxelwise uncertainty in WET and WET based warped dose. Additionally, we assessed the relationship between mismatch in HUs and dose uncertainty using linear regression and the correlation coefficient. The mismatch in HUs between each vCT and the corresponding rCT was defined in terms of the normalized cross correlation (NCC) between the images. Only voxels with dose larger than 20% of the prescribed dose were considered when computing the dose differences and similarity between images.

3. Results

3.1. Summary of the results for all CBCT datasets

Fig. 2 provides a qualitative example of the differences between CBCT datasets used in this study in terms of image quality; the color overlay with the rCT highlights the differences in setup. Note that the vCBCT showed a greater similarity with the rCBCT in terms of image quality, while the sCBCT showed greater similarity with the rCT in terms of setup.

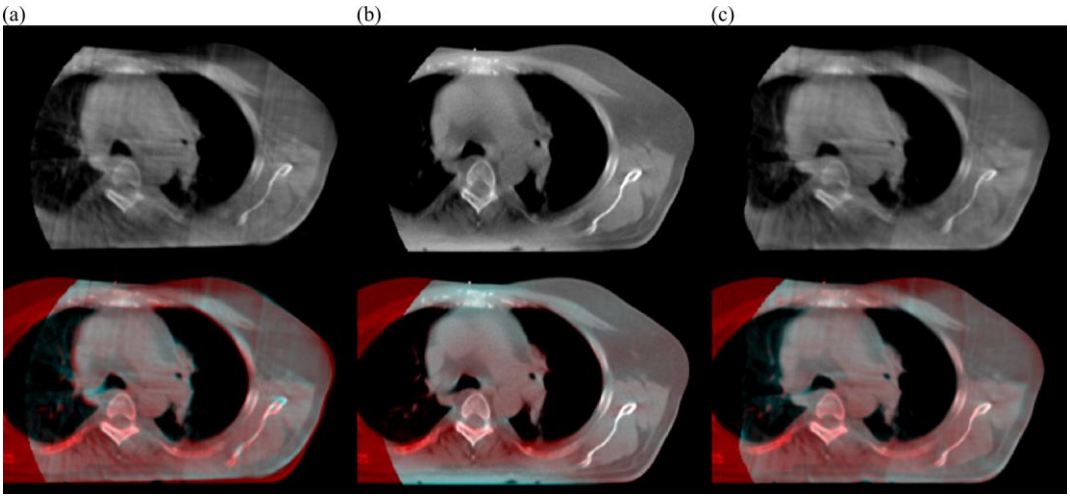

Fig. 2 - Example of (a) rCBCT, (b) sCBCT, and (c) vCBCT for one of the patients included in this study (top row), and corresponding CBCT (cyan) color overlay with the rCT (red).

The CBCT FoV is smaller than that of CT imaging due to hardware limitations, which causes truncation of the data. This can remove information of the anatomy in the beam path and also cause severe artefacts in the vCT where information is missing; however, with accurate positioning online most of the truncated anatomy can be positioned outside the proton beam path and thus minimize its importance (Veiga *et al* 2016).

The overall uncertainties in WET and dose associated with the vCTs were computed using the three CBCTs. Table 1 presents the difference in WET calculation between the vCT and the rCT on the distal and proximal surfaces, and within the PTV, as well as the comparison between warped dose and the forward dose calculation using the rCT. The dose difference (DD) results were normalized to the maximum dose (%mD). The results were averaged for all patients, fields and CBCT datasets.

Table 1- Overall uncertainty in WET (vCT vs rCT) within the planning target volume (PTV), and on the distal and proximal surfaces (mean value ± standard deviation), and overall uncertainty in voxelwise warped dose (vs forward calculation in rCT). The results were averaged over all patients, fields and CBCT datasets.

		mean	RMS	95% percentile
Uncertainty in WET	Distal surface	0.9±2.0 mm	3.6±1.8 mm	7±4 mm
	Proximal surface	0.4±1.7 mm	2.2±1.4 mm	4±2 mm
	PTV	0.7±2.0 mm	3.3±1.8 mm	7±4 mm

Uncertainty in dose	-0.5±0.9 %mD	6±2 %mD	11±6 %Md

Regarding the spatial distribution of the uncertainties, we found the absolute WET uncertainty to increase with depth which is in agreement with the fact that WET is a cumulative quantity. The larger dose uncertainties were found to occur mainly at the distal edge of the fields, and not in the homogeneous region of dose distribution, due to the sharp fall-off of the proton dose distributions We also note that the total uncertainty in dose results not only from the registration errors but also from to the approximations considered in the dose warping process itself.

While Table 1 presents results for individual fields, each patient was actually treated with two fields. When considering the total dose instead of individual field doses, the dose uncertainty (RMS) decreased from the 6±2 reported in Table 1, to 4±2 %mD. Dosimetric parameters were therefore less sensitive to registration errors when multiple fields were used since summing the contribution of each field averaged out the differences.

3.2. Effect of dataset used to generate the vCTs

The unique features of the vCT generated by each CBCT dataset were used to decouple the different sources of variation in the WET and dose warping uncertainty (Table 2). The vCT_{rCBCT} exhibited the largest reported uncertainties, while vCT_{vCBCT} exhibited the lowest. The results obtained for vCT_{sCBCT} and vCT_{vCBCT} were however comparable. When assessing the overall similarity between each vCT and the corresponding rCT, we found values for the NCC of 0.942±0.028, 0.976±0.019 and 0.980±0.019 for vCT_{rCBCT}, vCT_{sCBCT} and vCT_{vCBCT}, respectively. This is in agreement with the trends found for WET and dose uncertainties. A linear regression between NCC and uncertainty in dose gives an insight in understanding the relationship between mismatch in HUs and dose uncertainty (Fig. 3. If the vCT perfectly matches the rCT, then one expects the dose uncertainty to be zero; i.e., the higher the similarity in HUs between vCT and rCT the smaller the dose uncertainty. However, we only found a moderate negative correlation between NCC and dose uncertainty RMS (ρ=-0.53, Pearson's correlation coefficient) and we hypothesize this to be related with the properties of the individual proton fields. Characteristics of the beam may (1) cause some proton beams to be more robust or sensitive than others to registration errors of similar magnitude and (2) the dose warping process itself has higher uncertainties for specific beams. Therefore, choice of beam direction, range and modulation may impact the overall uncertainty in dose.

Table 2 - Uncertainty in WET (vCT vs rCT) within the planning target volume (PTV), and on the distal and proximal surfaces generated by each CBCT dataset and rCT, and corresponding uncertainty in warped dose (vs forward calculation in rCT) (mean value ± standard deviation). The results were averaged over all patients and fields.

	Uncertainty in WET:(mm)									Uncertainty in dose (%mD)		
	mean			RMS			95%					
	Dist	Prox	PTV	Dist	Prox	PTV	Dist	Prox	PTV	mean	RMS	95%
rCBCT	0.5±2.8	-0.1±2.6	0.2±2.8	4.6±1.8	3.3±1.5	4.3±2.1	9±4	6±3	8±4	-0.4±1.0	7.3±2.5	14±7
sCBCT	1.6±1.3	1.1±0.7	1.5±1.2	3.3±1.6	1.8±0.8	2.9±1.4	7±4	3.4±1.7	6±3	-0.6±0.8	5.6±1.9	10±5
vCBCT	0.6±1.4	0.3±0.8	0.6±1.3	2.9±1.6	1.4±0.9	2.6±1.5	6±4	2.8±1.7	6±3	-0.4±0.7	5.5±1.9	9±5
rCT	0.7±1.5	0.3±0.8	0.6±1.5	2.8±1.7	1.3±0.9	2.5±1.6	6±4	2.6±1.8	5±4	-0.4±0.7	5.4±1.9	9±5

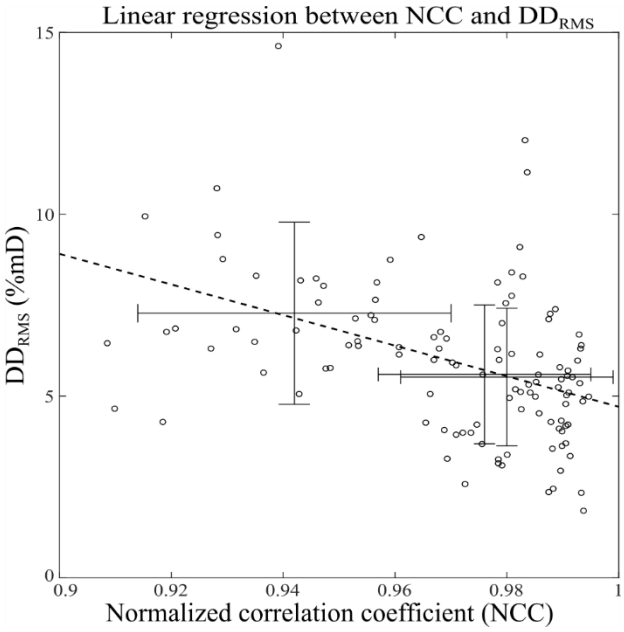

Fig. 3 - Linear regression of similarity between vCT and rCT (expressed as the normalized correlation coefficient, NCC) and dose uncertainty (expressed as RMS, DD_{RMS}). Each point plotted corresponds to a different vCT generated (i.e., three points plotted per patient). The crosses correspond to the average NCC and DD_{RMS} (±standard deviation) for vCT_{rCBCT}, vCT_{sCBCT} and vCT_{vCBCT} (from left to right).

As none of the CBCT datasets is ideal, the results of the similarity between vCT_{rCT} and rCT bring additional confidence on where the true uncertainty lies (Veiga *et al* 2015). The residual uncertainty of up to 1 mm in WET can be attributed to both the poorer quality of the CBCT compared to CT and to residual differences in setup for the different CBCT datasets.

4. Discussion

CBCT imaging has an important role in image-guided proton therapy, particularly as a substitute for verification scans. In this work we investigated validation methods of the use of CBCT and DIR for adaptive lung proton therapy radiotherapy and quantified the uncertainties inherent to the workflow. This study is a companion to a recently published work from our group, where the vCT method validated here is applied

retrospectively as part of a clinical adaptation workflow. A complete discussion of the clinical implications of the inaccuracies investigated here can be found in its clinical counterpart (Veiga *et al* 2016).

The accuracy of vCT for WET estimation and consequently the "dose of the day" for lung cancer patients was investigated in this work using regular and synthetic CBCT datasets in comparison with rescan CTs. The rCBCTs are not anatomically identical to the reference rCT, hence the reason why vCBCTs and sCBCTs were created to circumvent this limitation. An overall uncertainty of 3.6±1.8 mm (RMS) in the WET was found at the targets' distal surface. While the results taken together offer enough evidence for quantifying the uncertainties associated with DIR and CBCT imaging for the proposed application, each type of CBCT has its own advantages as well as limitations to validate in-house CBCT based ART techniques. Additionally, deforming the pCT directly to the rCT brings additional confidence of where the true uncertainty lies. Validation of new technologies before clinical translation is of utmost importance; however, we appreciate that not all proton facilities have the resources to perform such a lengthy and comprehensive validation. We recommend to other centres validating their in-house or commercial tools to choose whichever datasets are more readily available as part of their clinical routine and software tools available to validate their method, while recognising the limitations of the chosen approach when analysing their results. The limitations and advantages of each of the CBCT datasets are discussed in further detail in the following paragraphs.

The rCBCT reports the highest uncertainties due to setup differences; because the CBCT and rCT are not acquired simultaneously, there are generally differences due to variations in setup and respiratory pattern. When using the rCBCT for validation purposes, one must assume that it will provide a worst-case estimate of the uncertainties of the vCT. An important point to discuss regarding using rCBCT for validation of DIR is the reasoning behind acquiring a CBCT and rCT in the same day. A rCBCT does not provide any additional clinical information compared to a rCT acquired on the same day, unless the patient was flagged for replanning due to dramatic changes. As part of routine clinical work either a CBCT or a rCT is acquired for plan verification. Acquiring an additional scan poses additional and unnecessary dose to the patient and must be properly justified. For example, in our study this additional scan was acquired as part of the commissioning of the newly-installed CBCT system.

Using the vCBCT to generate the vCT results in the lowest reported uncertainties in the WET and dose calculation. Of all the CBCT datasets, this has the best balance between agreement in geometry with the rCT and agreement in intensities with the rCBCT. However, one can argue that the vCBCT can help the DIR workflow to appear better than it really is. While it is unlikely to occur, it is possible for the errors in registering the pCT to the vCBCT to be the exact opposite of the errors when generating the vCBCT, and hence using vCBCT for validation purposes may locally underestimate the real uncertainties (Veiga *et al* 2015). One advantage of using the vCBCT resides in the fact that it could be used when rCBCT are not

available so close in time to the rCT (Veiga *et al* 2014). Thus, the vCBCT is a good method to validate DIR for facilities that already routinely use CBCT for image-guidance and want to implement ART workflows clinically. It is also versatile as it benefits from a wide variety of open-source and friendly DIR tools to generate the datasets.

While other publications have used the concept of rCBCT (Peroni *et al* 2012, Landry *et al* 2015b) and vCBCT (Veiga *et al* 2014, Veiga *et al* 2015) for validation of CBCT-based ART, to the best of our knowledge no other studies have proposed a sCBCT as we did here. The sCBCT removes any geometric uncertainty from the validation workflows; however, the main disadvantage of the proposed methodology as currently implemented was not reproducing all the features in the CBCT imaging exactly. For example, scatter artifacts appeared more severe when farther away from the central region of the volume, cupping artifacts were slightly different, and artifacts and distortions caused by the couch, from neglecting cardiac and respiratory motion, and other contributions (Yao and Leszczynski 2009) were not included in the simulation. In the sCBCTs we found that the scatter was overestimated in the interface between lung and tissue/bone. The deterioration of the contrast resulted in systematic registration errors. This was most likely the main source of discrepancy in WET calculation between the vCBCT and sCBCT datasets. These differences in the intensity of the CBCT datasets are particularly important if the DIR uses monomodal similarity measures. In spite of the limitations of the current implementation, the sCBCT method has the potential to become an ideal platform to validate CT-to-CBCT DIR using patient data. This may be achievable with further optimization of the scatter and noise parameters to tailor for the specificities of the individual CBCT systems, and introduction of 4DCT phase information to reproduce motion artefacts. However, due to the significant amount of work necessary to develop the sCBCT to include motion we consider this improvement to be out of the scope of this work. In fact, it was in light of the perceived individual limitations of each CBCT dataset that we used several datasets to strengthen the validation.

DIR errors lead to inaccuracies in WET and dose computation. The cohort of lung patients included in this study included a broad range of clinical situations that can reduce the accuracy of the registrations achievable. It is paramount to validate the workflow on a cohort representative of the challenges of the particular site. The main limitation of DIR is related with the mathematical basis of the most common and popular DIR algorithms, whose deformations include only translation, expansion and contraction. This is not specific to the Morphons algorithm as we have also performed a similar validation study with a B-Spline Free Form based algorithm (Modat *et al* 2010) and the hybrid free-form registration algorithm in RayStation (version 4.5, RaySearch Laboratories, Stockholm, SE). The accuracy achieved with each algorithm was comparable (supplementary material) and all algorithms struggled to reproduce non-deformable and/or large anatomical changes. Changes in topology, such as the appearance or disappearance of tissue, cannot be reproduced without introducing singularities in the deformation fields. Since the intensities are mapped from the CT to the CBCT space, variations in the density of the tissue between

images cannot be reproduced (Palma *et al* 2011). Additionally, the volume of rigid bodies is not necessarily preserved and the displacement between ribs and lung may not be accurately modelled without specialized algorithms (Kim *et al* 2013, Risser *et al* 2012. There is therefore a large scope to continue developing DIR algorithms and/or post-processing methods that account for site-specific issues.

The ART workflow was implemented as a prototype in REGGUI which is a research platform, and therefore not optimized for fast online use. In its proposed form the workflow can be implemented more efficiently: the most computationally demanding processes (DIR, vCT correction and dose warping) can be performed in several minutes using the GPU. In terms of improving the accuracy, other approaches to estimate the "dose of the day" may be also be implemented in the future; while dose warping can be computed rapidly it is still only an approximation. In fact, the uncertainty in dose results from two different processes: the uncertainty in WET resulting from the DIR process, and the uncertainty of the dose warping process itself. The method here presented is therefore more appropriate to verify the range than to accurately calculate dose distributions, as it can identify if beam under or over-ranging results in loss of coverage or OAR doses above tolerance levels (Veiga *et al* 2016). Incorporating in the workflow a fast GPU-Monte Carlo calculation (Jia *et al* 2012), or by incorporating the current workflow as part of treatment planning systems are rival and more accurate approaches to assess dose on a vCT. Direct dose calculation on the CBCT is a possible way to remove the errors introduced by DIR. However, while feasible in conventional photon treatments (Richter *et al* 2008, Yang *et al* 2007, Guan and Dong 2009) for proton therapy the usability of CBCT after simple intensity rescaling methods remains debatable (Veiga 2015, Kurz 2015) and other groups are working on correction techniques that use DIR and scatter correction to improve the HU accuracy of CBCT directly (Park *et al* 2015).

5. Conclusions

We described procedures for in-house validation of CBCT and DIR-based adaptive radiotherapy workflows using regular and synthetic CBCT datasets. The uncertainties in WET and dose warping were quantified in the context of an adaptive proton therapy for lung cancer patients. The vCTs generated were found to provide similar WET and dosimetric information as repeat CT.

Acknowledgments

The authors have no relevant conflicts of interest to disclose. C. V. is funded by Fundação para a Ciência e a Tecnologia (FCT) grant SFRH/BD/76169/2011, co-financed by FSE, POPH/QREN and EU.

References

Bernard G, Verleysen M, and Lee J A, "Incremental classification of objects in scenes: Application to the delineation of images," J. Neurocom. 152, 45-57 (2015).

Britton K R, Starkschall G, Tucker S L, Pan T, Nelson C, Chang J Y, Cox J D, Mohan R, and Komaki R, "Assessment of gross tumor volume regression and motion changes during radiotherapy for non-small-cell lung cancer as measured by four-dimensional computed tomography," Int. J. Radiat. Oncol. Biol. Phys. 68(4) 1036-1046 (2007).

Brousmiche S, Seabra J, Labarbe R, Vila M, Rit S, Wikler D, Lee J, Teo K, Orban de Xivry J, Macq B PO-0783: Design of cone-beam CT for proton therapy gantry. Radiother. Oncol. 111, S52-53 (2014).

Cousty J, Bertrand G, Najman L, and Couprie M, "Watershed cuts: Minimum spanning forests and the drop of water principle," IEEE Trans. Pattern Anal. Mach. Intell. 31(8), 1362-1374 (2009).

DeSantis C E, Lin C C, Mariotto A B, Siegel R L, Stein K D, Kramer J L, Alteri R, Robbins A S, and Jemal A, "Cancer treatment and survivorship statistics, 2014," CA: A Cancer Journal for Clinicians 64(40), 252-271 (2014).

Feldkamp L A, Davis L C, and Kress J W, "Practical cone-beam algorithm," J. Opt. Soc. Am. A 1, 612-619 (1984).

Freund Y and Schapire R E, "Large margin classification using the perceptron algorithm," Mach. Learn. 37(3), 277-296 (1999).

Gallant S I, "Perceptron-based learning algorithms," IEEE Trans. Neural Netw. 1(2), 179-191 (1990).

Gomez D R, and Chang J Y "Adaptive radiation for lung cancer," J. Oncol. 2011, e898391 (2011).

Guan H and Dong H, "Dose calculation accuracy using cone-beam CT (CBCT) for pelvic adaptive radiotherapy," Phys. Med. Biol. 54(20), 6239-6250 (2009).

Guckenberger M, Wilbert J, Richter A, Baier K, and Flentje M, "Potential of adaptive radiotherapy to escalate the radiation dose in combined radiochemotherapy for locally advanced non-small cell lung cancer," Int. J. Radiat. Oncol. Biol. Phys. 79(3), 901-908 (2011).

Hui Z, Zhang X, Starkschall G, Li Y, Mohan R , Komaki R, Cox J D, and Chang J Y, "Effects of interfractional motion and anatomic changes on proton therapy dose distribution in lung cancer," Int. J. Radiat. Oncol. Biol. Phys. 72(5), 1385-1395 (2008).

Jan S *et al*, "GATE: a simulation toolkit for PET and SPECT " Phys. Med. Biol. 49(1):4543-4561 (2004).

Janssens G, Jacques L, Orban de Xivry J, Geets X, and Macq B, "Diffeomorphic registration of images with variable contrast enhancement, diffeomorphic registration of images with variable contrast enhancement," Int. J. Biomed. Imaging 2011, e891585 (2011).

Jia X, Schumann J, Paganetti H, and Jiang S B, "GPU-based fast Monte Carlo dose calculation for proton therapy," Phys, Med. Biol. 57(23), 7783 (2012).

Kim J, Matuszak M M, Saitou K, and Balter J M, "Distance-preserving rigidity penalty on deformable image registration of multiple skeletal components in the neck," Med. Phys 40(12), 121907 (2013).

Knutsson H and Andersson M, "Morphons: segmentation using elastic canvas and paint on priors," in IEEE International Conference on Image Processing 2, II-1226-9 (2005).

Kurz C, Dedes G, Resch A, Reiner M, Ganswindt U, Nijhuis R, Thieke C, Belka C, Parodi K, and Landry G, "Comparing cone-beam CT intensity correction methods for dose recalculation in adaptive intensity-modulated photon and proton therapy for head and neck cancer," Acta Oncol, 54(9), 1651-1657 (2015).

Kurz C, Nijhuis R, Rener M, Ganswindt U, Thieke C, Belka C, Parodi K, and Landry G, "Feasibility of automated proton therapy plan adaptation for head and neck tumors using cone beam CT images", Radiation Oncology 11(1) (2016).

Landry G *et al*, "Phantom based evaluation of CT to CBCT image registration for proton therapy dose recalculation," Phys. Med. Biol. 60(2), 595 (2015a).

Landry G *et al*, "Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation," Med. Phys. 42(3), 1354-1366 (2015b).

Modat M, Ridgway G R, Taylor Z A, Lehmann M, Barnes J, Hawkes D J, Fox N C, and Ourselin S, "Fast free-form deformation using graphics processing units," Comput Methods Programs Biomed. 98(3), 278-284.

Moteabbed M, Sharp G C, Wang Y, Trofimov A, Efstathiou J A, and Lu H M, "Validation of a deformable image registration technique for cone beam CT-based dose verification," Med. Phys. 42(1), 196-205 (2015).

Ohnesorge B, Flohr T, and Klingenbeck-Regn K, “Efficient object scatter correction algorithm for third and fourth generation CT scanners,” Eur. Radiol. 9(3):563—569 (1999).

Ozhasoglu C and Murphy M J, “Issues in respiratory motion compensation during external-beam radiotherapy," Int. J. Radiat. Oncol. Biol. Phys. 52(5), 1389-1399 (2002).

Palma D A, van Sornsen de Koste J, Verbakel W F A R, Vincent A, and Senan S, "Lung density changes after stereotactic radiotherapy: a quantitative analysis in 50 patients," Int. J. Radiat. Oncol. Biol. Phys. 81(4), 974-978 (2011).

Park P C, Cheung J, Zhu X R, Sahoo N, Court L, and Dong L, “Fast range-corrected proton dose approximation using prior dose distribution,” Phys. Med. Biol. 57(11):3555-3569 (2012).

Park Y-K, Sharp G C, Phillips J, and Winey B A, “Proton dose calculation on scatter-corrected CBCT image: Feasibility study for adaptive proton therapy,” Med. Phys. 42(8), 4449-4459 (2015).

Peroni M, Ciardo D, Spadea M F, Riboldi M, Comi S, Alterio D, Baroni G, and Orecchia R, "Automatic segmentation and online virtual CT in head-and-neck adaptive radiation therapy," Int. J. Radiat. Oncol. Biol. Phys. 84(3), e427-433 (2012).

Qin Y, Zhang F, Yoo D S, Kelsey C R, Yin F-F, and Cai J, “Adaptive stereotactic body radiation therapy planning for lung cancer," Int. J. Radiat. Oncol. Biol. Phys. 87(1), 209-215 (2013).

Richter A, Hu Q, Steglich C, Baier K, Wilbert J, Guckenberger M, and Flentje M, "Investigation of the usability of conebeam CT data sets for dose calculation," Radiat. Oncol. 3(1), 42 (2008).

Risser L, Heinrich M, Matin T, and Schnabel J, "Piecewise-diffeomorphic registration of 3D CT/MR pulmonary images with sliding conditions," 9th IEEE International Symposium on Biomedical Imaging (ISBI), 1351-1354 (2012).

Rit S, Oliva M V, Brousmiche S, Labarbe R, Sarrut D, and Sharp G C, "The reconstruction toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the insight toolkit (ITK)," J. Phys: Conf. Ser. 489(1), 012079 (2014).

Schneider U, Pedroni E, and Lomax A, "The calibration of CT hounsfield units for radiotherapy treatment planning," Phys. Med. Biol. 41(1), 111-124 (1996).

Seabra J, Brousmiche S, Labarbe R, Vila M, Rit S, Wikler D, Lee J, Teo K, Orban de Xivry J, and Macq B, "Design of cone-beam CT for proton therapy gantry," in Belgian Hospital Physicists Association Symposium, (Louvain-La-Neuve, Belgium) (2014).

Stock M, Pasler M, Birkfellner W, Homolka P, Poetter R, and Georg D, "Image quality and stability of image-guided radiotherapy (IGRT) devices: A comparative study," Radiother. Oncol. 93(1), 1-7 (2009).

Sun M, and Star-Lack J M, "Improved scatter correction using adaptive scatter kernel superposition," Phys. Med. Biol. 55(2):6695-6720 (2010).

van Zwienen M, van Beek S, Belderbos J, van Kranen S, Rasch C, van Herk M, and Sonke J, "Anatomical changes during radiotherapy of lung cancer patients," Int. J. Radiat. Oncol. Biol. Phys. 72(1) S111 (2008).

Veiga C, McClelland J, Moinuddin S, Lourenço A, Ricketts K, Annkah J, Modat M, Ourselin S, D'Souza S, and Royle G, "Toward adaptive radiotherapy for head and neck patients: Feasibility study on using CT-to-CBCT deformable registration for "dose of the day" calculations," Med. Phys. 41(3), 031703 (2014).

Veiga C, Alshaikhi J, Amos R, Lourenço A M, Modat M, Ourselin S, Royle G, and McClelland J R, "Cone-beam computed tomography and deformable registration-based "dose of the day" calculations for adaptive proton therapy," Int. J. Particle Ther. 2(2), 404-414 (2015).

Veiga C, Janssens G, Teng C L, Baudier T, Hotoiu L, McClelland J R, Royle G, Lin L, Yin L, Metz J, Solberg T D, Tochner Z, Simone CB 2nd, McDonough J, and Teo B K, "First clinical investigation of CBCT and deformable registration for adaptive proton therapy of lung cancer," Int. J. Radiat. Oncol. Biol. Phys. 95(1) 549-559 (2016).

Wink K C J, Roelofs E, Solberg T, Lin L, Simone C B, Jakobi A, Richter C, Lambin P, and Troost E G C, "Particle therapy for non-small cell lung tumors: Where do we stand? A systematic review of the literature," Front Oncol. 4, 292 (2014).

Yang Y, Schreibmann E, Li T, Wang C, and Xing L, "Evaluation of on-board kV cone beam CT (CBCT)-based dose calculation," Med. Phys. 52(3), 685-705 (2007).

Yao W and Leszazynski K, “Analytically derived weighting factors for transmission tomography cone beam projections,” Phys. Med. Biol. 54(3), 513-533 (2009).

Zhang T, Chi Y, Meldolesi E, and Yan D, "Automatic delineation of on-line head-and-neck computed tomography images: Toward on-line adaptive radiotherapy," Int. J. Radiat. Oncol. Biol. Phys. 68(2), 522-530 (2007).