Physica Medica 61 (2019) 103–111

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Comparison of five dose calculation algorithms in a heterogeneous media using design of experiment

M. Vangvichith[a], D. Autret[a], T. Tiplica[b], M. Barreau[b], S. Dufreneix[a,*]

[a] Institut de Cancérologie de l'Ouest, Centre Paul Papin, 49055 Angers, France
[b] Institut de Sciences et Techniques de l'Ingénieur d'Angers, 49000 Angers, France

ARTICLE INFO

Keywords:
Design of experiment
Heterogeneous media
Algorithm
Dose calculation

ABSTRACT

Purpose: Design of experiments (DoE) provides a methodology to reveal the influence of input values on the measured output with a limited number of trials. The purpose of this study was to describe how DoE can be used to evaluate the performances of several dose calculation systems in heterogeneous media, including algorithms like Pencil Beam (PB), Anisotropic Analytical Algorithm (AAA), Acuros XB (AXB), Monte Carlo (MC) and Collapsed Cone Volume (CCV).

Method: This study was carried out using a CIRS Model 002LFC IMRT Thorax Phantom customized with a water-equivalent heterogeneity inside the lung. The calculated dose distributions were compared to Gafchromic® EBT3 film measurements. The beam configurations were selected using DoE to study the influence of five parameters simultaneously (energy, collimator angulation, gantry angulation, X and Y jaws) and to optimize the number of experiments. An analysis of variance was performed over the entire irradiation field and over various regions of interest (tumour, shadow of tumour and lungs).

Results: DoE enabled to quantify and determine the statistically significant factors, leading to an evaluation of the dose calculation systems in the lung case. The resulting scoring could be as follow (from best to worst): AXB_D_m, CCV, AXB_D_w, XVMC_D_m, XVMC_D_w, AAA and last PB. Differences between the algorithms were specially observed in the tumour and the shadow regions.

Conclusion: DoE is a robust statistical method to compare several dose calculation systems. The various analyses lead to the conclusion that AXB handled more accurately most of the situations investigated in heterogeneous media.

1. Introduction

Design of experiments (DoE) is a body of statistical techniques for the effective and efficient collection of data for a number of purposes. Two significant ones are the investigation of research hypotheses and the accurate determination of the relative effects of the many different factors that influence the quality of a product or process. DoE can be employed in both the product design phase and production phase [1,2]. The principle of DoE is to perform experiments by varying some inputs for each trial, instead of varying inputs one by one. Furthermore, by carefully choosing the input combinations to be used in experiments, it is possible to diminish the number of trials needed to reveal the influence of input values on the measured output [1,3].

In external radiotherapy, heterogeneities can significantly influence the dose distribution calculated inside the patient. The choice of the algorithm is particularly important in the presence of heterogeneities such as lungs, air and bones [4]. In the case of stereotactic treatment of lung lesions, Fogliata *et al.* [5] summarized the dose calculation accuracy depending on their degree of complexity in the management of charged particle transport: algorithms belonging to the type "a" class do not consider the changes in electron transport, while the type "b" takes into consideration the electron transport [6]. In type "c" algorithms, modeling of secondary electron transport is improved, taking into account the physics generating the dose absorption process [7]. Also, the dose deposition can be calculated in biological tissues and in presence of high-Z materials.

Many studies evaluated the performances of various algorithms compared to Monte Carlo and/or measurements: in one dimension (dose point) [8,9] or in two-dimensions (percent depth dose (PDD), profiles, and dose distributions) [6,8,10–12] on heterogeneous or anthropomorphic phantoms. However, no comparison based on design of experiment to study the performance of several algorithms clinically

* Corresponding author.
E-mail address: stephane.dufreneix@ico.unicancer.fr (S. Dufreneix).

https://doi.org/10.1016/j.ejmp.2019.04.014
Received 29 October 2018; Received in revised form 12 April 2019; Accepted 17 April 2019
Available online 03 May 2019

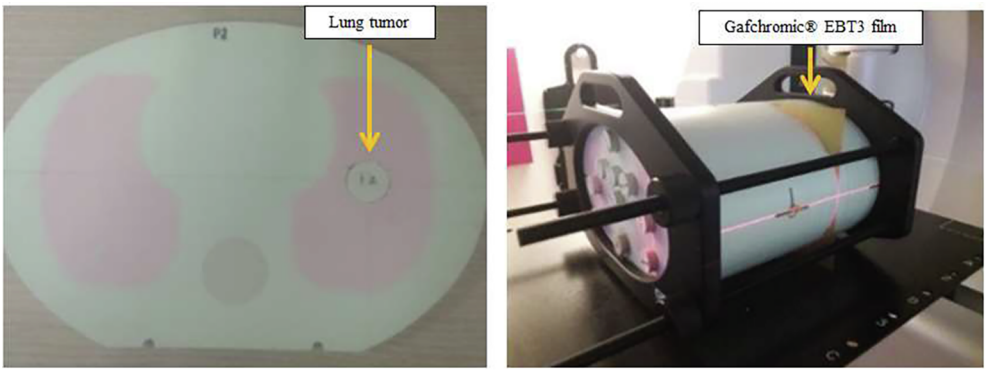

Fig. 1. On the left, one of the sections where the insert was added to simulate the lung tumor in the CIRS phantom. On the right, the set-up of the phantom and the Gafchromic® EBT3 film.

available on anthropomorphic phantoms was presented with the same methodology for clinical configuration encountered in the lung case.

The aim of this paper was to explore the methodology of DoE and apply it in medical physics for the comparison of the performances of the dose calculation systems with five algorithms in various clinical beams for tumors located in the lung. Dose distributions computed by five dose calculation algorithms commercially available were compared to film measurements in an anthropomorphic phantom. The use of a design of experiment made it possible to select only 18 tests in order to cover various clinical beams and to highlight potential statistically influent factors. The accuracy of the algorithms was estimated by subtracting computed dose distributions and film measurements. A statistical analysis was conducted and a summary of the performances of the dose calculation systems was suggested.

2. Materials and methods

2.1. Design of experiment

Design of experiments is a multivariate optimization technique, used in various fields, such as biology [13], space industry [14] or radiology [15] and more recently medical physics [16,17]. Statistical DoE were introduced in the 1920s, in order to efficiently reveal the influence of some inputs (referred to as factors) on outputs of interest [18]. Usually, in order to investigate the output variance with different input values, experiments are performed by varying inputs one by one for each trial. Therefore, when having many inputs, the number of experiments quickly increases greatly. Efficient experimental design uses the best combination of factors levels (i.e. input values) in order to diminish the number of trials. They are referred to as fractional factorial experiments, since they use only a portion of the total possible combinations of factors levels to estimate the main factor effects and some of the interactions [1,19].

Dr. G. Taguchi developed a family of highly fractional factorial experiments, known as orthogonal arrays. Orthogonality means that factors can be evaluated independently of one another; estimating the effect of one factor is statistically independent of the estimation of the effect of another factor [3,20]. These designs can be used to estimate main effects using only a few experimental runs. The orthogonal arrays proposed by Taguchi give the best combination of factors levels to use in the experiments, in order to insure the independency of effects estimation [3,20]. Since Taguchi arrays are fractional factorial experiments, the choice of these combinations is very important in order to correctly estimate the effects of the factors on the measured output. Combinations are tabled in arrays which can be found in the literature [1,3,20] or on the internet. The choice of an orthogonal array is made with regard to the number of factors and interactions to be evaluated and the number of levels considered for each factor [1,3]. Taguchi designs allow investigating main effects for mixed level experiments where the factors included do not have the same number of levels.

The factors' effects are estimated for each level of each factor separately and their significance is interpreted by analysis of variance [18]. DoE allows having more values of the measured output for each level of a factor, diminishing thus the uncertainty in the estimation of the effects of the factors taken into account [18].

Our use of DoE consisted in optimizing the organization of experiments to maximize the information with a minimum of beam configurations and to highlight the influence of factors on the dose deviation between dose distribution calculations and measurements.

2.2. Film measurements in an anthropomorphic phantom

The algorithm comparison was realised in a heterogeneous phantom for various irradiation beams. A CIRS Model 002LMFC Thorax Phantom representing an average human torso in proportion and density was used. It represents a thorax with the presence of heterogeneities: lung (ρ = 0.21 g/cm^3), bone (ρ = *1.60* g/cm^3) and muscle (ρ = *1.04* g/cm^3). In order to simulate a lung tumour, a water-equivalent insert (ρ = *1.10* g/cm^3) was inserted in the lung region for two slices (Fig. 1). Gafchromic® EBT3 films were inserted between these two slices. A tumour was thus modelled with a total length of 2 cm and a diameter of 2 cm. The plan treatment was performed on the anthropomorphic phantom CT scan, with the film placed between two slices as the one described in Fig. 1.

A film calibration was performed in a water-equivalent phantom under reference conditions. Films were scanned 24 h after irradiation using an Epson Expression 10000 XL [US Epson, Long Beach, CA, USA], with transmission mode, 48 bits RGB (16 bits per channel color) and a resolution of 200 dpi (0.35 mm/pixel). The methodology described in reference [21] was followed and self-attenuation in the film neglected. Measurements were repeated 3 times for each selected beam configuration in order to take into account the uncertainty originating from the measured response of the film to radiation which is expected to be the main contributor to the total uncertainty [22].

2.3. Dose calculation algorithms

Five dose algorithms commercially available in different Treatment Planning System (TPS) were included in this study. The classification summarized by Fogliata *et al.* [5] (type "a", "b" or "c" depending on the way electron transport is taken into account) is reported below. The types "a", "b" were the first time suggested by Knöös et al [6] and the type "c" was added the first time as an extension of the coding by Ojala et al. [7].

- Varian Eclipse TPS:
  - AAA: Anisotropic Analytical Algorithm (version 13.7, type "b"). This algorithm is a pencil-beam, convolution/superposition model

described in [23,24]. The AAA accounts for heterogeneities by correction of the photon scatter kernel according to the three-dimensional anisotropic changes in electron density in the neighborhood of an interaction point [12].

- AXB: Acuros XB (version 13.7, type “c”). This algorithm solves the Linear Boltzman Transport Equation numerically (LBTE) [25,26]. The calculation method of Acuros XB requires mass densities and atomic composition to evaluate the macroscopic interaction cross-sections for each voxel (unlike the AAA dose calculation that used relative electron density). The determination of the material composition uses a method based on the conversion of Hounsfield Unit (HU) values into mass density from the CT calibration curve which can be configured by the users for their specific CT scanner. Once mass density is known in a voxel, the material is determined based on a library of materials stored in the system. This algorithm supports two dose reporting option: dose-to-medium (D_m) and dose-to-water (D_w). When D_m is calculated, the energy dependent response function is based on the material properties of that voxel. When D_w is calculated, in non-water materials, the dose calculation is similar to the one received by a volume of water. For both reporting modes the energy dependent electron fluence is calculated based on the material properties of the patient. The dose to medium or to water is then computed using the specific electron energy deposition cross sections and densities.

- BrainLab iPlan TPS:
  - PB: Pencil Beam (version 4.5.5, type “a”) described in [27]. This algorithm calculates dose by scaling pencil beam dose distribution kernels in water to take into account tissue heterogeneities, but this method showed limited accuracy in these regions [4] because the lateral scattering is not taken into account in the dose calculation of this algorithm. Differences between calculated and measured dose distributions are primarily due to changes in electron transport in the lung which are not adequately taken into account by the simple tissue inhomogeneity correction algorithms. As a consequence, PB may over-estimate the dose to the tumor.
  - XVMC: X-ray Voxel Monte Carlo (version 4.5.5, type “c”). Details of this Monte Carlo algorithm adapted for clinical use can be found in [28,29]. The algorithms based on Monte Carlo are statistical methods for solving the LBTE by simulating all particle interactions in detail, taking into account heterogeneities accurately. The method for material assignment uses a function relating a Compton cross-section ratio versus mass density for all materials of ICRU report 46. This function is used by XVMC to calculate the Compton cross-section and cannot be modified by the user. The mass density is derived from the CT HU calibration curve, providing a HU to mass density mapping function. This algorithm allows the calculation of 2 different dose types: dose-to-medium (D_m) and dose-to-water (D_w). The defaults setting is the D_m which represents the energy absorbed in a small tissue element divided by the mass of the tissue element. D_w means energy absorbed in a small cavity of water divided by the mass of that cavity. D_w is obtained by multiplying D_m by the stopping power ratio for water to that for the medium.

- Mobius Medical System:
  - CCV: Collapsed Cone Volume (version 2.0.1, type “b”). Mobius uses a collapsed cone (CC) convolution/superposition (CS) algorithm developed by the manufacturer. The dose calculations are conceptually similar to the algorithm developed by Philips in the Pinnacle3 software [30]. Both are rooted in the principles of the original collapsed cone [31,32] and have been updated with enhancements based on subsequent research [33,34].

Dose calculations were performed for a Varian Novalis TrueBeam STx linear accelerator using the smallest calculation grid available: 1 mm for the Varian and Mobius algorithms; 2 mm for the iPlan algorithms.

2.4. Validation of algorithms in a water phantom

All algorithms were first validated in a water phantom using experimental designs: computed dose was compared to measurements with an ionization chamber (CC04, IBA) in various configurations. The following factors were selected: energy (2 levels: 6 MV and 16 MV), presence of MLC (2 levels: “yes”: field is defined by the MLC and “no”: field is defined by the jaws), depth (3 levels: z_{max}, 10 cm and 20 cm), X jaws (3 levels: 4 cm, 10 cm and 22 cm), Y_1 jaw (3 levels: 2 cm, 5 cm and 10 cm), Y_2 jaw (3 levels: 2 cm, 5 cm and 10 cm), off-axis (3 levels: on the axis, at 50% and 150% of the field size defined by Y_1). Details of the methodology and results for the AAA algorithm can be found in the literature [16].

Considering the number of factors and levels for this study, a Taguchi L_{36} ($2^{11} \times 3^{12}$) array was chosen. This array is a mixed level design, giving the level combinations to be used in trials (by the assignment of factors to the columns of the array), in order to independently estimate the effects. It can be used to investigate the effects of up to 11 two-levels factor (2^{11}) and a maximum of 12 three-levels factors (3^{12}) [3,33]. 36 tests provide enough information to estimate the main effects of the factors.

The L_{36} Taguchi array previously used was adjusted to validate FFF beam models which were validated in a second time. Factors and associated levels were defined as follows: presence of MLC (2 levels: “yes”: field is defined by the MLC and “no” field is defined by the jaws), depth (3 levels: z_{max}, 10 cm and 20 cm), X jaws (3 levels: 2 cm, 5 cm and 10 cm), Y_1 jaw (3 levels: 1 cm, 2.5 cm and 5 cm), Y_2 jaw (3 levels: 1 cm, 2.5 cm and 5 cm), off-axis (2 levels: on the axis and at 50% of the field size defined by Y_1).

For each trial, the following quantity of interest was used:

$$\delta[\%] = \left(\frac{D_{cal} - D_{meas}}{D_{meas}}\right) \times 100$$

where D_{cal} is the absorbed dose calculated by the TPS and D_{meas} is the measured absorbed dose.

2.5. Comparison of the performances in a heterogeneous media

Considering the potential clinical configurations encountered in the case of lung tumours, the following factors were selected:

- Gantry angulation (3 levels): 0°, 90° and 160°. This last angulation was chosen in order to evaluate the table attenuation estimation.
- Collimator angulation (2 levels): 0° and 90°
- Energy (3 levels): 6MV, 6 MV FFF and 16 MV
- X jaws (3 levels): 2, 5 and 8 cm
- Y jaws (3 levels): 2, 5 and 8 cm.

For each factor, the choice of the different levels was made according to the clinical configurations most often encountered in our clinical practice (gantry angulation, collimator angulation, X and Y jaws), or the available parameter (energy). All beams were static beams with no intensity modulation. The size of the smallest field in the X and Y directions was clinically significant for stereotactic pulmonary treatments. It should also be recalled that the tumour had a total length of 2 cm and a diameter of 2 cm. Similarly to clinical practice, the MLC defined the shape of the beam. Jaws were positioned 0.8 cm (respectively 0.2 cm) behind the extremity of the leaves in the X direction (respectively Y direction). For each test and algorithm, 300 monitor units were planned in order to deliver approximately 2 Gy at the isocenter.

Considering the number of factors and levels for this study, a Taguchi L_{18} ($2^1 \times 3^7$) design of experiment was chosen, requiring 18 tests. Since one two-level factor and four three-level factors were selected, if all level combinations were to be used, a total of 162 trials would have to be conducted. The description of the 18 trials is given in

Table 1
Description of the 18 trials defined by design of experiment.

Trials	Collimator	Gantry	Energy	X jaws	Y jaws
#1	0°	0°	6 MV	2 cm	2 cm
#2	0°	0°	6 MV FFF	5 cm	5 cm
#3	0°	0°	16 MV	8 cm	8 cm
#4	0°	160°	6 MV	8 cm	8 cm
#5	0°	160°	6 MV FFF	2 cm	2 cm
#6	0°	160°	16 MV	5 cm	5 cm
#7	0°	90°	6 MV FFF	5 cm	8 cm
#8	0°	90°	16 MV	8 cm	2 cm
#9	0°	90°	6 MV	2 cm	5 cm
#10	90°	0°	16 MV	5 cm	2 cm
#11	90°	0°	6 MV	8 cm	5 cm
#12	90°	0°	6 MV FFF	2 cm	8 cm
#13	90°	160°	6 MV FFF	8 cm	5 cm
#14	90°	160°	16 MV	2 cm	8 cm
#15	90°	160°	6 MV	5 cm	2 cm
#16	90°	90°	16 MV	2 cm	5 cm
#17	90°	90°	6 MV	5 cm	8 cm
#18	90°	90°	6 MV FFF	8 cm	2 cm

Table 1.

2.6. Analysis

For each trial, the evaluation of the performances of the dose calculation systems was assessed by calculating the difference between the calculated dose and the measured dose. The dose differences (δ) were calculated for each pixel of the film image using home-made software based on ImageJ [35], according to the equation given in Section 2.4. Calculated dose was interpolated to match the film resolution.

For each trial and each of the three films, signal-to-noise ratio (S/N) was calculated as:

$$S/N = \frac{\text{Mean dose difference inside the ROI}}{\text{Standard deviation of dose difference inside the ROI}}$$

The ROI was first defined as the intersection between the irradiation field and the phantom (Fig. 2). The beam and phantom intersection was then segmented into three regions of interest (Fig. 2): the tumour, the shadow of the tumour and the intersection between the lung region and the irradiation field (minus the shadow region).

For each region of interest, mean dose differences, mean S/N ratio and their associated standard deviation for the 18 trials were reported. The S/N ratio metric reflects the variability in the response [36]. It is a relative quantity which can be used for comparison purposes to evaluate the statistical significance of the difference between the target and the samples' mean. Details on the S/N ratio are provided in Supplementary data 1. Note that in this study and contrary to what is usually done in the frame of signal processing, the signal is a dose difference. As a consequence, the aim is not to maximize the S/N but to reach the expected value (zero).

To evaluate the reproducibility of the films, the measurements of the 18 tests were performed 3 times. The performances of each algorithm could thus be synthetized by a S/N ratio, derived from the mean S/N ratio of the 18 trials and the associated standard deviation ($\sigma_{S/N}$), divided by $\sqrt{3}$. The meaning of $\sqrt{3}$ is the following: in statistics it is well known that if samples of size "n" are randomly taken from a population with a standard deviation σ, then the standard deviation of the means of the samples is $\frac{\sigma}{\sqrt{n}}$. In our case, n = 3 repetitions (3 films) were made and for each film the S/N ratio was calculated for each trial. Then, the standard deviation of the mean of the S/N ratios is $\frac{\sigma_{S/N}}{\sqrt{3}}$, where $\sigma_{S/N}$ is the standard deviation of the S/N ratios for the various trials. These two metrics took into account the reproducibility of the film measurements and the dispersion of dose differences inside the intersection region.

An analysis of variance (ANOVA) was performed on the intersection between the irradiation field and the phantom to identify factors that could influence the results. This analysis was based on the calculation of the mean S/N ratio for the trials sharing the same value of a specific factor. For the factor at 2 levels (collimator angulation), each level (0° and 90°) grouped 9 of the 18 tests and was characterized by the mean S/N ratio and associated standard deviation for the 9 tests divided by $\sqrt{3}$. For the 4 factors at 3 levels (gantry angulation, energy, X jaws and Y jaws), each level included 6 of the 18 trials and was represented by the average of the S/N ratio of these 6 trials and the associated standard deviation divided by $\sqrt{3}$. If the mean value of the S/N ratio for a confidence interval of 95% intercepts the value 0, the corresponding parameter is not considered significant. Conversely, if the S/N ratio value does not intercept 0 within one standard deviation, the corresponding parameter is considered significant.

3. Results

3.1. Validation of algorithms in a water phantom

For the 6 MV and 16 MV beam models, mean dose deviations were 0.2 ± 0.5% for the AAA, −0.4 ± 0.6% for the AXB_D_m, 0.5 ± 1.8% for the PB, 0.7 ± 0.9% for the XVMC_D_m, and 0.3 ± 1.2% for the CCV algorithms. All dose deviations were smaller than 2%. For all algorithms except the CCV, the field size defined by Y_2 was statistically significant with a systematic overestimation of the computed dose for

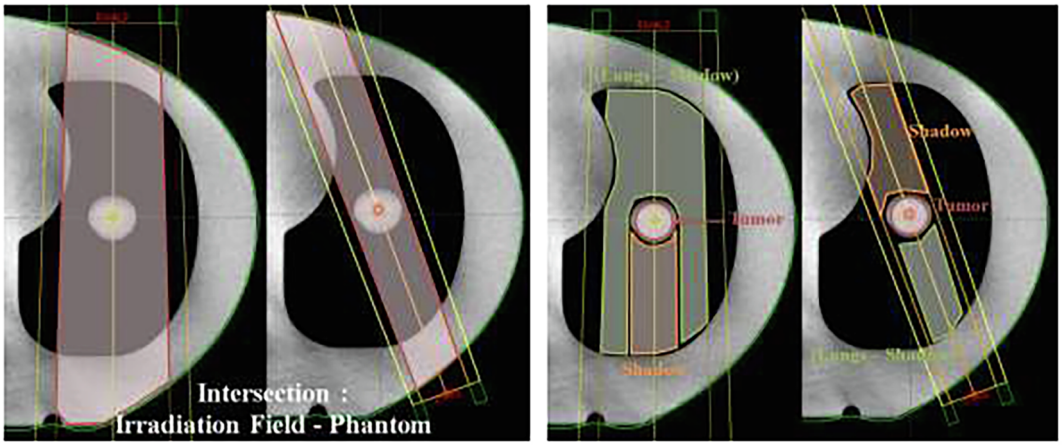

Fig. 2. Left: representation of the intersection between the irradiation field and the phantom for two trials. Right: segmentation of the region of analysis for two trials (Red: the tumor, Green: shadow of the tumor and Orange: (Lungs – Shadow)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Mean difference doses and signal-to-noise ratio for all algorithms, between the irradiation field and the heterogeneous phantom. Significant values are in bold.

	Mean difference on image [%]	S/N	Number of significant levels
AAA	2.2	**0.4 ± 0.3**	10
AXB_D_m	−0.6	−0.1 ± 0.4	1
AXB_D_w	0.5	0.2 ± 0.5	5
PB	19.4	**1.5 ± 0.1**	14
XVMC_D_m	2.0	0.5 ± 0.5	6
XVMC_D_w	1.2	0.4 ± 0.6	7
CCV	−0.2	0.0 ± 0.4	4

smaller field sizes. Dose computed on the axis were systematically overestimated for all algorithms except CCV and AXB whereas out of field doses were underestimated.

Mean dose deviation for the 6 MV FFF beam models were −0.7 ± 0.5% for the AAA, −0.1 ± 0.8% for the AXB_D_m, 1.6 ± 1.5% for the PB, 1.7 ± 0.8% for the XVMC_D_m and −1.3 ± 1.5% for the CCV algorithms. For AAA, AXB and XVMC algorithms, dose deviations were smaller than 4%. Larger deviations (up to 7%) were observed for the PB and CCV algorithms. Compared to 6 MV and 16 MV beam models, the agreement between computed and measured doses is deteriorated for the FFF beam, especially for the PB, XVMC and CCV algorithms. The adjustment of the beam models and investigations conducted with the manufacturer gave no better results.

3.2. Analysis on the intersection between the irradiation field and the heterogeneous phantom

The mean dose differences between measurements and calculations, the mean signal-to-noise ratio (S/N) were reported in Table 2, for the 18 trials and the five algorithms. The number of significant levels is also reported. The levels are considered not significant when the mean value of the S/N ratio intercepts the value 0, for a confidence interval of 95%. Conversely, if the S/N ratio does not intercept 0 within one standard deviation, the corresponding parameter is considered significant. Considering the factors studied (one factor -energy- with 2 levels and 4 factors -gantry angulation, collimator angulation, X and Y jaws- with 3 levels) a maximum number of 14 significant levels can be obtained.

The dose distributions calculations obtained with AXB and CCV algorithms were closest to the measurements, followed by the XVMC, AAA and finally PB.

3.3. Analysis of variance

An ANOVA was carried out to identify the influential factors for the different algorithms. The number of significant levels for which the mean S/N ratio was larger than its associated standard deviation is given in Table 2. Figs. 3–5 graphically represents the mean S/N ratio for some of the factors studied.

The AXB_D_m showed one significant negative S/N ratio for the smallest field size in the Y direction highlighting a 3.4% under-estimation of the dose. Interestingly, the smallest field size in the X direction was not found significant, possibly because of a difference in the

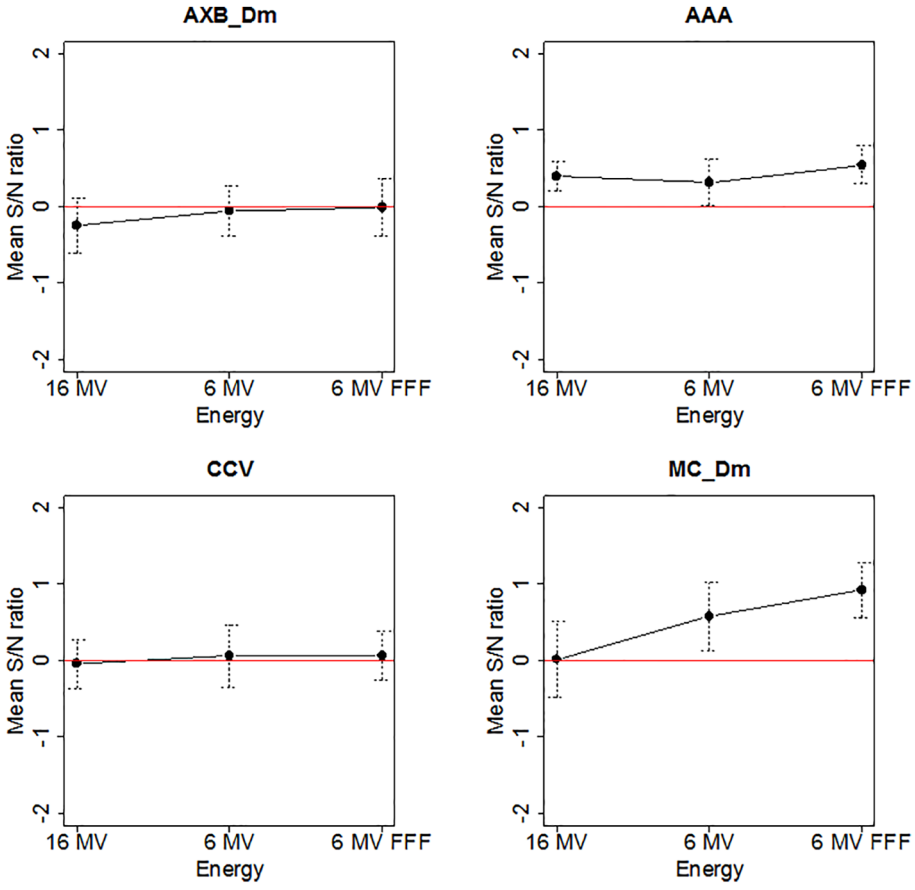

Fig. 3. Variation of the mean S/N ratio of the energy factor in the case of various algorithms.

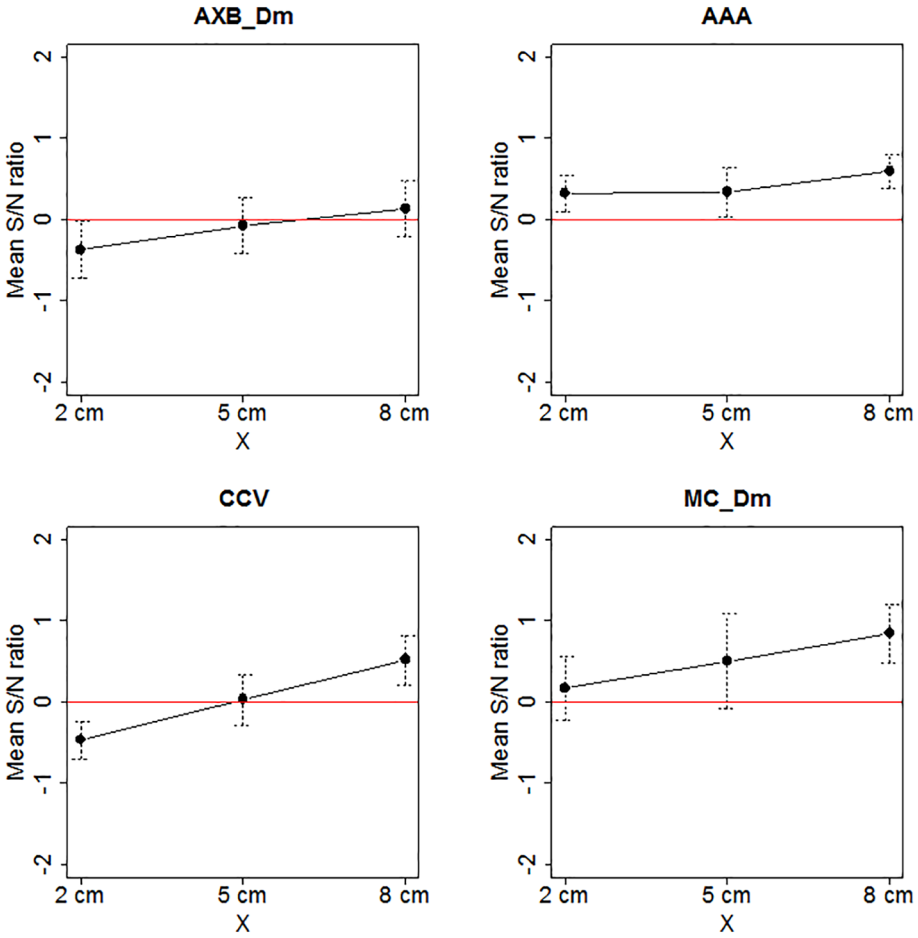

Fig. 4. Variation of the mean S/N ratio of the X jaw factor in the case of various algorithms.

penumbra and scattering modelling. The AXB_D_w revealed 5 levels statistically significant: gantry at 0°, X at 8 cm and Y at 2 cm, 5 cm and 8 cm. The difference between the two modalities could be explained by the energy dependent response function, in non-water materials, which is based on the material properties of each voxel (D_m) or on the small volume of water which has a minimal impact on this parameter (D_w).

The CCV revealed 4 levels statistically significant: X at 2 cm, 8 cm and Y at 2 cm and 5 cm. A very simplified modeling was used for the CCV algorithm which could explain these results: the beam model only requires 30 relative dose points per energy in various conditions. Conclusions regarding the performances of the CCV algorithm must therefore be balanced with the simplistic beam model used by Mobius.

Because the PB over-estimated the dose of nearly 20%, the S/N ratio for all the studied levels was positive. The analysis of the AAA revealed that 8 levels out of 14 induced a statistically significant deviation: collimator at 0° and 90°, gantry at 0 and 90°, energy at 6 FFF and 16 MV, X at 8 cm and Y at 5 and 8 cm. The overall over-estimation of 2.2% reported in Table 2 could be explained by the over-estimation of the 6 FFF and 16 MV beam models (+3.0% and +2.4% respectively against +1.3% for the 6 MV). As a consequence, the S/N ratio values did not oscillate around 0 and the fluctuations of the levels of a given factor. Based on this ANOVA, one might conclude that the best agreement between dose calculations and measurements for the AAA algorithm could be achieved at 6 MV for a field size of 5 cm in the X direction and 2 cm in the Y direction and with a gantry angle of 160°.

The analysis of XVMC_D_m showed that 6 levels had a statistically significant deviation: collimator 0°, gantry at 0°, energy at 6 FFF, X at 8 cm and Y at 5 cm and 8 cm. The over-estimation raised to 2.0% as reported in Table 2. As for XVMC_D_w, 7 levels revealed a statically significant deviation. The additional factor compared to the modality D_m was the Y field size at 2 cm. A strong over-estimation of the dose was observed for the 6 FFF (+4.5% for the D_m modality, +3.5% for the D_w modality) strengthening the overestimation already observed in water.

3.4. Analysis by region of interest (ROI)

The comparison on the intersection of the entire field provided an initial assessment of algorithm performances. However, the Fig. 6 shows large dose differences within this intersection. An analysis over the entire intersection might mask these differences due to possible compensation for underestimations in one area with overestimations in another area. Thus, an analysis by region of interest seemed more appropriate for a detailed assessment of the performances of the algorithms.

Mean dose differences and mean signal-to-noise ratio (S/N) were reported in Table 3, for the 18 trials and the 3 regions of interest studied.

According to the results obtained for all algorithms except PB, the performances were identical in medium such as lung (Lungs – Shadow). For all algorithms except PB, the mean S/N ratio were found similar and the mean dose differences varied from −1.4% to 1.4%.

For all regions of interest, the PB overestimates the dose compared to the measurements. Mean difference doses range from 10.5% (in the tumor region) to 29.7% (for the other two regions of interest).

In the tumor region, only the XVMC algorithm obtained a statistically significant mean S/N ratio with an average dose difference up to 3.7%. For the shadow of the tumor, the mean S/N ratio was statistically significant for the AAA and CCV, with mean difference doses of 4.4%

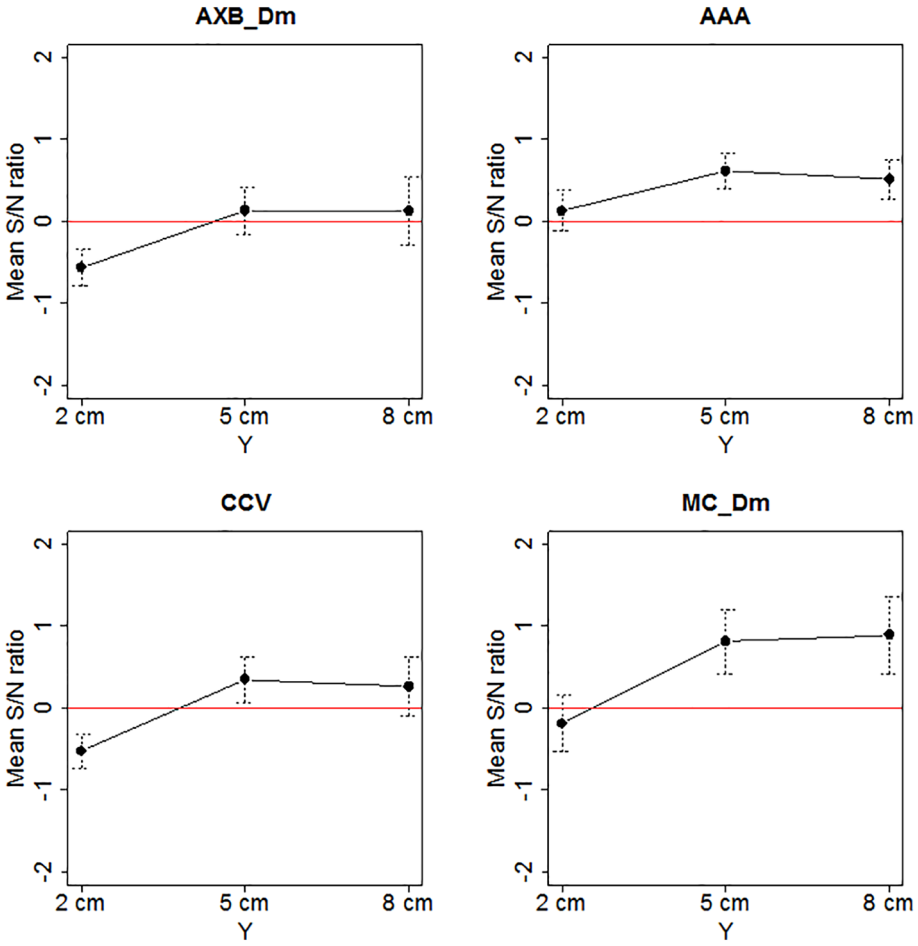

Fig. 5. Variation of the mean S/N ratio of the Y jaw factor in the case of various algorithms.

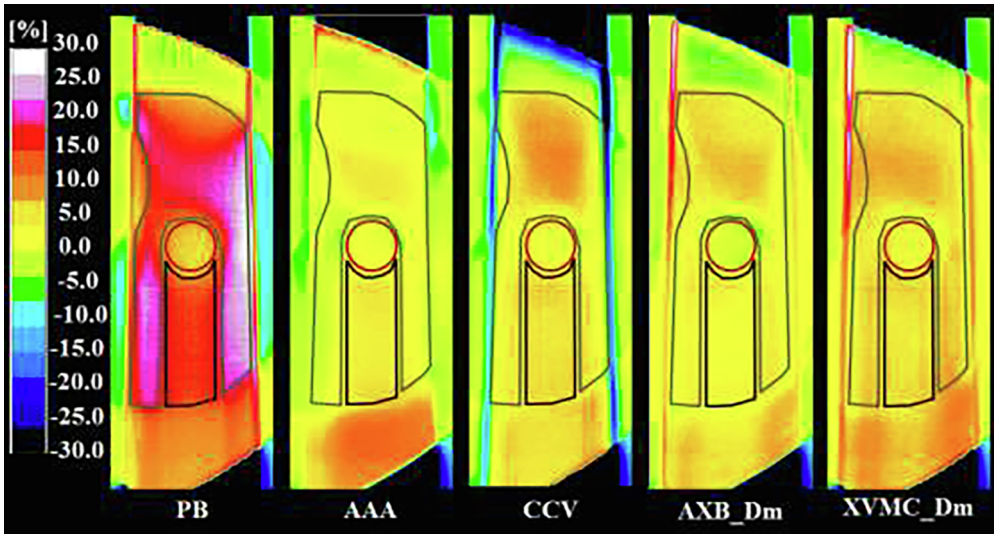

Fig. 6. Dose distribution differences for trial #2, for all algorithms. The color scale shows relative dose differences, compared to the film measurements, from −30% up to +30% (local normalization).

Table 3
Mean difference doses for different regions, for 5 algorithms. In bold: mean values larger than the standard deviation.

	Tumor		Shadow		Lungs – Shadow	
	Mean difference [%]	S/N	Mean difference [%]	S/N	Mean difference [%]	S/N
AAA	2.0	0.9 ± 0.9	4.4	**1.7 ± 0.9**	1.4	0.3 ± 0.4
AXB_D_m	0.0	−0.1 ± 0.8	0.0	0.4 ± 1.0	−1.4	−0.2 ± 0.6
AXB_D_w	1.3	0.6 ± 0.7	0.0	0.7 ± 1.0	0.8	0.1 ± 0.8
PB	10.5	**4.7 ± 1.0**	27.1	**6.1 ± 0.7**	29.7	**3.3 ± 0.4**
XVMC_D_m	3.7	**2.0 ± 1.0**	2.4	1.4 ± 1.3	0.5	0.3 ± 0.9
XVMC_D_w	3.1	**1.8 ± 0.9**	1.7	1.2 ± 1.6	0.0	0.1 ± 1.1
CCV	1.6	0.9 ± 1.4	2.8	**1.6 ± 1.3**	1.0	0.4 ± 0.6

and 2.8%, respectively.

Despite the results for the entire irradiation field close to zero for the CCV and AXB algorithms, the difference in the dose deposition was visible in the segmentation by region of interest: the tumor and the shadow of the tumor. The results obtained for the AXB (all modalities) were closer to 0 than the CCV.

4. Discussion

Predictably, the PB algorithm (type “a”) showed significant deviations from film measurements. The PB over-estimated the dose relative to measurements of 19.4%. The most important approximation in this model concerns the modelling of the effects due to heterogeneities encountered in the crossed medium: each “pencil” was corrected for heterogeneity along its main axis. However, the “pencil” does not consider the lateral scattering. This algorithm is thus strongly discouraged for dose calculation in heterogeneous media [4,37].

Dose deviations reported in this study were consistent with other studies were algorithms were compared either to Monte Carlo or measurements: mean deviations were reported smaller than 5% for AAA [8,38], 1.1% for AXB [8,39,40], 16% for path-length algorithms like PB [11,41], 3.1% for XVMC [40,42] and 0.5% for CCV [43].

Results of the present analysis showed that AXB (type “c”) and CCV (type “b”) were almost equivalent. The differences in dose for XVMC (type “c”) and AAA (type “b”) showed a dose difference compared to films smaller than 2.5%. These results were in agreement with the literature. Indeed, Dunn *et al.* [8] obtained an average dose difference by ionization chamber of 2.9% ± 1.2% in an anthropomorphic phantom, for the AAA algorithm. In the same way, Tsuruta *et al.* [40] showed that the XVMC values agreed with the film measurements within 2.2% for PDD evaluation, in heterogeneous phantoms.

International Commission on Radiation Units and Measurements (ICRU) [44] recommended an overall dose accuracy within 5%. Other recommendations [4,45] advocated the overall uncertainty in the computed dose distribution should be less than 3%. With this in mind, the mean dose differences observed in this study were acceptable for clinical use for the AXB (all modalities) and CCV. The clinical use of the XVMC (all modalities) and AAA algorithms remained acceptable. As a matter of fact, at the intersection between the irradiation field and the phantom, their mean differences were smaller than 3% or even 2% for the XVMC (all modalities).

In the present study, differences were noted between AXB and XVMC although they can be considered equivalent in terms of dosimetric performances. The source modeling and statistical errors derived from the calculation process within the body might explain the significant difference. Approximations were carried out with the XVMC calculation process compared to a real Monte Carlo in order to make it faster and clinically usable. However, these approximations are unknown and could possibly explain the dose differences compared to film measurements. Another source of deviation between the two type “c” algorithms could be the material assignment. Indeed, as mentioned in Section 2.2, from the same CT scanner calibration curve, the assignment of materials is performed by different approaches. Moreover, only human tissues (lung, muscle, …) were assigned for dose calculations while measurements were performed in a film sheet placed inside two slices of plastic material which can introduce a bias in the results. Physical density and corresponding material are automatically assigned to the anthropomorphic phantom, based on a CT calibration curve. Unfortunately, with plastic material, it is only possible to fine tune the density (HU). Thus, this material could not be forced with the different densities composing the anthropomorphic phantom. Zavan *et al.*[46] faced a similar issue when comparing AXB_Dm calculations to measurements with films placed in 3D printed low-density phantoms. They estimated to 0.5% the uncertainty arising from overlapping dosimetric quantities (dose-to-water/dose-to-film) and considered dose-to-water and dose to most soft tissues clinically insignificant (within 2%). As mentioned by Andreo [47], the key issue for Monte Carlo Treatment Planning is the determination of the tissue composition and their mean excitation energies which need to be improved in future TPS development.

Interesting to note is that influencing factors (like the field size defined by Y_2) highlighted by the experimental design conducted in water are not revealed by the experimental design conducted in the thorax phantom. One possible explanation is that the deviations observed in water are compensated by the heterogeneity correction of the algorithms. However, separating the uncertainties coming from the algorithm configuration from those arising from the dose calculation in inhomogeneous medium remains delicate in this case because factors, levels and dosimeter differed between the experimental designs conducted in both media. Linac conditions (like jaws and MLC calibrations for instance) at the time of measurements in water and of measurements in the heterogeneous phantom were also different which could bias the comparison of the results in both media. Dose differences observed in the heterogeneous phantom encompass differences that could possibly be observed in water (for example in small fields) and cannot be exclusively attributed to differences in the heterogeneity correction method.

A scoring of the algorithms can be suggested using the number of significant levels. This approach based on 18 beams gathering various clinical situations is complementary to the known theoretical classifications suggested by Fogliata *et al.* [5]. Based on these results in lung heterogeneities, the scoring could be as follow (from best to worst): AXB_D_m, CCV, AXB_D_w, XVMC_D_m, XVMC_D_w, AAA and last PB.

The segmentation analysis allowed localizing spatially the dose differences: in the tumor for XVMC and in the shadow for AAA and CCV. Significant differences were found between the CCV and AXB algorithms, particularly in the region following the tumor (shadow) whereas these two algorithms showed similar results in Table 2. These results confirmed that an analysis over the entire intersection masked differences due to compensations between underestimations and overestimations in different areas.

Experimental designs allowed studying the influence of 5 factors with a total of 14 levels in only 18 trials and enabled covering various configurations often encountered in clinical practice. If all level

combinations of the five factors considered were to be tested, 162 trials would have been required. Up to 7 factors with 3 levels and 1 factor with 2 levels can be studied simultaneously with the Taguchi L_{18} ($2^1 \times 3^7$) array and if necessary other arrays could be used. Designs of experiments showed to be a powerful tool to minimize the number of trials and at the same time to keep the possibility to conduct a detailed statistical analysis.

5. Conclusion

DoE is a robust statistical method that was used in order to compare several algorithms. The interest of using DoE is, on one hand, reducing the number of trials to be done in order to estimate the main effects of the factors and, on the other hand, improving the accuracy of the estimation. It was shown in this study that it is possible with just 18 tests (instead of 162 trials if all level combinations of the five factors considered were to be tested) to compare the performances of dose calculation systems with five different algorithms, by taking into account five factors (gantry and collimator angulations, energy, X and Y jaws) at various levels. The influential factors for each algorithm were highlighted and a classification of the algorithms from a clinical point of view was suggested for the lung case. Dose differences were also spatially localized. The analysis led to the conclusion that Acuros XB is the best of the five algorithms studied for dose calculations in heterogeneous media for the tests investigated.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmp.2019.04.014.

References

[1] Montgomery DC. Design and analysis of experiments. Wiley; 2017.
[2] Dean A, Voss D. Design and analysis of experiments. Springer; 1999.
[3] Taguchi G, Chowdhury S, Wu Y. Quality engineering: The Taguchi Method. Taguchi's Quality Engineering Handbook. Wiley-Interscience; 2007. p. 56–123.
[4] Papanikolaou N, Battista JJ, Boyer AL, Kappas C, Klein E, Mackie TR, et al. Tissue inhomogeneity corrections for megavoltage photon beams. AAPM Report NO 852004.
[5] Fogliata A, Cozzi L. Dose calculation algorithm accuracy for small fields in non-homogeneous media: the lung SBRT case. Phys Med 2016.
[6] Knoos T, Wieslander E, Cozzi L, Brink C, Fogliata A, Albers D, et al. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. Phys Med Biol. 2006;51:5785–807.
[7] Ojala JJ, Kapanen MK, Hyodynmaa SJ, Wigren TK, Pitkanen MA. Performance of dose calculation algorithms from three generations in lung SBRT: comparison with full Monte Carlo-based dose distributions. J Appl Clin Med Phys 2014;15:4662.
[8] Dunn L, Lehmann J, Lye J, Kenny J, Kron T, Alves A, et al. National dosimetric audit network finds discrepancies in AAA lung inhomogeneity corrections. Phys Med 2015;31:435–41.
[9] Kawai D, Takahashi R, Kamima T, Baba H, Yamamoto T, Kubo Y, et al. Variation of the prescription dose using the analytical anisotropic algorithm in lung stereotactic body radiation therapy. Phys Med 2017;38:98–104.
[10] Bragg CM, Wingate K, Conway J. Clinical implications of the anisotropic analytical algorithm for IMRT treatment planning and verification. Radiother Oncol 2008;86:276–84.
[11] Gagne IM, Zavgorodni S. Evaluation of the analytical anisotropic algorithm in an extreme water-lung interface phantom using Monte Carlo dose calculations. J Appl Clin Med Phys 2006;8:33–46.
[12] Aarup LR, Nahum AE, Zacharatou C, Juhler-Nottrup T, Knoos T, Nystrom H, et al. The effect of different lung densities on the accuracy of various radiotherapy dose calculation methods: implications for tumour coverage. Radiother Oncol 2009;91:405–14.
[13] Maheshwari V, Rangaiah GP, Lau T, Samavedham L. Application of design of experiments in hemodialysis: optimal sampling protocol for β2-microglobulin kinetic model. Chem Eng Sci 2015;131:84–90.
[14] Khanna N, Davim J. Design-of-experiments application in machining titanium alloys for aerospace structural components. Measurement 2014.
[15] Makary MS, Shah SH, Warhadpande S, Vargas IG, Sarbinoff J, Dowell JD. Design-of-experiments approach to improving inferior vena cava filter retrieval rates. J Am Coll Radiol 2017;14:72–7.
[16] Dufreneix S, Legrand C, Di Bartolo C, Bremaud M, Mesgouez J, Tiplica T, et al. Design of experiments in medical physics: application to the AAA beam model validation. Phys Med 2017;41:26–32.
[17] Miki K, Saito A, Nakashima T, Murakami Y, Kimura T, Nishibuchi I, et al. Evaluation of optimization workflow using design of experiment (DoE) for various field configurations in volumetric-modulated arc therapy. Phys Med 2018;54:34–41.
[18] Fisher RA. The design of experiments. Edinburgh: Macmillan Pub Co; 1937.
[19] Eriksson L, Johansson E, Kettaneh-Wold N, Wikström C, Wold S. Design of experiments, principles and applications. UMetrics Academy; 2008.
[20] Roy RK. Design of Experiments Using the Taguchi Approach: 16 Steps to Product and Process Improvement. Wiley; 2001.
[21] Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. Med Phys 2011;38:2523–34.
[22] Devic S. Reference radiochromic film dosimetry: Review of technical aspects. Phys Med 2016;32:541–56.
[23] Tillikainen L, Helminen H, Torsti T, Siljamaki S, Alakuijala J, Pyyry J, et al. A 3D pencil-beam-based superposition algorithm for photon dose calculation in heterogeneous media. Phys Med Biol 2008;53:3821–39.
[24] Sievinen J, Ulmer W, Kaissl W. AAA Photon Dose Calculation Model in Eclipse™ Varian Medical Systems.
[25] Vassiliev ON, Wareing TA, McGhee J, Failla G, Salehpour MR, Mourtada F. Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. Phys Med Biol 2010;55:581–98.
[26] Failla G, Wareing T, Archambault Y, Thompson S. Acuros®XB advanced dose calculation for the Eclipse™ treatment planning system. Varian Med Syst 2015.
[27] Mohan R, Chui CS. Use of fast Fourier transforms in calculating dose distributions for irregularly shaped fields for three-dimensional treatment planning. Med Phys 1987;14:70–7.
[28] Fippel M. Fast Monte Carlo dose calculation for photon beams based on the VMC electron algorithm. Med Phys 1999;26:1466–75.
[29] Fragoso M, Wen N, Kumar S, Liu D, Ryu S, Movsas B, et al. Dosimetric verification and clinical evaluation of a new commercially available Monte Carlo-based dose algorithm for application in stereotactic body radiation therapy (SBRT) treatment planning. Phys Med Biol 2010;55:4445–64.
[30] McNutt T. Pinnacle3® White paper: dose calculations - collapsed cone convolution superposition and delta pixel beam. Philips Med Syst 2002.
[31] Mackie TR, Scrimger JW, Battista JJ. A convolution method of calculating dose for 15-MV X rays. Med Phys 1985;12:188–96.
[32] Ahnesjo A. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. Med Phys 1989;16:577–92.
[33] Chen Q, Chen M, Lu W. Ultrafast convolution/superposition using tabulated and exponential kernels on GPU. Med Phys 2011;38:1150–61.
[34] Childress C, Stevens E, Eklund D, Zhang M. Mobius3D White Paper: Dose Calculation Algorithm. 2012.
[35] ImageJ: Image Processing and Analysis in Java. National Institutes of Health. https://imagej.nih.gov/ij/index.html [accessed 10 April 2019].
[36] Fowlkes WY, Creveling CM. Engineering methods for robust product design. Boston: Addison-Wesley Publishing Company; 1995.
[37] ICRU. ICRU Report 91: Prescribing, Recording, and Reporting of Stereotactic Treatments with Small Photon Beams. Journal of the ICRU. 2017;14.
[38] Stathakis S, Esquivel C, Quino LV, Myers P, Calvo O, Mavroidis P, et al. Accuracy of the small field dosimetry using the Acuros XB dose calculation algorithm within and beyond heterogeneous media for 6 MV Photon Beams. Int J Med Phys Clin Eng Radiat Oncol 2012;01(03):10.
[39] Han T, Mikell JK, Salehpour M, Mourtada F. Dosimetric comparison of Acuros XB deterministic radiation transport method with Monte Carlo and model-based convolution methods in heterogeneous media. Med Phys 2011;38:2651–64.
[40] Tsuruta Y, Nakamura M, Miyabe Y, Nakata M, Ishihara Y, Mukumoto N, et al. Use of a second-dose calculation algorithm to check dosimetric parameters for the dose distribution of a first-dose calculation algorithm for lung SBRT plans. Phys Med 2017;44:86–95.
[41] Dobler B, Walter C, Knopf A, Fabri D, Loeschel R, Polednik M, et al. Optimization of extracranial stereotactic radiation therapy of small lung lesions using accurate dose calculation algorithms. Radiat Oncol 2006;1:45.
[42] Verma TR, Painuly NK, Mishra SP, Singh N, Bhatt MLB, Jamal N, et al. Evaluation of dose calculation accuracy of various algorithms in lung equivalent inhomogeneity: comparison of calculated data with Gafchromic film measured results. J Cancer Res Ther 2017;13:1007–14.
[43] McDonald DG, Jacqmin DJ, Mart CJ, Koch NC, Peng JL, Ashenafi MS, et al. Validation of a modern second-check dosimetry system using a novel verification phantom. J Appl Clin Med Phys 2017;18:170–7.
[44] ICRU. ICRU Report 24: Determination of absorbed dose in a patient irradiated by beams of X or gamma rays in radiotherapy procedures. Journal of the ICRU. 1976;os13:76.
[45] Smilowitz JB, Das IJ, Feygelman V, Fraass BA, Kry SF, Marshall IR, et al. Medical physics practice guideline 5.a.: commissioning and QA of treatment planning dose calculations - megavoltage photon and electron beams. J Appl Clin Med Phys 2015;16:14–34.
[46] Zavan R, McGeachy P, Madamesila J, Villarreal-Barajas JE, Khan R. Verification of Acuros XB dose algorithm using 3D printed low-density phantoms for clinical photon beams. J Appl Clin Med Phys 2018;19:32–43.
[47] Andreo P. Dose to 'water-like' media or dose to tissue in MV photons radiotherapy treatment planning: still a matter of debate. Phys Med Biol 2015;60:309–37.