

PAPER • OPEN ACCESS

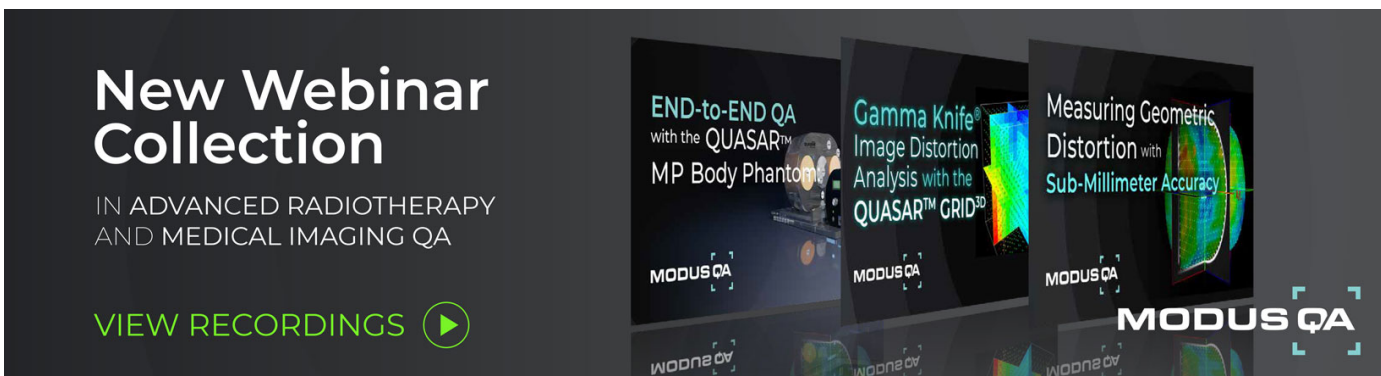
## Efficient uncertainty quantification for Monte Carlo dose calculations using importance (re-)weighting

To cite this article: P Stammer *et al* 2021 *Phys. Med. Biol.* **66** 205003


View the [article online](#) for updates and enhancements.

You may also like

- [SRP Annual General Meeting: The Changing Role of the Radiation Protection Professional \(Cardiff, 16-17 April 2002\)](#)  
W Jones
- [The effect of systematic set-up deviations on the absorbed dose distribution for left-sided breast cancer treated with respiratory gating](#)  
A Edvardsson and S Ceberg
- [Implementation of random set-up errors in Monte Carlo calculated dynamic IMRT treatment plans](#)  
S Stapleton, S Zavgorodni, I A Popescu et al.



**New Webinar Collection**  
IN ADVANCED RADIOTHERAPY AND MEDICAL IMAGING QA

[VIEW RECORDINGS](#) 

**MODUS QA**

The banner features three overlapping video thumbnails with the following titles:  
1. END-to-END QA with the QUASAR™ MP Body Phantom  
2. Gamma Knife® Image Distortion Analysis with the QUASAR™ GRID<sup>3D</sup>  
3. Measuring Geometric Distortion with Sub-Millimeter Accuracy



## PAPER

## Efficient uncertainty quantification for Monte Carlo dose calculations using importance (re-)weighting

## OPEN ACCESS

## RECEIVED

1 July 2021

## REVISED

13 September 2021

## ACCEPTED FOR PUBLICATION

20 September 2021

## PUBLISHED

7 October 2021

P Stammer<sup>1,2,3</sup> , L Burigo<sup>2,4</sup>, O Jäkel<sup>2,3,4,5</sup> , M Frank<sup>1,3</sup> and N Wahl<sup>2,4</sup> <sup>1</sup> Karlsruhe Institute of Technology, Steinbuch Centre for Computing, Karlsruhe, Germany<sup>2</sup> German Cancer Research Center—DKFZ, Department of Medical Physics in Radiation Oncology, Heidelberg, Germany<sup>3</sup> HIDS4Health—Helmholtz Information and Data Science School for Health, Karlsruhe/Heidelberg, Germany<sup>4</sup> Heidelberg Institute for Radiation Oncology (HIRO), Heidelberg, Germany<sup>5</sup> Heidelberg Ion Beam Therapy Center—HIT, Department of Medical Physics in Radiation Oncology, Heidelberg, GermanyE-mail: [pia.stammer@kit.edu](mailto:pia.stammer@kit.edu)

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



**Keywords:** proton therapy, uncertainty, setup error, range error, Monte Carlo, importance sampling, intensity modulated particle therapy (IMPT)

## Abstract

**Objective.** To present an efficient uncertainty quantification method for range and set-up errors in Monte Carlo (MC) dose calculations. Further, we show that uncertainty induced by interplay and other dynamic influences may be approximated using suitable error correlation models. **Approach.** We introduce an importance (re-)weighting method in MC history scoring to concurrently construct estimates for error scenarios, the expected dose and its variance from a single set of MC simulated particle histories. The approach relies on a multivariate Gaussian input and uncertainty model, which assigns probabilities to the initial phase space sample, enabling the use of different correlation models. Through modification of the phase space parameterization, accuracy can be traded between that of the uncertainty or the nominal dose estimate. **Main results.** The method was implemented using the MC code TOPAS and validated for proton intensity-modulated particle therapy (IMPT) with reference scenario estimates. We achieve accurate results for set-up uncertainties ( $\gamma_{2\text{ mm}/2\%} \geq 99.01\%$  ( $E[\mathbf{d}]$ ),  $\gamma_{2\text{ mm}/2\%} \geq 98.04\%$  ( $\sigma(\mathbf{d})$ )) and expectedly lower but still sufficient agreement for range uncertainties, which are approximated with uncertainty over the energy distribution. Here pass rates of 99.39% ( $E[\mathbf{d}]$ )/ 93.70% ( $\sigma(\mathbf{d})$ ) (range errors) and 99.86% ( $E[\mathbf{d}]$ )/ 96.64% ( $\sigma(\mathbf{d})$ ) (range and set-up errors) can be achieved. Initial evaluations on a water phantom, a prostate and a liver case from the public CORT dataset show that the CPU time decreases by more than an order of magnitude. **Significance.** The high precision and conformity of IMPT comes at the cost of susceptibility to treatment uncertainties in particle range and patient set-up. Yet, dose uncertainty quantification and mitigation, which is usually based on sampled error scenarios, becomes challenging when computing the dose with computationally expensive but accurate MC simulations. As the results indicate, the proposed method could reduce computational effort while also facilitating the use of high-dimensional uncertainty models.

## 1. Introduction

Monte Carlo (MC) methods are considered the gold standard for dose calculation in radiotherapy treatment planning due to their accuracy (Weng *et al* 2003, Paganetti 2012). However, the accuracy of a simulated compared to a delivered dose is not only determined by the chosen dose engine, but also compromised by treatment uncertainties in water-equivalent path length, patient set-up and anatomy. Especially in proton and carbon-ion therapy, the high dose localization in the Bragg-peak usually does not allow for uncertainty quantification and mitigation using approximations known for photon therapy, such as the static dose cloud (Lomax 2008a, 2008b).

Consequently, particle therapy demands personalized robustness analyses and mitigation. Such techniques may be based on explicit propagation of input uncertainties using probabilistic methods and statistical analysis

(Bangert *et al* 2013, Kraan *et al* 2013, Park *et al* 2013, Perkó *et al* 2016, Wahl *et al* 2017, 2020) or worst-case estimates (Casiraghi *et al* 2013, McGowan *et al* 2015, Lowe *et al* 2016). Most of these methods then further translate to robust and probabilistic optimization to extend the conventional, generic margin approach to uncertainty mitigation (Sobotta *et al* 2010, Fredriksson 2012, Liu *et al* 2012, Unkelbach *et al* 2018).

The additional computational effort of robustness analyses and robust optimization techniques, however, clashes with the long computation times of MC dose calculation. The use of faster, less accurate deterministic pencil-beam dose calculation algorithms instead is not always feasible, because their accuracy is low in particularly heterogeneous anatomies like lung (Taylor *et al* 2017), which at the same time show high sensitivity to uncertainties in range and set-up.

More efficient uncertainty quantification approaches for MC methods, developed for example by the radiative transport community (e.g. Hu and Jin 2016, Poëtte 2018), often do not demonstrate an application to realistic patient data and it is not clear how well the results transfer. Also in many cases, more sophisticated methods are intrusive, which limits the applicability when using proprietary MC simulation engines.

In this paper, we introduce a simple, minimally-intrusive method for uncertainty quantification in MC dose computations. It is based on (re-)weighting a single set of MC simulated particle histories. Histories can either be weighted during the simulation, using multiple scoring routines, or post simulation given the storage of individual histories. The weighting, which can also be represented as multiplications of a weight vector with a history dose matrix, replaces simulations of different dose scenarios. The method enables uncertainty propagation during the simulation, making it possible to estimate the dose uncertainty induced by range and set-up errors from nominal dose calculations. In contrast to the conventional approach of simulating different scenarios separately, our method significantly reduces the required computational effort. We demonstrate the application of this method to specifically approximate expected value and variance of dose, given a respective uncertainty model for set-up and range errors, which includes the choice of different beam and pencil beam correlation scenarios.

The remainder of this paper is organized as follows: in section 2, we introduce basic definitions and notation, derive a direct computation of the expected value before introducing the concept of importance (re-)weighting for set-up and range uncertainty models. Section 3 then compares estimated expected doses and corresponding standard deviations to reference computations based on scenario sampling. The accuracy and convergence of the proposed method are discussed in section 4. Discussion and conclusion follow in sections 5 and 6, respectively.

## 2. Materials and methods

### 2.1. The MC method for dose computation

First, we briefly recapitulate the basic principles of the MC method for radiotherapy. This serves the purpose of establishing notation and parameters used to introduce our method and simplifying the illustration of later adaptations. For a more detailed description we refer to other sources, such as Paganetti (2012), Fippel and Soukup (2004), Ma *et al* (2002), Bielajew (1994), Mackie (1990), among many others.

The MC method is a numerical integration technique, based on random sampling. When used for dose calculations, a set of particles is created with properties including position, momentum and energy, which evolve dynamically over the course of a simulation. The initial values of these properties constitute the random input parameters of the MC simulation and are sampled from a known probability distribution function. On this basis, the trajectories of each primary particle and its secondaries are simulated and the deposited dose is aggregated, by sampling interactions such as scattering and energy loss according to physical laws and material properties. While this appears to be an intuitive simulation of the actual physical process, it is essentially a statistical method to solve the linear Boltzmann transport equation and therefore compute the expected value of a model with random input.

Let  $\xi$  be the vector of random input parameters of the dose simulation.  $\Phi_0(\xi)$  is the joint density of these parameters and is assumed to be known. For our purposes, which will not interfere with the simulation itself, we assume that the trajectory of a primary particle is given by the ‘black box’ simulation engine, yielding the dose deposited in voxel  $i$  within an individual particle’s history  $h_i(\xi)$ .

The nominal dose  $d_i$  in voxel  $i$  is given by the expected value  $\mathbb{E}_{\Phi_0}[h_i(\xi)]$ . The associated integration problem can be solved numerically using the MC method, which is equivalent to computing the mean over a sample of histories corresponding to realizations of  $\xi$ :

$$d_i = \mathbb{E}_{\Phi_0}[h_i(\xi)] = \int_{\mathcal{D}_{\Phi_0}} h_i(\xi) \cdot \Phi_0(\xi) d\xi \approx \frac{1}{H} \sum_{p=1}^H h_i(\xi_p), \quad (1)$$

where  $H$  is the sample size (number of computed primary particle histories) and  $\Xi_p, p = 1, \dots, H$  are realizations of the primary particle properties  $\xi \sim \Phi_0$ . In the following, we write  $\Xi_p \leftarrow \Phi_0$  when  $\Xi_p$  is a realization of the random variable  $\xi \sim \Phi_0$ .

Here we omit the dependence on random factors within the simulation, such as particle scattering, as well as their probability distribution. Particle histories  $h_i(\Xi_p)$ , for input realizations  $\Xi_p$ , implicitly also include realizations of these random parameters. For a large number of histories, their effect on the dose estimates can however be assumed to be constant.

## 2.2. Beam model

The initial state of each particle is represented by a point in the seven-dimensional phase space, which encompasses the particle position  $\mathbf{r} = (r_x, r_y, r_z)$ , momentum  $\mathbf{p} = (p_x, p_y, p_z)$  and energy  $E$ . We assume a Gaussian emittance model, i.e. the parameters within each pencil beam are multivariate normal distributed with

$$\Phi_0^b(\xi) = \Phi_0^b(\mathbf{r}, \varphi, E) = \mathcal{N}(\mu_\xi^b, \Sigma_\xi^b), \text{ for pencil beams } b = 1, \dots, B. \text{ Here, } \varphi = (\varphi_x, \varphi_y) = \left( \frac{dp_x}{dp_z}, \frac{dp_y}{dp_z} \right) \text{ describes}$$

the transverse divergence of the momentum direction from the axial beam direction.

The joint density over all pencil beams is then defined by a Gaussian mixture model

$$\Phi_0(\xi) = \sum_{b=1}^B w_b \Phi_0^b(\xi), \quad (2)$$

where  $w_b$  are the pencil beam weights.

To introduce our method, we will initially assume a simplified phase space where  $\varphi_x = \varphi_y = 0$ . Results including a distribution in the momentum direction can be found in the appendix.

## 2.3. Uncertainties

Among the most important sources of uncertainty in proton therapy are errors in the patient set-up  $\delta_r = (\delta_{r_x}, \delta_{r_y}, \delta_{r_z})$  and the proton range  $\delta_\rho$  (comp. Lomax 2008a, 2008b, Liu *et al* 2012, Park *et al* 2013, Perkó *et al* 2016). While these errors are random variables with, in principle, unknown probability distributions, we follow the common approach of assuming normally distributed errors (Unkelbach *et al* 2007, Fredriksson *et al* 2011, Bangert *et al* 2013, Perkó *et al* 2016, Wieser *et al* 2020).

Set-up errors directly affect the primary particle positions in an additive way, such that the actual position  $\mathbf{r}_\delta$  of a primary particle under uncertainty is given by its position  $\mathbf{r}$  according to the emittance model, plus the error  $\delta_r$ .

Uncertainties in the particle range are caused by a variety of factors, ranging from the conversion of Hounsfield units to stopping powers and imaging artifacts, over changes in the patient geometry to biological effects and inaccuracies in physics models (Unkelbach *et al* 2007, Lomax 2008a, Paganetti 2012, McGowan *et al* 2013). Here, we focus on calculational uncertainties, such as conversion errors, and model these by scaling the complete tissue density with the random factor  $\delta_\rho$  (comp. Lomax 2008a, Malyapa *et al* 2016, Souris *et al* 2019). Since the density is assumed to be deterministic, the error is not directly linked to a random input parameter. In section 2.7, we however present an approximation which models range errors using the initial energy distribution.

Sampling-based uncertainty quantification approaches, similar to Park *et al* (2013) or Kraan *et al* (2013), rely on repeated dose calculations for different realizations  $\Delta_k, k = 1, \dots, K$  of the error vector. For an individual error scenario  $\Delta_k$ , the dose is computed as

$$d_i^{\Delta_k} = \mathbb{E}_{\Phi(\xi, \Delta_k)}[h_i(\xi, \Delta_k)] \approx \frac{1}{H} \sum_{p=1}^H h_i(\Xi_p), \quad \Xi_p \leftarrow \Phi(\xi, \Delta_k). \quad (3)$$

In the case of set-up uncertainties  $\Phi(\xi, \Delta_k)$  for example corresponds to the nominal parameter density  $\Phi_0(\xi)$ , where all particle positions are shifted by  $\Delta_{r,k}$ . Due to its accuracy, this procedure is later used to obtain reference values to validate our results (see section 2.9). It is however extremely computationally expensive, since it requires numerous runs of the complete MC dose simulation.

## 2.4. Direct computation of the expected value

When the distribution  $\Psi(\xi_\delta)$  of the initial parameters under uncertainty can be explicitly defined, it is possible to compute the expected dose directly by replacing the nominal parameter distribution  $\Phi_0$  with  $\Psi$  in the MC dose simulation as follows:

$$E(d_i) = \mathbb{E}_{\Psi(\xi_\delta)}[h_i(\xi_\delta)] \approx \frac{1}{H} \sum_{p=1}^H h_i(\Xi_p), \quad \Xi_p \leftarrow \Psi(\xi_\delta). \quad (4)$$

For example when the error is additive, i.e.

$$\xi_\delta = \xi + \delta \quad (5)$$

and  $\xi \sim \mathcal{N}(\mu_\xi, \Sigma_\xi)$ , as well as  $\delta \sim \mathcal{N}(\mu_\delta, \Sigma_\delta)$ , the distribution of  $\xi_\delta$  is the convolution  $\Psi = \mathcal{N}(\mu_\xi + \mu_\delta, \Sigma_\xi + \Sigma_\delta)$ . For  $\mu_\delta = 0$ , this is just a wider Gaussian distribution.

## 2.5. Importance (re-)weighting

We now consider the dose deposited by histories  $h(\xi, \delta)$ , which are a function of the random input parameters  $\xi \sim \Phi_0(\xi)$  and random error vector  $\delta \sim p_\delta$ . In the following we focus on computing estimates for the dose expected value and standard deviation, the method can however be analogously applied to the computation of several worst case scenarios.

We propose a replacement of the dose calculations for different error scenarios by a more efficient weighting of particle histories  $h$ . For this, we adopt the concept of importance sampling (Kahn 1950, Hastings 1970). Instead of sampling primary particles from  $\Phi(\xi, \Delta_k)$  for different error scenarios, we sample from a different density function—e.g. the nominal parameter distribution  $\Phi_0(\xi)$ . Then, the dose for all scenarios can be estimated using histories from the nominal dose calculation:

$$\begin{aligned} d_i^{\Delta_k} &= \mathbb{E}_{\Phi(\xi, \Delta_k)}[h_i(\xi, \Delta_k)] \\ &\approx \frac{1}{H} \sum_{p=1}^H h_i(\Xi_p) \frac{\Phi(\Xi_p, \Delta_k)}{\Phi_0(\Xi_p)}, \quad \Xi_p \leftarrow \Phi_0(\xi). \end{aligned} \quad (6)$$

Thus, scenario computation reduces to a scoring problem. Equation (6) can either be applied directly during the simulation, making use of weighted dose scoring, or as a re-weighting step using stored history information. We collectively refer to both approaches as importance (re-)weighting. The dose expectation and variance can now be computed through the sample mean and variance over the respectively obtained scenarios

$$\text{Var}(d_i) \approx \frac{1}{K-1} \sum_{k=1}^K (d_i^{\Delta_k} - E[d_i])^2 \quad (7)$$

$$E[d_i] \approx \frac{1}{K} \sum_{k=1}^K d_i^{\Delta_k}. \quad (8)$$

Instead, when it is possible to derive the distribution  $\Psi$  of the uncertain parameters, the expected value can also be computed as discussed in section 2.4. In this case, the importance (re-)weighting estimate does not require a discretization of the error distribution or the computation of several scenarios, but rather determines the expected dose directly from  $\Psi$ :

$$E[d_i] \approx \frac{1}{H} \sum_{p=1}^H h_i(\Xi_p) \frac{\Psi(\Xi_p)}{\Phi_0(\Xi_p)}, \quad \Xi_p \leftarrow \Phi_0(\xi). \quad (9)$$

## 2.6. Modeling set-up uncertainties

Set-up uncertainties correspond to a shift of the patient position or equivalently the positions of primary particles relative to the patient. While errors occur in three dimensional space, shifts along the beam axis do not affect the dose distribution. In the Gaussian model, set-up errors can hence be assumed to follow a bivariate normal distribution for each pencil beam  $b = 1, \dots, B$ :

$$\delta_r^b = (\delta_{r_x}^b, \delta_{r_y}^b) \sim \mathcal{N}(\mu_{\delta_r}^b, \Sigma_{\delta_r}^b), \quad (10)$$

with  $\mu_{\delta_r}^b \in \mathbb{R}^2$  and  $\Sigma_{\delta_r}^b \in \mathbb{R}^{2 \times 2}$ .

Particles are initialized in a 2D plane, thus the primary particle positions follow a bivariate Gaussian mixture (2.2)

$$\Phi_{0;r}(\mathbf{r}) = \sum_{b=1}^B w_b \Phi_{0;r}^b(\mathbf{r}), \quad \Phi_{0;r}^b = \mathcal{N}(\mathbf{r}; \mu_r^b, \Sigma_r^b). \quad (11)$$

Here  $\mu_r^b$  is the mean lateral position of initial particles in pencil beam  $b$  in beam's eye view i.e. in the 2D plane perpendicular to the central beam axis. Then, according to 2.3 the initial position  $\mathbf{r}_\delta$  of a particle under uncertainty is determined by

$$\mathbf{r}_\delta = \mathbf{r} + \boldsymbol{\delta}_r \quad (12)$$

and  $\mathbf{r}_\delta$  is distributed with the convolution function

$$\Psi_r = \sum_{b=1}^B w_b \Psi_r^b, \quad \Psi_r^b = \mathcal{N}(\boldsymbol{\mu}_r^b + \boldsymbol{\mu}_{\delta_r}^b, \boldsymbol{\Sigma}_r^b + \boldsymbol{\Sigma}_{\delta_r}^b). \quad (13)$$

An individual error realization  $\boldsymbol{\Delta}_{r,k} \leftarrow p_{\delta_r}$  then formally just corresponds to a shift of the original primary particle positions, which now follow the distribution

$$\Phi_r(\mathbf{r}, \boldsymbol{\Delta}_k) = \sum_{b=1}^B w_b \cdot \mathcal{N}(\mathbf{r}; \boldsymbol{\mu}_r^b + \boldsymbol{\Delta}_{r,k}^b, \boldsymbol{\Sigma}_r^b), \quad (14)$$

corresponding to the nominal distribution shifted by  $\boldsymbol{\Delta}_{r,k}$ .

The above distributions can be directly used with (6), (7) and (9) to obtain the expected dose and variance for set-up uncertainties.

## 2.7. Modeling range uncertainties

The proposed approach could be analogously applied to any type of uncertainty directly affecting input parameters of the simulation, which have an a-priori probability distribution. Range uncertainties, however, modify the density values, which are deterministic and can thus not be directly modeled within the proposed framework.

To still approximate our quantities of interest, we exploit that the largest dose uncertainty is induced near the range of a beam (Bortfeld 1997), although the uncertain density variation affects the whole trajectory. Range can be expressed in terms of the initial energy of particles, using the Bragg–Kleemann rule

$$R = \alpha \cdot E_0^p, \quad (15)$$

where  $R$  is the range,  $E_0$  is the initial energy and  $\alpha$  and  $p$  are application-specific parameters. For the case of the slow-down of therapeutic protons in water, values of  $\alpha = 0.022 \text{ mm/MeV}^p$  and  $p = 1.77$  can be chosen (Ulmer and Matsinos 2011).

The initial energy spectrum of a scanned pencil beam at the exit of the nozzle can be approximately represented by a Gaussian (Bortfeld 1997, Soukup *et al* 2005, Tourovsky *et al* 2005, Kimstrand *et al* 2007). We can use this to model range uncertainties through random variations of the initial energy (compare treatment of range straggling in Pedroni *et al* (2005)).

Let us assume range uncertainties are normally distributed, i.e.  $\delta_p \sim \mathcal{N}(0, \sigma_R^2)$  (comp. Lomax 2008a, Yang *et al* 2012). With a Taylor approximation (order 1 for the mean and 2 for the variance) around  $X = E[X]$ , we can determine the parameters  $\mu_{E_0}$ ,  $\sigma_{E_0}^2$  of the energy distribution due to range uncertainties

$$E_0 = \left(\frac{1}{\alpha} R\right)^{\frac{1}{p}} =: g(E[R]) \quad (16)$$

$$\mu_{E_0} = E[g(E[R])] \approx g(E[R]) = \left(\frac{1}{\alpha} E[R]\right)^{\frac{1}{p}} \quad (17)$$

$$\sigma_{E_0}^2 = \text{Var}(g(E[R])) \approx g'(E[R])^2 \text{Var}(E[R]) \quad (18)$$

$$= \left(\frac{1}{p\alpha} \left(E[R] \frac{1}{\alpha}\right)^{\frac{1}{p}-1}\right)^2 \sigma_R^2.$$

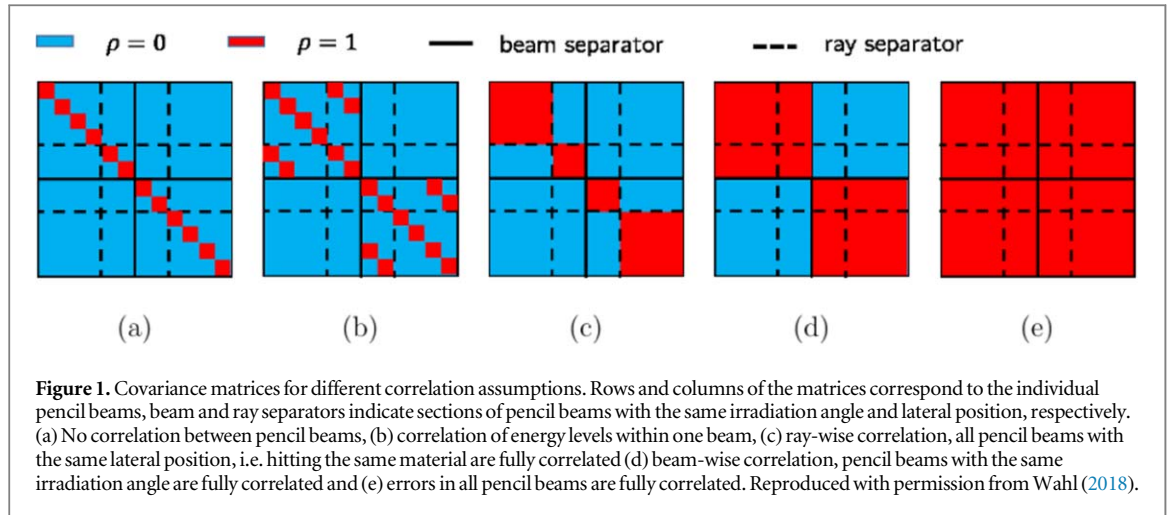
Thus the randomness in range is approximated through an energy distribution  $E_0 \sim \mathcal{N}(\mu_{E_0}, \sigma_{E_0}^2)$  and the expected dose and variance can be computed by (re-)weighting histories analogously to 2.6. This can again be extended to multiple pencil beams using Gaussian mixtures.

Note that, again, the expected value can be directly computed from simulations with an energy spectrum convolved with the Gaussian uncertainty kernel. Alternatively to the nominal energy distribution, this convolved distribution can also be used to obtain the required histories. In this case, the nominal distribution is replaced by the convolved distribution in (6)–(9).

## 2.8. Correlation models

In the previous sections, the distributions of different types of phase space parameters were considered independently. Note that the derived distributions are however all marginals of the joint multivariate Gaussian mixture spanning the complete phase space and all pencil beams (see also 2.2).

Similarly, the univariate normal distributions of errors of different types and in different pencil beams can be connected using a joint multivariate Gaussian distribution. This framework in principle allows for the definition of arbitrary correlation models for uncertainties between pencil beams. For the dose variance computation,



these correlations can be easily implemented using the covariance matrix of this joint distribution, since the samples for the weighted scenarios are directly drawn from the respective multivariate normal distribution.

Mathematically, the expected value is independent of correlation assumptions. While our results could still vary slightly due to the stochastic nature of the MC method, these random fluctuations are not related to uncertainties and thus no separate considerations of correlation assumptions are necessary for the expected dose.

Using the example of set-up uncertainties, if we define the errors in each beam by  $\delta_r = (\delta_r^1, \dots, \delta_r^B)^T$ , the multivariate Gaussian  $\mathcal{N}(\mu, C)$  would be parametrized with

$$\mu = \begin{pmatrix} \mu_{\delta_r}^1 \\ \mu_{\delta_r}^2 \\ \vdots \\ \mu_{\delta_r}^B \end{pmatrix},$$

$$C = \begin{pmatrix} \Sigma_r^1 & \rho_{xx}^{12} & \rho_{xy}^{12} & \dots & \rho_{xx}^{1B} & \rho_{xy}^{1B} \\ & \rho_{yx}^{12} & \rho_{yy}^{12} & & \rho_{yx}^{1B} & \rho_{yy}^{1B} \\ \rho_{xx}^{21} & \rho_{xy}^{21} & \Sigma_r^2 & & & \\ \rho_{yx}^{21} & \rho_{yy}^{21} & & & & \\ \vdots & & & \ddots & & \vdots \\ \rho_{xx}^{B1} & \rho_{xy}^{B1} & & & \Sigma_r^B & \\ \rho_{yx}^{B1} & \rho_{yy}^{B1} & \dots & & & \end{pmatrix}, \quad (19)$$

where  $\rho_{xy}^{ab}$  is the covariance between set-up errors in the  $x$ -direction in pencil beam  $a$  and errors in the  $y$ -direction in pencil beam  $b$ .

A few simple examples for correlation models are shown in figure 1, more can be found in literature (Pflugfelder *et al* 2008, Unkelbach *et al* 2009, Bangert *et al* 2013).

In case the correlation matrix is singular (perfect correlation between some pencil beams), the dimension of the uncertain vector can be reduced and one joint error can be sampled for the respective perfectly correlated pencil beams.

More complex correlation models are possible.

## 2.9. Implementation

For the proof-of-concept in this work, the weighting method was implemented as a post-processing routine in Matlab. Radiation plans were generated with matRad (Wieser *et al* 2017) and exported to the MC simulation engine TOPAS (Perl *et al* 2012) for dose calculations. The required particle histories  $h(\xi)$  are stored during the simulation using a custom extension for TOPAS. On average, this requires around 500 bytes of storage per history (when using double precision), divided into 34 bytes per primary particle and 12 bytes for each event in the particle's trajectory.

For both the references and the estimates using the proposed (re-)weighting method, the expected dose is determined directly according to the approach introduced in section 2.4. The variance on the other hand is approximated based on random error scenarios. Thus, for the expected value a direct simulation of  $\Psi$  in TOPAS is compared to a reconstruction using 9. In case of the variance, numerous scenarios are sampled from the error

**Table 1.** Overview of uncertainties investigated for each patient/test case. Beam irradiation angles are given as (couch angle, gantry angle) in degrees and error values refer to the standard deviations of the corresponding normal distributions.

		Patient Angles	Water phantom (0°, 0°)	Prostate (0°, 90°/270°)	Liver (0°, 315°)
Correlation	Type				
Full	Set-up		3 mm	3 mm	3 mm
	Range		3%	—	3%
	Both		3 mm/3%	—	3 mm/3%
Beam	Set-up		—	3 mm	—
Ray	Set-up		—	3 mm	—
Energy	Both		—	3 mm/3%	—
None	Set-up		—	3 mm	-

**Table 2.** Overview of simulated plans and error scenarios per patient.

Patient	Water phantom	Liver	Prostate	
Irradiation angles	(0°, 0°)	(0°, 315°)	(0°, 90°)	(0°, 270°)
Number of pencil beams	147	1 378	1 375	1 383
Number of histories	2 566 453	13 528 430	16 992 193	16 748 034
Number of error scenarios	100	100	100	

distribution and the dose for each error scenario is determined using TOPAS or 14, for the reference and (re-) weighting estimate respectively. The variance can then be estimated using the sample variance formula in both cases.

To reduce the number of required realizations, both for the reference computation and the (re-)weighting steps, a quasi-MC approach was used to sample the random parameters (see e.g. Caflisch 1998). Here, the random numbers in the classic MC method are replaced by low-discrepancy sequences to ensure a more uniform coverage of the random space. This results in a faster convergence rate and typically less required samples for the same accuracy, compared to the MC method. However, quasi-random samples are not independent, which can affect the mathematical properties of estimates.

## 2.10. Patient data

We evaluate the proposed method on three different patient cases of varying complexity: a water phantom, as well as a liver and prostate patient obtained from the open CORT (common optimization for radiation therapy) dataset (Craft *et al* 2014).

Table 1 shows the irradiation angles and table 2 the number of histories and pencil beams for each considered patient. Note, that the number of histories per pencil beam were determined based on weights from the optimized radiation plan (see 2.9), where around  $10^5$  histories are computed for the pencil beams with the highest weights.

## 2.11. Investigated uncertainties

In the following, we consider range and set-up uncertainties, as well as a combination of both. For the set-up uncertainties, we assume a symmetric, bivariate normal distribution with zero mean (no systematic errors) and a standard deviation of 3 mm (comp. Perkó *et al* 2016, Wahl 2018). For range uncertainties, in the reference computations we scale the density with a normally distributed factor, where the mean is equal to the nominal density and the standard deviation is 3% (as recommended in Lomax (2008b), Yang *et al* (2012)). The corresponding parameters of the energy distribution, used to approximate range errors in the importance (re-) weighting estimate, are determined based on this distribution as detailed in 2.7. Table 1 provides an overview of which uncertainty models were computed for which patient.

The number of error scenarios computed for the importance (re-)weighting estimates can be found in table 2.

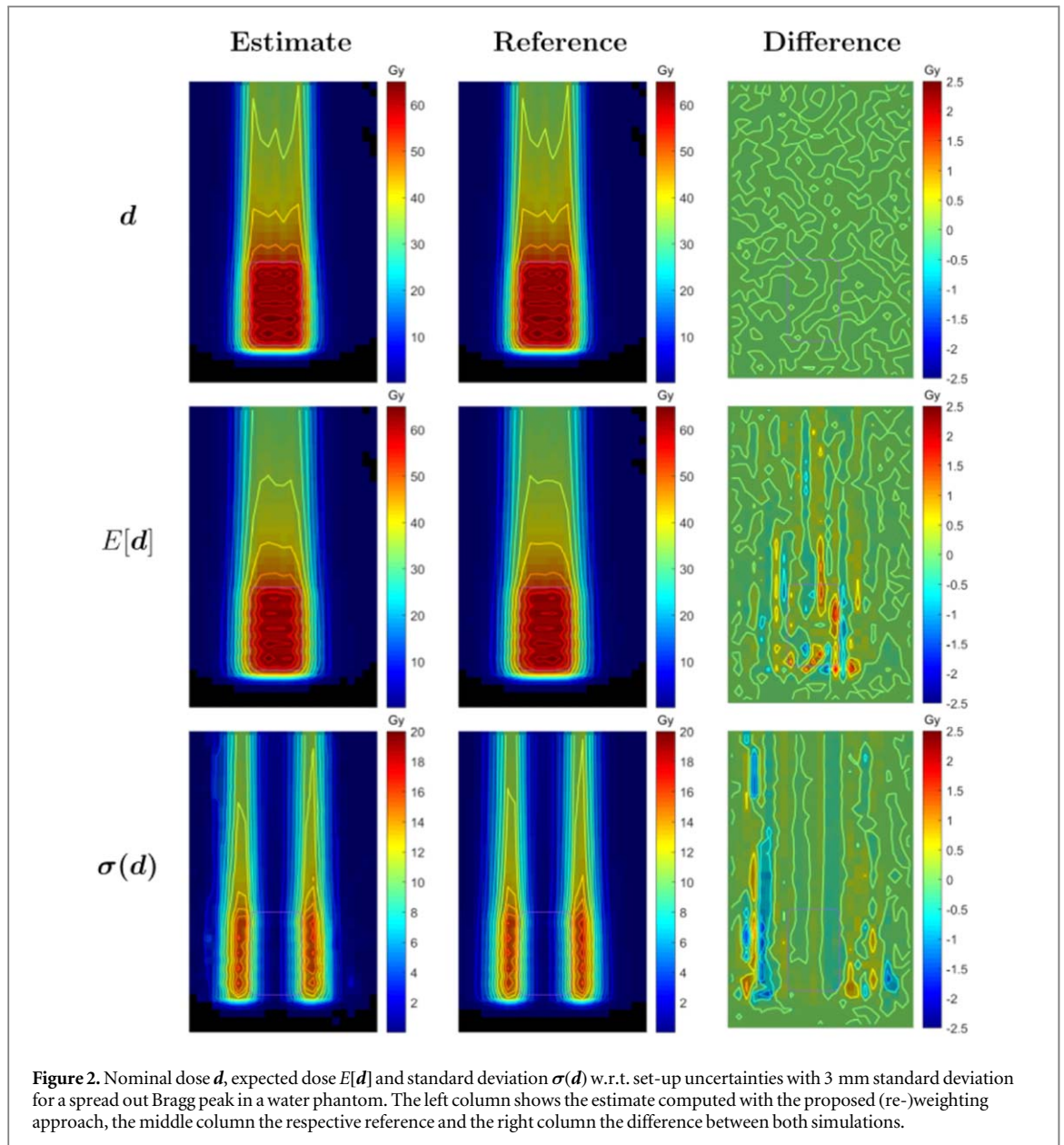
## 2.12. Evaluation criteria

To compare our results to the respective reference computations, we plot two-dimensional slices of the dose cubes as well as a difference map and employ a global three-dimensional  $\gamma$ -analysis. For the difference maps, we compute

$$\text{diff}_i(d_i^{\text{ref}}, d_i^{\text{est}}) = d_i^{\text{ref}} - d_i^{\text{est}}, \quad (20)$$

for each voxel  $i$  in the reference result  $d^{\text{ref}}$  and (re-)weighting estimate  $d^{\text{est}}$ . For the  $\gamma$ -analysis, we use the matRad implementation based on Low *et al* (1998), with a distance to agreement of 2 mm and a dose difference criterion of 2%.





### 3. Results

In the following we present results for the cases given in table 1. Unless specified otherwise, results were computed on the basis of histories from nominal dose calculations, i.e. with phase space parameters sampled from  $\Phi_0$  (see 2.2). The references computed for nominal and expected dose stem from MC dose calculations with the respective phase space distributions  $\Phi_0$  and  $\Psi$  (see 2.4), the reference standard deviation is derived using numerous such MC simulations for different error scenarios sampled from the joint error distribution. Therefore, the importance (re-)weighting estimate for the nominal dose only differs from the reference by round-off errors introduced in post-processing, as can be seen in figure 2. It is omitted in the following section, as is the reference.

#### 3.1. Set-up errors

Figure 2 displays the nominal dose, expected value and standard deviation estimates for a water phantom, computed using the (re-)weighting approach in comparison to the respective references. While we see some minor deviations in the difference maps for the expected dose and standard deviation, they do not appear systematic.

The distance-to-agreement analysis using the  $\gamma$ -criterion supports this quantitatively (table 3(a)), with a  $\gamma_{2\%}^{2\text{mm}}$ -pass rate of 99.95% for the expected dose and 98.04% for the standard deviation. Figure 3 demonstrates that this is transferable to the more complex patient cases. With overall  $\gamma_{2\%}^{2\text{mm}}$ -pass rates of 99.82% (prostate

**Table 3.**  $\gamma_{2\%}^{2\text{mm}}$ -pass rates in volumes of interest (VOI) of the water phantom, liver and prostate patient for (a) set-up errors, (b) range errors and (c) set-up and range errors. All estimates were computed from the nominal distribution  $\Phi_0$  and in (b) and (c) also compared against such from the expected distribution  $\Psi$ .

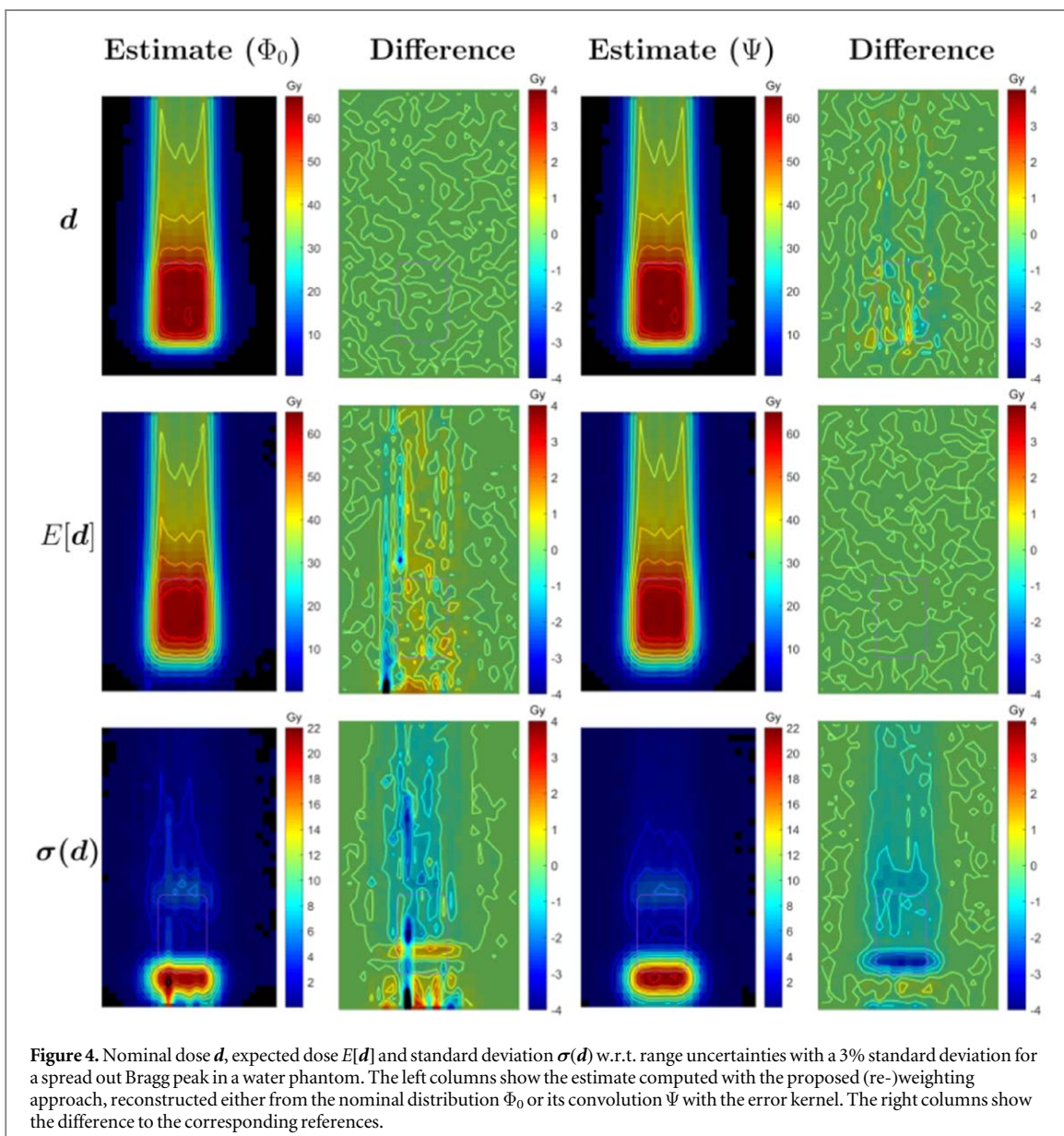
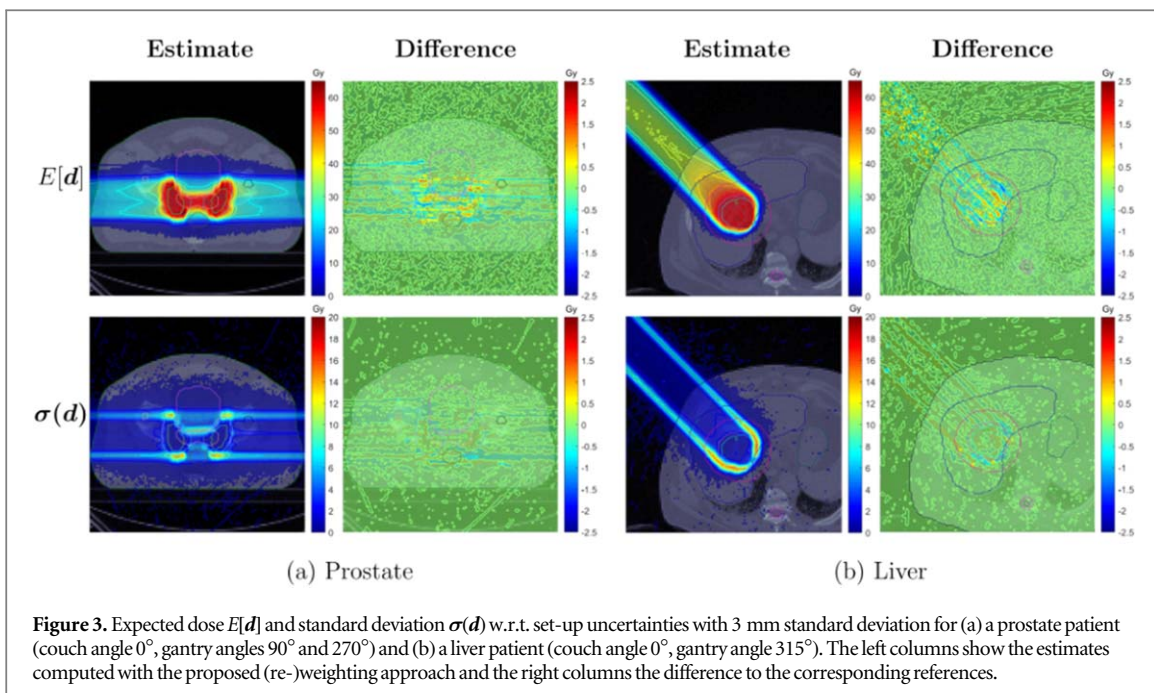
(a) Set-up errors											
Patient	VOI										
Waterbox	Total	Body	Target								
	$E[\mathbf{d}]$	99.95	99.95	99.81							
	$\sigma(\mathbf{d})$	98.04	98.04	98.44							
Liver	Total	GTV	Liver	Heart	CTV	Contour	PTV				
	$E[\mathbf{d}]$	99.01	98.07	97.71	99.83	94.99	97.83	96.59			
	$\sigma(\mathbf{d})$	99.81	100	99.79	99.67	100	99.72	99.49			
Prostrate	Total	Rectum	Penile bulb	Lymph nodes	Femoral heads	Prostate bed	PTV 68	PTV 56	Bladder	Body	
	$E[\mathbf{d}]$	99.99	100	100	99.75	100	99.71	99.85	99.88	99.96	
	$\sigma(\mathbf{d})$	99.82	99.85	95.92	100	100	100	100	99.36	99.94	
(b) Range errors											
Distribution	Patient	VOI									
		Waterbox	Total	Body	Target						
$\Phi_0$	$E[\mathbf{d}]$		99.58	99.58	99.44						
	$\sigma(\mathbf{d})$		91.63	91.63	92.58						
$\Psi$	$\mathbf{d}$		99.98	99.98	100						
	$\sigma(\mathbf{d})$		93.55	93.55	87.88						
		Liver	Total	GTV	Liver	Heart	CTV	Contour	PTV		
$\Phi_0$	$E[\mathbf{d}]$		99.91	99.71	99.91	100	99.74	99.93	99.81		
	$\sigma(\mathbf{d})$		73.32	79.47	70.68	83.38	66.49	72.16	57.38		
$\Psi$	$\mathbf{d}$		99.93	99.90	99.96	100	99.85	99.85	99.91		
	$\sigma(\mathbf{d})$		93.70	99.05	92.17	95.51	90.06	93.80	86.22		
(c) Set-up and range errors											
Distribution	Patient	VOI									
		Waterbox	Total	Body	Target						
$\Phi_0$	$E[\mathbf{d}]$		99.39	99.39	99.81						
	$\sigma(\mathbf{d})$		95.10	95.10	82.76						
$\Psi$	$\mathbf{d}$		99.93	99.93	99.81						
	$\sigma(\mathbf{d})$		99.50	99.50	97.04						
		Liver	Total	GTV	Liver	Heart	CTV	Contour	PTV		
$\Phi_0$	$E[\mathbf{d}]$		99.86	99.71	99.86	99.93	99.61	99.91	99.67		
	$\sigma(\mathbf{d})$		95.56	96.76	89.87	87.77	95.85	91.74	91.89		
$\Psi$	$\mathbf{d}$		99.87	100	99.94	100	99.77	99.97	99.86		
	$\sigma(\mathbf{d})$		96.64	100	93.39	84.96	98.67	92.50	92.58		

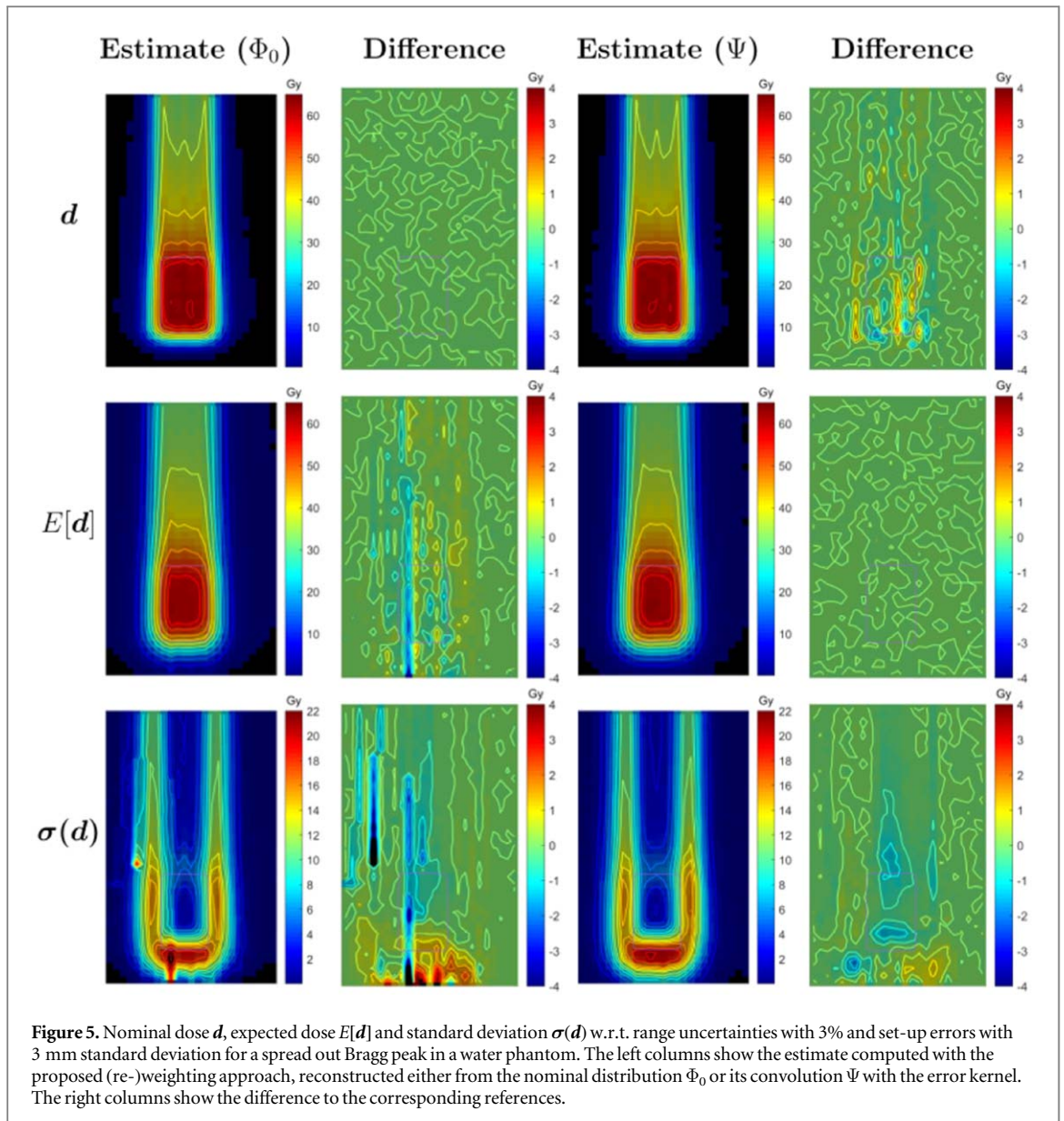
patient) and 99.81% (liver patient), the standard deviation agrees as well with the reference computations as the expected value, with 99.99% and 99.01%, respectively (table 3(a)).

### 3.2. Range errors

In contrast to the set-up errors, for which dose estimates can also be shown to be mathematically accurate, range errors can only be modeled through an approximation introduced in 2.7. Figures 4 and 5 display results for range errors as well as a combination of range and set-up errors in the water phantom, respectively.

The difference maps for both expected value and standard deviation show that the deviations when including range errors are expectedly higher. We observe a systematic bias primarily at the distal edge, where our



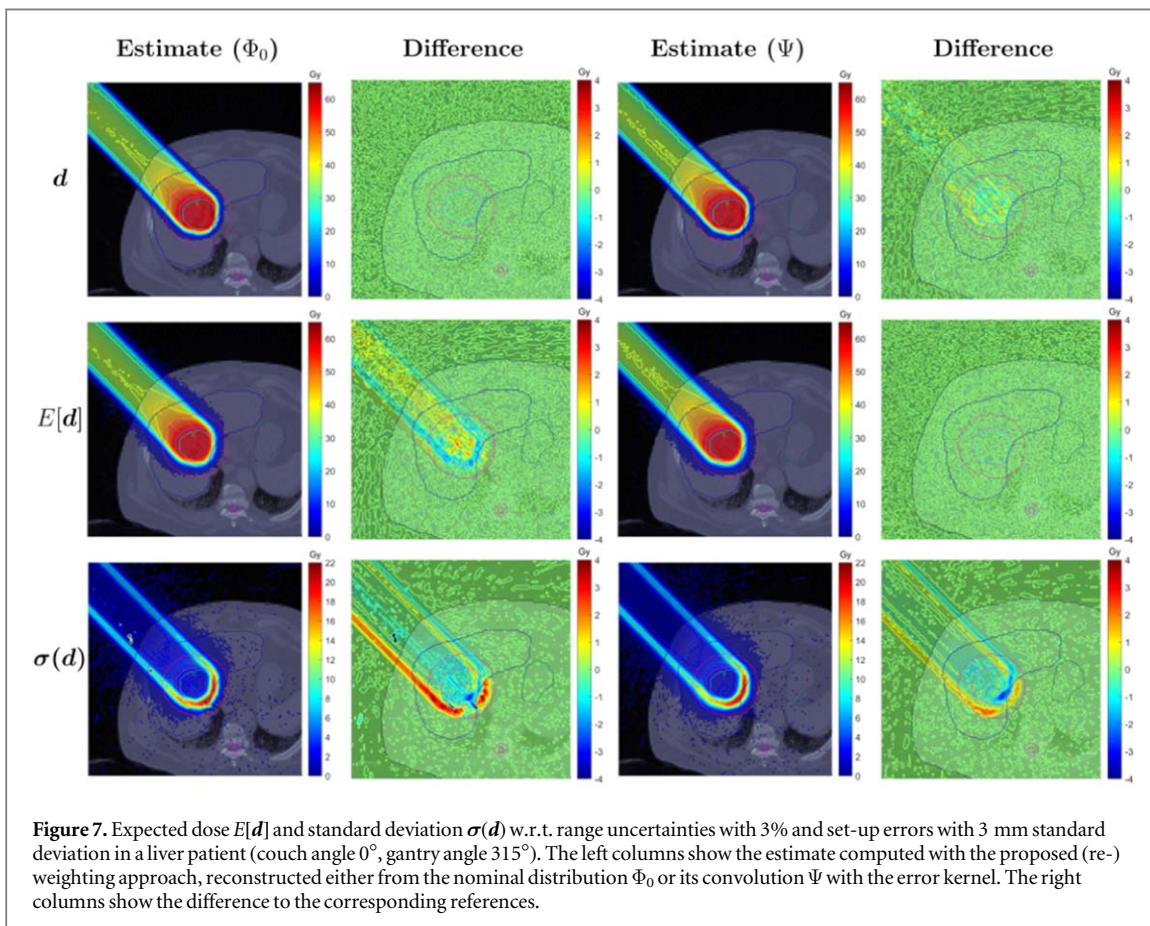
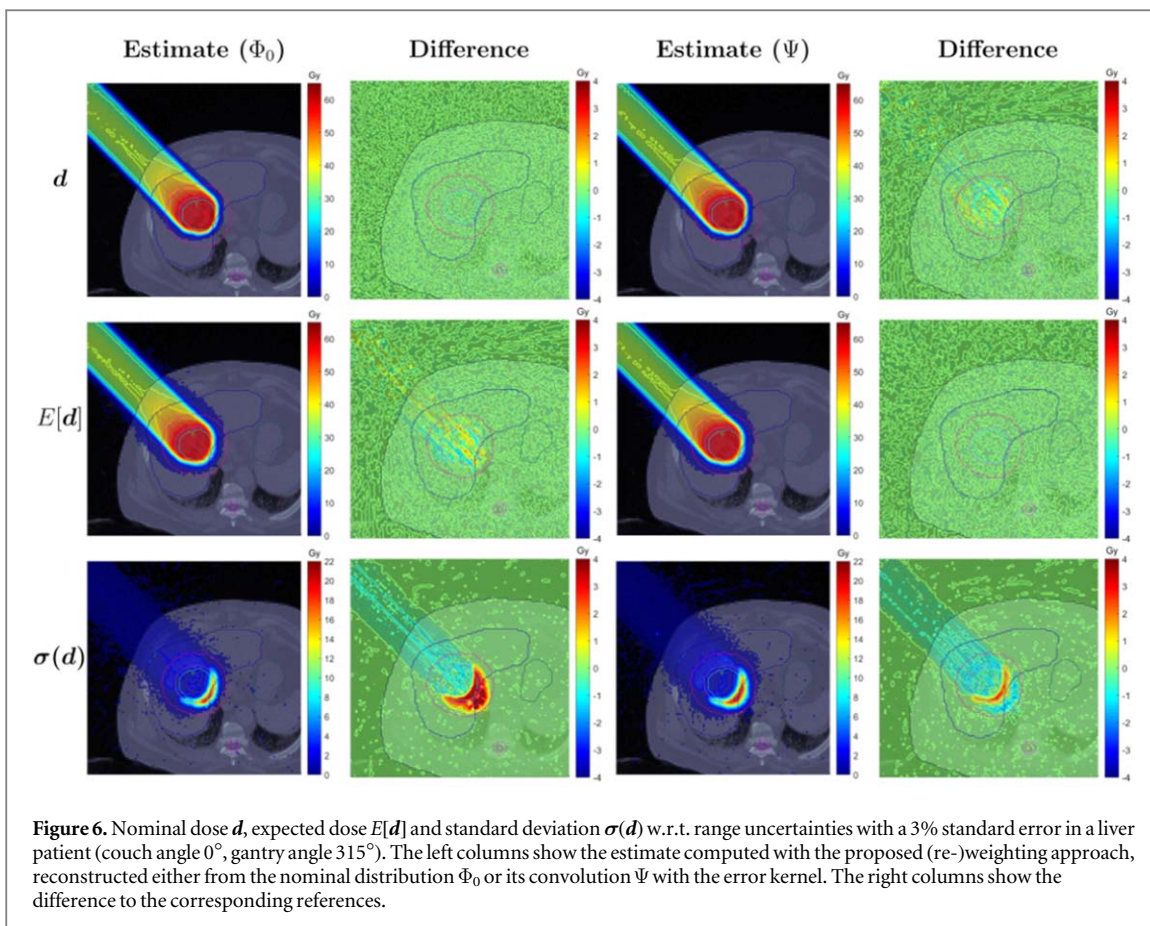


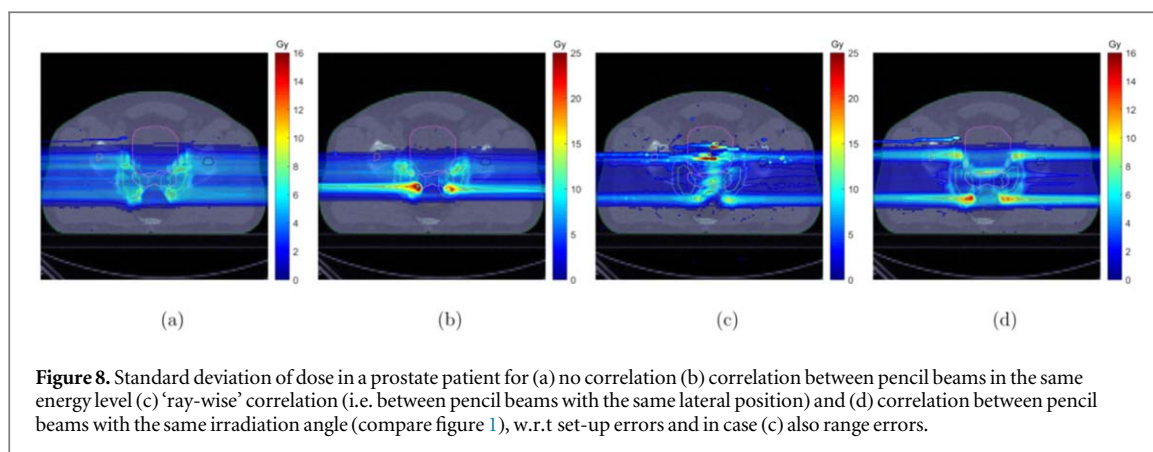
method seems to consistently underestimate the variance. The standard deviation estimate using our importance weighting method also expresses strong local artifacts, as evident in the difference maps (compare figures 4 and 5). This is an indicator of too little statistical mass, i.e. computed particle trajectories, in the original simulation. For more extreme error realizations, relatively high weights are assigned to a small number of particles, thereby amplifying single realizations or errors. Especially in case of a relatively small beam energy spread in the original simulation (here 1%), compared to the range error of 3%, such artifacts are likely to appear. In order to prevent this, one could either compute a larger number of particle histories in the simulation or sample the particles from a different distribution which has more density mass in its outer regions or tails.

To underline the explanation for the appearance of the artifacts above, we recomputed the estimates using the (re-)weighting method based on a direct computation of the expected value, which can be calculated using the convolution  $\Psi$  of the Gaussian error kernel with the nominal phase space parameter distribution (compare 2.4). Figures 4 and 5 show that this alleviates the discrepancy from the references, causing artifacts to disappear and also reducing the overall amount of deviation displayed in the difference maps.

Thereby we can conclude that the irregularities in the solution can be attributed to the lack of statistical support in certain areas. Contrary to this, parts of the systematic differences remain and are thus most likely a result of the model approximations.

Figures 6 and 7 validate these observations for a liver patient. The difference maps for estimates computed based on the expected distribution  $\Psi$ , have less severe artifacts and systematic deviations. The  $\gamma_{2\%}^{2\text{ mm}}$ -pass rate also consistently increases for both the liver patient and water phantom (see table 3(b), (c)).





Also, it has to be noted, that using  $\Psi$  to sample the initial particles leads to an expected dose estimate which is equivalent to the reference computations (compare 9), but a nominal dose estimate which now shows deviations from a nominal standard MC reference computation in the order of magnitude that we could previously observe for the expected dose (see table 3(b), (c)). This is due to the fact that the importance sampling error depends on the similarity of the sampling and target distribution. One can however expect a slightly lower error when constructing the nominal dose from a simulation of  $\Psi$ , since here the deviations of the target distribution occur in a region with a higher probability mass.

### 3.3. Correlation models

So far, we have only shown results for the case of fully correlated pencil beams, meaning one global shift of the patient position or scaling factor for the beam range. One of the advantages of the proposed method is, however, the high flexibility in changing the uncertainty model. In figure 8 we therefore present the standard deviation estimate for four examples of different error correlation models discussed in section 2.8.

The results indicate, that different correlation assumptions have a crucial impact on the standard deviation of dose distributions. While it is in principle possible to define arbitrary correlations within the proposed framework, estimates can be prone to artifacts due to a lack of statistical information, especially for the ray-wise correlation model. When sampling error realizations independently for smaller beam components, the reconstruction depends solely on the particle histories associated with these components. For rays with small weights, only very few histories are computed, therefore we observe similar artifacts as encountered in the above range uncertainty computations (3.2).

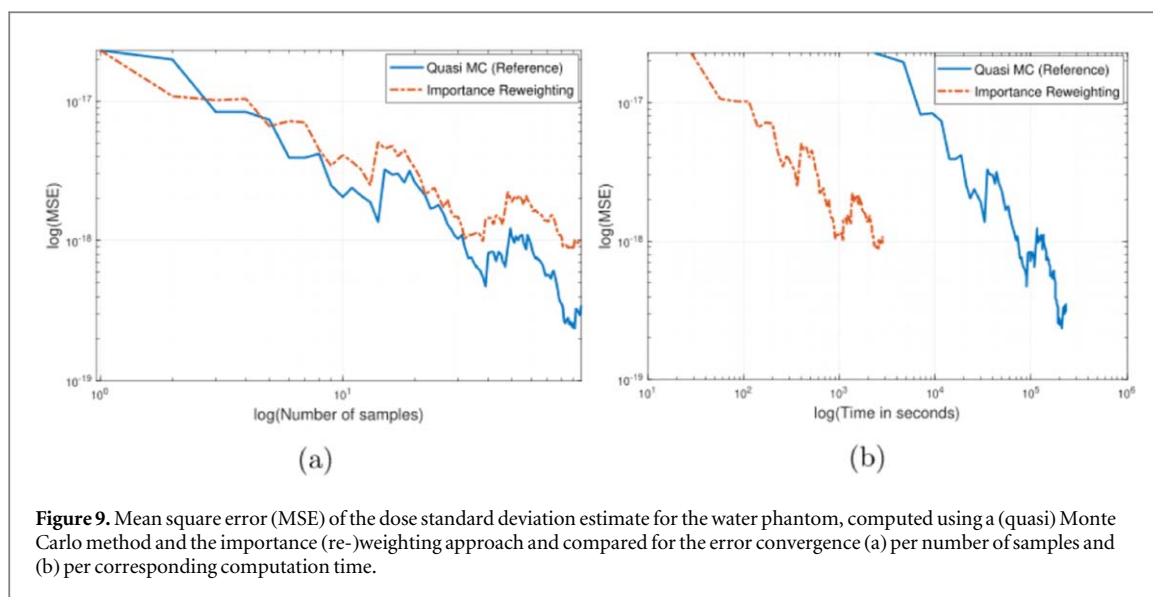
## 4. Accuracy and convergence

Mathematically, it can be shown, that the expected and nominal dose estimates are unbiased in the case of set-up errors. This also holds for the doses corresponding to each individual error realization. While it does not generally apply for the variance, the  $\gamma_{2\%}^{2\text{mm}}$ -pass rates of the variance estimates are not significantly lower than those of the expected dose (compare table 3). Considering the additional effect of the use of discrete error scenarios, this indicates that the bias does not have a significant impact on the quality of variance estimates.

For quicker convergence we used quasi-random numbers throughout the whole comparisons, both for the reference computation and the importance (re-)weighting approach. Note, that the combination of importance sampling with quasi-MC methods has been shown to be not only possible, but advantageous and preserves the convergence properties of quasi-MC (Caflisch 1998, Ökten 1999, Schürer 2004, Hörmann and Leydold 2005). Since the procedure mimics a (quasi-) MC method for uncertainty quantification, where the repeated simulation runs are replaced by (re-)weighting steps, the convergence of the variance per computed error realization is identical. However, due to the lower cost of the (re-)weighting steps, the convergence per time is much faster (see figure 9).

For run-time comparisons, the reference computations using TOPAS and the (re-)weighting approach, implemented as post-processing in Matlab, were run on the same virtual machine<sup>6</sup>. We observe reduced CPU times by a factor of 80, 32 and 23 for the water phantom, liver and prostate patient, respectively (see table 4).

<sup>6</sup>Virtual machine including 64 CPUs with 1.995 GHz and 200 GB RAM.



**Figure 9.** Mean square error (MSE) of the dose standard deviation estimate for the water phantom, computed using a (quasi) Monte Carlo method and the importance (re-)weighting approach and compared for the error convergence (a) per number of samples and (b) per corresponding computation time.

**Table 4.** CPU time comparison for the reference versus (re-)weighting approach applied to different patients and computed on the same machine. All values are given in seconds. Note that the times for 100 realizations include the initialization times, while the time for a single realization only refers to the dose computation time.

		Reference	(re-)weighting
Water phantom	Initialization	2.35	61.53
	One realization	2331.30	28.51
	100 realizations	233 126.93	2912.53
Liver	Initialization	2.44	2038.75
	One realization	39 066.44	1198.74
	100 realizations	3906 650.90	121 912.75
Prostate	Initialization	4.26	4867.75
	One realization	58 762.40	2479.07
	100 realizations	5876 253.86	252 774.75

## 5. Discussion

In this paper, we introduce an efficient approach for uncertainty quantification in MC dose calculations using history (re-)weighting. We demonstrate how particle histories from one simulation can be scored to construct estimates for error scenarios, the expected dose and standard deviation, for set-up and range errors in intensity modulated proton therapy. As demonstration example, Gaussian range and set-up uncertainties, with 3% and 3 mm standard deviation respectively, were considered for a water phantom, a liver patient and a prostate patient.

For set-up uncertainties, we observed good agreement of at least 99.01% for the expected dose and 98.04% for the dose standard deviation in the  $\gamma_{2\text{ mm}/2\%}$ -criterion for all quantities of interest. Range error propagation could be approximated by transforming the assumed range uncertainty into energy uncertainty via the range-energy relationship. High pass rates of at least 99.39% in the  $\gamma_{2\text{ mm}/2\%}$ -criterion indicate that the error caused by this model approximation does not noticeably affect the accuracy of the expected dose estimate. Lower pass rates are to be expected for the standard deviation, since it exhibits stronger inhomogeneity compared to the expected or nominal dose. The standard deviation estimates are, however, also sensitive to the number of histories and usage of the nominal Gaussian pencil beam width or the convolved distribution. Agreement within the  $\gamma_{2\text{ mm}/2\%}$ -criterion significantly increases and visible artifacts in the standard deviation estimate can be partly eliminated, by simulating the initial phase-space parameters using the convolved beam parameterization. Areas of systematic deviations remain, as evidenced by the low pass rates in particular volumes of interest, especially in the heterogeneous patient cases. Therefore, when computing the dose standard deviation in the presence of range uncertainties, the proposed method can currently only be recommended if no exact estimate is required. However, these local pass rates have to be interpreted with caution when considering organs which are far

removed from the beam center, such as the heart in the liver patient case. Here only a low number of particles passing through a small percentage of the voxels contribute to the result and therefore random fluctuations are probable, especially considering the error inherent in the reference values based on 100 samples. Since areas of high uncertainty are still correctly identified, a use for robust optimization is, for example, plausible. Also, parts of the inaccuracies can be attributed to the comparably large magnitude of the assumed range errors in relation to the initial energy spread. Improvements in the actual range accuracy, for example through dual energy CTs, might soon allow for the realistic assumption of relative range errors significantly lower than 3%.

As the case of range uncertainties demonstrates, the use of different distribution functions for the initial phase space parametrization of the MC simulation can substantially impact the results of the reconstructions. We initially presented the approach and results based on a nominal dose calculation, which is especially of interest for use cases such as dose verification. Here, the accuracy of the nominal dose is not compromised and one can gain some additional insight into uncertainties. However, the lack of statistical mass in parts of the domain impedes the dose computation for more extreme error realizations. Therefore, the use of a wider distribution, such as the convolution of the nominal parameter density with the error kernel, can improve the expected value and standard deviation estimate and be beneficial for applications in planning or robust optimization. Other configurations, such as a mixture of several Gaussians, are also thinkable and could be specifically tuned with regard to uncertainty quantification. Thus, the choice of parameter distributions gives the user a degree of flexibility in putting the focus on either retaining accuracy in the nominal dose computation or trading it against better accuracy of the uncertainty estimate according to the specific use case.

We also demonstrate the use of different pencil beam correlation models within the framework. It is clear that the choice of correlation model has a significant impact on the standard deviation estimate. Therefore, it is particularly convenient that the (re-)weighting method allows for the definition of principally arbitrary correlation matrices to put into the underlying multivariate Gaussian error model. These could possibly be extended to simulate interplay effects or other dynamic influences in the context of 4D treatment planning. Since the applied correlation models are not only experimental but also difficult and time-consuming to evaluate in scenario sampling, we did not quantitatively compare them to reference computations. Further studies could explore whether they agree with other methods computing such correlations based on an analytical probabilistic dose engine (Bangert *et al* 2013, Wahl *et al* 2017, Wieser *et al* 2020).

Compared to the reference scenario estimates, which rely on performing full MC dose calculations repeatedly, the CPU-time for standard deviation estimates could be reduced by more than an order of magnitude using our method in combination with a quasi-MC approach. This is achieved by reducing the costs of repeated expensive simulations to those of scoring based on matrix-vector multiplications. Consequently, it has to be noted, that the time reduction depends largely on the proportion of computational overhead of the initialization and simulation steps in the MC engine. Therefore the factor of speed increase varies strongly between different test-cases and, most likely, implementations. But even then, our method holds two performance advantages: first, it can directly compute the expected dose by using the convolved phase space parameterization (2.4) in *one* standard simulation. Second, multiple uncertainty models with different correlation patterns and magnitude can be reconstructed from the same set of histories. This could for example be used to investigate the impact of fractionation effects, using the framework proposed by Wahl *et al* (2018) or to consider a number of (worst case) scenarios besides the expected dose and standard deviation.

A combination of our approach with other efficient uncertainty quantification approaches, which rely on scenario computations could lead to run-time improvements. For instance a polynomial chaos expansion as introduced in Perkó *et al* (2016) could be adjusted such that the evaluations are computed by (re-)weighting histories instead of the usual dose calculations.

Further, we argue that computational performance can be improved through a more efficient, performance-tuned implementation. The computational complexity is inherently reduced compared to scenario computations, since no physics simulations are required in addition to scoring. We therefore expect performance benefits also with faster MC dose calculation algorithms, given an equivalently tuned implementation. In particular, when one chooses to apply importance weighting on-the-fly using multiple sub-scoring routines, performance gains should be achievable since the existing efficient structures of the MC code could be directly used. However, in this case the number of error scenarios has to be predetermined, causing a loss in flexibility, and is also limited with regard to memory constraints.

Last but not least, the method is not inherently limited to the discussed application in proton therapy; a calculation of uncertainty estimates using the (re-)weighting approach would also be feasible for other intensity-modulated particle therapy modalities like carbon ions but also photons. In its current description, it is however limited to uncertainties which can be modeled in terms of variations of phase space parameters with a prior probability distribution (this excludes biological parameters, like RBE). Application to, for example, pre-simulated phase spaces might also be feasible using numerical convolution techniques. Also, a disproportionately high magnitude of uncertainties in relation to this probability distribution can compromise



the accuracy of results. Furthermore, it needs to be mentioned, that the current computational speed, especially for the standard deviation, might still not be sufficient for optimization purposes, where a full dose influence matrix needs to be computed. Due to the simplicity of the process and the high flexibility in post-processing at virtually no cost for the original simulation, we are confident that the approach has the potential for further development and use.

## 6. Conclusion

Dose distributions in intensity modulated proton therapy are known to be sensitive to uncertainties. The computational efforts in estimating such uncertainties become particularly evident when MC dose calculation is used. We showed how the concept of importance sampling can be adapted to estimate the expected dose and its variance using histories from only a single MC simulation. Set-up uncertainties can be efficiently modeled and exhibit almost exact agreement with reference computations. The inclusion of range uncertainties, by modeling them as energy uncertainty via the range-energy relationship, yields comparably high agreements with respect to the expected dose. While the standard deviation estimate is substantially less accurate, it can give a rough insight into the dose variations, which can be sufficient for some application purposes. Further, the physical simulation of particles is completely decoupled from uncertainty quantification, thereby allowing for the incorporation of arbitrary correlation assumptions and the comparison of different scenarios, at no additional cost to the nominal dose calculation. Therefore, the presented approach has several benefits over classic non-intrusive methods and is a step towards reconciling efficient uncertainty quantification and, in the future, robust optimization based on MC dose calculations.

## Acknowledgments

The present contribution is supported by the Helmholtz Association under the joint research school HIDSS4Health—Helmholtz Information and Data Science School for Health.

## Appendix A. Importance sampling

Importance sampling is a method most frequently used for variance reduction (Kahn 1950, Hastings 1970). Assume the integral  $I(g) = \int g(x)p(x)dx$  is to be computed for  $x \sim p(x)$  without directly sampling from  $p(x)$ . The importance sampling estimate is defined as

$$I(g) = \int g(x)p(x)dx = \int g(x)\frac{p(x)}{q(x)}q(x)dx \\ \approx \sum_{i=1}^N g(X_i)\frac{p(X_i)}{q(X_i)}, \quad X_i \leftarrow q(x). \quad (\text{A.1})$$

Thus, the integral of  $g(x)$  with the probability distribution  $x \sim p(x)$  can be reconstructed from samples of a suitable density function  $q(x)$ .

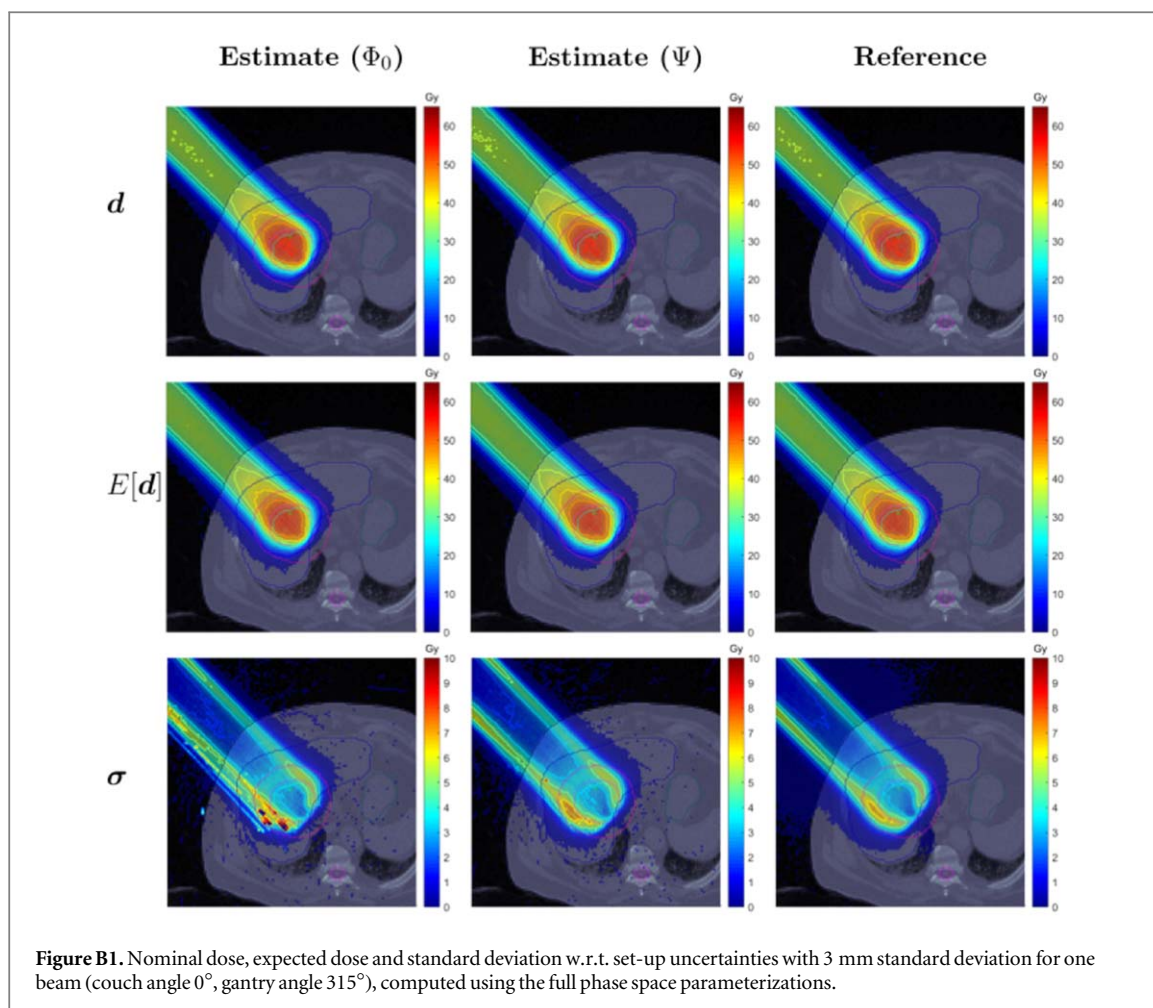
## Appendix B. Full phase space model

The simplified model used in the main paper can be extended to include distributions in the momentum direction  $\varphi = (\varphi_x, \varphi_y, \varphi_z) \neq 0$ . The additional variables are included in the Gaussian mixture model:

$$\Phi_0(\xi) = \sum_{b=1}^B w_b \Phi_0^b(\xi), \quad \Phi_0^b(\xi) = \Phi_0^b(\mathbf{r}, \varphi, E) = \mathcal{N}(\boldsymbol{\mu}_\xi^b, \boldsymbol{\Sigma}_\xi^b), \quad (\text{B.1})$$

where the respective entries in the covariance matrix  $\boldsymbol{\Sigma}_\xi^b$  can be chosen  $\neq 0$  to model randomness in the momentum direction as well as correlations of the momentum directions with primary particle positions.

Figure B1 presents results for the nominal dose, expected dose and standard deviation in a liver patient, for set-up uncertainties with 3 mm standard deviation, 0.2 standard deviation in the momentum direction and 0.3 correlation between  $\varphi_v$  and  $r_v$ ,  $v \in \{x, y\}$ . Estimates were computed based on the convolution function  $\Psi$  of the error and beam parameter densities, as well as the nominal parameter density  $\Phi_0$ . The corresponding global  $\gamma$ -analysis pass rates can be found in table B1.



**Table B1.**  $\gamma_{2\%}^{2\text{mm}}$ -pass rates in volumes of interest (VOI) of the liver patient computed using the full phase space parameterizations.

Error type	From $\Phi_0$			From $\Psi$		
	Set-up					
	$d$	$E[d]$	$\sigma(d)$	$d$	$E[d]$	$\sigma(d)$
Liver						
<b>Overall</b>	100	99.86	95.88	99.85	100	96.32
GTV	100	99.52	95.09	99.23	100	97.21
Liver	100	99.96	94.63	99.93	100	95.32
Heart	100	100	95.10	100	100	97.04
CTV	100	99.79	92.78	99.67	100	94.22
Contour	100	99.99	95.89	99.97	100	96.30
PTV	100	99.89	90.92	99.78	100	89.93

## ORCID iDs

P Stammer  <https://orcid.org/0000-0002-5595-0439>

O Jäkel  <https://orcid.org/0000-0002-6056-9747>

N Wahl  <https://orcid.org/0000-0002-1451-223X>

## References

- Bangert M, Hennig P and Oelfke U 2013 Analytical probabilistic modeling for radiation therapy treatment planning *Phys. Med. Biol.* **58** 5401–19
- Bielajew A F 1994 Monte Carlo modeling in external electron-beam radiotherapy—why leave it to chance *Proc. 11th ICCR (Manchester, UK)* pp 2–5 ([https://www.researchgate.net/publication/44064216\\_Monte\\_Carlo\\_modeling\\_in\\_external\\_electron-beam\\_radiotherapy\\_-\\_why\\_leave\\_it\\_to\\_chance](https://www.researchgate.net/publication/44064216_Monte_Carlo_modeling_in_external_electron-beam_radiotherapy_-_why_leave_it_to_chance))

- Bortfeld T 1997 An analytical approximation of the Bragg curve for therapeutic proton beams *Med. Phys.* **24** 2024–33
- Caflich R E 1998 Monte Carlo and quasi-Monte Carlo methods *Acta Numerica* **7** 1–49
- Casiraghi M, Albertini F and Lomax A 2013 Advantages and limitations of the ‘worst case scenario’ approach in IMPT treatment planning *Phys. Med. Biol.* **58** 1323–39
- Craft D, Bangert M, Long T, Papp D and Unkelbach J 2014 Shared data for intensity modulated radiation therapy (IMRT) optimization research: The CORT dataset *GigaScience* **3** 2047–217X-3-37
- Fippel M and Soukup M 2004 A Monte Carlo dose calculation algorithm for proton therapy *Med. Phys.* **31** 2263–73
- Fredriksson A 2012 A characterization of robust radiation therapy treatment planning methods-from expected value to worst case optimization *Med. Phys.* **39** 5169–81
- Fredriksson A, Forsgren A and Hårdemark B 2011 Minimax optimization for handling range and setup uncertainties in proton therapy *Med. Phys.* **38** 1672–84
- Hastings W K 1970 Monte Carlo sampling methods using markov chains and their applications *Biometrika* **57** 97–109
- Hörmann W and Leydold J 2005 Quasi importance sampling *Preprint Series 57, Department of Applied Statistics and Data Processing, Vienna University of Economics and Business, Vienna* (<http://epub.wu.ac.at/1394/>)
- Hu J and Jin S 2016 A stochastic Galerkin method for the Boltzmann equation with uncertainty *J. Comput. Phys.* **315** 150–68
- Kahn H 1950 Random sampling (Monte Carlo) techniques in neutron attenuation problems-I *Nucleonics (US) Ceased Publication* **6** 27 (<https://www.osti.gov/biblio/4399718>)
- Kimstrand P, Traneus E, Ahnesjö A, Grusell E, Glimelius B and Tilly N 2007 A beam source model for scanned proton beams *Phys. Med. Biol.* **52** 3151–68
- Kraan A C, van de Water S, Teguh D N, Al-Mamgani A, Madden T, Kooy H M, Heijmen B J M and Hoogeman M S 2013 Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors *Int. J. Radiat. Oncol., Biol., Phys.* **87** 888–96
- Liu W, Zhang X, Li Y and Mohan R 2012 Robust optimization of intensity modulated proton therapy *Med. Phys.* **39** 1079–91
- Lomax A J 2008a Intensity modulated proton therapy and its sensitivity to treatment uncertainties: I. The potential effects of calculational uncertainties *Phys. Med. Biol.* **53** 1027–42
- Lomax A J 2008b Intensity modulated proton therapy and its sensitivity to treatment uncertainties: II. The potential effects of inter-fraction and inter-field motions *Phys. Med. Biol.* **53** 1043–56
- Low D A, Harms W B, Mutic S and Purdy J A 1998 A technique for the quantitative evaluation of dose distributions *Med. Phys.* **25** 656–61
- Lowe M, Albertini F, Aitkenhead A, Lomax A J and MacKay R I 2016 Incorporating the effect of fractionation in the evaluation of proton plan robustness to setup errors *Phys. Med. Biol.* **61** 413–29
- Ma C-M, Li J S, Pawlicki T, Jiang S B, Deng J, Lee M C, Koumrian T, Luxton M and Brain S 2002 A Monte Carlo dose calculation tool for radiotherapy treatment planning *Phys. Med. Biol.* **47** 1671–89
- Mackie T 1990 Applications of the Monte Carlo method in radiotherapy *Dosimetry of Ionizing Radiation* ed K R Kase, B E Bjärngard and F H Attix III (San Diego: Academic Press) pp 541–620
- Malyapa R, Lowe M, Bolsi A, Lomax A J, Weber D C and Albertini F 2016 Evaluation of robustness to setup and range uncertainties for head and neck patients treated with pencil beam scanning proton therapy *Int. J. Radiat. Oncol. \*Biol. \*Phys.* **95** 154–62
- McGowan S E, Albertini F, Thomas S J and Lomax A J 2015 Defining robustness protocols: A method to include and evaluate robustness in clinical plans *Phys. Med. Biol.* **60** 2671–84
- McGowan S E, Burnet N G and Lomax A J 2013 Treatment planning optimisation in proton therapy *Br. J. Radiol.* **86** 20120288–20120288
- Ökten G 1999 Error reduction techniques in quasi-monte carlo integration *Math. Comput. Modell.* **30** 61–9
- Paganetti H 2012 Range uncertainties in proton therapy and the role of Monte Carlo simulations *Phys. Med. Biol.* **57** R99–117
- Park P C et al 2013 Statistical assessment of proton treatment plans under setup and range uncertainties *Int. J. Radiat. Oncol., Biol., Phys.* **86** 1007–13
- Pedroni E, Scheib S, Böhringer T, Coray A, Grossmann M, Lin S and Lomax A 2005 Experimental characterization and physical modelling of the dose distribution of scanned proton pencil beams *Phys. Med. Biol.* **50** 541–61
- Perkó Z, van der Voort S R, van de Water S, Hartman C M H, Hoogeman M and Lathouwers D 2016 Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion *Phys. Med. Biol.* **61** 4646–64
- Perl J, Shin J, Schumann J, Faddegon B and Paganetti H 2012 TOPAS: an innovative proton Monte Carlo platform for research and clinical applications *Med. Phys.* **39** 6818–37
- Pflugfelder D, Wilkens J J and Oelfke U 2008 Worst case optimization: a method to account for uncertainties in the optimization of intensity modulated proton therapy *Phys. Med. Biol.* **53** 1689–700
- Poëtte G 2018 A gPC-intrusive Monte-Carlo scheme for the resolution of the uncertain linear Boltzmann equation *J. Computat. Phys.* **385** 135–162
- Schürer R 2004 Adaptive quasi-Monte Carlo integration based on MISER and VEGAS *Monte Carlo and Quasi-Monte Carlo Methods 2002* ed H Niederreiter (Berlin: Springer) pp 393–406
- Sobotta B, Söhn M and Alber M 2010 Robust optimization based upon statistical theory *Med. Phys.* **37** 4019–28
- Soukup M, Fippel M and Alber M 2005 A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations *Phys. Med. Biol.* **50** 5089–104
- Souris K, Montero A B, Janssens G, Perri D D, Sterpin E and Lee J A 2019 Technical Note: Monte Carlo methods to comprehensively evaluate the robustness of 4D treatments in proton therapy *Med. Phys.* **46** 4676–84
- Taylor P A, Kry S F and Followill D S 2017 Pencil beam algorithms are unsuitable for proton dose calculations in lung *Int. J. Radiat. Oncol., Biol., Phys.* **99** 750–6
- Tourovsky A, Lomax A J, Schneider U and Pedroni E 2005 Monte Carlo dose calculations for spot scanned proton therapy *Phys. Med. Biol.* **50** 971–81
- Ulmer W and Matsinos E 2011 Theoretical methods for the calculation of Bragg curves and 3D distributions of proton beams *Eur. Phys. J. Spec. Top.* **190** 1–81
- Unkelbach J, Bortfeld T, Martin B C and Soukup M 2009 Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning *Med. Phys.* **36** 149–63
- Unkelbach J, Chan T C and Bortfeld T 2007 Accounting for range uncertainties in the optimization of intensity modulated proton therapy *Phys. Med. Biol.* **52** 2755–73
- Unkelbach J et al 2018 Robust radiotherapy planning *Phys. Med. Biol.* **63** 22TR02
- Wahl N 2018 Analytical models for probabilistic inverse treatment planning in intensity-modulated proton therapy *PhD Thesis Ruprecht Karl University of Heidelberg* (<https://doi.org/10.11588/heidok.00025127>)

- Wahl N, Hennig P, Wieser H-P and Bangert M 2017 Efficiency of analytical and sampling-based uncertainty propagation in intensity-modulated proton therapy *Phys. Med. Biol.* **62** 5790–807
- Wahl N, Hennig P, Wieser H-P and Bangert M 2018 Analytical incorporation of fractionation effects in probabilistic treatment planning for intensity-modulated proton therapy *Med. Phys.* **45** 1317–28
- Wahl N, Hennig P, Wieser H-P and Bangert M 2020 Analytical probabilistic modeling of dose-volume histograms *Med. Phys.* **47** 5260–73
- Weng X, Yan Y, Shu H, Wang J, Jiang S and Luo L 2003 A vectorized Monte Carlo code for radiotherapy treatment planning dose calculation *Phys. Med. Biol.* **48** N111–20
- Wieser H-P, Karger C P, Wahl N and Bangert M 2020 Impact of Gaussian uncertainty assumptions on probabilistic optimization in particle therapy *Phys. Med. Biol.* **65** 145007
- Wieser H-P *et al* 2017 Development of the open-source dose calculation and optimization toolkit matRad *Med. Phys.* **44** 2556–68
- Yang M, Zhu X R, Park P C, Titt U, Mohan R, Virshup G, Clayton J E and Dong L 2012 Comprehensive analysis of proton range uncertainties related to patient stopping-power-ratio estimation using the stoichiometric calibration *Phys. Med. Biol.* **57** 4095–115