Fast and accurate sensitivity analysis of IMPT treatment plans using Polynomial Chaos Expansion

Perkó, Zoltá; Van Der Voort, Sebastian R.; Van De Water, Steven; Hartman, Charlotte M.H.; Hoogeman, Mischa; Lathouwers, Danny

DOI
10.1088/0031-9155/61/12/4646

Publication date
2016

Published in
Physics in Medicine and Biology

Fast and Accurate Sensitivity Analysis of IMPT Treatment Plans Using Polynomial Chaos Expansion

Zoltán Perkó*[1], Sebastian R. van der Voort†[1,2], Steven van de Water‡[2], Charlotte M. H. Hartman[1,2], Mischa Hoogeman§[2], and Danny Lathouwers¶[1]

[1]Delft University of Technology, Department of Radiation, Science and Technology, Section Nuclear Energy and Radiation Applications, Mekelweg 15, Delft, The Netherlands

[2]Erasmus MC Cancer Institute, Department of Radiation Oncology, Rotterdam, The Netherlands

2016.03.23.

Abstract

The highly conformal planned dose distribution achievable in Intensity Modulated Proton Therapy (IMPT) can severely be compromised by uncertainties in patient setup and proton range. While several robust optimization approaches have been presented to address this issue, appropriate methods to accurately estimate the robustness of treatment plans are still lacking. To fill this gap we present Polynomial Chaos Expansion (PCE) techniques which are easily applicable and create a meta-model of the dose engine by approximating the dose in every voxel with multidimensional polynomials. This Polynomial Chaos (PC) model can be built in an automated fashion relatively cheaply and subsequently it can be used to perform comprehensive robustness analysis. We adapted PC to provide among others the expected dose, the dose variance, accurate probability distribution of dose-volume histogram (DVH) metrics (e.g. minimum tumor or maximum organ dose), exact bandwidths of DVHs, and to separate the effects of random and systematic errors. We present the outcome of our verification experiments based on 6 head-and-neck (HN) patients, and exemplify the usefulness of PCE by comparing a robust and a non-robust treatment plan for a selected HN case. The results suggest that PCE is highly valuable for both research and clinical applications.

Keywords: Proton Therapy, Uncertainty, Sensitivity, Setup Error, Range Error, Robust Optimization, Polynomial Chaos.

*Corresponding author: zperko@mgh.harvard.edu. Currently at Massachusetts General Hospital, Department of Radiation Oncology, Physics Research Group, Boston, Massachusetts, US.

†S.vanderVoort@erasmusmc.nl

‡S.vandeWater@erasmusmc.nl

§M.Hoogeman@erasmusmc.nl

¶D.Lathouwers@tudelft.nl

1 Introduction

Intensity Modulated Proton Therapy (IMPT) has the potential to surpass traditional radiotherapy due to its ability to better conform the dose distribution to the exact shape of the tumor and thereby better spare the surrounding healthy tissues. While this is highly desirable, the finite range of protons that makes dose conformity possible is also the source of a major drawback: the significantly higher sensitivity of IMPT treatments to uncertainties [1, 2, 3], which are always present due to inaccuracies in patient setup, uncertainties in the conversion from Computer Tomography (CT) Hounsfield Unit into stopping power, anatomical changes, etc. Clearly it is vital to assure that treatment plans are sufficiently robust against these uncertainties.

In recent years considerable efforts have been made to address this issue. The two most widely used methods are robust and probabilistic treatment planning. During robust planning a few worst case error scenarios are taken into account (e.g. a maximum error in patient setup in one of the directions or in proton range) and plans are optimized such that all constraints are respected in all of these scenarios [4, 5, 6, 7]. In the probabilistic approach the errors are assumed to follow a certain distribution and the optimization problem is formulated in a stochastic sense, subsequently the solution minimizes the expected value of the objective function while satisfying all constraints with a given probability only [4, 5, 8]. Theoretically both these strategies can provide superior treatment plans compared to current planning methods under the realistic, uncertain conditions. In practice however several proton therapy centers still use simpler techniques, such as the Single Field Uniform Dose (SFUD) approach, partially because the daily application of such advanced optimization strategies is far less understood.

One reason for this is that although the above methods do allow the uncertainties to be taken into account, they do not provide an accurate estimate of the actual sensitivities and the true robustness of the generated plans. Appropriate tools for such evaluation after treatment planning are also lacking, since unlike in traditional radiotherapy (where high energy photons are used and hence the dose distribution is relatively invariant under small errors), in proton therapy the dose has to be recalculated for every error scenario [9]. This makes the regular use of well-known Monte Carlo sampling techniques impractical, since they require several hundreds or even thousands of dose calculations for sufficient accuracy [10].

Our paper demonstrates that Polynomial Chaos Expansion (PCE) methods [11] can fill this gap by allowing truly comprehensive, yet fast robustness analysis of IMPT plans. PCE represents a meta-model of the dose engine for a given plan and can be constructed by calculating the dose in a limited number of predefined error scenarios. Once obtained, it provides among others the expected dose, the variance of the dose, and makes it possible to calculate the full dose distribution in any given error scenario. Consequently, for the purpose of comprehensive robustness analysis our work focuses on three main aspects. First, we adapted existing Polynomial Chaos (PC) techniques for the analysis of radiotherapy treatments. Second, we verified that PC can provide acceptable accuracy for real clinical IMPT plans with affordable computational costs. Third, we extended the existing PC methodology to better suit radiotherapy specific requirements. As a result the effects of random and systematic setup errors can be separated, thousands of full dose distributions can be calculated in a matter of seconds, and the full probability density function (PDF) of Dose Volume Histogram (DVH) metrics can be derived. This enables the accurate estimation of the probability of adequate Clinical Target Volume (CTV) coverage or the exact bandwidths of DVHs for instance. In conclusion, PCE allows a wide variety of estimates for plan robustness to be derived (practically all measures previously introduced in the literature) and can be a highly valuable tool in research as well as in clinical practice.

2 Methods and Materials

2.1 The Theory of Polynomial Chaos Expansion

The basic idea of Polynomial Chaos Expansion is to approximate the output of a model (e.g. the dose distribution calculated by the dose engine in our case) as a polynomial function of the stochastic input variables (e.g. errors in patient setup and proton range) [11]. Once this function - the PCE of the output - is available, it constitutes a meta-model of the original system, which - among others - can be used to quickly calculate the values of the response corresponding to different realizations of the uncertain inputs. To perform PCE one has to follow three steps. First, the stochastic input variables and their distribution

have to be specified, and appropriate polynomial basis functions have to be chosen (Section 2.1.1). Next, the computational model has to be run with specific realizations of the input parameters to calculate the corresponding outputs and to determine the expansion coefficients in the PCE (Section 2.1.2). Finally, the results have to be post-processed (Section 2.1.3). For brevity only the most essential aspects of these steps are detailed here, for all technical details see Section 1 of the Supplementary Materials (SM).

2.1.1 Input Uncertainties and Polynomial Chaos Basis Vectors

In proton therapy the most important uncertainties are the inaccuracies in patient setup (in the x, y and z directions) and proton range. Other uncertainties - such as internal organ motion for example - are also relevant, but our study does not cover these. Setup uncertainties have a systematic component (which is constant for a given patient during the course of the treatment) and a random one (which is different in each fraction). Range errors are usually assumed to be systematic and can have an absolute and a relative component, however in this work only the latter one was taken into account (including absolute range errors is straightforward though, they simply represent an additional input variable). These all together constitute $N = 7$ distinct sources of uncertainties, which can be described by a stochastic input vector $\underline{\xi} \sim (\delta\rho, \delta x, \delta y, \delta z, \delta X, \delta Y, \delta Z)$, where $\delta\rho$ is the range error, δx and δX are the systematic and random setup errors in the x direction respectively, etc. The systematic and random setup errors are however additive and can be handled together as combined setup errors, i.e. as $\Delta x = \delta x + \delta X$ for the x direction, etc. As a result the uncertainties can be described by an only $N = 4$ dimensional stochastic input vector $\underline{\xi} \sim (\delta\rho, \Delta x, \Delta y, \Delta z)$ (and consequently a computationally cheaper, but equally accurate "compressed" PCE can be obtained, for more details see Section 1.4.1 of the SM).

Next the distribution of the input variables, i.e. their $p_{\underline{\xi}}\left(\underline{\xi}\right)$ joint PDF has to be specified. In this work it was assumed that the uncertainties are independent, therefore the joint PDF is simply the product of the PDFs of the individual parameters, i.e. $p_{\underline{\xi}}\left(\underline{\xi}\right) = \prod_{j=1}^{N} p_{\xi_j}\left(\xi_j\right)$. Furthermore all uncertainties were supposed to be Gaussian with zero mean and various standard deviations (see later in Section 2.2.3).

Last, the orthogonal basis of the PCE is specified. The $\Psi_k\left(\underline{\xi}\right)$ PC basis vectors are multidimensional polynomials constructed by the tensorization of one-dimensional polynomials, i.e. as the product of univariate polynomials each depending on a single input variable ξ_j. The type of the polynomial family corresponding to parameter ξ_j depends on its distribution and follows the Wiener-Askey scheme [12]. For our normally distributed parameters probabilists' Hermite polynomials He_o are appropriate, consequently all PC basis vectors are of the form of $\Psi_k\left(\underline{\xi}\right) = \prod_{j=1}^{N} He_{o_{k,j}}\left(\xi_j\right)$, where $o_{k,j}$ in the subscripts denotes the order of the polynomial corresponding to PC basis vector k and input parameter j. A traditional choice is to use a full PC basis set, where all multidimensional polynomials up to a certain combined order O are included (i.e. where $\sum_{j=1}^{N} o_{k,j} \leq O$). In this work a computationally more efficient sparse basis set was used, where some of the high order multivariate basis vectors (representing strong interactions between the input variables) were excluded. For further details see Section 1.1 of the SM.

Once the basis vectors are specified the Polynomial Chaos Expansion of a response R depending on the $\underline{\xi}$ uncertain inputs as $R\left(\underline{\xi}\right)$ takes the form of

$$R\left(\underline{\xi}\right) = \sum_{k=0}^{P} r_k \Psi_k\left(\underline{\xi}\right), \tag{1}$$

where r_k are the PC expansion coefficients (PCCs) corresponding to the $\Psi_k\left(\underline{\xi}\right)$ PC basis vectors and $P + 1$ is the number of vectors in the basis set [11].

2.1.2 Polynomial Chaos Coefficients and Their Calculation Using Sparse Grids

Due to the orthogonality of the PC basis set (see Section 1.1.3 of the SM) the r_k expansion coefficients in Equation 1 can be determined as a spectral projection, i.e. as

$$r_k = \frac{\left\langle R\left(\underline{\xi}\right), \Psi_k\left(\underline{\xi}\right)\right\rangle}{\left\langle \Psi_k\left(\underline{\xi}\right), \Psi_k\left(\underline{\xi}\right)\right\rangle} = \frac{\int\int ... \int R\left(\underline{\xi}\right) \Psi_k\left(\underline{\xi}\right) p_{\underline{\xi}}\left(\underline{\xi}\right) \mathrm{d}\xi_1 \mathrm{d}\xi_2 ... \mathrm{d}\xi_N}{\left\langle \Psi_k\left(\underline{\xi}\right), \Psi_k\left(\underline{\xi}\right)\right\rangle}. \tag{2}$$

The main issue in Equation 2 is the evaluation of the multidimensional integral in the numerator, since it contains the unknown dependence of the response on the input parameters. In this work the Non-Intrusive Spectral Projection (NISP) approach is followed and a sparse grid based numerical integration

technique is used for this purpose [13]. The biggest advantage of this method is that it relies on using the original computer code (that calculates $R\left(\underline{\xi}\right)$ corresponding to different values of $\underline{\xi}$, in our case the dose engine) as a black box and needs no modification of its source code (hence the non-intrusive adjective). Therefore constructing the PCE of the dose distribution only needs the repeated execution of the dose calculation algorithm to compute the dose in each voxel (which are the $R\left(\underline{\xi}\right)$ responses) in a limited number of error scenarios. For all further technical details see Section 1.2 of the SM.

2.1.3 Post-processing the Polynomial Chaos Meta-model

Once the PCE is available it can be used to perform comprehensive sensitivity and uncertainty analysis of the output. The mean and the variance of the response are readily available as $\mu_R = \int R\left(\underline{\xi}\right) p_{\underline{\xi}}\left(\underline{\xi}\right) \mathrm{d}\underline{\xi} = r_0$ and $\sigma_R^2 = \int \left(R\left(\underline{\xi}\right) - \mu_R\right)^2 p_{\underline{\xi}}\left(\underline{\xi}\right) \mathrm{d}\underline{\xi} = \sum_{k=1}^{\infty} r_k^2 h_k^2 \approx \sum_{k=1}^{P} r_k^2 h_k^2$ respectively, which follow from the orthogonality of the PC basis set and the choice of the constant term as the zeroth term ($\Psi_0\left(\underline{\xi}\right) = 1$). Classic sensitivity analysis (i.e. determining the $\partial R/\partial \xi_j$ derivatives of the response) and variance decomposition (e.g. calculating Sobol sensitivity indices to identify the main sources of the uncertainty [14]) can be done analytically, as PCE is a simple function containing only polynomials. The full probability density function of R (and more generally the PDF of any computationally cheap function depending on the response) can also be obtained by sampling the PCE and using histograms or kernel density estimators (as we only need to evaluate polynomials and potentially a simple function, even millions of samples can be calculated in a matter of seconds for a single response, for more details see Section 1.3 of the SM). To better satisfy the needs of our radiotherapy application we also extended the general PCE methodology by devising two ways to construct new PCEs from an existing one. The first makes it possible to separate the effects of additive Gaussian variables (resulting in an "extended" PCE), the second to take the expected value over a subset of the inputs (making an "expected" PCE, both detailed in Section 1.4 of the SM). These are particularly important for radiotherapy, as they allow a convenient handling of random and systematic setup errors, such that the dependence of the expected dose (used to approximate the cumulative dose blurred out by the random setup errors during the course of a fractionated treatment, see in Section 2.5 of the SM) on the systematic errors can be investigated. Consequently the probability of CTV underdosage or the exact bandwidths of DVHs corresponding to any confidence level can be determined (see later in Section 2.3.2 and Section 3.3).

2.2 Radiotherapy Application of Polynomial Chaos Methods

2.2.1 Dose Calculation and Treatment Planning Tools

For our study all treatment plans were generated with Erasmus-iCycle [15, 16], an in-house developed multicriteria optimization method, in which instead of combining the various objectives with different weights into a single objective function, they are given priorities. The algorithm successively optimizes for the objective with the highest priority, then sets it as a constraint and proceeds by optimizing for the objective with the next priority. Erasmus-iCycle uses the dose engine of the ASTROID code developed at Massachusetts General Hospital [17], and it enables robust optimization using the minimax approach [6] with nine planning scenarios (nominal, 6 for worst case setup errors along the three principal axes and 2 for range [18, 19]). Dose recalculation for a given setup error is done by shifting the pencil beams, whereas range errors are taken into account by scaling the values of the CT image.

2.2.2 Polynomial Chaos Tools and Their Use

All Polynomial Chaos related calculations were done using a modified version of the Fully Adaptive Non-Intrusive Spectral Projection (FANISP) algorithm [13, 20]. FANISP is a versatile PC tool coded in Matlab, which can be coupled to external codes and is capable of building the PCE of responses using traditional sparse grids with a fixed PC basis set, as well as doing this adaptively (i.e. using adaptive sparse grids with a PC basis set constructed on-the-fly).

For the present work FANISP was coupled with Erasmus-iCycle as the external code. No adaptivity was used, sparse grids were chosen a priori as Smolyak or Extended Smolyak sparse grids of different grid orders (detailed in Section 1.2.2 and in Section 1.2.3 of the SM) and the corresponding dose calculations were done with Erasmus-iCycle. To reduce memory requirements without compromising accuracy only

those voxels were included in the PC models which had a dose exceeding 0.01 Gy in at least one of the 9 planning scenarios described in Section 2.2.1. The PCEs were built this way even in the cases where the treatments themselves were planned without robust optimization.

To evaluate the accuracy of PC a detailed cost-accuracy analysis, i.e. a systematic evaluation of the accuracy of different PC meta-models was done. Smolyak sparse grids with grid orders of $2, 3$ and 4 were used to build sparse PCEs using $O = 2, 3, 4$ respectively. After analyzing these results the extended Smolyak sparse grids were created by replacing the 4^{th} order one-dimensional grids with 5^{th} order ones while increasing the total polynomial order to $O = 5$. In this way the optimal set of PC parameters could be found that provided satisfactory results at minimal computational costs.

2.2.3 Investigated Patients and Uncertainties

For our initial exploratory investigations and the cost-accuracy analysis a head-and-neck (HN) patient was chosen with a unilateral tumor. For treatment planning a simultaneous integrated boost technique was used. The dose prescription was 66 Gy to the high dose CTV (consisting of the primary tumor and the positive neck levels) and 54 Gy to the low dose CTV (containing the elective neck levels). The treatment was planned for 30 fractions using 3 beam directions having angles of 60, 180 and 300 degrees. Plans were generated both without and with robust optimization, for the latter using the 9 planning scenarios described in Section 2.2.1 (nominal scenario, 6 scenarios with setup errors of ± 3 mm, 2 scenarios with ranger errors of $\pm(3\%+1$ mm)). The cost-accuracy analysis was done for both plans, along with a detailed study of the dependence of the voxel doses and other parameters on the errors. Based on the results the PC basis set and the required sparse grids were chosen and fixed. This set of parameters provides PC meta-models with $O = 5^{\text{th}}$ order sparse PC basis sets containing $P+1 = 73$ PC basis vectors and requires 217 evaluation with the dose engine according to 4^{th} order extended Smolyak sparse grids (the grids themselves need 209 calculations, 8 additional ones are used for the selection of the included voxels as described in Section 2.2.2). For all subsequent calculations (and all calculations presented in this paper) these settings were used. Further details of the cost-accuracy analysis can be found in Section 2.2 and Section 2.1 of the SM.

For a larger scale, clinical verification of PC, 6 head-and-neck patients (3 bi- and 3 unilateral cases, the latter including the example patient as well) were investigated with non-robustly optimized treatment plans (robustly optimized plans showed very similar accuracy in terms of PCE performance). The dose prescription was the same as for the example patient described above. In all cases the PCE of the dose was constructed by taking into account 4 uncertainty sources ($N = 4$): 3 setup errors and 1 range error. All were supposed to follow a Gaussian distribution with zero mean and various variances. The setup errors represented the combined systematic and random components (see Section 1.4.1 of the SM) and had $\sigma_x = \sigma_y = \sigma_z = 2.828$ mm (since for both components a 2 mm standard deviation was assumed and $\sqrt{2^2 + 2^2} \approx 2.828$). The relative range uncertainty was $\sigma_r = 2\%$ [21].

2.3 Clinical Verification and Illustrative Use of PC

2.3.1 Verification Methods

Both during the cost-accuracy analysis and the clinical verification, the accuracy of PC was tested in several error scenarios. These test scenarios all had different errors and were chosen such that they covered 99% of the input phase space. They were generated by uniformly sampling each of the 4 errors, making all possible combinations and retaining the ones within the 99% confidence ellipsoid [22] (for more details see Appendix A). This procedure resulted in 681 test scenarios for each patient. For each scenario the dose distribution was calculated by the dose engine (exact calculation) as well as by using the PC meta-model(s) and the two were quantitatively compared.

Three metrics were used for the comparison. For each patient and each test scenario first a Γ-evaluation was done [23, 24], with strict dose difference and distance-to-agreement criteria of $\Delta D = 0.1$ Gy and $\Delta d = 1$ mm respectively. These results were plotted in cumulative histograms similar to DVHs, showing the proportion of scenarios that have at least a given fraction of accepted voxels from the Γ-evaluation. To have a more quantitative measure of the dose errors the D_2 of the dose difference between PC and the dose engine was also analyzed. This is the minimal difference among the 2% of the voxels having the largest dose difference and is denoted by ΔD_2 in the rest of the paper. Finally, for each patient the dose difference distribution (i.e. the fraction of voxels as a function of the dose difference)

averaged over the simulated test scenarios was plotted, as well as the dose difference distribution in the scenarios with the lowest fraction of accepted voxels by the Γ-evaluations (these are the scenarios in which PC performs the worst).

Due to our method of choosing the test scenarios many of them lay on or very close to the surface of the 99% confidence ellipsoid and therefore represent unlikely cases (see Figure 7b in Appendix A). For example 20% of the them are beyond the 98% confidence level and only 50% are within the 95% level. This - combined with the strict Γ-evaluation criteria - makes our evaluation method very rigorous, as these improbable scenarios are taken into account with the same weight as the other more realistic ones.

2.3.2 Illustration of Comprehensive Robustness Analysis with PC

To demonstrate the usefulness of PC a comprehensive robustness analysis was performed for both the robust and the non-robust plan of the example HN patient. The same uncertainties were assumed as described in Section 2.2.3 and "extended" and "expected" PCEs were used to obtain various metrics of robustness. Separately most of these methods have already been published in the literature, however PC allows all of them to be derived and provides much higher accuracy than previously possible. The applied metrics are as follows.

DVH clouds A simple way to investigate plan robustness is to plot several DVHs corresponding to different error scenarios in the same figure [25, 21]. Such DVH clouds give a "feeling" of the variability of the DVH, however they only provide information about the specific scenarios that are investigated. The advantage of PCE here is that it allows a large number of scenarios to be investigated, and Section 3.3.1 highlights this by comparing the DVH clouds of a 100 scenarios and that of only the 9 planning scenarios. For a fair comparison in these calculations both the random and the systematic errors were sampled (by sampling the "extended" PCEs), therefore each of these DVHs can be thought of as a possible realization of a single fraction treatment (or a fraction dose if scaled properly).

Dose Standard Deviation The standard deviation of the dose is a well defined metric, giving the width of the PDFs of the dose in the individual voxels. Its distribution can be visualized on the CT image, as well as summarized in Standard Deviation Volume Histograms (SDVH) [8]. The benefit of PCE here is that it can provide the standard deviation computationally much cheaper than statistical estimators (see in Section 2.3 of the SM for details). For all results in Section 3.3.2, as well as for the rest of the metrics the "expected" PCEs were used and only the systematic errors were sampled when needed (these PCEs hence represent the cumulative dose during the course of a fractionated treatment).

Under- and Overdosage Probability The dose uncertainty can also be visualized as target underdosage or OAR overdosage probability [26, 10]. With PCE it is easy to derive these probabilities accurately, since we can draw a large number of samples of the dose distribution (depending on the size of the structure even up to 10000 - 100000) and simply count how many are above/below the defined value. In Section 3.3.3 we illustrate such a use by plotting the probability of each voxel in the high and low dose CTV on a representative slice of the CT not meeting their planning constraint ($D \geq 0.95 \cdot D_{pr}$).

Dose Population Histograms Dose Population Histograms (DPHs) conveniently summarize the spatial distribution of the dose uncertainty in a similar way as DVHs do for the dose itself [27]. A DPH gives the probability that - or the fraction of the population in which - the given parameter is bigger than a certain value. In Section 3.3.4 DPHs are used to investigate CTV underdosage and OAR doses. Just like in the previous case, without PCE the accurate calculation of these probabilities would be computationally much more expensive and far more complicated.

Dose Volume Histogram Distributions For a comprehensive description of the uncertainties DVH distributions (DVHDs) were derived using $\alpha - DVH$-s [28, 29]. An $\alpha - DVH(D)$ gives the volume V_D that receives a dose higher than D with a probability $1 - \alpha$, and can be obtained by sampling the PCE of the dose, producing the DVH for each sample on the same dose grid, then determining the $1 - \alpha$ percentiles. Subsequently $\alpha - DVH$ pairs for α and $1 - \alpha$ can be used to visualize the confidence intervals around the DVH for confidence level $1 - 2\alpha$ (Section 3.3.5). Once again PCE allows a sufficiently high

number of samples in a matter of seconds, from which the accurate estimates of the confidence intervals can be obtained.

3 Results

3.1 Clinical Verification of PCE For Head-and-Neck Patients

In Figure 1a the results of the gamma evaluations can be seen for the 6 investigated head-and-neck patients. 82% of the voxels in all patients pass the test in all 681 test scenarios, and the dose is correctly predicted by the PC model in more than 90% of the voxels in 95% of the scenarios. The pass rate rises to 94% when looking at only 80% of the scenarios, representing the more probable error scenarios closer to the nominal case (roughly the ones within the 98% confidence ellipsoid). To further investigate the accuracy of PC the explicit differences between PCE and the dose engine (i.e. the exact calculation) were examined. Figure 1b shows the ΔD_2 parameter among the scenarios. In all patients the ΔD_2 is less than 3.5 Gy, furthermore in roughly 80% of the scenarios it is less than 1.5 Gy.

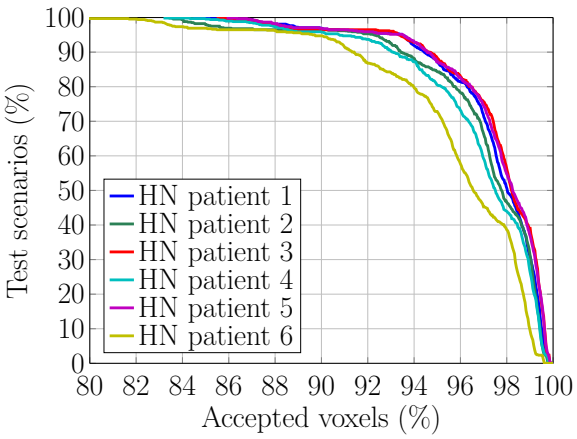
Naturally by looking at ΔD_2 we lose some information as there are voxels that have a higher dose difference than 3.5 Gy. Therefore the distribution of the dose difference is also depicted for the 6 patients, both averaged over the 681 test scenarios (Figure 1c) as well as for the worst performing scenarios (Figure 1d). As can be seen on average PC is very accurate, all voxels have an error smaller than 0.8 Gy in all patients. Even for the worst performing scenarios the largest dose difference is around 4 Gy, and 90% of the voxels are accurate up to 2 Gy (these worst case scenarios were always the ones with an extreme error in one of the directions or in the range, i.e. scenarios on the surface of the 99% confidence ellipsoid which are the least likely to occur in clinical practice).

3.2 Computational Cost Assessment

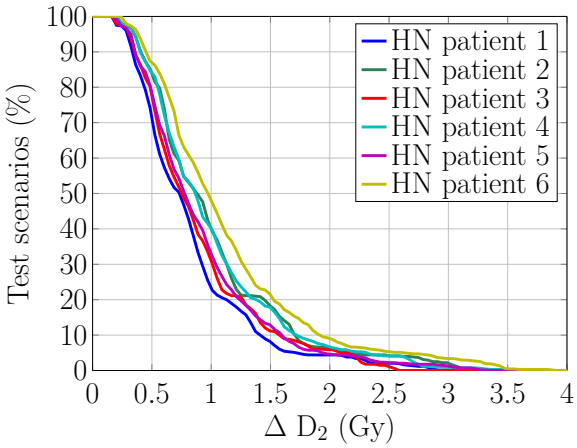
Table I shows the required total time to construct the PCE for the different patients as well as the time that was necessary to calculate the dose distribution corresponding to the 9 planning and a 1000 random error scenarios. The time for the 9 scenarios was chosen instead of the time for a single one since in Erasmus-iCycle multiple scenarios can always be calculated faster as a batch than one-by-one due to the CT image scaling and data input-output issues, hence it was considered a more representative measure. On average it took roughly 80 min to derive the PCE for the head-and-neck patients. This included both the time to run the 217 dose calculations as well as to extract the necessary data and calculate the PCCs. The computational overhead for all PC related calculations is small (typically it was around 5 minutes), most of the computational effort lies with having to recalculate the dose in various error scenarios. The true power of PCE is best shown by being able to calculate the dose in a 1000 random error scenarios within seconds, meaning that on average dose calculation with PCE takes less than 0.01 s.

Table I. Overview of the time needed to calculate the dose distribution for a given scenario and to construct the PCEs for the different patients. PCE results are representative for a modern desktop PC (2 core 64 bit Intel(R) Xeon(R) X5675 processor and 24 GB memory).

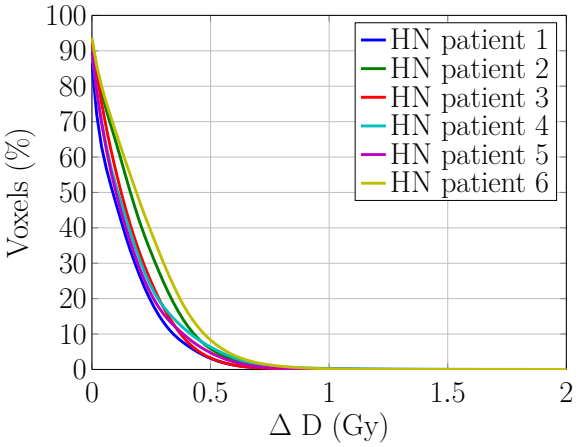
Patient	Patient group	Number of included voxels	Calculation time for 9 planning scenarios PCE	Exact	Calculation time for 1000 random scenarios (PCE)	PCE construction time
HN 1	unilateral	994265	0.45 s	366 s	2.89 s	67 min
HN 2	bilateral	1516041	0.62 s	500 s	4.00 s	82 min
HN 3	bilateral	1647850	0.68 s	556 s	4.62 s	83 min
HN 4	unilateral	1018986	0.41 s	310 s	2.67 s	53 min
HN 5	unilateral	1137738	0.45 s	351 s	3.10 s	72 min
HN 6	bilateral	2066136	0.86 s	823 s	5.96 s	110 min



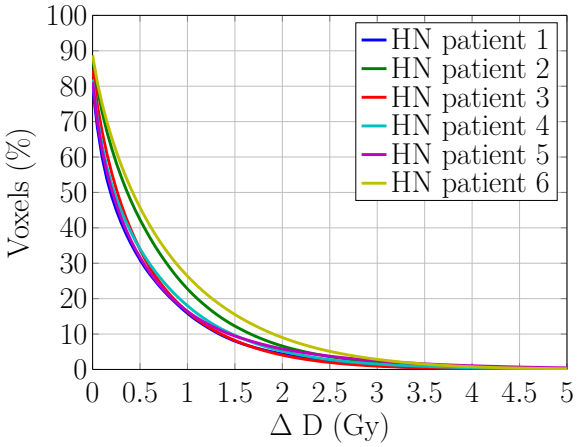
(a) Gamma evaluation in different scenarios for 6 head-and-neck patients.



(b) The minimum dose difference in the least accurate 2% of the volume for the 6 head-and-neck patients.



(c) The distribution of the dose difference in the HN cases averaged over the 681 test scenarios.



(d) The distribution of the dose difference in the HN cases for the worst case scenarios.

Figure 1. Evaluation of the accuracy of PC for head-and-neck patients. For the vast majority of test scenarios performance is excellent, and even in the most unrealistic ones doses are predicted with 4 Gy accuracy in even the worst performing voxels.

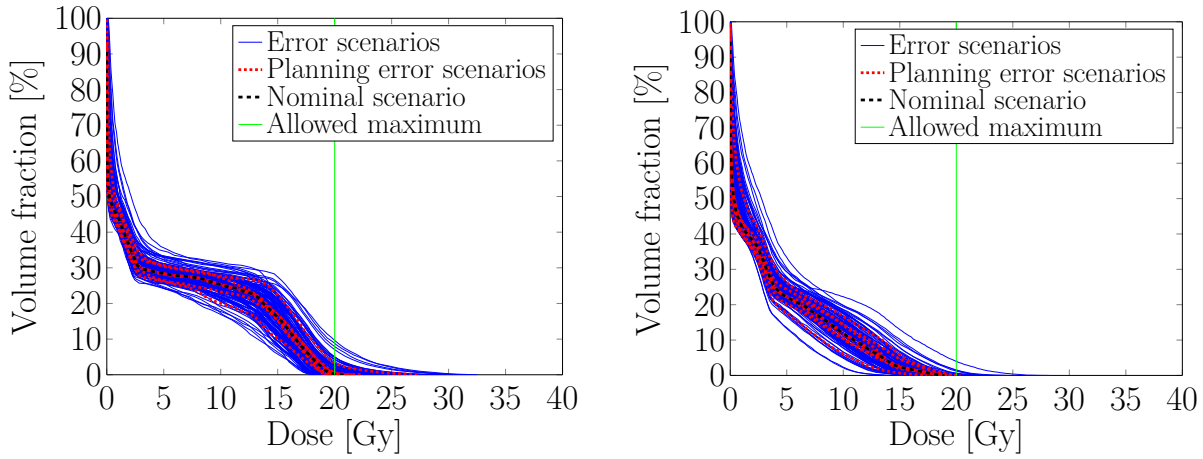
3.3 Illustrative Examples of Comprehensive Robustness Analysis with PCE

To exemplify the usefulness of PCE we present different methods to analyze plan robustness. Throughout this section the nominal scenario indicates not having any errors (i.e. neither random nor systematic) and delivering exactly the planned dose. The planning error scenarios refer to the 8 error scenarios used in robust planning (6 for setup and 2 for range). For the best realistic scenario a zero systematic uncertainty was assumed, therefore it provides the most optimistic realistically achievable treatment, where the cumulative dose is only blurred out by the random uncertainties. For the DVH clouds the "extended", for all other metrics the "expected" PCEs were used, hence for the former both the random and systematic errors, whereas for the latter only the systematic errors were sampled (when needed).

3.3.1 DVH clouds

Figure 2 shows the DVH of the cord in a robust and a non-robust plan for 100 different error scenarios, as well as the nominal scenario and the 8 planning error scenarios used in robust planning. Figure 2a shows that in the non-robust plan the nominal scenario naturally satisfies the constraint on the maximum cord dose (20 Gy), however even perturbations in the range or in the position along one of the axes alone can result in significantly higher maximum dose. Furthermore there is a large number of error scenarios

that violate the constraint. Figure 2b displays the same graph for the robust plan, highlighting that it successfully satisfies the constraint in all the 8 considered planning error scenarios. However when all errors are included there were still 11 scenarios among the 100 where the optimization constraint was violated. Including extra scenarios with simultaneous errors (i.e. errors in all 3 directions and the range at the same time) in addition to the 9 planning scenarios as proposed by [21] produces very similar DVH clouds as the planning scenarios alone, therefore would also fail to identify the overdosage of the cord compared to the optimization constraint (for a detailed discussion see Section 2.4 of the SM).



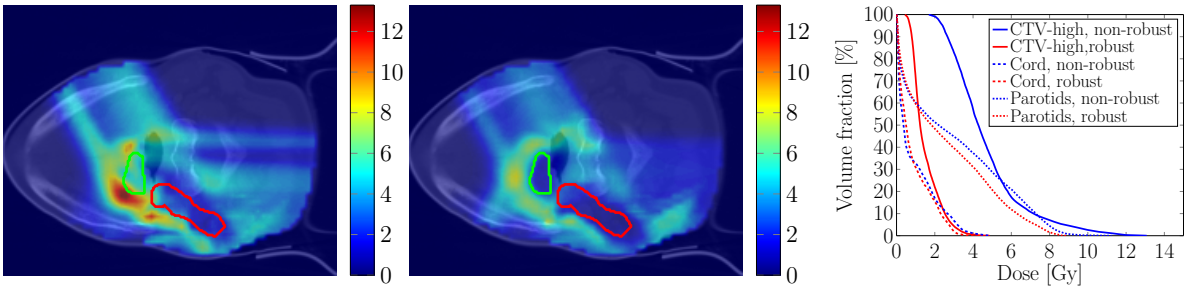
(a) Non-robust plan

(b) Robust plan

Figure 2. DVH cloud for the cord in the non-robust and the robust plan, together with the DVH for the nominal scenario and 8 error scenarios used in the robust optimization. Computation time: 0.03 s.

3.3.2 Standard Deviation Volume Histograms

The standard deviation of the cumulative dose (directly available from the PCCs) is shown in Figure 3. Figure 3a and Figure 3b display the spatial distribution on a CT slice in the middle of the CTV for the non-robust and the robust plans respectively. In the non-robust plan the standard deviation is much higher than in its robust counterpart, especially in the high dose CTV, reaching values up to 10 Gy. While in the robust plan large variations in the dose are mainly seen around the CTV (where they are expected due to the large dose gradients), in the non-robust plan there is a significant variability inside the CTV as well, indicating that parts of it may be underdosed. Figure 3c showing the SDVH for the high dose CTV, the cord and the parotid in the two plans further highlights that in the non-robust plan standard deviations in the CTV are $2-4$ Gy larger than in the robust plan. For the OARs however little difference is seen in this case.



(a) Dose standard deviation in the non-robust plan.

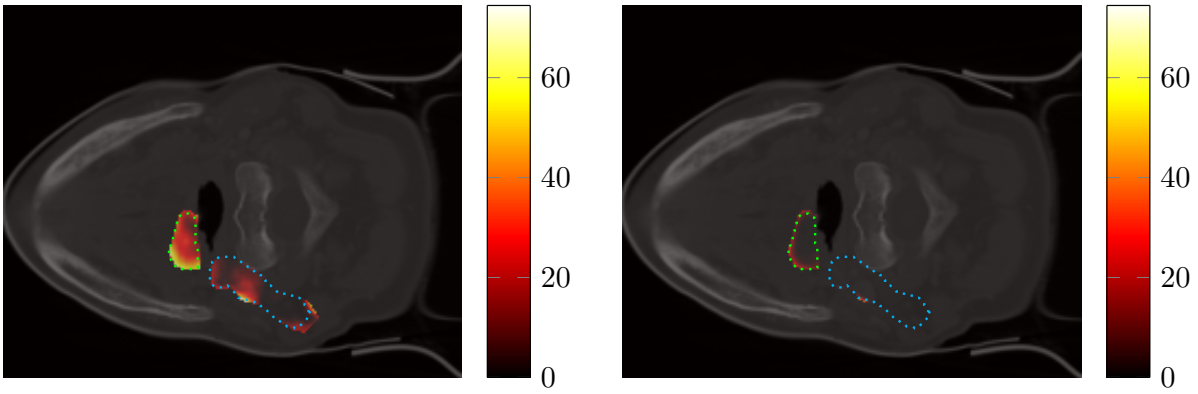
(b) Dose standard deviation in the robust plan.

(c) Standard deviation volume histograms.

Figure 3. The standard deviation of the dose (in Gy) on a CT slice in the middle of the CTV (highlighted in green and red) and standard deviation volume histograms of the example HN patient.

3.3.3 Visualization of Dose Uncertainty

The probability of CTV underdosage derived from the "expected" PCEs using 100000 error scenarios is shown in Figure 4. While in the non-robust plan there is a high, up to 70% chance of underdosing certain parts the CTV this is markedly reduced in the robust plan. The latter still has a non-negligible chance for cold spots at the edges of the target volumes, but this is drastically worse in the non-robust plan: basically all of the high dose CTV and large parts of the low dose CTV too have failure probabilities above 20%. Since the planning constraint requires 98% of the target volumes to receive 95% of their prescribed dose we can already suspect that non-robust plan has little chance to meet these constraints when the uncertainties are taken into account.



(a) Non-robust plan. Computational time 27.02 s.

(b) Robust plan. Computational time 24.10 s.

Figure 4. Probability of underdosage in the high and low dose CTV (green and blue dotted contours respectively) on a slice in the middle of the CTV (in %). Input uncertainties were Gaussian with 2.828 mm standard deviation for the setup and 2% for the range errors.

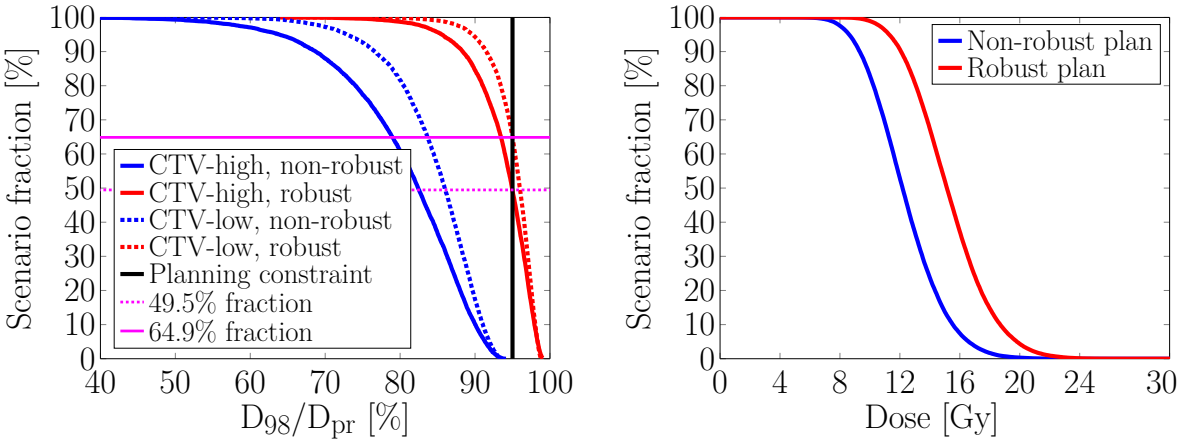
3.3.4 Dose-Population Histograms

Figure 5 shows the DPH for the CTVs and the parotid glands. Figure 5a reveals that - as suspected from Figure 4 - the non-robust plan is critically compromised by the uncertainties: there is practically 0 probability that the (near) minimum dose (D_{98}) in either the high or the low dose CTV would meet its planning constraint. This clearly improves by robust planning, however the satisfactoriness of the plan is still questionable, as there is only 49.5% and 64.9% chance of adequate high and low dose CTV coverage. The example well highlights practical issues with robust planning: the resulting plans always satisfy the planning constraints in the investigated error scenarios (the 9 planning scenarios in our case), however there can be a lot of other scenarios where they do not. Choosing which error scenarios to include in the optimization therefore is much more a guesswork currently than a rigorous process, since the connection between the considered scenarios and the general robustness of a plan is mostly unknown [30].

Figure 5b shows the DPH for the mean dose in the parotid glands. The mean is always between 6 Gy and 23 Gy, and the graph illustrates that the price of robustness in the CTVs is a markedly higher dose to the parotid (the parotid is the OAR with the highest priority on the wishlist of the optimization [30]).

3.3.5 Dose-Volume Histogram Distributions

Figure 6 shows the DVHs together with their confidence intervals for the high and low dose CTVs, as well as the cord, both for the non-robust and the robust plan, derived from the "expected" PCEs using 10000 error scenarios. The nominal scenario and the best realistic scenario (the plan with zero systematic errors) are also shown. Comparing Figure 6a and Figure 6b reinforces our conclusion from Figure 5a: in the non-robust plan, although the nominal scenario satisfies the constraint that 98% of the volume should receive 95% of the prescribed dose, this is practically never achieved in reality. Even in the best realistic scenario without any systematic error the random errors blur out the dose such that the planning constraint is violated. In the vast majority of cases the D_{98} is significantly, up to $5-10$ Gy lower than



(a) D_{98} of high and low dose CTV. Computation time: 0.46 s and 1.85 s for CTV-high and low respectively.

(b) Mean dose of parotid glands. Computation time: 0.29 s

Figure 5. Dose Population Histograms based on 10000 error scenarios.

supposed to be, and at the same time the maximum doses get very high, up to 90 Gy. In the robust plan the confidence intervals for the same confidence levels are drastically smaller. As a result several of the scenarios do satisfy the CTV coverage constraint and overdosage is greatly reduced. However there is still a significant (50%, see in Figure 5a) chance of underdosing the CTV. Figure 6c and Figure 6d show magnified views of the DVH distributions of the low dose CTV, further highlighting the inadequacy of the non-robust plan (even the best realistic scenario fails to satisfy the planning constraint).

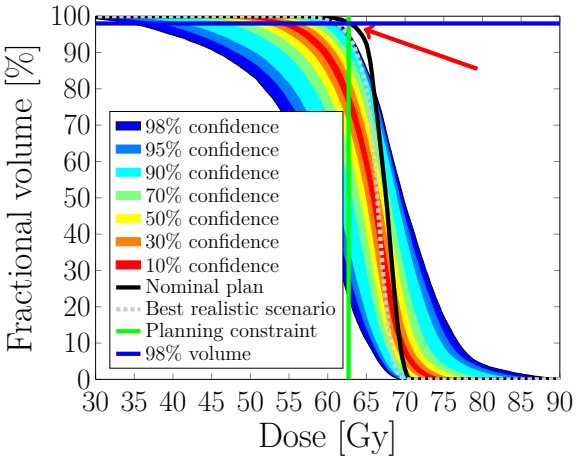
The DVH distributions for the cord are shown in Figure 6e and Figure 6f. As can already be expected from Figure 2a, there is a high, 50% probability of violating the planning constraint in the non-robust plan (Figure 6e). This is significantly reduced in the robust plan to around 15%, signaled by the much shorter tail of the DVH distribution (Figure 6f).

4 Discussion

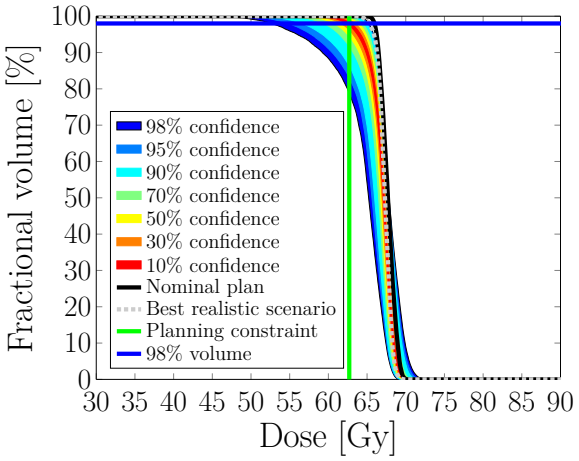
Accuracy From Figure 1 we can conclude that the accuracy of PC for head-and-neck patients is excellent. The most convincing argument is presented in Figure 1d showing that even in the most extreme (and therefore least likely) test scenarios the dose can be reconstructed with 4 Gy accuracy, with 90% of the voxels being 2 Gy accurate. This corresponds to 6% and 3% of the prescription dose, respectively. Furthermore for the vast majority of the scenarios the accuracy is substantially higher as exemplified by Figure 1a.

The performance of PC essentially depends on how the dose in the individual voxels changes with the various errors. This in turn is influenced by the anatomical details, the beam arrangement, the level of intensity modulation, the beam spot size, and several other factors, hence the computational cost associated with obtaining a PCE of desired accuracy is expected to vary with all of these. We have already seen such dependences when we tested PC on 5 prostate (PR) patients (see Section 3 of the SM for an extensive discussion). Using the same 217 scenarios and the same error distributions as for the HN cases PC only had an acceptance rate of 88% from the Γ-evaluation in 50% of the test scenarios (vs. 96% of the voxels for HN). A major reason for this worse performance is that PR patients had a higher prescription dose (74 Gy vs. 66 Gy), therefore the same 0.1 Gy dose agreement criterion in the Γ-evaluation is relatively stricter for them. Others include the smaller number of beams (2 vs. 3), the longer beam lengths, the larger effect of range errors and the stronger intensity modulation of the prostate plans, all detailed in Section 3 of the SM. Moreover the 50% acceptance rate covers 95% of the possible scenarios (see Section 2.3.1 and Appendix A), for which the accuracy is 88% from the Γ-evaluation and 2 Gy from the D_2 of the dose difference (corresponding to 3% of the prescription dose).

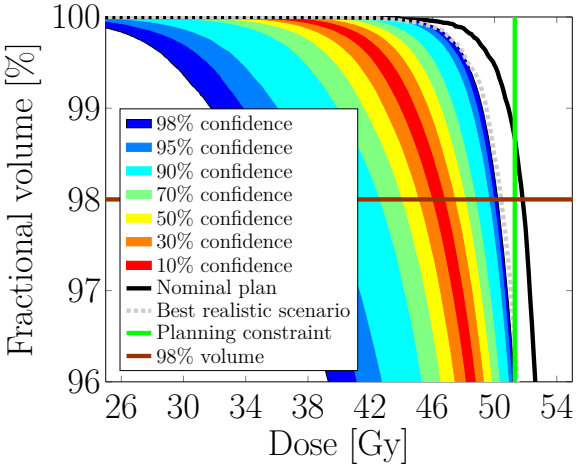
Using clinically typical 3 mm/3% criteria in the Γ evaluations for both the HN and PR cases (instead of our 1 mm/0.1 Gy values) leads to 99.94% and 99.4% of the voxels being accepted in the patient



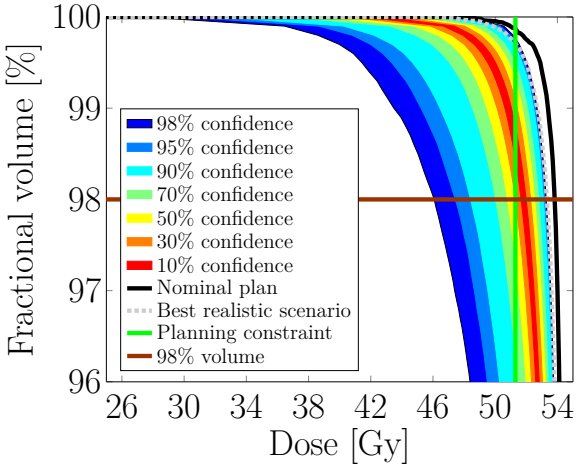
(a) DVH distribution of the high-dose CTV for the non-robust plan



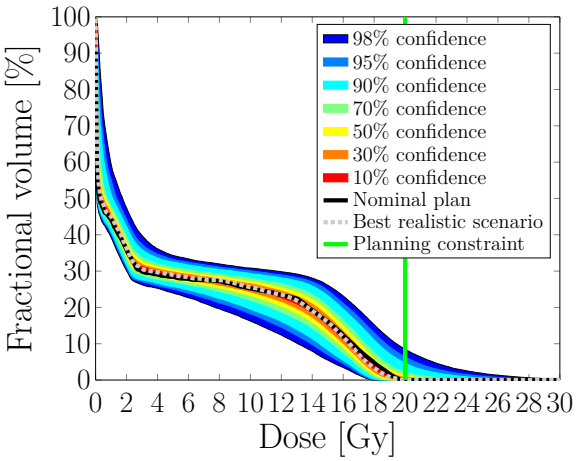
(b) DVH distribution of the high-dose CTV for the robust plan.



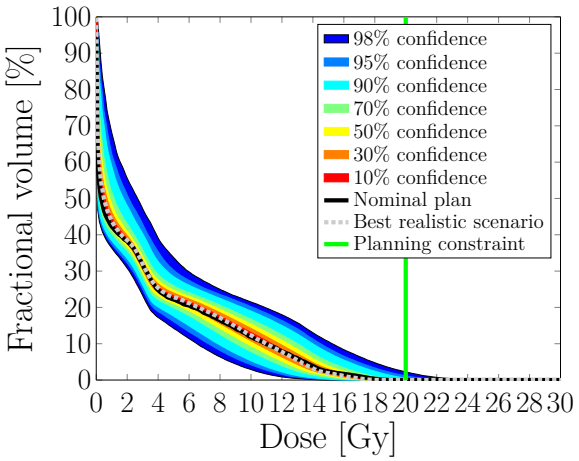
(c) DVH distribution of the low-dose CTV for the non-robust plan.



(d) DVH distribution of the low-dose CTV for the robust plan.



(e) DVH distribution of the cord for the non-robust plan. Computation time: 0.29 s.



(f) DVH distribution of the cord for the robust plan. Computation time: 0.34 s.

Figure 6. The distribution of the DVH of the CTVs and the cord in the example HN patient for the non-robust and the robust plan. Colored bandwidths signal different confidence levels for the DVH to be in. The nominal plan represents a treatment without errors, the best realistic scenario a fractionated treatment with infinite fractions and zero systematic errors.

groups respectively, in all 681 test scenarios (see Figure 10 of the SM). These results suggest that the accuracy of PC is significantly higher than what was found for other meta-models of the exact dose engine: for example the fast range-corrected dose approximation method proposed in [31] only had a 93% acceptance rate for an 8 mm unidirectional setup error in the examined lung patient (the most extreme unidirectional setup error among our 681 test scenarios was 7.3 mm). Considering that the 3 mm/3% Γ-evaluation criteria used in [32] to evaluate the accuracy of pencil beam algorithms compared to Monte Carlo dose calculation (without uncertainties) resulted in acceptance rates as low as 93% we can safely state that PCE well satisfies current clinical needs (at least for the examined HN and PR cases).

Computational Costs The computational cost associated with building an appropriate PC model is clearly site and requirement specific. The good performance achieved with 217 dose calculations is limited to the presented head-and-neck (and prostate) cases, the computational costs for other sites, error distributions, accuracy specifications, etc. can be both higher and lower. For example bigger spot size is expected to decrease cost, whereas anatomical complexity (e.g. metal implants) to increase it for similar accuracy. This implies that the cost-accuracy analysis has to be done separately for each disease site and intended application. However both the HN and PR cases suggest that there is little inter-patient variability for the same site, beam arrangement and plan objectives, therefore this is only expected to be necessary once for a given patient group. Applying adaptive PC methods [13, 20] could further alleviate this problem by automating the cost-accuracy analysis.

Comparison With Statistical Methods The most similar studies (e.g. [3, 10]) applied traditional Monte Carlo sampling to assess the effects of range and setup uncertainties of IMPT treatment plans. In these works the authors used a relatively large number of dose calculations to model fractionated treatments: [3] simulates 40 different systematic errors and 30 fractions (1200 calculations), [10] employs 60 samples with 10 fractions (600 calculations). [3] gives no justification how the number of samples was chosen, whereas in [10] this is determined by checking the convergence of the expected values of the DVHs (but not that of the standard deviations for example).

The quoted sample sizes are already higher than what PCE needs, at the same time they are far from being sufficient for accurately estimating more complicated DVH metrics, such as PDFs or under/overdosage probabilities. As Section 2.3 of the SM shows PCE drastically surpasses statistical methods even for calculating the (relatively simple) expected dose distribution: in the example HN patient PCE provided the same accuracy with 217 calculations as Monte Carlo sampling with almost 12000 dose calculations. In our experience this is true in general, PC methods systematically and decisively outperform statistical methods [33, 13, 20] (unless dealing with a large - higher than 50 - number of input variables). As a result, given a certain computational cost, PC allows far more accurate estimates than statistical methods. This makes its use worthwhile for radiotherapy applications even if exact dose engines become much faster than today [34].

Drawbacks and Limitations One drawback of our work is that the mathematical details may seem complicated at first sight (they are certainly more elaborate than evaluating a few worst case scenarios or using basic statistical methods). From a user point of view however there is a huge advantage of Polynomial Chaos techniques: appropriate tools are available to perform all the needed calculations [13, 20, 35], and one only has to ensure that the dose engine can be run with different errors and the dose distribution can be given back to the PC algorithm. If these interfaces are available PC can be applied without having to invest time and effort into the PC method itself.

A second limitation is that PCE relies on being able to parametrize the uncertainties in the problem by a finite set of random variables having given distributions. This makes it particularly useful for investigating the effects of patient setup, proton range, couch and gantry angle errors, however complicates its extension to cases where it is difficult (or even impossible) to parametrize the uncertainty (which is the case for example for anatomical changes). Moreover the higher the number of the stochastic inputs gets, the more costly the construction of the PC model becomes, hence it should always be carefully investigated what accuracy is needed and what computational cost is affordable.

As was shown in Table I with a current typical dose engine (Erasmus-iCycle) it takes roughly 60-80 minutes to run 217 dose calculations and to obtain a PCE. This excludes the instantaneous use of PCE

right after treatment planning, however it is a very reasonable time for all the advantages it provides, furthermore all necessary calculations can completely be automated (and run overnight for example).

Usefulness Appropriate sensitivity and robustness analysis are critical for the safe and widespread application of IMPT treatments. We have not encountered any technique available in the literature that can be as comprehensive as Polynomial Chaos Expansion can. Throughout Section 3.3 we have shown that PCE can practically yield all previously introduced sensitivity and uncertainty metrics, such as expected DVHs and the standard deviation of the DVH [10], DVH bandwidths for arbitrary confidence levels [28, 36, 29, 25], dose standard deviation distribution and Standard Deviation Volume Histograms [5, 8], PDFs of point doses and dose interval probabilities [26, 10], Dose Population Histograms [27, 37], error volume histograms [21], etc. What makes PCE unique is the speed and accuracy with which it can provide all these, and in particular its ability to enable the derivation of the spatial distribution of the probability of under- and overdosage, exact Dose-Population Histograms and Dose-Volume Histogram Distributions, which give a far more comprehensive description of dose uncertainties than previously presented methods. Furthermore PCE allows practically real time dose calculations, making it promising for regular clinical use. Naturally all these advantages come at a higher computational cost than some simpler methods (e.g. [1, 21, 25]), however these all rely on certain assumptions which are difficult to check, provide far less information, furthermore can result both in too conservative estimates, as well as in a false sense of safety (see Section 2.4 of the SM for further details).

Our adaptation of PCE also allows separating the effects of random and systematic setup errors, furthermore provides accurate DPHs. These combined enable the derivation of "robustness recipes". Such recipes establish a relationship between the magnitude of the setup and range errors (i.e. the standard deviations of their Gaussian distributions) and the error scenarios one should include in robust optimization (e.g. in a 9 scenario based optimization) to achieve a predefined CTV coverage in a given fraction of the patients. Consequently their use can make robust treatment planning much more quantitative than today, when the included error scenarios are mostly chosen ad-hoc. We have already demonstrated such a use for HN patients in [30] and are planning to extend this work to other sites.

5 Conclusions

The results indicate that Polynomial Chaos Expansion (PCE) is an attractive approach for performing the sensitivity and uncertainty analysis of intensity modulated proton therapy treatment plans. It allows the true robustness of treatments to be investigated in a quantitative way and therefore represents a valuable tool for both researchers and clinicians, filling the gap where such comprehensive robustness analysis methods are lacking. Possible uses include instantaneous visualization of the dose distribution and any dose related parameter for specified errors (e.g. to check the acceptability of a given fraction based on the daily imaging on-the-fly), automatic computation of whole treatment parameters (e.g. target underdosage, OAR overdosage, expected DVHs or DVH distributions), thorough robustness comparison of different modalities (e.g. IMPT versus passive scattering), derivation of robustness recipes, etc. Ultimately we plan to use PC methods during treatment planning itself.

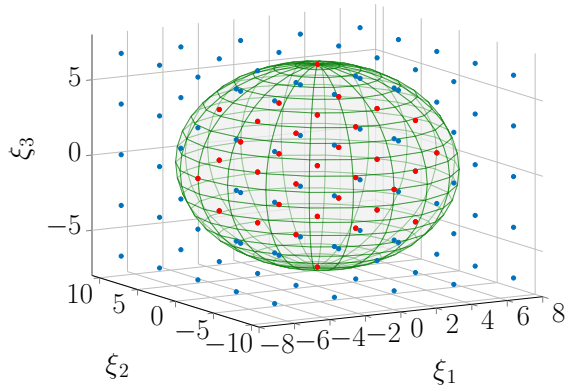
A Test Scenarios for the Evaluation of the PC Meta-models

For a multivariate normal distribution the contour encompassing $(1-\alpha)\cdot 100\%$ of the probability is given by the multidimensional ellipsoid of

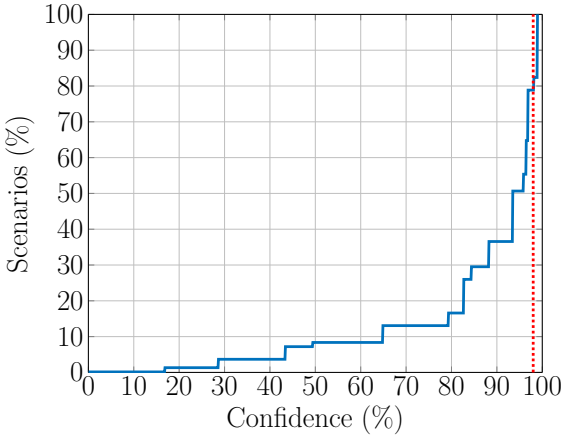
$$\left(\underline{\xi}-\underline{\mu}\right)^T \underline{\underline{\Sigma}}^{-1}\left(\underline{\xi}-\underline{\mu}\right) \leq \chi_p^2(\alpha), \tag{3}$$

where $\chi_p^2(\alpha)$ is the upper (100α)th percentile of the chi-square distribution with p degrees of freedom, $\underline{\mu}$ is the mean and $\underline{\underline{\Sigma}}$ is the covariance matrix [22]. In our case $p = N = 4$, α was chosen as $\alpha = 0.01$ to include 99% of the possible errors, $\underline{\mu} = (0,0,0,0)$ and $\underline{\underline{\Sigma}}$ is a diagonal matrix with the variances on the diagonal. The setup and range errors for the test scenarios were sampled in a way that they covered the 99% confidence volume in a grid-like structure. Along each axis a number of equally distanced points were chosen such that the two outermost ones always lay on the ellipsoid given by Equation 3. Then all possible combinations of the points along the axes were made, and finally all that fall outside the contour

were discarded. This procedure was done using 5, 6 and 7 points along the single dimensions and the combinations resulted in 681 included scenarios (as an illustration of the method in three dimensions using 5 points along the axes see Figure 7a). Half of the generated scenarios fall between the 95% and 99% confidence ellipsoids (see Figure 7b) and therefore represented unlikely cases (which are more difficult to model with PC). Consequently a 50% acceptance ratio for the scenarios accounts for the most probable 95% of the possible cases.



(a) Illustration of the error scenarios with three Gaussian variables with standard deviations of 2, 3 and 2. Points are uniformly positioned along the single dimensions (5 in this case) and the endpoints always lie on the surface of the ellipsoid. All combinations are made (blue points) and only the ones within the 99% (green) confidence ellipsoid are kept (red points).



(b) The distribution of the error scenarios as a function of the confidence level. Half of the scenarios are beyond the 95% level, and 20% of them are highly unlikely scenarios between the 98% (red dotted line) and 99% confidence ellipsoids.

Figure 7. Illustration of the error scenarios in which the performance of Polynomial Chaos Expansion was evaluated

References

[1] A. J. Lomax. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Physics in medicine and biology*, 53(4):1027–1042, 2008.

[2] A. J. Lomax. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 2: the potential effects of inter-fraction and inter-field motions. *Physics in medicine and biology*, 53(4):1043–1056, 2008.

[3] Aafke C. Kraan, Steven Van De Water, David N. Teguh, Abrahim Al-Mamgani, Tom Madden, Hanne M. Kooy, Ben J M Heijmen, and Mischa S. Hoogeman. Dose uncertainties in IMPT for oropharyngeal cancer in the presence of anatomical, range, and setup errors. *International Journal of Radiation Oncology Biology Physics*, 87(5):888–896, 2013.

[4] Millie Chu, Yuriy Zinchenko, Shane G Henderson, and Michael B Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Physics in medicine and biology*, 50(23):5463–5477, 2005.

[5] Jan Unkelbach, Timothy C Y Chan, and Thomas Bortfeld. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Physics in medicine and biology*, 52(10):2755–2773, 2007.

[6] Albin Fredriksson, Anders Forsgren, and Björn Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical physics*, 38(3):1672–1684, 2011.

[7] Wei Liu, Xiaodong Zhang, Yupeng Li, and Radhe Mohan. Robust optimization of intensity modulated proton therapy. *Medical Physics*, 39(2):1079, 2012.

[8] Jan Unkelbach, Thomas Bortfeld, Benjamin C Martin, and Martin Soukup. Reducing the sensitivity of IMPT treatment plans to setup errors and range uncertainties via probabilistic treatment planning. *Medical physics*, 36(1):149–163, 2009.

[9] S. E. McGowan, F. Albertini, S. J. Thomas, and A. J. Lomax. Defining robustness protocols: a method to include and evaluate robustness in clinical plans. *Physics in Medicine and Biology*, 60(7):2671–2684, 2015.

[10] Peter C. Park, Joey P. Cheung, X. Ronald Zhu, Andrew K. Lee, Narayan Sahoo, Susan L. Tucker, Wei Liu, Heng Li, Radhe Mohan, Laurence E. Court, and Lei Dong. Statistical assessment of proton treatment plans under setup and range uncertainties. *International Journal of Radiation Oncology Biology Physics*, 86(5):1007–1013, 2013.

[11] Olivier P. Le Maître and Omar M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, 2010.

[12] Dongbin Xiu and George Em Karniadakis. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[13] Zoltán Perkó, Luca Gilli, Danny Lathouwers, and Jan Leen Kloosterman. Grid and Basis Adaptive Polynomial Chaos Techniques for Sensitivity and Uncertainty Analysis. *Journal of Computational Physics*, 260:54–84, dec 2014.

[14] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55:271–280, 2001.

[15] Sebastiaan Breedveld, Pascal R. M. Storchi, Peter W. J. Voet, and Ben J. M. Heijmen. iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and non-coplanar IMRT plans. *Medical Physics*, 39(2):951, 2012.

[16] S. van de Water, A. C. Kraan, S. Breedveld, W. Schillemans, D. N. Teguh, H. M. Kooy, T. M. Madden, B. J. M. Heijmen, and M. S. Hoogeman. Improved efficiency of multi-criteria IMPT treatment planning using iterative resampling of randomly placed pencil beams. *Physics in medicine and biology*, 58(19):6969–83, 2013.

[17] Hanne M. Kooy, Benjamin M. Clasie, Hsiao Ming Lu, Thomas M. Madden, Hassan Bentefour, Nicolas Depauw, Judy a. Adams, Alexei V. Trofimov, Denis Demaret, Thomas F. Delaney, and Jacob B. Flanz. A Case Study in Proton Pencil-Beam Scanning Delivery. *International Journal of Radiation Oncology Biology Physics*, 76(2):624–630, 2010.

[18] Harald Paganetti. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in Medicine and Biology*, 57(11):R99–R117, 2012.

[19] Wei Liu, Steven J Frank, Xiaoqiang Li, Yupeng Li, Peter C Park, Lei Dong, X Ronald Zhu, and Radhe Mohan. Effectiveness of robust optimization in intensity-modulated proton therapy planning for head and neck cancers. *Medical physics*, 40(5):051711, 2013.

[20] Zoltán Perkó, Danny Lathouwers, Jan Leen Kloosterman, and Tim van der Hagen. Large scale applicability of a Fully Adaptive Non-Intrusive Spectral Projection technique: Sensitivity and uncertainty analysis of a transient. *Annals of Nuclear Energy*, 71:272–292, sep 2014.

[21] F. Albertini, E. B. Hug, and A. J. Lomax. Is it necessary to plan with safety margins for actively scanned proton therapy? *Physics in medicine and biology*, 56(14):4399–4413, 2011.

[22] Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson Eduction, Inc., Upper Saddle River, New Jersey, sixth edition, 2007.

[23] Daniel A. Low, W. B. Harms, S. Mutic, and J. A. Purdy. A technique for the quantitative evaluation of dose distributions. *Medical physics*, 25(5):656–661, 1998.

[24] Daniel A. Low and James F. Dempsey. Evaluation of the gamma dose distribution comparison method. *Medical physics*, 30(9):2455–2464, 2003.

[25] Alexei Trofimov, Jan Unkelbach, Thomas F. DeLaney, and Thomas Bortfeld. Visualization of a variety of possible dosimetric outcomes in radiation therapy using dose-volume histogram bands. *Practical Radiation Oncology*, 2(3):164–171, 2012.

[26] D Maleike, J Unkelbach, and U Oelfke. Simulation and visualization of dose uncertainties due to interfractional organ motion. *Physics in medicine and biology*, 51(9):2237–2252, 2006.

[27] Marcel Van Herk, Peter Remeijer, Coen Rasch, and Joos V. Lebesque. The probability of correct target dosage: Dose-population histograms for deriving treatment margins in radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 47(4):1121–1135, 2000.

[28] Francisco Cutanda Henríquez and Silvia Vargas Castrillón. A Novel Method for the Evaluation of Uncertainty in Dose-Volume Histogram Computation. *International Journal of Radiation Oncology Biology Physics*, 70(4):1263–1271, 2008.

[29] Francisco Cutanda Henriquez and Silvia Vargas Castrillon. Confidence intervals in dose volume histogram computation. *Medical Physics*, 37(4):1545–1553, 2010.

[30] Sebastian van der Voort, Steven van de Water, Zoltán Perkó, Ben Heijmen, Danny Lathouwers, and Mischa Hoogeman. Robustness recipes for minimax robust optimization in intensity-modulated proton therapy for oropharyngeal cancer patients. *International Journal of Radiation Oncology*Biology*Physics*, feb 2016.

[31] Peter C. Park, Joey Cheung, X. Ronald Zhu, Narayan Sahoo, Laurence Court, and Lei Dong. Fast range-corrected proton dose approximation method using prior dose distribution. *Physics in Medicine and Biology*, 57(11):3555–3569, jun 2012.

[32] Jan Schuemann, Drosoula Giantsoudi, Clemens Grassberger, Maryam Moteabbed, Chul Hee Min, and Harald Paganetti. Assessing the Clinical Impact of Approximations in Analytical Dose Calculations for Proton Therapy. *Radiation Oncology Biology*, 92(5):1157–1164, 2015.

[33] Luca Gilli, Danny Lathouwers, Jan Leen Kloosterman, and T. H. J. J. van der Hagen. Performing Uncertainty Analysis of a Nonlinear Point-Kinetics/Lumped Parameters Problem Using Polynomial Chaos Techniques. *Annals of Nuclear Energy*, 40(1):35–44, feb 2012.

[34] Joakim da Silva, Richard Ansorge, and Rajesh Jena. Sub-second pencil beam dose calculation on GPU for adaptive proton therapy. *Physics in Medicine and Biology*, 60(12):4777–4795, 2015.

[35] Stefano Marelli and Bruno Sudret. UQLab: A Framework for Uncertainty Quantification in Matlab. In *Vulnerability, Uncertainty, and Risk*, pages 2554–2563, 2014.

[36] F Cutanda Henríquez and S Vargas Castrillón. The effect of the different uncertainty models in dose expected volume histogram computation. *Australasian Physical and Engineering Sciences in Medicine*, 31(3), 2008.

[37] Marcel van Herk. Errors and margins in radiotherapy. *Seminars in Radiation Oncology*, 14(1):52–64, jan 2004.