

# Robust radiotherapy planning with spatially-based uncertainty sets

Noam Goldberg<sup>a</sup> , Mark Langer<sup>b</sup> , and Shimrit Shtern<sup>c</sup> 

<sup>a</sup>Department of Management, Bar-Ilan University, Ramat Gan, Israel; <sup>b</sup>School of Medicine, Indiana University, Indianapolis, IN, USA; <sup>c</sup>Faculty of Data and Decision Sciences, Technion - Israel Institute of Technology, Haifa, Israel

## ABSTRACT

Radiotherapy treatment planning is a challenging large-scale optimization problem plagued by uncertainty. Following the robust optimization methodology, we propose a novel, spatially based uncertainty set for robust modeling of radiotherapy planning, producing solutions that are immune to unexpected changes in biological conditions. Our proposed uncertainty set realistically captures biological radiosensitivity patterns that are observed using recent advances in imaging, while its parameters can be personalized for individual patients. We exploit the structure of this set to devise a compact reformulation of the robust model. We develop a row-generation scheme to solve real, large-scale instances of the robust model. This method is then extended to a relaxation-based scheme for enforcing challenging, yet clinically important, dose–volume cardinality constraints. The computational performance of our algorithms, as well as the quality and robustness of the computed treatment plans, are demonstrated on simulated and real imaging data. Based on accepted performance measures, such as minimal target dose and homogeneity, these examples demonstrate that the spatially robust model achieves almost the same performance as the nominal model in the nominal scenario, and otherwise, the spatial model outperforms both the nominal and the box-uncertainty models.

## ARTICLE HISTORY

Received 29 May 2023  
Accepted 14 May 2024

## KEYWORDS

Radiotherapy planning;  
robust optimization;  
biomarker uncertainty; row  
and column generation

## 1. Introduction


Radiotherapy treatment, and in particular Intensity-Modulated Radiation Therapy (IMRT), involves setting up an array of beams with variable intensities to irradiate a tumor. Planning such a treatment entails selecting the intensity of each beam or beam unit (called a beamlet). To facilitate such planning, the effect of the beam intensities – that is, the resulting radiation dose (whose unit of measure is  $1 \text{ Gy} = 1 \text{ J/kg}$ ) – on both the tumor and the healthy surrounding organs must be analyzed. The dose is measured over a three-dimensional grid of spatial cube units called voxels. The size of a voxel, and the granularity of the partition into voxels, are determined by the radiation beamlet size and the imaging resolution. Recent technology allows for voxel edge lengths of 2–3 mm. The dose at each voxel is generally assumed to be linearly related to the chosen beamlet intensities. In general, this is the theme of Radiotherapy Treatment Planning (RTP), which may prescribe the treatment over both time and space. The particular optimization problem for setting up the beamlet intensities is known as Fluence Map Optimization (FMO). A variety of formulations have been proposed for this problem. Typically, such formulations require that the tumor voxels are sufficiently irradiated, while healthy tissue voxels are spared from receiving an unhealthy dose. The formulations differ in the

way that these requirements are handled. Additional clinical requirements can also be factored into the model. Some formulations aim to maximize the average tumor voxel dose (Preciado-Walters *et al.*, 2006), while constraining the dose received by healthy organs. Other formulations minimize a measure of the dose received by healthy tissue voxels, or by all voxels, subject to ensuring that target (i.e., tumor) voxels receive at least the prescribed dose (Romeijn *et al.*, 2006).

Our approach is to maximize the tumor’s “weakest link”: the voxels receiving the minimum dose of all tumor voxels. This is a widely accepted objective for evaluating radiotherapy planning (Deasy, 1997; Yan *et al.*, 1997; Preciado-Walters *et al.*, 2004; Anon, 2010), subject to constraints on the healthy organ dose and on tumor dose homogeneity. In particular, the objective of maximizing the minimum dose objective is argued to better support biological optimization, as “a conservative surrogate for biological effect”, compared with minimization of deviations from prescribed doses (Deasy, 1997). The importance of combining this objective with homogeneity constraints is outlined by Yan *et al.* (1997), Allen Li *et al.* (2012), and Khairi *et al.* (2021).

Advances in imaging have exposed uncertainties in the effect of the prescribed radiation dosage. It is now known that tumors are not homogeneous, but rather, the sensitivity of cells varies across the tumor and changes over time (Titz

CONTACT Noam Goldberg  [noam.goldberg@biu.ac.il](mailto:noam.goldberg@biu.ac.il)

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/24725854.2024.2363316>.

Copyright © 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

and Jeraj, 2008; Bentzen and Gregoire, 2011; Saka *et al.*, 2014; Nohadani and Roy, 2017). In an effort to improve patient outcomes, there is a need for the model to account for the uncertain problem data. Modeling uncertainty in an optimization problem usually results in added complexity and scale. Uncertainty is involved in all radiotherapy treatment planning stages, starting with the target volume definition, but also in relation to the dose required to control the tumor, healthy tissue tolerances, and the actual dose delivered to the patient. Uncertainty in radiotherapy planning has been extensively addressed in the literature, mainly in the context of geometric uncertainty that results from the patient's movement and from imaging inaccuracies; see, for example, Bortfeld *et al.* (2008), the recent survey by Unkelbach *et al.* (2018), and the references therein.

Approaches for dealing with uncertainty in FMO include stochastic programming, robust optimization and distributionally robust optimization. The stochastic approach assumes that the distribution of the uncertainty is known and its parameters can be reasonably estimated. In practice, estimating the exact distribution is difficult, and incorrectly estimating the distribution may lead to solutions that are not clinically viable. In contrast, the robust optimization methodology (Ben-Tal *et al.*, 2009) does not assume any knowledge of the distribution, but rather, defines a set that contains all possible uncertainty realizations that need to be hedged against. By making few, or no, assumptions about the distribution, the robust approach optimizes the treatment for the worst-case realization of the uncertainty. Motivated by motion uncertainty, Chu *et al.* (2006) propose a robust formulation with uncertain influence matrices. The distributionally robust approach combines the stochastic and robust approaches – it protects against the worst-case distribution within a family of possible probability distributions of the uncertain parameters. For example, Bortfeld *et al.* (2008) model motion uncertainty by defining families of distributions that correspond to possible motion patterns of patients. Unkelbach *et al.* (2007) consider a (discrete) uncertainty set for beamlet range uncertainties in intensity modulated proton therapy (IMPT). This uncertainty set is constructed so that beamlets of the same beam must reach the same range. In this work they consider the objective of minimizing a weighted sum of the deviations from a prescribed dose. However, the model being solved may be overly conservative, as the worst case is applied to each of the voxels separately, rather than to the objective. A similar scenario-based approach to account for range uncertainty and setup uncertainty in IMPT is also used in Pflugfelder *et al.* (2008).

Another source of uncertainty, also mentioned in Unkelbach *et al.* (2018), arises from the biological conditions of the patients, which the authors consider a new frontier of robust RTP optimization. These conditions may affect the radiosensitivity of the tumor to the dose. The resulting effective dose is referred to as the biologically-adjusted dose. We note that this adjustment is different from the biological effective dose often modeled by the linear-quadratic model (Jones *et al.*, 2001), which aims to estimate the biological effectiveness in terms of cell-kill probability for a given

physical dose. In contrast, the biologically adjusted dose corrects the physical dose to account for the impact of a biological condition on the effective dose at a particular voxel. For the case of oxygen levels (or the lack thereof also known as hypoxia), this is described by the oxygen enhancement ratio, which is the ratio between doses required to obtain the same biological effect (Hill *et al.*, 2015; McMahon, 2018), or its normalized variant known as the Oxygen Modification Factor (OMF) (Titz and Jeraj, 2008). Information about a living tissue's (e.g., a tumor's) biological conditions is available through bio-markers and imaging data. Specifically, the tumor's oxygenation level, which strongly affects its radiosensitivity, is measured by imaging techniques such as fluoromisonidazole (FMISO) (Toma-Dasu *et al.*, 2012). Translating the imaging data into radiosensitivity values is a crude estimation that is subject to uncertainty. Further, the oxygenation level, and accordingly the radiosensitivity, may change between the time of the image scan and the commencement of treatment. In fact, biological uncertainty in radiation therapy, and the effect of oxygenation on radiosensitivity in particular, is an active area of research in the medical physics and other research communities (Titz and Jeraj, 2008; Toma-Dasu *et al.*, 2012; Shang *et al.*, 2021). Biological conditions and radiosensitivities are considered in the context of FMO by Saka *et al.* (2014), but their proposed methods do not attempt to model and protect from the associated uncertainty. Robust static (e.g., cumulative) dose planning under biological uncertainty has been proposed by Li *et al.* (2015) and Ten Eikelder *et al.* (2020). Ten Eikelder *et al.* (2020) propose a tumor control probability-based model assuming voxel independence and simple box uncertainty for its parameters. Simple box uncertainty here implies conservative solutions where all voxels are assumed to be simultaneously insensitive. Li *et al.* (2015) propose a dose-based model with budgeted uncertainty (see Bertsimas and Sim (2004)) on the deviation of the voxels' prescribed doses from their nominal value. In this model, the conservatism is controlled by a single budget parameter that controls how many voxels have prescribed doses that (maximally) deviate from their nominal values. However, in this model, the budget is allocated linearly to the change in the prescribed doses, so it might not accurately capture measurement errors whose effect on radiosensitivity may be highly nonlinear (Titz and Jeraj, 2008). Also, this model does not limit the differences in radiosensitivities among neighboring voxels, and thus, may consider unrealistic scenarios where the dose in one voxel radiosensitivity is significantly lower than its neighboring voxels.

Recent papers that consider radiosensitivity uncertainty in adaptive radiation therapy planning include Nohadani and Roy (2017), Roy *et al.* (2022), and Jeyakumar *et al.* (2023). These papers assume a prescribed dose approach where the objective is to minimize the total dose administered while satisfying simple lower and upper dose bound constraints on target voxels and organs at risk, respectively. Consequently, the biological uncertainty in this formulation appears only on the right-hand side. Nohadani and Roy (2017) utilize simple box uncertainty sets containing the

measured radiosensitivities, and model the evolution of this uncertainty through time. In the current article, we also formulate and solve a robust FMO that accounts for biological uncertainty at the voxel level where the main focus is to model the spatial connectivity or dependence of the voxel uncertainties. Both Roy *et al.* (2022) and Jeyakumar *et al.* (2023) propose models similar to Nohadani and Roy (2017). Roy *et al.* (2022) studies an extension of the model to a multistage adaptive problem, whereas Jeyakumar *et al.* (2023) consider adaptivity to uncertainty in influence matrix resulting in a semidefinite program.

In contrast with previous work, our approach is to model and explicitly account for the spatial interdependence among individual voxel radiosensitivity uncertainties. Accounting for such interdependence is expected to improve over simpler robust approaches that tend to result in overly conservative solutions. Spatial dependence has been considered in other applications; for example, by estimating covariance matrices and applying these using ellipsoid uncertainty sets in transportation (Chassein *et al.*, 2019), or general polyhedral uncertainty sets in energy (Lorca and Sun, 2014). These previous approaches may be less appropriate for biological uncertainty in radiotherapy because there may not be sufficient data to reliably estimate the correlations between voxel radiosensitivities. Also, applying both types of uncertainty sets to large-scale radiotherapy optimization problems, may be computationally intractable in practice. In radiotherapy, the scenario-based approaches to range and setup uncertainties in IMPT appear to resemble robust optimization with spatial uncertainty (for example Unkelbach *et al.* (2007) and Pflugfelder *et al.* (2008) discussed above). However, the uncertainty applies to the beamlets rather than to voxels and it is addressed by scenario-based approaches, so the dependency is not explicitly modeled. In contrast, the current paper focuses on spatial uncertainty modelling, reformulation of the corresponding robust FMO problem and solution methods for solving this problem. Dividing the administered dose into fractions and accounting for treatment sessions over time (also known as fractions) may involve a more elaborate uncertainty model over both space and time that is left for future work.

To make our proposed model applicable to clinical settings, we also consider important dose–volume constraints. These constraints limit the percentage of healthy-organ voxels that can receive more than a specified radiation dose. These are hard combinatorial constraints, which are virtually impossible to handle exactly given the problem size. Hence, many solution approaches have been proposed to approximately solve such constraints in the literature (Ferris *et al.*, 2003; Preciado-Walters *et al.*, 2006; Romeijn *et al.*, 2006). However, effectively solving models that incorporate such constraints remains a challenging problem, especially in the context of the large-scale problems considered herein.

To summarize our current contribution, we propose a novel FMO formulation that accounts for voxel radiosensitivity and its associated uncertainty, in both the objective and in the homogeneity constraints. To this end, Section 2 presents our novel spatially dependent uncertainty set with

justification based on publicly available real biomarker (FMISO imaging) data. In Section 3, we develop a scalable constraint generation algorithm for solving the resulting robust formulation. Section 4 presents a method for relaxing the dose–volume constraints – a method that can be solved and tuned in order to determine a feasible solution to the robust formulation when augmented with a dose–volume constraint. Finally, Section 5 includes a numerical study that demonstrates the scalability of our method, shows how to personalize the treatment plan, and presents a study of the effectiveness of our method under different scenarios. Conclusions are drawn in Section 6 along with proposals for future work.

## 2. Modeling robust radiotherapy planning

We now present our model for robust radiotherapy planning with biological uncertainty. In Section 2.1, the nominal model, which is the planning optimization model in the absence of uncertainty, is discussed. In Section 2.2, the robust approach to hedging against uncertainty is presented. This approach includes the development of a novel biological uncertainty set and its incorporation into the nominal problem.

### 2.1. Model introduction - nominal model

In the following, we use the notation  $[n] = \{1, \dots, n\}$ , for any integer  $n \geq 1$ . Additionally, for a vector  $a \in \mathbb{R}^n$  and index set  $I \subset [n]$ ,  $a_I \in \mathbb{R}^{|I|}$  denotes the subvector of  $a$  with coordinate set  $I$ . The set of all voxel indices is  $[m]$  for some positive integer  $m$ , which is partitioned into the Planning Target Volume (PTV) index set  $T \subset [m]$ , containing tumor voxel indices, and index sets corresponding to the  $K$  Organs At Risk (OARs),  $H_k \subset [m]$ , for each  $k \in [K]$ , such that  $[m] = T \cup \bigcup_{k=1}^K H_k$ .

FMO involves determination of the  $n$  beamlet intensities. For  $i \in [n]$ ,  $x_i$  is beamlet  $i$ 's intensity decision variable. Given the intensity vector  $\mathbf{x} \in \mathbb{R}^n$  and an influence matrix  $\mathbf{D} \in \mathbb{R}^{m \times n}$ , for each voxel  $v \in [m]$ ,  $d_v(\mathbf{x}) = \sum_{i=1}^n D_{vi}x_i$  is the physical dose applied to voxel  $v$ .

The objective is to maximize a concave function of the dose associated with the PTV voxels,  $f(\mathbf{d}_T)$ , while constraining the OAR dose. For convenience, throughout most of this article, we will consider a simple OAR bound of the form  $d_v(\mathbf{x}) \leq \bar{d}_k$  for each  $k \in [K]$  and  $v \in H_k$ , where  $\bar{d}_k$  is a predefined upper bound on the dose for OAR  $k \in [K]$ . In Section 4, we consider more elaborate dose–volume constraints. An important aspect of the problem that affects the definition of the function  $f$ , which is also the main focus of the model considered in the current article, is that the effect of the dose on a given PTV voxel  $v \in T$  depends on biological factors, such as the PTV's oxygenation level. This information is modeled by a radiosensitivity vector  $\phi \in [0, 1]^{|T|}$ . Accordingly, the biologically adjusted dose (or simply adjusted dose) of voxel  $v$  is given by  $\phi_v d_v$ . For example, if considering the effect of oxygen levels then  $\phi$  would be a vector of OMF values, as discussed in Section 1. The radiosensitivity is usually computed

based on biomarker and imaging data. Notably,  $\phi$  is highly uncertain, due to imaging and conversion inaccuracies and variation in the oxygen level over time. Let, the radiotherapy planning goal is generally to administer as much adjusted dose to the PTV voxels as possible, which is accomplished here by setting  $f(\mathbf{d}) = \min_{v \in T} \phi_v d_v(x)$ .

We also consider homogeneity constraints that limit the dose range across the PTV (Preciado-Walters *et al.*, 2004). This is intended to prevent the creation of “hot spots” in the tumor tissue, which cause damage to the healthy tissue in which the tumor is embedded. Specifically, our homogeneity constraints are of the form  $\mu f(\mathbf{d}) - \phi_v d_v \geq 0$ , for all  $v \in T$ , where  $\mu > 1$  is a given homogeneity parameter. Homogeneity has been traditionally applied to the physical dose (Allen Li *et al.*, 2012; Kataria *et al.*, 2012; Khairi *et al.*, 2021). The possibility of applying it to the adjusted dose has been enabled by recent technological advances in estimating necessary biological information, specifically the radiosensitivity vector  $\phi$ . The main goal of the current article is to address the challenge of modeling the uncertainty in  $\phi$  and using this information to improve radiotherapy treatment planning. The optimization problem of determining the intensity of the beams given the radiosensitivity parameter  $\phi$  is formulated as

$$\underset{\mathbf{x} \in \mathbb{R}_+^n, \mathbf{d} \in \mathbb{R}}{\text{maximum}} \quad \underline{d} \quad (1a)$$

$$\text{subject to} \quad \phi_v d_v(x) \geq \underline{d}, \quad v \in T \quad (1b)$$

$$\mu \underline{d} - \phi_v d_v(x) \geq 0, \quad v \in T \quad (1c)$$

$$d_v(x) \leq \bar{d}_k, \quad k \in [K], v \in H_k, \quad (1d)$$

which will be referred to as *the nominal formulation*.

Note that  $\phi$  is derived from imaging data through a series of crude transformations. Thus, it is potentially highly inaccurate. Ignoring this inaccuracy in practice may lead to poor clinical outcomes in terms of controlling the adjusted dose and adjusted homogeneity. The next section extends the model to address the uncertainty in  $\phi$ .

## 2.2. Robust optimization model with spatially bound uncertainty set

We now present a modeling scheme to deal with the uncertainty in the parameter vector  $\phi$ . The radiosensitivity uncertainty is caused by multiple factors, including imaging noise, biomarker conversion error, and variation over time. Inferring a distribution of the radiosensitivity parameter based on limited patient-specific (personalized) data is non-trivial and prone to error. Moreover, maximizing the clinical outcome (in our model, maximizing the minimum PTV adjusted dose) while using a high-dimensional multivariate probability distribution to model the biological uncertainty, would result in chance-constrained optimization problem (Luedtke *et al.*, 2010), which is notorious for being computationally hard. Instead, our approach to modeling uncertainty follows the *Robust Optimization (RO)* methodology. In this modeling framework, the uncertain parameters are

assumed to lie in a predetermined *uncertainty set* denoted by  $\mathcal{U}$ . An RO problem formulation is designed to protect its solutions against all realizations in the uncertainty set. In particular, an optimal solution  $(\mathbf{x}, \underline{d})$  must remain feasible for every possible realization  $\phi \in \mathcal{U}$  while maximizing the worst-case (minimum) PTV adjusted dose.

Formulating the robust counterpart of (1), in which variable  $\underline{d}$  now depends on the uncertain  $\phi$  subject to uncertain homogeneity constraints (1c), would lead to an adjustable RO problem. Such problems are computationally intensive (Ben-Tal *et al.*, 2004), and although it may be possible to solve these adjustable formulations when the problem is moderately sized, we find it more effective to avoid the adjustable robust model by reformulating the homogeneity constraints according to the equivalence

$$\begin{aligned} \mu \phi_u d_u - \phi_v d_v \geq 0, \quad \text{for all } v, u \in T &\iff \\ \mu f(\mathbf{d}) - \phi_v d_v \geq 0, \quad \text{for all } v \in T, & \end{aligned} \quad (2)$$

at the cost of increasing the number of constraints by a factor of  $O(|T|)$ . The resulting robust problem is given by

$$\underset{\mathbf{x} \in \mathbb{R}_+^n, \mathbf{d} \in \mathbb{R}}{\text{maximum}} \quad \underline{d} \quad (3a)$$

$$\text{subject to} \quad \max_{\phi \in \mathcal{U}} \phi_v d_v(x) \geq \underline{d}, \quad v \in T \quad (3b)$$

$$\max_{\phi \in \mathcal{U}} \{ \phi_v d_v(x) - \mu \phi_u d_u(x) \} \leq 0, \quad u, v \in T \quad (3c)$$

$$d_v(x) \leq \bar{d}_k, \quad k \in [K], v \in H_k. \quad (3d)$$

Constructing the uncertainty set poses some challenges, in particular, trying to satisfy and balance the following set of requirements for  $\mathcal{U}$ :

- (i) Personalized for the patient. This requirement relies on patient-specific information, such as imaging data acquired at the start of the treatment.
- (ii) Computationally tractable. The nominal problem may already be computationally intensive given the large number of voxels. The computational effort of solving the robust problem should not be excessive; in particular, it should not significantly increase the running time relative to that required to solve the nominal problem.
- (iii) Not too conservative. A goal of our modeling approach is to determine a set that captures the essence of the uncertainty, thereby providing “reasonable” protection, while not being too conservative, and thus tackling one of the major criticisms of RO.

In the next section, we present a novel uncertainty set that models the spatial dependence between voxels while addressing these requirements.

### 2.2.1. The spatially bound uncertainty set

Given a radiosensitivity vector  $\hat{\phi}$  that is estimated based on a patient’s imaging data at the start of the treatment, the goal is to construct an uncertainty set that satisfies our three

modeling requirements, described in Section 2.2. Perhaps the simplest uncertainty set that can be applied is a box (in particular, a hypercube) uncertainty set, which allows the radiosensitivity of each PTV voxel  $v$  to deviate from its calculated value  $\hat{\phi}_v$  by at most  $\delta > 0$ , where  $\delta$  is a given parameter. That is,  $\mathcal{U}_B = \{\phi \in \mathbb{R}^{|T|} : |\phi_v - \hat{\phi}_v| \leq \delta, \forall v \in T\}$ . This uncertainty set, used for example in Nohadani and Roy (2017), admits a straightforward robust formulation. However, if  $\delta$  is chosen to be sufficiently large to capture the actual uncertainty in tumor's biological conditions, then the resulting set may be overly conservative, thereby violating one of our key modeling requirements (see Section 2.2). To restrict the uncertainty further, a key observation to be made is that voxels that are situated close to each other should have similar radiosensitivities. In Figure 1, both plots demonstrate that the radiosensitivity differences are small for neighboring voxels and increase with the distance.

Incorporating this observation into the uncertainty model leads to a *spatially bound (SB) uncertainty set*; in particular, we define

$$\mathcal{U}_{SB} = \{\phi \in \mathbb{R}^{|T|} : |\phi_v - \hat{\phi}_v| \leq \delta, |\phi_v - \phi_u| \leq \gamma_{vu}, \forall v, u \in T\}, \quad (4)$$

where  $\gamma_{vu}$  are parameters of the uncertainty set that may depend on characteristics of the voxel pair  $v$  and  $u$ . Specifically, the  $\gamma_{uv}$  parameters are assumed to depend on the pairwise distances, and in particular, they are given by a metric-preserving function of the distance.

**Assumption 1.** For any  $u, v \in T$ ,  $\gamma_{uv} = \Gamma(\text{dist}(u, v))$ , where  $\text{dist}(u, v)$  is the distance in voxels between  $u$  and  $v$ , and  $\Gamma : \mathbb{R} \rightarrow [0, 1]$  is a nondecreasing subadditive function satisfying  $\Gamma(x) = 0$  if and only if  $x = 0$ .

Defining  $\gamma_{uv}$  as a function of the distance between  $u$  and  $v$  in this way captures the interdependence of the radiosensitivities of nearby voxels and limits the conservatism associated with the box uncertainty set. To the best of our knowledge, such spatially dependent uncertainty sets have not been previously investigated in the context of robust radiotherapy planning, nor it seems in the context of RO in general.

The  $\gamma_{uv}$  parameters for  $u, v \in T$  can be personalized to a particular patient's data, thereby addressing requirement (i).

Requirement (iii) may be satisfied by imposing the spatial constraints associated with the parameters  $\gamma_{uv}$ , for  $u, v \in T$ , to augment the box constraints in the definition of the spatially bound uncertainty set. In the absence of these spatial constraints, the uncertainty set would reduce to the overly conservative box uncertainty set. The following section addresses the question of whether optimal solutions of (3) for uncertainty set  $\mathcal{U}_{SB}$  can be computed efficiently in practice (requirement (ii)).

### 2.2.2. Compactly reformulating the robust problem

In this section, we present an exact reformulation of (3) for the case where  $\mathcal{U} = \mathcal{U}_{SB}$  with a polynomial number of constraints with respect to the size of the nominal formulation (1).

The following result, which will be used in the proofs that follow, establishes that  $\gamma_{uv}$  is a metric.

**Lemma 2** (Corazza, 1999; Proposition 3.2). *Suppose that Assumption 1 holds. Then,  $\gamma_{\cdot} = \Gamma(d(\cdot)) : T \times T \rightarrow [0, 1]$  is a metric.*

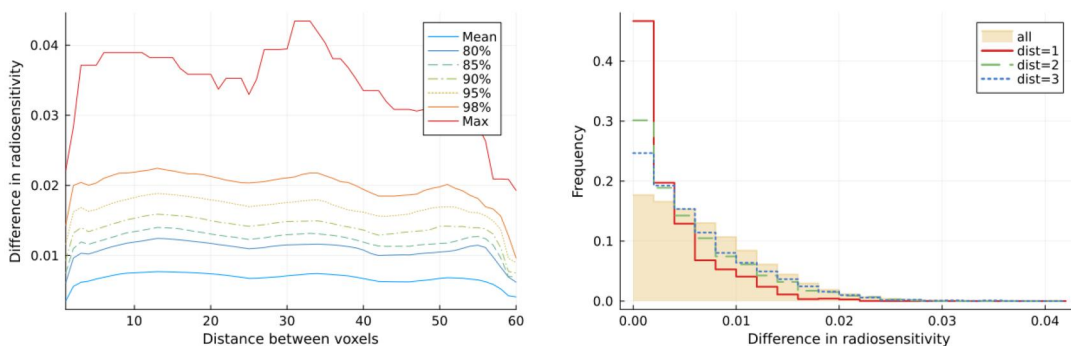
For  $S \subseteq \mathbb{R}^n$ , denote the projection of set  $S$  onto index set  $I \subseteq [n]$  by  $P_I(S)$ , so that  $P_I(S) = \{v_I \mid v \in S\}$ . For simplicity, for the singleton  $I = \{i\}$ , the projection will be denoted by  $P_i(S)$ . The following is a straightforward observation that is used in reformulating (3).

**Observation 3.** *Let  $\mathcal{U} \subseteq \mathbb{R}^n$  be a compact set. Then, for any continuous function  $f : \mathbb{R}^{|I|} \rightarrow \mathbb{R}$ ,  $\min_{\mathbf{u} \in \mathcal{U}} f(\mathbf{u}_I) = \min_{\mathbf{v} \in P_I(\mathcal{U})} f(\mathbf{v})$ .*

Problem (3) is now specialized and reformulated for the uncertainty set  $\mathcal{U}_{SB}$ . Henceforth, it is assumed that  $\mathcal{U}_{SB} \neq \emptyset$ . Starting with the the first set of constraints (3b), and using Observation 3, for each  $v \in T$ , the inner minimization in this constraint is given by

$$\min_{\phi \in \mathcal{U}_{SB}} \phi_v d_v(x) = \min_{\phi_v \in P_v(\mathcal{U}_{SB})} \phi_v d_v(x). \quad (5)$$

Since matrix  $\mathbf{D}$  and vector  $\mathbf{x}$  are both non-negative, the solution of (5) is given by the minimal value of  $\phi_v$ . For convenience, in the following, for each  $u \in T$ , define  $\underline{\phi}_u^0 = \max\{0, \hat{\phi}_u - \delta\}$ ,  $\bar{\phi}_u^0 = \min\{1, \hat{\phi}_u + \delta\}$ ,  $\underline{\phi}_u = \min_{\phi \in \mathcal{U}_{SB}} \phi_u$ , and  $\bar{\phi}_u =$



(A) Pairwise radiosensitivity difference percentiles vs. pairwise distances. (B) Histogram of the pairwise radiosensitivity differences for pairwise distances of 1, 2, 3, and all.

**Figure 1.** Relationship between radiosensitivity difference and pairwise PTV voxel distance in the first-visit data of Patient 1 from the TCIA brain dataset (Clark et al., 2013).

$\max_{\phi \in \mathcal{U}_{\text{SB}}} \phi_u$ . The following proposition characterizes the one-dimensional projection of  $\mathcal{U}_{\text{SB}}$ .

**Proposition 4.** For every  $v \in T$ :

- (i)  $P_v(\mathcal{U}_{\text{SB}}) = [\underline{\phi}_v, \bar{\phi}_v]$ .
- (ii)  $\underline{\phi}_v \geq \max_{u \in T \setminus \{v\}} \{\phi_u - \gamma_{uv}\}$   
and  $\bar{\phi}_v \leq \min_{u \in T \setminus \{v\}} \{\bar{\phi}_u + \gamma_{uv}\}$ .
- (iii) Suppose that *Assumption 1* holds. Then,

$$\underline{\phi}_v = \max_{u \in T} \{\underline{\phi}_u^0 - \gamma_{uv}\} \text{ and } \bar{\phi}_v = \min_{u \in T} \{\bar{\phi}_u^0 + \gamma_{vu}\}. \quad (6)$$

*Proof.* Part (i):  $P_v(\mathcal{U}_{\text{SB}}) = [\underline{\phi}_v, \bar{\phi}_v]$  follows from that fact that  $\mathcal{U}_{\text{SB}}$  is a convex and compact set and from Observation 3, which together imply that the projection of  $\mathcal{U}_{\text{SB}}$  onto  $v$  is a closed interval  $[\underline{\phi}_v, \bar{\phi}_v]$ .

Part (ii): For the sake of contradiction, suppose that  $\underline{\phi}_v < \underline{\phi}_u - \gamma_{uv}$  for some  $u \in T \setminus \{v\}$ , and let  $\phi^*$  be a minimizer (vector) that attains  $\underline{\phi}_v$ . Then, by optimality of  $\phi^* = \phi_v$ ,  $\phi_v^* - \phi_u^* \leq \underline{\phi}_v - \underline{\phi}_u < -\gamma_{uv}$ , thereby contradicting the feasibility of  $\phi^*$ . The proof of the upper bound on  $\bar{\phi}_v$  is similar.

Part (iii): Let  $\phi^* \in \operatorname{argmax}_{\phi \in \mathcal{U}_{\text{SB}}} \{\phi_v\}$ , so that  $\bar{\phi}_v = \phi_v^*$ . First, observe that by the optimality of  $\phi_v^*$ , there is at least one binding constraint,  $\bar{\phi}_v \leq \bar{\phi}_u^0$  or  $\bar{\phi}_v - \phi_{u_1}^* \leq \gamma_{vu_1}$ , for some  $u_1 \in T$ . Thus,  $\bar{\phi}_v = \min\{\bar{\phi}_u^0, \phi_{u_1}^* + \gamma_{vu_1}\}$  for some  $u_1 \neq v$ . If  $\bar{\phi}_v = \bar{\phi}_u^0$  then  $\bar{\phi}_v = \bar{\phi}_u^0 + \gamma_{vu}$  from Lemma 2, and the claim follows. Otherwise,  $\bar{\phi}_v = \phi_{u_1}^* + \gamma_{vu_1}$ . Applying this argument repeatedly to components of  $\phi^*$ , it follows that there exist  $u_2, \dots, u_k \in T$ , for some  $k \leq |T| - 1$ , satisfying

$$\phi_{u_l}^* = \phi_{u_{l+1}}^* + \gamma_{u_l u_{l+1}}, \text{ for } l = 1, \dots, k-1, \text{ and } \phi_{u_k}^* = \bar{\phi}_{u_k}^0. \quad (7)$$

**Lemma 2** (triangle inequality) implies that  $\bar{\phi}_v = \bar{\phi}_{u_k}^0 + \sum_{l=1}^{k-1} \gamma_{u_l u_{l+1}} \pm \gamma_{vu_1} \geq \bar{\phi}_{u_k}^0 + \gamma_{vu_k}$ . Next, from the feasibility of  $\phi^*$  and (7),  $\bar{\phi}_v = \phi_v^* \leq \phi_{u_k}^* + \gamma_{vu_k} = \bar{\phi}_{u_k}^0 + \gamma_{vu_k}$ , so it actually holds as an equality. A similar line of reasoning is used to prove the formula for  $\underline{\phi}_v$ .  $\square$

The one-dimensional projection established by Proposition 4 is graphically depicted by Figure 2. Note that following Proposition 4,  $\mathcal{U}_{\text{SB}}$  can be alternatively written as

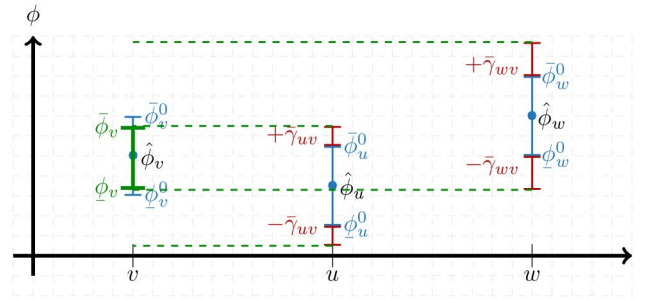
$$\mathcal{U}_{\text{SB}} = \{\phi \in \mathbb{R}^{|T|} : \underline{\phi}_v \leq \phi_v \leq \bar{\phi}_v, \forall v \in T, |\phi_v - \phi_u| \leq \gamma_{vu}, \forall v, u \in T\}.$$

Two-dimensional examples of  $\mathcal{U}_{\text{SB}}$  are graphically illustrated in Figure 3. The two-dimensional case will bear a special importance compactly reformulating the robust problem (3). Next, we reformulate the second set of constraints (3c). Again, due to Observation 3, for each  $v, u \in T$ , the inner maximization is

$$\max_{\phi \in \mathcal{U}_{\text{SB}}} \{\phi_v d_v(x) - \mu \phi_u d_u(x)\} = \max_{(\phi_v, \phi_u) \in P_{\{v, u\}}(\mathcal{U}_{\text{SB}})} \{\phi_v d_v(x) - \mu \phi_u d_u(x)\}. \quad (8)$$

The next proposition establishes that the projection onto a pair of indices  $u, v \in T$ ,  $P_{\{v, u\}}(\mathcal{U}_{\text{SB}})$ , can be expressed as a polyhedral set defined by six constraints.

**Proposition 5.** Suppose some  $v, u \in T$ ,  $\underline{\phi} \in \mathbb{R}^{|T|}$ , and  $\bar{\phi} \in \mathbb{R}^{|T|}$  whose components are given by (6), and that *Assumption 1* holds. Then,



**Figure 2.** An illustration of the one-dimensional projection result of Proposition 4. Assuming only three voxels,  $v$ ,  $u$ , and  $w$ , the bounds on the radiosenstivity of voxel  $v$  are determined by tightening its original interval  $[\phi_v^0, \bar{\phi}_v^0]$  by the intersection with  $[\underline{\phi}_u^0 - \gamma_{uv}, \bar{\phi}_u^0 + \gamma_{uv}]$  and  $[\underline{\phi}_w^0 - \gamma_{vw}, \bar{\phi}_w^0 + \gamma_{vw}]$ , imposed by voxels  $u$  and  $w$ , respectively.

$$P_{\{v, u\}}(\mathcal{U}_{\text{SB}}) = \mathcal{U}_{u, v}$$

$$\equiv \left\{ (\phi_v, \phi_u) : \begin{array}{l} \underline{\phi}_v \leq \phi_v \leq \bar{\phi}_v, \quad \underline{\phi}_u \leq \phi_u \\ \leq \bar{\phi}_w, \quad |\phi_v - \phi_u| \leq \gamma_{vu} \end{array} \right\}.$$

*Proof.* Let  $u, v \in T$ , and for convenience, define also  $\tilde{\mathcal{U}}_{u, v} = \{(\phi_v, \phi_u) : -\gamma_{vu} \leq \phi_v - \phi_u \leq \gamma_{uv}\}$ , to show that  $P_{\{u, v\}}(\mathcal{U}_{\text{SB}}) \subseteq \mathcal{U}_{u, v}$ . The contractive property of the projection onto a plane, as a convex set, implies that  $P_{\{u, v\}}(\mathcal{U}_{\text{SB}}) \subseteq P_u(\mathcal{U}_{\text{SB}}) \times P_v(\mathcal{U}_{\text{SB}})$ . Then, by Proposition 4-(i) and the definition of  $\mathcal{U}_{\text{SB}}$ , it follows that

$$\begin{aligned} P_{\{u, v\}}(\mathcal{U}_{\text{SB}}) &\subseteq (P_u(\mathcal{U}_{\text{SB}}) \times P_v(\mathcal{U}_{\text{SB}})) \cap \tilde{\mathcal{U}}_{uv} \\ &= ([\underline{\phi}_u, \bar{\phi}_u] \times [\underline{\phi}_v, \bar{\phi}_v]) \cap \tilde{\mathcal{U}}_{uv} = \mathcal{U}_{u, v}. \end{aligned}$$

To show the reverse inclusion, let  $(\tilde{\phi}_u, \tilde{\phi}_v) \in \mathcal{U}_{u, v}$ . Consider extending this two-dimensional vector to a  $|T|$ -dimensional vector in  $\mathcal{U}_{\text{SB}}$  in order to prove that  $(\tilde{\phi}_u, \tilde{\phi}_v) \in P_{\{u, v\}}(\mathcal{U}_{\text{SB}})$ . Now, consider some  $w \in T \setminus \{u, v\}$  and define  $\phi_w = \max\{\tilde{\phi}_v - \gamma_{vw}, \tilde{\phi}_u - \gamma_{uw}, \underline{\phi}_w\}$ . This choice of  $\phi_w$  satisfies the lower bounds imposed by  $\tilde{\phi}_u, \tilde{\phi}_v$  and  $\underline{\phi}_w$  in the definition of  $\mathcal{U}_{\text{SB}}$ . Also, to prove the corresponding upper bounds, first note that

$$\tilde{\phi}_w \leq \max\{\tilde{\phi}_v - \gamma_{vw}, \tilde{\phi}_u + \gamma_{vu} - \gamma_{uw}, \tilde{\phi}_v + \gamma_{vw}\} \leq \tilde{\phi}_v + \gamma_{vw},$$

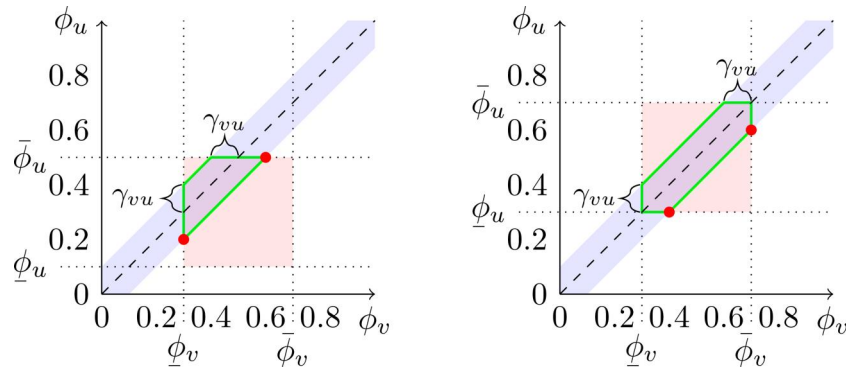
where the first inequality is due to  $(\tilde{\phi}_u, \tilde{\phi}_v) \in \mathcal{U}_{u, v}$  and  $\tilde{\phi}_v \geq \underline{\phi}_v \geq \underline{\phi}_w - \gamma_{vw}$  (the latter inequality following from Proposition 4-(ii)) and the second inequality is due to Assumption 1 and Lemma 2 (triangle inequality). Similarly,  $\tilde{\phi}_w \leq \max\{\tilde{\phi}_u + \gamma_{vu} - \gamma_{vw}, \tilde{\phi}_u - \gamma_{uw}, \tilde{\phi}_u + \gamma_{uw}\} \leq \tilde{\phi}_u + \gamma_{uw}$ . Finally, the fact that  $(\tilde{\phi}_u, \tilde{\phi}_v) \in \mathcal{U}_{u, v}$ , together with the definitions of  $\underline{\phi}$  and  $\bar{\phi}$ , imply that  $\tilde{\phi}_w = \max\{\tilde{\phi}_v - \gamma_{vw}, \tilde{\phi}_u - \gamma_{uw}, \underline{\phi}_w\} \leq \max\{\bar{\phi}_v - \gamma_{vw}, \bar{\phi}_u - \gamma_{uw}, \underline{\phi}_w\} \leq \bar{\phi}_w$ . Next, consider (arbitrary)  $w' \in T \setminus \{u, v, w\}$  in order to verify that  $\tilde{\phi}_w$  and similarly defined  $\tilde{\phi}_{w'}$  together satisfy

$$-\gamma_{ww'} \leq \tilde{\phi}_w - \tilde{\phi}_{w'} \leq \gamma_{ww'}. \quad (9)$$

Starting with the lower bound,

$$\begin{aligned} \tilde{\phi}_w - \tilde{\phi}_{w'} &= \max\{\tilde{\phi}_v - \gamma_{vw}, \tilde{\phi}_u - \gamma_{uw}, \underline{\phi}_w\} - \max\{\tilde{\phi}_v - \gamma_{vw'}, \tilde{\phi}_u - \gamma_{uw'}, \underline{\phi}_{w'}\} \\ &\geq \min\{\gamma_{vw'} - \gamma_{vw}, \gamma_{uw'} - \gamma_{uw}, \underline{\phi}_w - \underline{\phi}_{w'}\}. \end{aligned}$$

By definition of  $\underline{\phi}$ ,  $\underline{\phi}_w - \underline{\phi}_{w'} \geq -\gamma_{ww'}$ . By Lemma 2 (triangle inequality) we also have that  $\gamma_{vw'} - \gamma_{vw}, \gamma_{uw'} - \gamma_{uw} \geq -\gamma_{ww'}$ . Proving the upper bound is analogous (by interchanging the roles



**Figure 3.** An illustration of two different two-dimensional projections of  $\mathcal{U}_{SB}$  for voxels  $v$  and  $u$ . The shaded square depicts the  $\delta$ -bounds on the deviation from the measured  $\bar{\phi}$ . The gray 45°-band depicts the  $\gamma_{vu}$  bound on the radiosensitivity difference between the voxels. The green polygons are the projected uncertainty sets. The red points indicate the vertices that potentially maximize (8).

of  $w$  and  $w'$  and since  $\gamma_{ww'} = \gamma_{w'w}$ . Thus, (9) holds for all  $w, w' \in T$ . Finally, we have established that  $\bar{\phi} \in \mathcal{U}_{SB}$ , and accordingly, it follows that  $\mathcal{U}_{u,v} \subseteq P_{\{u,v\}}(\mathcal{U}_{SB})$ .  $\square$

Following the results of Propositions 4 and 5, formulation (3) with  $\mathcal{U} = \mathcal{U}_{SB}$  can now be compactly reformulated with a finite number of constraints. In particular, in this reformulation, no variables are added and the number of constraints of this formulation is  $2|T|$  times that of the nominal problem (1), thereby remaining polynomial in the size of the nominal formulation. Consider the formulation

$$\text{maximum}_{x \in \mathbb{R}^n, d \in \mathbb{R}} \underline{d} \quad (10a)$$

$$\text{subject to } \underline{\phi}_v d_v(x) \geq \underline{d}, \quad v \in T \quad (10b)$$

$$\bar{\phi}_v d_v(x) - \mu \max\{\bar{\phi}_v - \gamma_{vu}, \underline{\phi}_u\} d_u(x) \leq 0, \quad (u, v) \in T^2 : u \neq v \quad (10c)$$

$$\min\{\underline{\phi}_u + \gamma_{vu}, \bar{\phi}_v\} d_v(x) - \mu \underline{\phi}_u d_u(x) \leq 0, \quad (u, v) \in T^2 : u \neq v \quad (10d)$$

$$d_v(x) \leq \bar{d}_k, \quad k \in [K], v \in H_k. \quad (10e)$$

The following theorem establishes the correctness of this formulation, and specifically, demonstrates that its optimal solution set coincides with that of (3).

**Theorem 6.** *Suppose that Assumption 1 holds and let  $\mathcal{U} = \mathcal{U}_{SB}$ . Then, the optimal solution sets of (3) and (10) coincide.*

*Proof.* The reformulation of constraints (3b) is a straightforward consequence of (5) together and Proposition 4. Constraints (3c) are reformulated based on (8) to a maximization over  $P_{\{v,u\}}(\mathcal{U}_{SB})$  in place of a maximization over  $\mathcal{U}_{SB}$ . Due to Assumption 1, this set is given by the result of Proposition 5. Further, since the objective is linear, and since, for each  $u, v \in T$ ,  $P_{\{u,v\}}(\mathcal{U}_{SB})$  is a bounded polyhedral set, by the fundamental theorem of linear programming, there is an optimal  $(\phi_u, \phi_v)$  that is an extreme point of the set  $P_{\{u,v\}}(\mathcal{U}_{SB})$ . Moreover, since  $\phi_v$  and  $\phi_u$  have, respectively, a positive and a negative coefficient in (3c), this implies that the points in  $\text{ext}(P_{\{u,v\}}(\mathcal{U}_{SB}))$  are dominated by the points in

$$\begin{aligned} \mathcal{U}_{u,v}^* &\equiv \{(\max\{\bar{\phi}_v - \gamma_{vu}, \underline{\phi}_u\}, \bar{\phi}_v), (\underline{\phi}_u, \min\{\underline{\phi}_u + \gamma_{vu}, \bar{\phi}_v\})\} \\ &\subset \text{ext}(P_{\{u,v\}}(\mathcal{U}_{SB})); \end{aligned}$$

that is, for each pair  $v, u \in T$ , constraint (3c) is satisfied if and only if its left-hand side is nonpositive for these two points in  $\mathcal{U}_{u,v}^*$ .  $\square$

In Section 3, we will address the computational aspect of solving (10), specifically, how to deal with the potentially very large number of homogeneity constraints (10c) and (10d).

### 3. Solving large-scale RO

The radiation therapy problems that we consider involve tens of thousands of tumor voxels and millions of healthy tissue voxels. Accordingly, formulation (10) may have hundreds of millions of constraints. Therefore, it may prove difficult to solve or even store this problem on most computers. In order to address this challenge, we propose an algorithm that dynamically generates constraints of (10) as needed.

A schematic description is outlined by Algorithm 1.

#### Algorithm 1: Constraint generation algorithm sketch

**Input:** Initial beamlet configuration  $x_0$ , number of initial constraints for each OAR  $n_0$ , number of homogeneity constraints per iteration  $n_s$ , thresholds  $\tau_{\text{Infs},1} > \tau_{\text{Infs},2}$  and  $\tau_z$

**Initialize:**  $\hat{H}_1 \subset H_1, \dots, \hat{H}_K \subset H_K$  each with  $n_0$  constraints violated by  $x_0$ ,  $S \leftarrow \emptyset$ ;

- 1 Phase I: Alternate between iteratively generating at most  $n_H$  constraints (10e) to each OAR until a feasibility tolerance  $\tau_{\text{Infs},1}$  is achieved and the objective function value does not decrease by more than a factor of  $\tau_z$ , and appending at most  $n_s$  homogeneity constraints (10c) (10d) to set  $S \subseteq T^2$ . If none of the latter constraints are generated switch to Phase II;
- 2 Phase II: The same as Phase I, with feasibility tolerance of  $\tau_{\text{Infs},2} < \tau_{\text{Infs},1}$  for constraints (10e). Exit once no additional constraints can be generated.

A detailed version of the algorithm in pseudocode is given as Algorithm 3 in Appendix A.1. The algorithm is initialized with beamlet configuration  $x_0$ , typically a part of a

previously computed solution or an all-ones vector multiplied by some normalized dose. Based on this initial configuration, each OAR is initialized with the most violated  $n_0$  constraints. This initialization, and the initial emphasis on generating OAR constraints (10e) in the algorithm, ensures that the optimization problem is bounded from the start or that it becomes bounded after a few iterations in phase I of the algorithm. The algorithm consists of two phases. Each one of them alternates between generation of OAR constraints (10e) and (10c) (10d). The first phase involves an OAR constraint infeasibility tolerance  $\tau_{\text{Infs},1}$  that is larger than the tolerance applied in the second phase,  $\tau_{\text{Infs},2}$ . Intuitively, the larger tolerance in the first phase it intended to ensure both that the subproblems become bounded fairly quickly and to delay generation of OAR constraints that may be inactive at optimality until the homogeneity constraints are nearly satisfied.

Next we consider additional constraints that pertain to healthy organs and we develop an extension of the proposed computational method to address such constraints.

#### 4. Dose–volume constraints

In many cases, clinical guidelines stipulate, in addition to or in place of strict bounds on the OAR voxel dose, *dose–volume constraints*. This refers to the practice of allowing some proportion  $\alpha \in (0, 1)$ , usually less than one half, of the voxels to receive a larger dose than the upper bound  $\bar{d}_K$  applying to  $1 - \alpha$  of the voxels. In its most general form, a planning optimization problem could involve several dose–volume constraints for each organ. We develop a method for approximately solving large-scale planning problems in the case of a single dose–volume constraint. The single dose–volume constraint setting is clinically important, as illustrated by the brain-tumor dataset examined in the computational section.

In general, dose–volume constraints can be formulated by introducing additional auxiliary binary variables into formulation (10), turning it into the mixed-integer program,

$$\max_{\mathbf{x} \in \mathbb{R}_+^n, \underline{\mathbf{d}} \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^{|H_K|}, \mathbf{w} \in \{0, 1\}^{|H_K|}} \underline{\mathbf{d}} \quad (11a)$$

$$\text{subject to} \quad (10b), (10c), (10d) \quad (11b)$$

$$d_v(x) \leq \bar{d}_k \quad k \in [K-1], v \in H_k \quad (11c)$$

$$d_v(x) - y_v \leq \bar{d}_K \quad (11d)$$

$$y_v \leq (\hat{d}_K - \bar{d}_K) w_v \quad v \in H_K \quad (11e)$$

$$\sum_{v \in H_K} w_v \leq \theta. \quad (11f)$$

Here, without losing generality, it is assumed that the dose–volume constraint applies to the  $K$ th OAR, auxiliary continuous deviation variables  $y_v$  are defined for each  $v \in H_K$ , and  $\hat{d}_K$  is an “absolute” upper bound on the dose of

each voxel in  $H_K$ . Setting  $\theta = \lfloor \alpha |H_K| \rfloor$  in (11f) would ensure compliance with a dose–volume guideline specified as a proportion  $\alpha$ .

However, solving such a formulation to optimality involves a hard combinatorial problem that is neither theoretically nor practically tractable for the high-dimensional problems considered in the current article. The standard continuous relaxation of (11) replaces the binary decision variables  $w \in \{0, 1\}^{|H_K|}$  by continuous ones  $w \in [0, 1]^{|H_K|}$ . Therefore, it is straightforward to see that replacing constraint (11f) with  $\sum_{v \in H_K} y_v \leq \theta \cdot (\hat{d}_K - \bar{d}_K)$ , and fixing  $w_v = 1$  in (11e), for all  $v \in H_K$ , maintains the optimality of all solutions that are optimal to the continuous relaxation of (11). Hence, fixing  $w_v = 1$  for all  $v \in H_K$ , for  $\Theta = \theta \cdot (\hat{d}_K - \bar{d}_K)$ , the formulation

$$\max_{\mathbf{x} \in \mathbb{R}_+^n, \underline{\mathbf{d}} \in \mathbb{R}, \mathbf{y} \in \mathbb{R}^{|H_K|}} \underline{\mathbf{d}} \quad (12a)$$

$$\text{subject to} \quad (11b), (11c), (11d) \quad (12b)$$

$$\sum_{v \in H_K} y_v \leq \Theta \quad (12c)$$

$$y_v \leq \hat{d}_K - \bar{d}_K \quad v \in H_K \quad (12d)$$

is a continuous relaxation of (11). As a continuous relaxation, solutions of this formulation (12) with  $\Theta = \theta \cdot (\hat{d}_K - \bar{d}_K)$  may provide bounds but tend to be infeasible for (11). On the other hand, solutions of (12) for sufficiently small values of  $\Theta$ , for example  $\Theta = 0$ , together with a corresponding indicator variable vector  $w$ , is feasible for (11). This leads us to consider parametric solutions of (12) and in particular solving an equivalent penalty-based formulation (similar to a Lagrangian relaxation of (12)) that relaxes constraint (12c), excluding constant terms from the objective. In particular, denote the optimal solution set of this penalty-based problem variant as

$$\mathcal{S}(\beta) \equiv \operatorname{argmax} \left\{ \underline{\mathbf{d}} - \beta \sum_{v \in H_K} y_v : (12b) - (12d) \right\} \quad (13)$$

Of course, for every value of  $\Theta$  in problem (12), there exists a corresponding value of  $\beta$  in problem (13) such that the optimal values of  $\underline{\mathbf{d}}$  for the two problems coincide. For every  $\beta \geq 0$ , the notations  $s(\beta) \equiv (x(\beta), \underline{\mathbf{d}}(\beta), \mathbf{y}(\beta))$  and  $z(\beta)$  are used, respectively, to denote an optimal solution (an arbitrary element of the optimal solution set  $\mathcal{S}(\beta)$ ) and the optimal objective value of (13) with penalty parameter  $\beta$ . The next lemma establishes that  $\underline{\mathbf{d}}(\beta)$  and  $\sum_{v \in H_K} \mathbf{y}(\beta)_v$  are monotonically non-increasing in  $\beta$  (see proof in Appendix B.1).

**Lemma 7.** *For every  $0 \leq \beta_1 < \beta_2 \leq \beta_w$ , every pair of optimal solutions  $(x, \underline{\mathbf{d}}, \mathbf{y}) \in \mathcal{S}(\beta_1)$  and  $(\tilde{x}, \tilde{\mathbf{d}}, \tilde{\mathbf{y}}) \in \mathcal{S}(\beta_2)$  satisfies:*

$$\sum_{v \in H_K} \tilde{y}_v \leq \sum_{v \in H_K} y_v, \tilde{\mathbf{d}} \leq \underline{\mathbf{d}}, \text{ and } \tilde{\mathbf{d}} - \beta_2 \sum_{v \in H_K} \tilde{y}_v \leq \underline{\mathbf{d}} - \beta_1 \sum_{v \in H_K} y_v.$$

Lemma 7, together with the fact that there exists a value of  $\beta > 0$  for which  $\sum_{v \in H_K} \mathbf{y}_v(\beta) = 0$ , implies that for every

value of  $\Theta \geq 0$ , it is possible to find a minimal value of  $\beta$  such that (12c) is satisfied. Motivating our method is both monotonicity of  $\|y(\beta)\|_1 = \sum_{v \in H_K} y(\beta)_v$ , following Lemma 7, and that the one-norm is a convex approximation of the zero-“norm” ( $\|y(\beta)\|_0$ ), which is the number of non-zero components of the deviation variable vector  $y$ . The one-norm is effectively used to promote sparsity in large-scale machine learning and compressed sensing applications (see, for example, Donoho and Elad (2003)). We expect that solving (13) with appropriate  $\beta$  will approximate the problem with dose-volume constraints (11f) reasonably well. Our goal is to find the minimal value of  $\beta$  for which the optimal solution of (13) satisfies (11f). To achieve this, Linear Program (13) is solved parametrically, starting from a lower bound on  $\beta$  and traversing optimal bases for the given subsets of constraints defining the reduced problem. This algorithm is presented as Algorithm 2.

An additional difficulty in implementing our algorithm lies in the fact that, in practice, (13) is often too large to solve with all constraints present, and the constraints are generated on the fly as in Algorithm 1. Consequently, some of the constraints and variables may be absent from the formulation, and accordingly the reduced cost of these variables, as well as the full basis are not readily available. However, both the reduced cost and full basis are needed for determining the minimal increase in  $\beta$  for which the current optimal basis will no longer be optimal, and consequently  $\|y\|_0$  may change. Our parametric algorithm takes this into account by possibly retracting and allowing the value of the parameter  $\beta$  to decrease if the computed solution is not adjacent to the preceding basic feasible solution in the polyhedron that incorporates the newly generated constraints.

---

**Algorithm 2:** Parametric penalty algorithm sketch
 

---

**Input:** Initial bounds  $0 \leq \beta_l < \beta_u$ , BFS optimal for lower bound  $\beta = \beta_l$ , initial constraint subsets  $(S, \hat{H}_1, \dots, \hat{H}_K)$

- 1 **while**  $\|y_{\hat{H}_K}\|_0 > \theta$  or current BFS is not adjacent to the previous one in the full polyhedron with constraints  $(T, H_1, \dots, H_K)$  **do**
- 2 If the current BFS satisfies (or violates)  $\|y_{\hat{H}_K}\|_0 > \theta$  then minimally increase (respectively decrease)  $\beta$  so that current BFS is no longer optimal;
- 3 Solve restricted LP  $(S, \hat{H}_1, \dots, \hat{H}_K, \beta)$  to obtain optimal BFS with updated constraint sets  $(S, \hat{H}_1, \dots, \hat{H}_K)$
- 4 **end while**

**Output:**  $\beta$  and corresponding BFS.

---

A detailed version of this algorithm is given in Appendix A.3. The algorithm applies row and column generation in Step 3 in order to solve  $\Phi(S, \hat{H}, \dots, \hat{H}_K, \beta)$ ; the details of this procedure are presented as Algorithm 4 in Appendix A.3. This algorithm is a subroutine for solving (13) that is similar to Algorithm 1 in that it solves the problem for a subset of the constraints (rows), but it also generates columns corresponding to  $y$  variables as needed (only those components

of  $y$  corresponding to constraints (11d) and (11e) that have already been generated).

Let  $\beta^* = \min\{\beta : \exists(x, \underline{d}, y) \in \mathcal{S}(\beta), \|y\|_0 \leq \lfloor \alpha |H_K| \rfloor\}$ ; that is,  $\beta^*$  is the minimal penalty parameter for which the optimal solution to (13) satisfies (11f). Bounds on  $\beta^*$  can be established using Lagrange multipliers of constraint (12c) for solutions that are optimal to (12) for different values of  $\Theta$ , specifically,  $\beta_l$  and  $\beta_u$  for  $\Theta = \lfloor \alpha |H_K| \rfloor (\bar{d}_K - \underline{d}_K) + \epsilon$  (for some  $\epsilon > 0$ ) and  $\Theta = 0$ , respectively. These bounds are proven in Appendix B.2. The following proposition establishes the correctness of Algorithm 2 given such bounds.

**Proposition 8.** Suppose that LP (13) has a unique optimal basis for all  $\beta \in [\beta_l, \beta_u]$ . Then, Algorithm 2 determines  $\beta^*$  in a finite number of steps.

The proof of the proposition is deferred to Appendix B.3.

## 5. Experiments

In our experiments, we focus on a particular case study based on real brain data of Patient 4 from the TCIA archive (Clark *et al.*, 2013), referred to as the Brain dataset in Table 1. Due to the very large number of constraints in (1), and especially in (10), for this dataset, we also use a smaller liver imaging dataset (Craft *et al.*, 2014) to evaluate the computational performance of our dynamic generation of OAR and homogeneity constraints in Algorithm 1. FMISO-PET images are not available for this smaller dataset, so for this particular dataset, the radiosensitivity data are created synthetically. Oxygen levels tend to decrease, and accordingly radiosensitivity values tend to decrease, as we approach the center of the PTV. Thus, to create sensible radiosensitivity values, we first define vector  $f$  such that, for each PTV voxel  $v \in T$ ,  $f_v = (2\pi)^{-3/2} \det(\Sigma)^{-1/2} e^{-\frac{1}{2}(a_v - \bar{a})^\top \Sigma^{-1}(a_v - \bar{a})}$ , where  $a_v$  is the vector of  $x$ - $y$ - $z$  coordinates of voxel  $v$ 's position,  $\bar{a} = \frac{1}{|T|} \sum_{v \in T} a_v$  is the mean voxel position, and the covariance is given by a  $(3 \times 3)$  scaled covariance matrix of the PTV voxel positions,  $\Sigma = \frac{50}{|T|} \sum_{v \in T} (a_v - \bar{a})(a_v - \bar{a})^\top$ . The radiosensitivity for voxel  $v \in T$  is then set to be  $\phi_v = 0.85 + 0.15 \times \frac{f_{\max} - f_v}{f_{\max} - f_{\min}}$ , where  $f_{\max} = \max_{u \in T} \{f_u\}$  and  $f_{\min} = \min_{u \in T} \{f_u\}$ . This choice of  $\phi_v$  implies that it is contained in  $[0.85, 1]$ , while as voxel  $v$ 's position is closer to the center of the PTV, the radiosensitivity,  $\phi_v$ , approaches 0.85. Details of both datasets are given in Table 1.

**Table 1.** Imaging data summary.

Dataset	PTV $ T $	OAR			
		$k$	Name	$ H_k $	$\bar{d}_k$ bounds <sup>1</sup>
Brain (Patient 4)	32,367	1	Exterior	1,870,179	100
		2	Brain	1,360,182	62
		3	Chiasm	2003	54
Liver	6954	1	Liver	77,657	60
		2	Heart	32,983	50
		3	Entrance	146,462	60

<sup>1</sup>Based on (International Commission on Radiation Units and Measurements 2010).

## 5.1. Computational performance

Our implementation uses the Julia programming language, version 1.6.2, the JuMP package (Lubin and Dunning, 2015), and the Gurobi solver, version 9.1.2. The experiments are run on 40 core Linux servers each equipped with Intel Xeon E5-2650, 25 MB cache, 2.3 GHz base frequency CPUs, and 128 GB of RAM. Algorithm 1 (appearing in greater detail as Algorithm 3 in Appendix A.1) is invoked for solving the robust model (10) and the nominal model (1), with parameter values  $\tau_{\text{Infs},1} = 10$ ,  $\tau_{\text{Infs},2} = 0$ ,  $\tau_z = 10^{-2}$ , and  $n_H = 400$ . The homogeneity parameter is set to  $\mu = 1.15$  for Liver instances and  $\mu = 1.1250$  for Brain instances for the nominal experiments, and  $\mu = 1.1875$ , which obtains a similar homogeneity and biologically-adjusted dose in the nominal scenario. Table 2 shows the running times achieved when solving the nominal model (1), which involves only  $|T|$  homogeneity constraints. The results show that dynamic constraint generation is not needed to solve the nominal model; however, with careful tuning of the constraint generation algorithm, it can improve on the running time required to solve the full model (also yielding significant savings in terms of memory consumption), especially for larger data instances such as Brain. Table 3 shows the running times when applying the robust model to both datasets. The parameters of the uncertainty set are set to  $\delta = 0.04$  and  $\gamma = 0.04$  for Liver, and  $\delta = 0.08$  and  $\gamma = 0.04$  for Brain. Here, it is evident that dynamic constraint generation is necessary in order to be able to solve both models when the number of homogeneity constraints is  $O(|T|^2)$ . However, dynamically generating all violated constraints may result in excessive running times. In the example shown in Table 3, dynamically generating all violated constraints in each iteration results in a running time of nearly 4 hours for the Brain instance. In preliminary experiments, with a

**Table 2.** Running time in seconds and number of constraints generated in solving the nominal model (1). Here, constraint generation, if indicated, implies that all violated constraints are generated in each iteration. Best running time for each instance is marked by bold text.

Hom cons generation	OAR cons initial - $n_0$	Liver		Brain	
		Time	# generated	Time	# generated
No	2000	126.3	19,908	<b>586.7</b>	70,737
No	8000	<b>82.3</b>	37,908	731.7	82,737
No	all	453.4	271,010	5207.2	3,297,098
Yes	2000	189.5	22,858	3235.1	72,160
Yes	8000	175.9	37,531	3458.7	77,730
Yes	all	563.7	265,977	5117.7	3,264,928

**Table 3.** Running time in seconds and number of constraints generated in solving (10) using Algorithm 1. Best running time for each instance is marked by bold text.

Hom cons per iter - $n_s$	OAR cons initial - $n_0$	Liver		Brain	
		Time	# generated	Time	# generated
2000	2000	<b>3718.6</b>	92,890	<b>4940.4</b>	82,160
2000	all	10,215.3	335,528	12,861.4	3,271,931
8000	2000	3857.8	130,343	5255.7	106,261
8000	all	10,033.0	385,506	12,029.5	3,286,449
all	all	8133.5	408,747	14,356.1	3,424,511

different choice of parameters than experimented with here, when generating all violated constraints, we have observed cases where the running times took more than 48 hours to solve, and cases memory consumption exceeded its limits.

## 5.2. Robust model evaluation

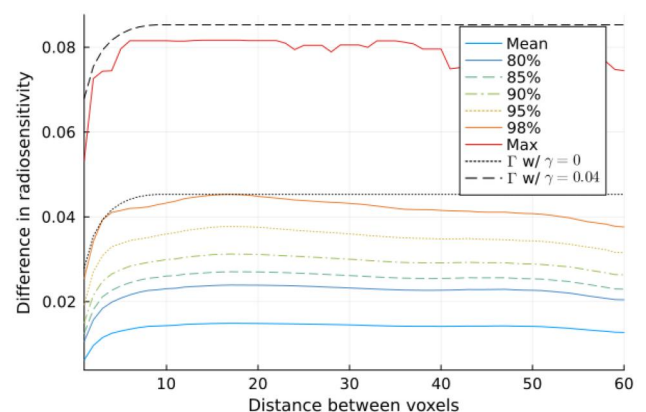
Next, the spatially bound uncertainty set ( $\mathcal{U}_{\text{SB}}$ ), introduced in Section 2.2, will be constructed and customized for our patient data. Subsequently, the benefits of both the robust model (10) based on this uncertainty set, and the dose-volume methodology of Section 4, will be examined and discussed.

### 5.2.1. Uncertainty set construction

The method of constructing our uncertainty set is justified in Section 2.2 using the data of brain cancer Patient 1 from the TCIA archive in Clark *et al.* (2013), illustrated in Figure 1. The graphs in Figure 4, which are based on the Brain data of Table 1 (Patient 4 from the archive), demonstrate a similar pattern to the one that emerges from Figure 1. Both figures clearly illustrate that the radiosensitivity distance increases in the geometric voxel distance, up to a geometric distance of approximately 10. To estimate the difference in radiosensitivity between voxel pairs as a function of the voxel distance  $\Delta$ , we adopt the following model with parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\alpha_2$ , and  $\gamma$  (initially 0),

$$\Gamma(\Delta) = \begin{cases} \gamma + \alpha_0 + \alpha_1 \Delta + \alpha_2 \log(\Delta) & \Delta \leq 10 \\ \Gamma(10) & \text{otherwise.} \end{cases}$$

The estimated  $\Gamma$  is personalized to Patient 4's PET data. It is estimated through a constrained least squares fit of the radiosensitivity difference  $p$ th percentile curve (see Figure 4), for voxel distances up to  $\Delta = 10$ . The least squares problem solved, is constrained in our case, so that the fitted curve upper bounds the 98th percentile graph. Having obtained the best fit curve, the constant  $\alpha_0$  is then slightly increased so that the curve upper bounds the same  $p$ th percentile for all  $\Delta \geq 1$ . The constant  $\gamma \geq 0$  is used to adjust the model to be even more conservative. The final estimated parameter values are  $\alpha_0 = 0.0292761$ ,  $\alpha_1 =$



**Figure 4.** Radiosensitivity difference vs. voxel distance for PTV voxel pairs of Patient 4 in the TCIA archive Brain dataset.

$-0.0013514$ , and  $\alpha_2 = 0.0128265$ . Evidently,  $\Gamma$  satisfies Assumption 1 for all  $\gamma \geq 0$ . Moreover, for  $\gamma = 0$ , by construction,  $\Gamma$  upper bounds the 98th percentile graph, and for  $\gamma = 0.04$ ,  $\Gamma$  upper bounds the graph of the maximum, as seen in Figure 4.

The magnitude of the uncertainty set parameter  $\delta$ , which bounds the deviation of a voxel's radiosensitivity from its calculated value can be estimated by a sensitivity analysis of underlying theoretical equations in the uncertain constants; see for example the discussion in Appendix C.

### 5.2.2. Robust formulation experiments

We now evaluate optimal solutions of *spatially robust* model (10) and compare them with optimal solutions of the nominal formulation (1) for the Brain dataset. We also compare with the *robust box* model that is a robust model (3) under a box uncertainty set ( $\mathcal{U}_B$ ). The solutions to the robust box model can also be obtained by solving (10) with  $\underline{\phi} = \underline{\phi}^0$  and  $\bar{\phi} = \bar{\phi}^0$  by setting  $\gamma_{uv}$  in the  $\mathcal{U}_{SB}$  uncertainty set to a large value (it suffices to set  $\gamma_{uv} = 1$ ) for all  $u, v \in T$ . Thus, the robust box case can be viewed as a special (simple) case of our suggested approach.

Optimal solutions of (10) (with  $\mathcal{U}_{SB}$ ) are computed for a range of uncertainty set parameter values,  $\delta = 0.02, 0.04, \dots, 0.14$  and  $\gamma = 0.01, \dots, 0.05$ , and a range of homogeneity parameter values,  $\mu = 1.0625, 1.0750, \dots, 1.4000$ . Note that for some combinations of uncertainty set and homogeneity parameters, formulation (10) has  $x=0$  as the only feasible solution; such solutions are omitted from the graphs. For the box uncertainty set ( $\mathcal{U}_B$ ), the solutions were computed with the same values for  $\delta$ , however, a larger range of homogeneity parameters,  $\mu = 1.0625, 1.0750, \dots, 1.4750$ , is explored to be able to obtain nonzero solutions for all  $\delta$  values. Moreover, in order to obtain a more accurate value of  $\mu$  below which there is no nonzero dose solution, for each  $(\gamma, \delta)$  pair, the  $\mu$  values between the first nonzero dose solution and the last zero dose solution were refined, first in increments of 0.0025, then in increments of 0.0005, and finally in increments of 0.0001. To evaluate a plan involving beamlet decision variable vector  $x^*$  under uncertainty set  $\mathcal{U}$ , which could be a singleton when considering the nominal case, empirical measures of worst-case dose and homogeneity are defined as follows

$$\hat{d} = \hat{d}(x^*, \mathcal{U}) = \min_{\hat{\phi} \in \mathcal{U}} \min_{v \in T} \hat{\phi}_v d_v(x^*) \quad \text{and}$$

$$\hat{\mu} = \hat{\mu}(x^*, \mathcal{U}) = \max_{\hat{\phi} \in \mathcal{U}} \max_{u, v \in T} \frac{\hat{\phi}_u d_u(x^*)}{\hat{\phi}_v d_v(x^*)};$$

$\hat{d}$  and  $\hat{\mu}$  are used as abbreviations when  $x^*$  and  $\mathcal{U}$  are unambiguous in the given context.

Before comparing the performance of solutions that are optimal to (10) under different scenarios, it should be noted that the optimal solution value of (10) decreases in the uncertainty set parameters  $\delta$  and  $\gamma$  (see Figure 13 in Appendix D). Note that combinations of  $\mu$ ,  $\delta$ , and  $\gamma$  resulting in a zero beamlet intensity configuration ( $x = \mathbf{0}$ ) are omitted from all figures and are interpreted as “infeasible” plans.

A “cost-of-robustness” can be measured by evaluating a robust solution in the nominal scenario where there is no uncertainty. Figures 5 and 6 illustrate this phenomenon. The nominal formulation (1) corresponds to setting  $\delta=0$  and  $\gamma$  sufficiently large in  $\mathcal{U}_{SB}$ ; for the particular dataset being considered (Patient 4), the nominal formulation is equivalent to a setting of  $\delta=0$  and  $\gamma \geq 0.04$  in formulation (10). On the other hand, uncertainty sets with  $\gamma \leq 0.03$  exclude the nominal scenario. Consequently, models with smaller  $\gamma$  values tend to perform poorly, at least in terms of the minimum biological dose, in the nominal scenario. Figure 5 shows the performance in the nominal scenario, for  $\delta = 0.14$ , with different values of  $\gamma$  and  $\mu$  in (10) (a curve is presented for each value of  $\gamma$ , obtained by varying the values of  $\mu$ ). In general, the “cost of robustness” appears more evident as  $\gamma$  decreases: the minimal  $\hat{\mu}$  that is attained with a nonzero dose solution increases, and  $\hat{d}$  somewhat decreases. This can be explained by the fact that the nominal scenario is excluded from the uncertainty set when  $\gamma \leq 0.03$ , and the distance between this scenario and the uncertainty set increases as  $\gamma$  decreases. The figure also shows that the robust box solution and the robust spatially bound solution with  $\gamma = 0.04$  exhibit fairly similar performance in the nominal scenario, but the robust box appears less stable and demonstrates a notable drop in dose for biological homogeneity values slightly larger than 1.1. Note that a lack of stability stems from the fact that the measured  $\hat{\mu}$  and  $\hat{d}$  values typically differ from the values of  $\mu$  and  $\hat{d}$ , respectively, that are used in optimizing (10). Consequently, even slight changes in the parameter  $\mu$  may in some cases induce a different optimal solution that is associated with significantly different values of the measured  $\hat{\mu}$  and  $\hat{d}$ . In this case, when increasing  $\mu$  from 1.3603 to 1.3607, the nominal performance of the robust box model solution improves with respect to both homogeneity and dose, from  $\hat{\mu} = 1.105$  and  $\hat{d} = 37.4$  to  $\hat{\mu} = 1.1$  and  $\hat{d} = 55.1$ .

Figure 6 shows nominal performance curves for  $\gamma = 0.02, 0.04, 1$ ,  $\delta = 0.02, 0.04, \dots, 0.14$ , and a range of  $\mu$  values (each curve is obtained by varying the value of  $\mu$ ). For each of the  $\gamma$  values, the “cost of robustness” becomes more evident as  $\delta$  increases: the curves for larger  $\delta$  values appear to be nearly dominated in the  $\hat{\mu} - \hat{d}$  space by the curves obtained for smaller  $\delta$  values. The robust box results appear much less stable than the corresponding SR model results for  $\gamma \in \{0.02, 0.04\}$ , especially for values of  $\hat{\mu} \leq$

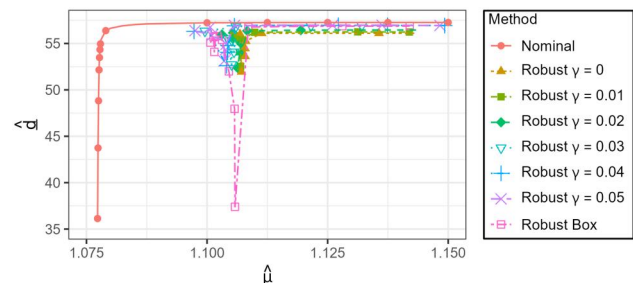
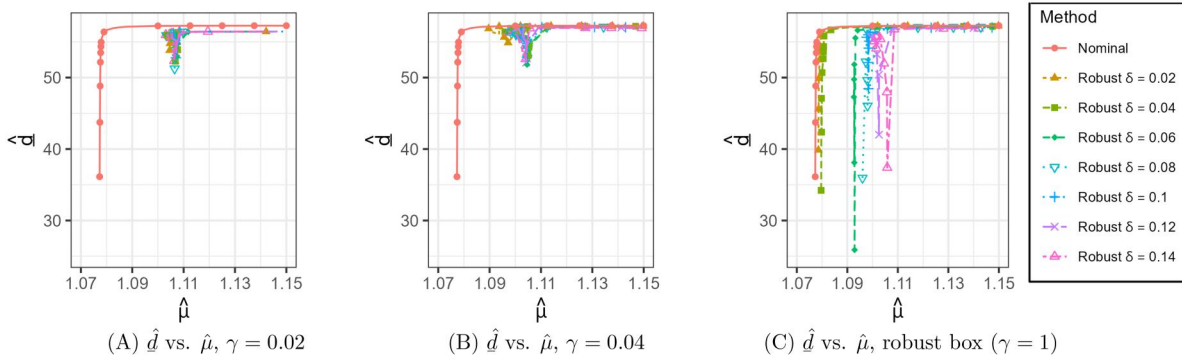
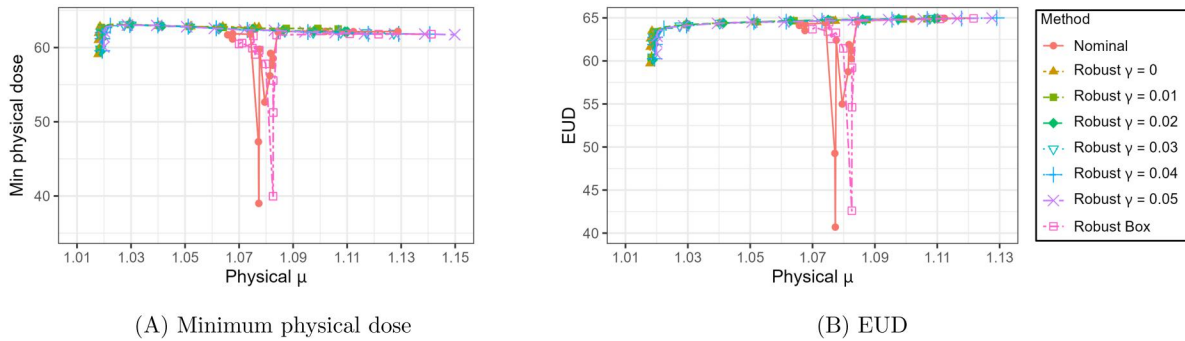


Figure 5. Nominal biologically-adjusted performance of solutions optimal to model (10) for  $\delta = 0.14$  and different values of  $\gamma$ , compared to the nominal solution.



**Figure 6.** Nominal biologically-adjusted performance of solutions optimal to models (1) and (10) (for  $\gamma = 0.02$ ,  $\delta = 0$  is infeasible.)



**Figure 7.** Physical performance of solutions optimal to (10) with  $\delta = 0.14$  and a range of  $\gamma$  values, compared with the nominal solution.

1.11, with doses 25 Gy lower than the optimal nominal solution compared to at most 5 Gy lower for  $\gamma = 0.02$  and 0.04. Recall that solutions of (10) with  $\gamma = 0.02$  are not designed to protect against the nominal realization (which is excluded from the uncertainty sets with  $\gamma < 0.04$ ). Hence, the solutions obtained in the case where  $\gamma = 0.02$  have a biologically-adjusted dose of approximately 0.81 Gy less than the corresponding optimal nominal dose for all  $\hat{\mu} > 1.11$ . When increasing  $\gamma$  to 0.04, the uncertainty set captures the nominal solution and so the dose obtained for  $\hat{\mu} \geq 1.11$  become equal to those of the nominal model, albeit the minimal  $\hat{\mu}$  for which there was a nonzero dose is higher than the nominal model, even for small values of  $\delta$ . This trend persists when moving to the robust box model, which results in almost optimal nominal performance for  $\delta \leq 0.04$ .

Next, we consider the physical dose performance of the obtained solutions. Although physical performance measures do not account for the variable radiosensitivity, and planning based on such measures may not utilize the information that has been recently made available by advanced biomarkers such as FMISO, they are still important as the ongoing and current basis for treatment guidelines and clinical practice. Keeping in mind that physical performance is not explicitly maximized in the optimization formulations (1) or (10), it may be interesting to examine how the resulting planning solutions fare with respect to the physical performance measures. Figure 7 compares solutions of (10) with  $\delta = 0.14$  and different values of  $\gamma$ , the robust box solution (also with  $\delta = 0.14$ ), and the nominal solution. Figure 7(a) depicts the physical homogeneity vs. minimal physical dose of the different models, and Figure 7(b) shows the EUD (Equivalent Uniform Dose). The EUD measure was

introduced as a single number representing the dose throughout the PTV (Niemierko, 1997; Choi and Deasy, 2002), and is given by the formula  $EUD = (\frac{1}{|T|} \sum_{v \in T} d_v^\alpha)^\frac{1}{\alpha}$ , where, as in Nohadani and Roy (2017) we use  $\alpha = -10$ . Surprisingly, both the optimal nominal solution and the robust box solution exhibit poor performance compared with the solutions of SR model (10) with respect to both physical homogeneity and physical minimal PTV dose, for physical  $\mu$  values in the interval (1.07, 1.10). Specifically, the solutions of the nominal and robust box models are less stable and, with both physical dose and EUD up to 20 Gy lower than for  $\gamma \leq 0.05$ , and a minimal physical homogeneity value of around 1.07 compares to 1.02 for  $\gamma \leq 0.05$ . Similar behavior can be observed for other values of  $\delta$  tested; see Figure 15 in Appendix D. A possible explanation for this behavior is that under both the nominal and the box uncertainty plans, the dose is allocated in a way that is focused mainly on the PTV voxels with the lowest radiosensitivities (following the max-min formulation). The spatially bound uncertainty may smooth out the PTV radiosensitivity uncertainty, resulting in a worst case that considers a greater number of voxels, including voxels for which the radiosensitivity is nominally higher, thus increasing the minimal physical dose.

Figures 8 and 9 display the physical dose obtained by the nominal solution, the robust box solution with  $\delta = 0.14$ , and the robust spatial solution of (10) with  $(\delta, \gamma) = (0.14, 0.04)$ . For each model, the value of the homogeneity parameter  $\mu$  is chosen so that it yields the highest minimal physical dose while maintaining physical homogeneity less than 1.1 (which is a standard clinical requirement). All three

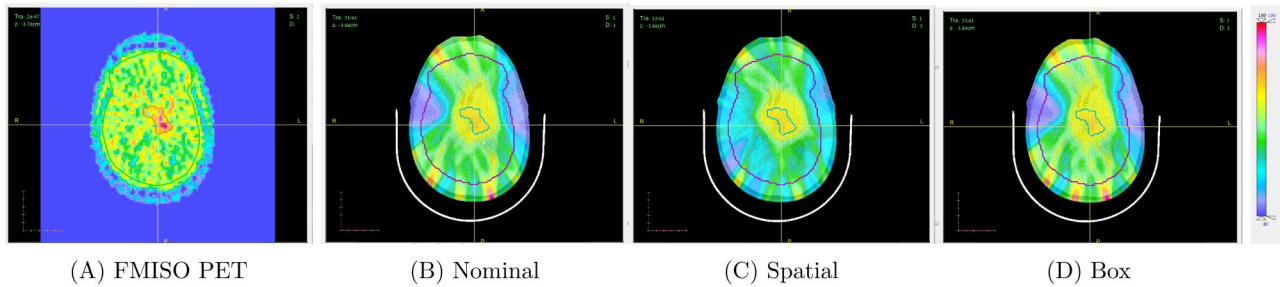


Figure 8. (a) – PET, pink and red levels indicating low oxygen. (b)–(d) – physical dose.

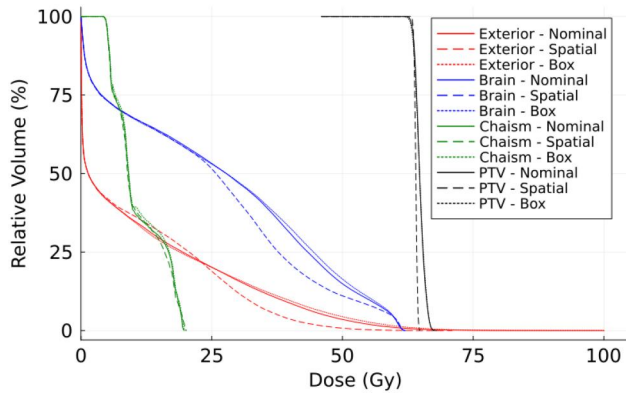


Figure 9. Cumulative dose–volume histograms.

dose images in Figure 8(b)–8(d) are aligned with the same PET/CT FMISO scan slice 8(a). The robust box and nominal solutions apply wider physical dose ranges to the PTV, 61.5–67.5 Gy for the robust box and 62–68 Gy for the nominal, compared to the robust spatial solution, with a range of 63–65 Gy. Furthermore, the nominal and robust box solutions create dose “hot spots” of approximately 100 Gy in the perimeter of the exterior OAR, while the robust spatial model solution applies a maximal physical dose of less than 71 Gy to this OAR. This may be due to the fact that the nominal and robust box solutions use fewer beams, with some of the beams set to a higher intensity, than the robust spatial solution. Thus, while the robust spatial model retains approximately the same adjusted dose as the nominal solution, it accounts for the variability and uncertainty in the radiosensitivities. It also appears to avoid excessively high OAR doses.

Finally, to evaluate the cost of misspecifying the uncertainty set parameters, Figures 10 and 11 depict the worst-case performance of the various solutions under the assumption that the correct uncertainty models differ with respect to the parameter values assumed in formulation (10). In Figure 10 the robust solutions use the “true value” of  $\delta = 0.1$ , so the misspecification is only in terms of  $\gamma$ . Specifically,  $\gamma = 0.02$  and  $\gamma = 0.04$  in Figure 10(a) and Figure 10(b), respectively. Note that while the worst-case performance deteriorates as the value of  $\gamma$  used to derive the solution deviates from the true value, the dose declines by at most (and is generally much less than) 1.1 Gy. Importantly, the spatially robust models may achieve homogeneity values that are close to the homogeneity of the correct model, improving the homogeneity by more than 0.05 relative to

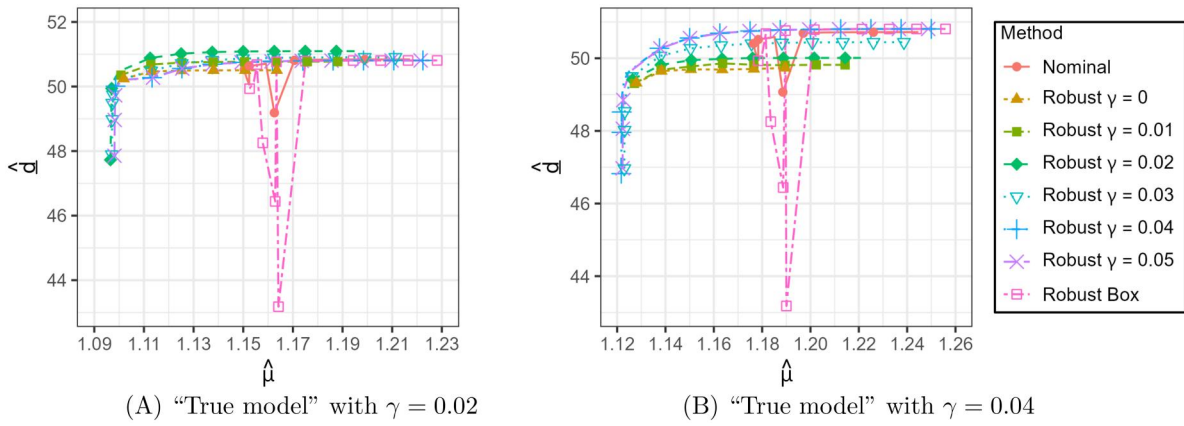
the nominal and robust box solutions. This is evident for most values of  $\delta$  (see Figures 16 and 17 in Appendix D). Moreover, although the robust box and nominal model appear to attain a similar biologically-adjusted dose as the true model for high values of  $\hat{\mu}$ , the performance of these models appears much less stable for desirable (lower) values of  $\hat{\mu}$ . Notably, the robust box obtains doses significantly lower than all models considered including the nominal one for some values of  $\hat{\mu}$ . Hence, even when  $\gamma$  is misspecified in (10), it may still yield nearly the same homogeneity,  $\hat{\mu}$ , perhaps at a slight cost to  $\hat{d}$ .

Figure 11 depicts experimental results for evaluating the effect of misspecifying  $\delta$  in the uncertainty model. In this set of experiments, the true uncertainty set, as well as the uncertainty sets assumed in solving (10), all share the same value of  $\gamma = 0.04$ , but differ in the values of  $\delta$ . While the misspecification of  $\delta$  seems to cause only a modest decrease in the value of  $\hat{d}$  (of up to 0.5 Gy) compared with the value obtained under the true  $\delta$ , while the homogeneity,  $\hat{\mu}$ , appears to be more significantly affected. Specifically, underestimating  $\delta$  increases  $\hat{\mu}$  of up to 0.03 (for a similar dose), while overestimating  $\delta$  may lead to little or no improvement in the homogeneity (a small decrease in  $\hat{\mu}$ ). Further, the  $\hat{\mu}$  values attained for the nominal model are significantly worse (larger than the spatially robust models by up to 0.06) for the higher true value of  $\delta = 0.14$ , compared to being approximately 0.03 larger than all spatially robust models when the true  $\delta = 0.02$ .

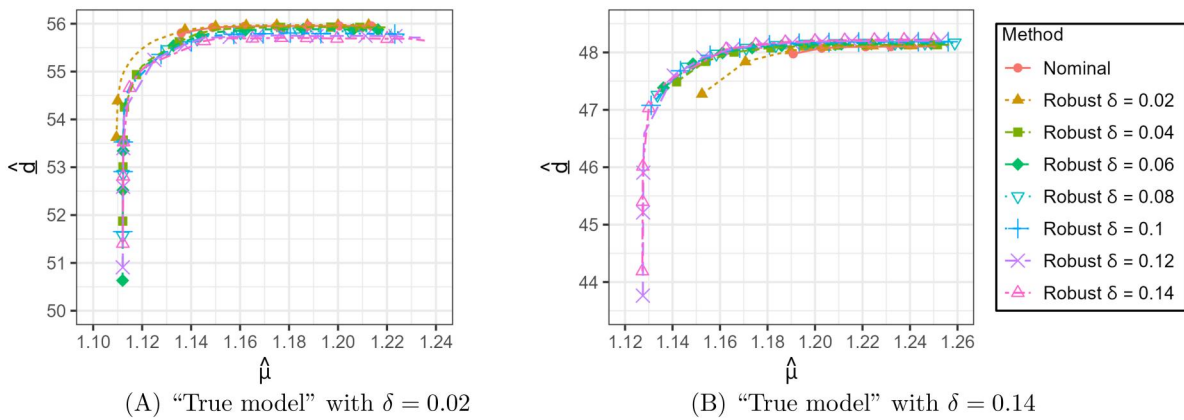
### 5.3. Dose-volume experiments

In this section, we conduct experiments using Algorithm 2, introduced in Section 4 for handling dose–volume constraints. The experiments in this segment were run using Julia 1.3.0, and Gurobi solver 9.0.0, on an Intel Xeon Gold 6254 CPU with a 3.10 GHz base frequency, 24.75 MB cache, and 376 RAM.

In order to speed up the computation, we may first narrow down the search interval further by performing a preliminary bisection search, and then proceed by running Algorithm 2 given this reduced interval. Note that invoking the bisection search may sacrifice some of the theoretical properties guaranteed by the initial lower bound, thus serving only as an effective heuristic. Further, in order to deal with possible degeneracy in Algorithm 2 we determine adjacency of basic feasible solutions only based on the  $x$  and  $y$  variables. In these experiments, we focus on the dose–volume constraint for the



**Figure 10.** Worst-case biologically-adjusted performance  $(\hat{\mu}, \hat{d})$  of solutions optimal to model (10) with misspecified  $\gamma$ , under the assumption that the “true”  $\mathcal{U}_{SB}$  has  $\delta = 0.1$  and  $\gamma$  as indicated. The performance of the nominal solution is also shown for comparison.



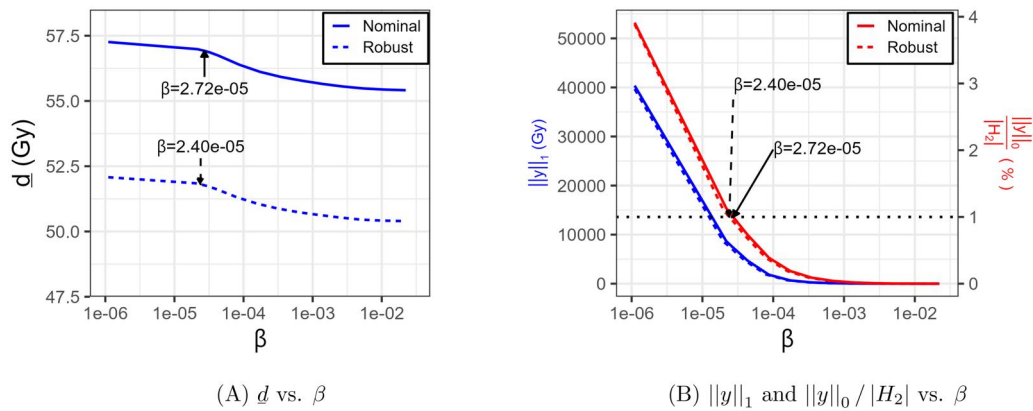
**Figure 11.** Worst-case biologically-adjusted performance  $(\hat{\mu}, \hat{d})$  of solutions optimal to (10) with misspecified  $\delta$ , under the assumption that the “true”  $\mathcal{U}_{SB}$  has  $\gamma = 0.04$  and  $\delta$  as indicated.

brain (OAR 2 of Patient 4). Specifically, following the ICRU guidelines (Anon, 2010), 99% of the brain should receive less than  $\bar{d}_2 = 60$  Gy while  $\bar{d}_2 = 62$  Gy; that is,  $\alpha = 0.01$  in constraint (11f). For these experiments we consider both the nominal and spatially robust settings with the same parameters as in Section 5.1. Recall that Algorithm 2 determines the minimal  $\beta$  value for which the solution satisfies this constraint. Intuitively, a small  $\beta$  and corresponding penalty term in the objective should correspond to a small change in the objective compared with the non-penalized problem (where  $\beta = 0$ ). Thus, the constraint is expected to be satisfied with a slight decline in the optimal objective value relative to that achieved when  $\beta = 0$ .

Figure 12 depicts the value of the adjusted dose  $\hat{d}$ , and the  $\ell_1$  and  $\ell_0$  norms of the deviation vector  $y$ , as a function of  $\beta$ , for the different values of  $\beta$  obtained by the algorithm for both cases. In order to run the algorithm, we first computed  $\hat{\beta} \approx 2.22 \times 10^{-2}$  for the nominal case, and  $\hat{\beta} \approx 2.58 \times 10^{-2}$  for the specially robust case. We used  $\hat{\beta} = 0$  in both cases, while the true lower bounds are  $\beta \approx 1.73 \times 10^{-6}$  for the nominal setting and  $\beta \approx 1.36 \times 10^{-6}$  for the robust setting. The algorithm determined that  $\beta^* = 2.72 \times 10^{-5}$  and  $\beta^* = 2.4 \times 10^{-5}$  are the smallest  $\beta$  values that satisfy  $\|y\|_0 \leq \lfloor \alpha |H_2| \rfloor = 13,601$  for the nominal and the spatially robust case, respectively (thus satisfying the dose-volume constraint).

We compared the solutions obtain by Algorithm 2 in both of the spatially robust and nominal settings, with a solution obtained from problem (3) with an additional CVAR constraint, as suggested by Romeijn *et al.* (2003). Table 4 shows these results, comparing also with the unrestricted case, where the dose-volume constraints are not imposed (having  $\beta = 0$  in (13)).

Algorithm 2 only reduces the dose by 0.59% and 0.52% compared with the unrestricted case, in the nominal and robust settings, respectively. Also, note that CVaR results in lower values of both minimal biologically-adjusted dose and EUD, compared with Algorithm 2. This is especially evident in the nominal setting. This is due to the CVaR constraint being a safe approximation of the dose-volume constraint, resulting in more conservative solutions to the planning problem. Indeed, the percent of brain voxels exceeding the  $\bar{d}_2 = 60$  Gy bound is approximately 0.4% using the CVaR approach, which is lower than the required 1%. Finally, the CVaR constraint requires a variable for each of the brain voxels, which may be computationally intensive; without initializing to the unrestricted solution, we found that CVaR would not solve even after 72 hours. With this particular initialization, we are able to solve for CVaR but it takes about 1.5 times as long as Algorithm 2, as displayed in Table 4. Thus, when Algorithm 2 can be applied, it provides an alternative to the CVaR approach



**Figure 12.** Minimum biologically-adjusted dose  $\underline{d}$  and deviation vector  $y$  norms vs.  $\beta$  values enumerated by Algorithm 2 for the nominal and spatially robust settings.

**Table 4.** Comparison of Algorithm 2 with adding a CVaR constraint.

Setting	Method	$\underline{d}$	EUD	% Violation	Running Time (sec)
Nominal	Unrestricted	57.26	64.85	3.91	2904
	Algorithm 2	56.92	64.50	1.00	55,784
	CVaR	56.46	64.03	0.41	87,192
Robust	Unrestricted	52.08	64.73	3.89	4327
	Algorithm 2	51.81	64.33	1.00	68,980
	CVaR	51.38	63.80	0.41	105,366

that is both superior in terms of performance and in terms of the running time.

## 6. Conclusions and future work

We proposed a novel spatially based uncertainty set to model uncertainty in tumor radiosensitivity of radiation treatment based on measured biomarkers. Naively incorporating this uncertainty set in robust voxel-based FMO models would lead to an unmanageable increase in the already large problem size. Alternatively, we compactly reformulate the robust problem by proving that at most a few uncertainty set extreme points need to be considered for each constraint. While theoretically tractable, the new compact formulation may still contain over a billion constraints. We first show that although the compact formulation cannot be solved directly, it can be solved within a reasonable running time and consumption of memory, by carefully generating constraints as needed. Extensive experiments demonstrate advantages of the proposed uncertainty set in terms of the adjusted homogeneity–dose tradeoff over the nominal and robust box-uncertain models.

In this work, we also develop a method for satisfying a single dose–volume constraint that outperforms the standard CVaR modeling approach to this problem. Although handling dose volume constraints remains computationally challenging for clinical use, this can be addressed by future research focused on voxel aggregation techniques for solving large scale instances within reasonable running times.

This work has focused on spatial aspects of biomarker-related uncertainty in a voxel-level planning model (also known as “dose painting”). In future work, we would like to extend the proposed model to include temporal aspects,

taking into account fractionation of treatments as well as a possible additional image scan midtreatment. This leads to a challenging two-stage adaptive robust problem, where the spatial uncertainty changes over time. Finally, it is worth noting that the study of robust and adaptive planning, under the condition that the uncertainty of neighboring elements is connected, has application beyond radiation therapy. Communicable disease, traffic congestion, and even agriculture are all examples of situations where uncertainty among neighbors in space and time is connected.

## Acknowledgments

The authors are grateful to Aditya P. Apte for his help in using the CERR software to process and export the imaging datasets used in this paper. Noam Goldberg acknowledges the computational resources provided by the Data Science Institute at Bar-Ilan University. Finally, the authors would like to thank the area editor and anonymous referees for many comments and corrections that improved the quality of the manuscript.

## Notes on contributors

**Noam Goldberg** is a senior lecturer in the Department of Management at Bar-Ilan University and an incoming associate professor at the Department of Industrial Engineering and Management at Ben-Gurion University. He completed a PhD in operations research at Rutgers University (RUTCOR), New Brunswick, NJ, in 2010. His graduate studies followed an early non-academic system and software engineering career in telecommunications. His PhD research on sparse data classification models was awarded a first prize in a contest of the INFORMS NJ chapter in 2009. He held several postdoctoral positions during 2009–2013 at the Technion-Israel Institute of Technology, Argonne National Laboratory and Carnegie Mellon University. He joined Bar-Ilan University in 2013 as a tenure-track lecturer, where he became a tenured senior lecturer since 2018. In fall 2019, he was a visiting associate professor at the Rutgers Business School. He has more than 30 peer reviewed publications including top journals in optimization and conference proceedings in computer science. In current work he continues to be motivated by high-impact applications of OR, data-driven optimization, game models and solution methods.

**Mark Langer, MD** is Professor of Clinical Radiation Oncology at Indiana University School of Medicine, Indianapolis, IN USA. His research interests lie in mathematical programming and optimization for radiation treatment planning and design, and clinical strategies to improve outcomes in head and neck cancers. He is the recipient of grants from the NIH, NSF and the Indiana 21st Century Fund and has

served on review panels for the NIH, NSF, BSF and the DOD. He served as chair for the Operations Research and Radiotherapy Collaborative Working Group (ORART) sponsored by the NIH and was a subawardee on the ORART Collaborative Working Group's Tool Kit program funded by the NSF. He has authored publications in *Ann. Operations Research*, *Mathematical Programming B*, *Linear Algebra and its Applications*, *J. Operat. Res. Soc.*, *Journal of Mechanisms and Robotics*, *Journal of Applied Clinical Medical Physics*, *Dysphagia*, *Int. J. Radiat. Oncol. Biol. Phys.*, *Radiotherapy and Oncology*, *IIE Transactions on Health Care Systems Engineering*, *Phys. Med. Biol.*, and *Medical Physics*.

**Shimrit Shtern** is a senior lecturer in the Faculty of Data and Decision Sciences at the Technion - Israel Institute of Technology. She completed a PhD in optimization at the Technion in 2015. She was a post-doctoral fellow at the Operation Research Center at MIT in 2015-2017 and joined the Technion as a faculty member in 2018. Her research focuses on theory and applications of optimization under uncertainty as well as development and convergence analysis of first order methods for structured optimization problems. Her work has been published in top operation research and optimization journals including *Management Science*, *Operations Research*, *Mathematics of Operations Research*, *SIAM Journal on Optimization*, and *Mathematical Programming*.

## ORCID

Noam Goldberg  <http://orcid.org/0000-0002-1340-7569>

Mark Langer  <http://orcid.org/0000-0002-1234-1908>

Shimrit Shtern  <http://orcid.org/0000-0001-8398-7178>

## Data availability statement

All data used in this paper is based on data available from public repositories whose references are indicated and with processing as described in Section 5. The processed data and code are available at <https://github.com/ShimShtern/RobustRTPStatic>.

## References

- Allen Li, X., Alber, M., Deasy, J.O., Jackson, A., Ken Jee, K.-W., Marks, L.B., Martel, M.K., Mayo, C., Moiseenko, V., Nahum, A.E., *et al.* (2012) The use and qa of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPM. *Medical Physics*, **39**(3), 1386–1409.
- International Commission on Radiation Units and Measurements. (2010) Report 83: Prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT). *Journal of the International Commission on Radiation Units and Measurements*, **10**(1), 1–106.
- Ben-Tal, A., El Ghaoui, L. and Nemirovski, A. (2009) *Robust Optimization*. Princeton University Press, Princeton, N.
- Ben-Tal, A., Goryashko, A., Guslitzer, E. and Nemirovski, A. (2004) Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, **99**(2), 351–376.
- Bentzen, S.M. and Gregoire, V. (2011) Molecular imaging-based dose painting: A novel paradigm for radiation therapy prescription, in *Seminars in Radiation Oncology*, volume 21, Elsevier, pp. 101–110.
- Bertsimas, D. and Sim, M. (2004) The price of robustness. *Operations Research*, **52**(1), 35–53.
- Bortfeld, T., Chan, T.C., Trofimov, A. and Tsitsiklis, J.N. (2008) Robust management of motion uncertainty in intensity-modulated radiation therapy. *Operations Research*, **56**(6), 1461–1473.
- Chassein, A., Dokka, T. and Goerigk, M. (2019) Algorithms and uncertainty sets for data-driven robust shortest path problems. *European Journal of Operational Research*, **274**(2), 671–686.
- Choi, B. and Deasy, J.O. (2002) The generalized equivalent uniform dose function as a basis for intensity-modulated treatment planning. *Physics in Medicine & Biology*, **47**(20), 3579.
- Chu, M., Zinchenko, Y., Henderson, S.G. and Sharpe, M.B. (2006) Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Physics in Medicine & Biology*, **50**(23), 5463.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L. and Prior, F. (2013) The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, **26**, 1045–1057.
- Corazza, P. (1999) Introduction to metric-preserving functions. *The American Mathematical Monthly*, **106**(4), 309–323.
- 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**(1).
- Deasy, J.O. (1997) Multiple local minima in radiotherapy optimization problems with dose-volume constraints. *Medical Physics*, **24**(7), 1157–1161.
- Donoho, D.L. and Elad, M. (2003) Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization. *Proceedings of the National Academy of Sciences*, **100**(5), 2197–2202.
- Ferris, M.C., Lim, J. and Shepard, D.M. (2003) Radiosurgery treatment planning via nonlinear programming. *Annals of Operations Research*, **119**(1), 247–260.
- Hill, R.P., Bristow, R.G., Fyles, A., Koritzinsky, M., Milosevic, M. and Wouters, B.G. (2015) Hypoxia and predicting radiation response. *Seminars in Radiation Oncology*, **25**(4), 260–272.
- Jeyakumar, V., Li, G., Woolnough, D. and Wu, H. (2023) Affinely adjustable robust optimization for radiation therapy under evolving data uncertainty via semi-definite programming. *Optimization*, **73**(6), 1807–1832.
- Jones, L., Hoban, P. and Metcalfe, P. (2001) The use of the linear quadratic model in radiotherapy: A review. *Australasian Physics & Engineering Sciences in Medicine*, **24**, 132–146.
- Kataria, T., Sharma, K., Subramani, V., Karrthick, K. and Bisht, S.S. (2012) Homogeneity index: An objective tool for assessment of conformal radiation treatments. *Journal of Medical Physics/Association of Medical Physicists of India*, **37**(4), 207–213.
- Khairi, Y., Omer, H., Sulieman, A., Deiab, N., Mokhtar, M.H., Abolaban, F.A., Alkhorayef, M. and Bradley, D. (2021) Radiation dose homogeneity and critical organs in radiotherapy treatment of prostate cancer. *Radiation Physics and Chemistry*, **178**, 109000.
- Li, H., Bissonnette, J.-P., Purdie, T. and Chan, T.C. (2015) Robust PET-guided intensity-modulated radiation therapy. *Medical Physics*, **42**(8), 4863–4871.
- Lorca, A. and Sun, X.A. (2014) Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind. *IEEE Transactions on Power Systems*, **30**(4), 1702–1713.
- Lubin, M. and Dunning, I. (2015) Computing in operations research using Julia. *INFORMS Journal on Computing*, **27**(2), 238–248.
- Luedtke, J., Ahmed, S. and Nemhauser, G.L. (2010) An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, **122**(2), 247–272.
- McMahon, S.J. (2018) The linear quadratic model: Usage, interpretation and challenges. *Physics in Medicine & Biology*, **64**(1), 01TR01.
- Niemierko, A. (1997) Reporting and analyzing dose distributions: A concept of equivalent uniform dose. *Medical Physics*, **24**(1), 103–110.
- Nohadani, O. and Roy, A. (2017) Robust optimization with time-dependent uncertainty in radiation therapy. *IISE Transactions on Healthcare Systems Engineering*, **7**(2), 81–92.
- Pflugfelder, D., Wilkens, J. and Oelfke, U. (2008) Worst case optimization: A method to account for uncertainties in the optimization of intensity modulated proton therapy. *Physics in Medicine & Biology*, **53**(6), 1689–1700.

- Preciado-Walters, F., Langer, M.P., Rardin, R.L. and Thai, V. (2006) Column generation for IMRT cancer therapy optimization with implementable segments. *Annals of Operations Research*, **148**, 65–79.
- Preciado-Walters, F., Rardin, R., Langer, M. and Thai, V. (2004) A coupled column generation, mixed integer approach to optimal planning of intensity modulated radiation therapy for cancer. *Mathematical Programming*, **101**, 319–338.
- Romeijn, H.E., Ahuja, R.K., Dempsey, J.F. and Kumar, A. (2006) A new linear programming approach to radiation therapy treatment planning problems. *Operations Research*, **54**, 201–216.
- Romeijn, H.E., Ahuja, R.K., Dempsey, J.F., Kumar, A. and Li, J.G. (2003) A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Physics in Medicine & Biology*, **48**(21), 3521.
- Roy, A., Dabadghao, S.S. and Marandi, A. (2022) Value of intermediate imaging in adaptive robust radiotherapy planning to manage radio-resistance. *Annals of Operations Research*, Advance online publication. <https://doi.org/10.1007/s10479-022-04699-z>
- Saka, B., Rardin, R.L. and Langer, M.P. (2014) Biologically guided intensity modulated radiation therapy planning optimization with fraction-size dose constraints. *Journal of the Operational Research Society*, **65**, 557–571.
- Shang, H., Pu, Y., Chen, Z., Wang, X., Yuan, C., Jin, X. and Liu, C. (2021) Impact of multiple beams on plan quality, linear energy transfer distribution, and plan robustness of intensity modulated proton therapy for lung cancer. *ACS Sensors*, **6**(2), 408–417.
- Ten Eikelder, S., Ferjančić, P., Ajdari, A., Bortfeld, T., den Hertog, D. and Jeraj, R. (2020) Optimal treatment plan adaptation using mid-treatment imaging biomarkers. *Physics in Medicine & Biology*, **65**(24), 245011.
- Titz, B. and Jeraj, R. (2008) An imaging-based tumour growth and treatment response model: investigating the effect of tumour oxygenation on radiation therapy response. *Physics in Medicine & Biology*, **53**(17), 4471–4488.
- Toma-Dasu, I., Uhrdin, J., Antonovic, L., Dasu, A., Nuyts, S., Dirix, P., Haustermans, K. and Brahme, A. (2012) Dose prescription and treatment planning based on FMISO-PET hypoxia. *Acta Oncologica*, **51**(2), 222–230.
- Unkelbach, J., Alber, M., Bangert, M., Bokrantz, R., Chan, T.C., Deasy, J.O., Fredriksson, A., Gorissen, B.L., Van Herk, M., Liu, W., *et al.* (2018) Robust radiotherapy planning. *Physics in Medicine & Biology*, **63**(22), 22TR02.
- Unkelbach, J., Chan, T.C. and Bortfeld, T. (2007) Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Physics in Medicine & Biology*, **52**(10), 2755.
- Yan, D., Vicini, F., Wong, J. and Martinez, A. (1997) Adaptive radiation therapy. *Physics in Medicine & Biology*, **42**(1), 123.