



## Fast and insightful bi-objective optimization for prostate cancer treatment planning with high-dose-rate brachytherapy



Ngoc Hoang Luong <sup>a,\*</sup>, Tanja Alderliesten <sup>b</sup>, Bradley R. Pieters <sup>b</sup>, Arjan Bel <sup>b</sup>, Yury Niatsetski <sup>c</sup>, Peter A.N. Bosman <sup>a</sup>

<sup>a</sup> Centrum Wiskunde & Informatica (CWI), Life Sciences and Health, Amsterdam, The Netherlands

<sup>b</sup> Amsterdam UMC, University of Amsterdam, Department of Radiation Oncology, Amsterdam, The Netherlands

<sup>c</sup> Elekta, Veenendaal, The Netherlands

### HIGHLIGHTS

- Challenges of the brachytherapy planning problem for prostate cancer treatment.
- Accuracy and efficiency issues in multi-objective prostate brachytherapy planning.
- A multi-resolution scheme to accelerate the optimization of brachytherapy planning.
- A multi-core CPU parallelization of our approach in brachytherapy planning.
- Comparing solutions created by our approach with clinically approved treatment plans.

### ARTICLE INFO

#### Article history:

Received 14 March 2019

Received in revised form 20 July 2019

Accepted 2 August 2019

Available online 8 August 2019

#### Keywords:

Multi-objective optimization

Accuracy

Performance

Multi-resolution scheme

Brachytherapy

Radiotherapy

Prostate cancer treatment planning

### ABSTRACT

**Purpose:** Prostate high-dose-rate brachytherapy (HDR-BT) planning involves determining the movement that a high-strength radiation stepping source travels through the patient's body, such that the resulting radiation dose distribution sufficiently covers tumor volumes and safely spares nearby healthy organs from radiation risks. The Multi-Objective Real-Valued Gene-pool Optimal Mixing Evolutionary Algorithm (MO-RV-GOMEA) has been shown to be able to effectively handle this inherent bi-objective nature of HDR-BT planning. However, in clinical practice there is a very restricted planning time budget (often less than 1 h) for HDR-BT planning, and a considerable amount of running time needs to be spent before MO-RV-GOMEA finds a good trade-off front of treatment plans (about 20–30 min on a single CPU core) with sufficiently accurate dose calculations, limiting the applicability of the approach in the clinic. To address this limitation, we propose an efficiency enhancement technique for MO-RV-GOMEA solving the bi-objective prostate HDR-BT planning problem.

**Methods:** Dose-Volume (DV) indices are often used to assess the quality of HDR-BT plans. The accuracy of these indices depends on the number of dose calculation points at which radiation doses are computed. These are randomly uniformly sampled inside target volumes and organs at risk. In available HDR-BT planning optimization algorithms, the number of dose calculation points is fixed. The more points are used, the better the accuracy of the obtained results will be, but also the longer the algorithms need to be run. In this work, we introduce a so-called *multi-resolution scheme* that gradually increases the number of dose calculation points during the optimization run such that the running time can be substantially reduced without compromising on the accuracy of the obtained results.

**Results and conclusion:** Experiments on a data set of 18 patient cases show that with the multi-resolution scheme, MO-RV-GOMEA can achieve a sufficiently good trade-off front of treatment plans after five minutes of running time on a single CPU core (4–6 times faster than the old approach with a fixed number of dose calculation points). When the optimization with the multi-resolution scheme is run on a quad-core machine, five minutes are enough to obtain trade-off fronts that are nearly as good as those obtained by running optimization with the old approach in one hour (i.e., 12 times faster). This leaves ample time to perform the selection of the preferred treatment plan from the trade-off front for the specific patient at hand. Furthermore, comparisons with real clinical treatment plans, which were manually made by experienced BT planners within 30–60 min, confirm that the plans obtained by our approach are superior in terms of DV indices. These results indicate that our proposed approach has the potential to be employed in clinical practice.

© 2019 Elsevier B.V. All rights reserved.

\* Corresponding author.

E-mail address: [Hoang.Luong@cwi.nl](mailto:Hoang.Luong@cwi.nl) (N.H. Luong).

## 1. Background

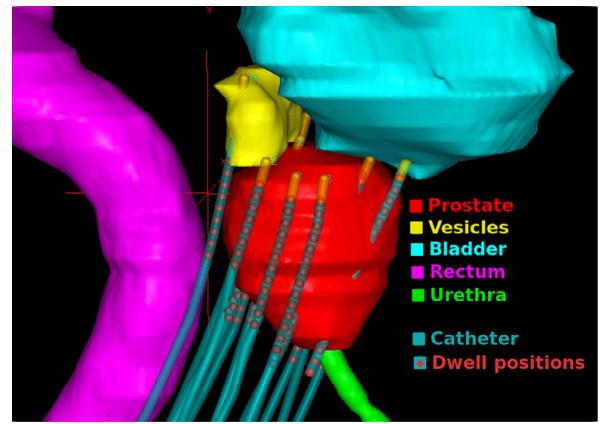
### 1.1. High-dose-rate brachytherapy

Radiotherapy (RT) involves the use of radiation to sterilize cancerous cells, preventing them from further growing and dividing [1]. While external-beam RT (EBRT) uses radioactive beams that target tumor regions from the outside of the patient's body, brachytherapy (BT) uses radiation sources that are delivered to the vicinity of tumor regions and irradiate from the inside of the patient's body. High-dose-rate (HDR) BT [2,3] is a cancer treatment modality, in which radiation sources of high strength are passed through the patient's body in a certain movement that yields a radiation dose distribution that is deemed clinically suitable for the specific patient. The local nature of HDR-BT helps the resulting dose distribution better conform to treatment targets, reducing radiation risks to surrounding healthy organs.

We here focus on HDR-BT planning for prostate cancer, but the methodology can be modified for application to other tumor locations, e.g., cervix, head and neck, or breast. In prostate HDR-BT, prostate and seminal vesicles are the typical treatment targets while bladder, rectum, and urethra are the organs at risk (OARs). An implant, consisting of a number, typically 14–20, of catheters, is inserted to the treatment targets through the patient's transperineal skin in the operation room. At our clinic (Amsterdam UMC), the implant is fixated, and the patient is transferred to the imaging room where 3D images (computed tomography or magnetic resonance imaging scans) of the patient's pelvic area are acquired. Next, these are loaded into a BT treatment planning system. Treatment targets and OARs are delineated, and the inserted catheters are reconstructed on the acquired medical images by the planners. In each catheter, a number of so-called *dwell positions* are defined, where the radiation source can reside for a certain amount of time (i.e., *dwell time*) when passing through the catheter. Fig. 1 shows an example of a prostate HDR-BT implant. The longer the dwell time at a dwell position, the more radiation dose will be delivered from that dwell position toward surrounding tissues. Different dwell time settings result in different radiation dose distributions, and BT planners need to determine how dwell times should be configured to yield a radiation dose distribution that is desired for the patient. An HDR-BT plan can thus be seen as a list of dwell times associated with the dwell positions. A plan that is approved by BT planners (i.e., radiation oncologists, RT technologists, and medical physicists) will be used to treat the patient. The inserted catheters, attached to flexible cables, are then connected to a so-called *afterloader* machine that contains a radiation source and controls the movement of that source through the connected catheters as specified in the approved plan. The afterloader retrieves the source from the catheters after the treatment is completed.

### 1.2. Clinical protocol

Radiation oncologists specify the HDR-BT *planning-aim dose*, which is a radiation dose level that is considered sufficient to treat the tumor under concern. The planning-aim dose for prostate HDR-BT at the hospital involved in this study (Amsterdam UMC, location AMC) is 13 Gy. The utopian plan, in which all treatment targets are sufficiently treated while no radiation is delivered to OARs, does not exist because the BT source irradiates the surrounding tissues from inside in all directions. In current practice, a treatment plan is iteratively adapted (i.e., often manually fine-tuned) and assessed based on a clinical protocol, which consists of *planning aims* that indicate how treatment plans should be developed. Planning aims are typically criteria for the so-called Dose-Volume (DV) indices:



**Fig. 1.** An example implant of catheters in HDR-BT treatment for prostate cancer.

**Table 1**

Prostate HDR-BT clinical protocol at Amsterdam UMC (location AMC).

Prostate	Bladder	Rectum	Urethra	Vesicles
$V_{100} > 95\%$	$D_{1cm^3} < 86\%$	$D_{1cm^3} < 78\%$	$D_{0.1cm^3} < 110\%$	$V_{80} > 95\%$
$V_{150} < 50\%$	$D_{2cm^3} < 74\%$	$D_{2cm^3} < 74\%$		
$V_{200} < 20\%$				

- $V_d^o$  index: the cumulative volume of an organ  $o$  that is covered by the radiation dose level  $d$ .
- $D_v^o$  index: the radiation dose level that covers the most-irradiated cumulative volume  $v$  of an organ  $o$ .

DV criteria specify the *aspiration values* for DV indices, which are often presented in relative terms to the planning-aim dose and organ volumes. For example, the criterion  $V_{100}^{prostate} > 95\%$  indicates that the cumulative prostate volume that receives at least 100% of the planning-aim dose is aspirated to be more than 95% of the total prostate volume. The criterion  $V_{200}^{prostate} < 20\%$  indicates that the cumulative prostate volume that receives at least 200% of the planning-aim dose be less than 20% of the total prostate volume (in order to prevent necrosis, i.e., unnatural cell death). Because seminal vesicles are considered to harbor microscopic disease, the criterion  $V_{80}^{vesicles} > 95\%$  indicates that 80% of the planning-aim dose should cover at least 95% of the seminal vesicles volume. Also, to avoid over-irradiation of the bladder, the criterion  $D_{1cm^3}^{bladder} < 86\%$  indicates that the most-irradiated 1 cm<sup>3</sup> of the bladder volume receive less than 86% of the planning-aim dose. Table 1 shows the DV criteria of the prostate HDR-BT clinical protocol currently employed at the Amsterdam UMC (location AMC).

The radiation dose distribution of a treatment plan is often approximated based on a sufficiently large number of random *dose calculation points*. Let  $D$  be the set of  $n_D$  dose calculation points and  $T$  be the set of all  $n_T$  dwell positions. Regarding the strength and the shape of the radiation source employed for the treatment, we can define an  $n_D \times n_T$  dose rate matrix  $\mathbf{R}$ , in which each dose rate  $R_{ij}$  is the amount of radiation released per second (i.e., Gy/s) to the dose calculation point  $i \in D$  from the radiation source when it resides at dwell position  $j \in T$ . Let  $\mathbf{t} = (t_1, t_2, \dots, t_{n_T})$  describe a treatment plan, i.e., a vector of dwell times at  $n_T$  dwell positions. This treatment plan  $\mathbf{t}$  yields a specific radiation dose distribution  $\mathbf{d} = (d_1, d_2, \dots, d_{n_D})$ , i.e., a vector of radiation dose at  $n_D$  dose calculation points, as follows.

$$\mathbf{d} = \mathbf{R}\mathbf{t} \quad (1)$$

where each  $d_i$  is the total dose received at dose calculation point  $i$  from all dwell positions. Let  $D_o$  be the set of  $n_D^o$  dose calculation

points inside organ  $o$ ,  $D_o \subset D$ ,  $|D_o| = n_D^o$ . Correspondingly, let  $\mathbf{d}_o$  be the dose distribution in organ  $o$ , i.e., a vector of radiation dose at  $n_D^o$  dose calculation points in organ  $o$ . An indicator function  $\chi(d_i, d)$  is defined as:

$$\chi(d_i, d) = \begin{cases} 1 & d_i \geq d \\ 0 & d_i < d \end{cases} \quad (2)$$

The value of a DV index  $V_d^o$  can then be computed as:

$$V_d^o = \frac{1}{n_D^o} \sum_{i \in D_o} \chi(d_i, d) \quad (3)$$

The value of a DV index  $D_v^o$  can be computed as:

$$D_v^o = \mathbf{d}_{\lfloor \frac{v}{v^o} \rfloor}^{o,s} \quad (4)$$

where  $\mathbf{d}^{o,s}$  is the descending sorted version of  $\mathbf{d}^o$ , and  $v^o$  is the volume (in  $\text{cm}^3$ ) represented by each dose calculation point inside organ  $o$ .

An HDR-BT plan normally consists of dwell times of the radiation source at about 100–300 dwell positions. The making of a plan is carried out between catheter insertion and the actual irradiation. To reduce the inconvenience for the patient, the planning time (not including the delineation of the organs and the implant) should be as short as possible, and should not exceed 1 h. It is not a trivial task to determine values for a few hundred dwell times within such a limited time budget so that the resulting radiation dose distribution (almost) satisfies all DV criteria in the clinical protocol. Therefore, HDR-BT planners employ computer software to assist the planning process. However, there exist certain computational challenges that need to be properly addressed.

### 1.3. Computational challenges in HDR-BT planning and related works

#### 1.3.1. Discontinuity

Computing the value of a DV index ( $V_d^o$  or  $D_v^o$ ) involves counting the number of dose calculation points where the radiation dose exceeds a certain value. The discrete nature of the counting introduces discontinuities in the search landscape of any optimization problem formulation that is based on DV criteria. Solving these problems by (classical) mathematical programming methods result in Mixed Integer Linear Programming (MILP) models, which could require many hours/days to obtain optimal solutions [4]. In available BT planning software, the problem is often simplified by employing linear models (solved by simulated annealing [5]) or by employing quadratic models (solved by the gradient-based Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [6]). The so-called *linear-penalty* formulation in [5] can also be solved by linear programming [7]. It has been shown that there exists only a weak correlation between the objective values of plans optimized under linear-penalty models and the quality of DV indices of those plans that are used to assess the treatment plan quality in clinical practice [8]. The optimal solutions of such simplified models, therefore, are not guaranteed to satisfy the original DV criteria-based models. In practice, BT planners need to run available software many times with different penalty factors (i.e., effectively creating new optimization models) or need to perform *graphical optimization* (i.e., directly modifying the spatial radiation dose distribution, and thereby adjusting dwell times) until satisfactory plans are obtained [4]. A few researches [9–11] have worked toward the automation of solving these linear models or quadratic models with multiple optimization runs, in which each run has a different setting of penalty factors, resulting in multiple treatment plans (300 in [9,10] and 1000 in [11]). The obtained plans are then filtered to identify the ones that satisfy all the DV criteria in the clinical protocol. However, such an approach still suffers from certain limitations. In particular:

- Because the optimization models are not based on DV indices, the obtained plans can still be further improved with respect to DV indices.
- It might happen that none of the obtained plans satisfies the clinical protocol. The more DV criteria are considered, the higher the probability that such a situation might happen. For example, the protocol considered in [9–11] has only four DV criteria while the one at our clinic has nine DV criteria (see Table 1). Different rounds of multiple optimization runs then need to be performed again until satisfactory plans are obtained.

Recently, our preliminary work [12] has shown that algorithms of the Gene-pool Optimal Mixing Evolutionary Algorithm (GOMEA) family can exploit the dependency structure among decision variables of the HDR-BT planning problem to efficiently handle DV criteria without the need for model approximations.

#### 1.3.2. Multi-objective optimization problem

When making a treatment plan, BT planners aim to maximize the coverage of the planning-aim dose in target volumes while minimizing undesired irradiation in OARs. BT planning, therefore, is an inherent multi-objective optimization problem. However, if each DV criterion is treated as a separate objective, such a clinical protocol, like the one in Table 1, will result in a many-objective problem (9 in our case). Such a many-objective optimization problem is not only challenging and time-consuming to solve, its trade-off results are also difficult to visualize and comprehend [13]. Several optimization models with 3–4 objectives for a simpler clinical protocol that contains only 3–4 DV criteria were proposed in [14], where the well-known Non-dominated Sorting Genetic Algorithm II (NSGA-II [15]) was employed as the solver. However, the problem modeling used in [14] does not scale for more complicated protocols that contain many more DV criteria (e.g., see Table 1), and it is not straightforward for BT planners to determine which treatment plans from the resulting trade-off front satisfy all criteria in the protocol at hand. In practice, optimization tools in BT planning software often handle the multi-objective aspect of HDR-BT planning by weighted-sum approaches, in which all DV indices are simplified and then aggregated together according to some weight vector, resulting in a single objective function to be optimized [5,6]. Such weighted-sum approaches can only yield a single treatment plan each time, which is not guaranteed to exhibit the desired trade-off between *target coverage* and *organ sparing* for the patient case under concern. If BT planners are not satisfied with the resulting plan, its dwell times must be re-optimized (by running optimization tools again with different weight vectors or performing manual graphical optimization), typically incurring 30–60 min of the planning time (at our clinic). A review on optimization for HDR-BT planning can be found in [16].

In our previous work [12], a bi-objective formulation was proposed with the *maximin* approach to effectively model the clinical protocol of 9 objectives (Table 1). Resulting trade-off fronts from solving the model are easy to visualize and investigate, intuitive to interpret, and no DV indices need to be omitted. The model is also shown to be more efficiently solved by the Multi-Objective Real-Valued GOMEA (MO-RV-GOMEA [17]), compared to many common MOEAs (i.e., NSGA-II [15], MOEA/D [18], and MAMAL-GaM [19]) [12]. A key reason for why the real-valued version of GOMEA (RV-GOMEA [20]) can outperform state-of-the-art real-valued EAs, such as Natural Evolution Strategy (NES [21]) or Covariance Matrix Adaptation Evolution Strategy (CMA-ES [22]), is that RV-GOMEA can straightforwardly employ partial evaluations, i.e., the fact that the leverage impact of changing only a few variables of a solution can be computed much more efficiently

to update its objective values than performing a full function evaluation. In addition, a clinical observer study performed on the results of using MO-RV-GOMEA to solve the bi-objective problem formulation using 20,000 dose calculation points (see next subsection) for 18 prostate cancer patients and with 3 physicians, showed that in 98% of cases, a MO-RV-GOMEA generated plan was preferred to a clinically generated plan using currently available methods [23].

### 1.3.3. The number of dose calculation points

The accuracy of computing DV index values depends on the number of dose calculation points. The more points are used, the more accurate DV index values are obtained, but also the more computation time is incurred. In the BT planning software Oncentra Brachy (Elekta AB, Stockholm, Sweden), the treatment plan assessment module (in the default setting) computes DV indices based on 500,000 points (i.e., 100,000 points per organ). The optimization modules of the software, however, employ much smaller sets of dose calculation points [5,6], about 100 times fewer, during each optimization run for the sake of speed. Mathematical programming approaches also employ similar amounts of dose calculation points to keep the models solvable by classical optimization methods [4,7]. In our earlier work [12], a larger set of dose calculation points (i.e., 20,000 points) was chosen to be used for the multi-objective optimization of HDR-BT planning but no analysis was provided to support the reason for this choice. The results in [12], obtained after 1 h of running MO-RV-GOMEA, were deemed to be good approximations of the Pareto trade-off front and contain clinically desirable solutions [23]. However, the whole time budget of 1 h for planning should not be used entirely for optimization and the optimization time should be minimized as much as possible without compromising too much on the quality of the plans. It would be favorable if the optimization can be run for at most 5 min, and the remaining time could then be allocated to the investigation of trade-off plans and the selection (and adaptation, if required) of the preferred treatment plan. However, as shown in [12], there exist substantial differences between the non-dominated fronts obtained after 10 min and the ones after 1 h. This indicates that further research needs to be done to reduce the amount of computing time.

### 1.4. Purpose & contributions

The purpose of our work is to develop an efficient computational technique to accelerate the performance of MO-RV-GOMEA on solving the bi-objective HDR-BT planning problem such that the approach is sufficiently efficient to be used in clinical practice.

Using clinical data, we first conduct experiments to show the need to use large sets of dose calculation points. We then propose a so-called multi-resolution scheme for adapting the size of the set of dose calculation points such that the optimization time can be substantially reduced without compromising on the quality of the resulting trade-off fronts. Finally, we show that the performance of our method can be further improved with straightforward parallelization of the MO-RV-GOMEA solver by using a multi-core CPU system that is nowadays commonly available in modern desktops.

The remainder of the article is organized as follows. In Section 2, we present our bi-objective formulation of the prostate HDR-BT planning problem. In Section 3, we give a brief description of the MO-RV-GOMEA for the self-containedness of this article, and propose the operation of the multi-resolution scheme. In Section 4, we describe the experiment settings and, in Section 5, we present and discuss the experimental results. Finally, we conclude the article in Section 6.

## 2. Problem formulation

The DV indices in Table 1 are divided into two groups:

- Target Coverage:  $V_{100}^{\text{prostate}}$  and  $V_{80}^{\text{vesicles}}$
- Organ Sparing:  $V_{200}^{\text{prostate}}$ ,  $V_{150}^{\text{prostate}}$ ,  $D_{1\text{cm}3}^{\text{rectum}}$ ,  $D_{2\text{cm}3}^{\text{rectum}}$ ,  $D_{1\text{cm}3}^{\text{bladder}}$ ,  $D_{2\text{cm}3}^{\text{bladder}}$ ,  $D_{0.1\text{cm}3}^{\text{urethra}}$

We exclude the two indices  $V_{150}^{\text{prostate}}$  and  $V_{200}^{\text{prostate}}$  from the Organ Sparing group and we use their criteria as hard constraints in our optimization model, i.e., candidate treatment plans must satisfy the two constraints  $V_{150}^{\text{prostate}} < 50\%$  and  $V_{200}^{\text{prostate}} < 20\%$  to be considered as feasible solutions. While we want to irradiate the prostate volume with the planning-aim dose as much as possible, we need to restrict the coverage of excessive dose levels (e.g., 150% and 200% of the planning-aim dose). An advantage of excluding  $V_{150}^{\text{prostate}}$  and  $V_{200}^{\text{prostate}}$  is that these two  $V_d^o$  indices have different units (in  $\text{cm}^3$  or percentage of the prostate volume) compared to  $D_v^o$  indices (in Gy or percentage of the planning-aim dose). Ensuring DV indices of a group having the same measure unit is beneficial for employing the maximin approach to formulate the two objective functions as follows. Note that it might not always be sensible to set some DV criteria as hard constraints. In such cases, a  $V_d^o$  index can be converted into a corresponding  $D_v^o$  index, and vice versa (see [24] for an example where all  $D_v^o$  indices in Table 1 are converted to  $V_d^o$  indices.).

For each  $V_d^o$  index in the Target Coverage group, the clinical protocol has a corresponding lower-bound aspiration value  $V_d^{o,\min}$ . The higher the value of such a  $V_d^o$  index than this aspiration value, the better the coverage of the required radiation dose level  $d$  on this organ  $o$ . The Least Coverage Index (LCI) of a candidate HDR-BT plan is defined as:

$$LCI(\mathbf{t}) = \min\{\delta_v(V_{100}^{\text{prostate}}), \delta_v(V_{80}^{\text{vesicles}})\} \quad (5)$$

where  $\delta_v(V_d^o) = V_d^o - V_d^{o,\min}$  measures the deviation of the  $V_d^o$  index from the aspiration value  $V_d^{o,\min}$ .

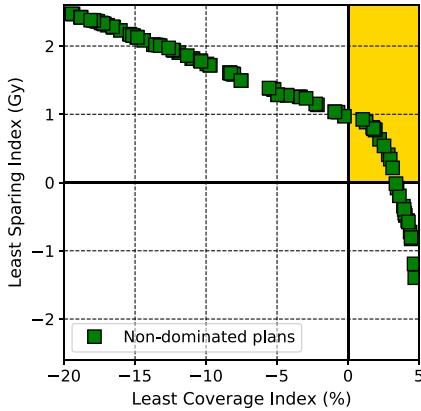
For each  $D_v^o$  index in the Organ Sparing group, the clinical protocol has a corresponding upper-bound aspiration value  $D_v^{o,\max}$ . The lower the value of such a  $D_v^o$  index than this aspiration value, the better the sparing of organ  $o$ . The Least Sparing Index (LSI) of a candidate HDR-BT plan is defined as:

$$\begin{aligned} LSI(\mathbf{t}) \\ = \min\{\delta_d(D_{1\text{cm}3}^{\text{bladder}}), \delta_d(D_{2\text{cm}3}^{\text{bladder}}), \delta_d(D_{1\text{cm}3}^{\text{rectum}}), \delta_d(D_{2\text{cm}3}^{\text{rectum}}), \delta_d(D_{0.1\text{cm}3}^{\text{urethra}})\} \end{aligned} \quad (6)$$

where  $\delta_d(D_v^o) = D_v^{o,\max} - D_v^o$  measures the deviation of the  $D_v^o$  index from the aspiration value  $D_v^{o,\max}$ .

In general, maximizing LCI corresponds to optimizing the coverage of the planning-aim dose on target volumes while maximizing LSI corresponds to optimizing the sparing of OARs from radiation risks. These two objectives are conflicting with each other and we thus have a bi-objective optimization model for the prostate HDR-BT planning problem.

Fig. 2 shows the (LCI, LSI) objective space of the bi-objective optimization model of the prostate HDR-BT planning problem based on the clinical protocol in Table 1. The LCI axis is calibrated in terms of percentage of a target volume. A plan  $\mathbf{t}$  with  $LCI(\mathbf{t}) = 0$  indicates that  $V_{100}^{\text{prostate}} = 95\%$  and  $V_{80}^{\text{vesicles}} \geq 95\%$ , or  $V_{100}^{\text{prostate}} \geq 95\%$  and  $V_{80}^{\text{vesicles}} = 95\%$ . A plan  $\mathbf{t}$  with  $LCI(\mathbf{t}) < 0$  indicates that  $V_{100}^{\text{prostate}} < 95\%$  or  $V_{80}^{\text{vesicles}} < 95\%$ , i.e., at least one target volume is under-irradiated according to its criteria. A plan  $\mathbf{t}$  with  $LCI(\mathbf{t}) > 0$  indicates that  $V_{100}^{\text{prostate}} > 95\%$  and  $V_{80}^{\text{vesicles}} > 95\%$ , i.e., all target volumes are irradiated sufficiently. The LSI axis is calibrated in terms of the radiation dose (Gy unit), which can be converted by  $LSI(\mathbf{t}) \times$  planning-aim dose, because such absolute values are



**Fig. 2.** The (LCI,LSI) objective space with an example trade-off front of HDR-BT treatment plans. Treatment plans in the golden-colored corner satisfy all DV criteria of the clinical protocol.

more convenient for planners to see the amount of excessive dose above an OAR sparing aspiration value. A plan  $\mathbf{t}$  with  $LSI(\mathbf{t}) = 0$  indicates that at least one sparing DV index is at its corresponding aspiration value while the other indices satisfy their criteria. A plan  $\mathbf{t}$  with  $LSI(\mathbf{t}) < 0$  indicates that at least one sparing DV index exceeds its aspiration value, i.e., an organ is over-irradiated. A plan  $\mathbf{t}$  with  $LSI(\mathbf{t}) > 0$  satisfies all DV criteria in the Organ Sparing group. Consequently, a plan  $\mathbf{t}$  with  $LCI(\mathbf{t}) > 0$  and  $LSI(\mathbf{t}) > 0$  satisfies all DV criteria of the clinical protocol. In Fig. 2, if a plan resides in the so-called *golden corner* of the (LCI,LSI) objective space, that plan satisfies the clinical protocol. Note that it is unnecessary to look for treatment plans that result in excessive over-irradiation of OARs or excessive under-irradiation of target volumes. Therefore, besides the two constraints that are presented above for  $V_{150}^{\text{prostate}}$  and  $V_{200}^{\text{prostate}}$ , we employ two additional constraints  $LCI(\mathbf{t}) \geq -0.2$  and  $LSI(\mathbf{t}) \geq -0.2$ .

It is not always possible to obtain treatment plans in the golden corner. The geometry of the inserted catheter implant and the organs determine the quality of the trade-off front of treatment plans that can be achieved. If the implant does not allow golden-corner plans to be achieved, a treatment plan that slightly violates some DV criteria can be chosen. Furthermore, BT planners might also approve a plan outside the golden corner for treatment if the trade-off between target coverage and organ sparing exhibited by that plan is more beneficial for the specific patient under concern, e.g., regarding the patient's health conditions. Such information, however, is difficult to be formulated for optimization. We argue that it would be more beneficial to solve the bi-objective model of HDR-BT planning presented here to obtain the best possible trade-off front of treatment plans. BT planners can then investigate the trade-off front to locate the plan that exhibits the desired trade-off. Such an *a posteriori* approach to the multi-objective HDR-BT planning can be facilitated by MO-RV-GOMEA, presented in Section 3.

### 3. Optimization algorithm, multi-resolution scheme, and multi-core CPU parallelization

#### 3.1. MO-RV-GOMEA

The bi-objective problem formulation for prostate HDR-BT planning can be more efficiently solved by MO-RV-GOMEA [17] than many common MOEAs [12]. A more detailed outline of MO-RV-GOMEA can be found in the Appendix A.

MO-RV-GOMEA employs an adaptive elitist archive [25] to keep track of a pre-defined number of non-dominated solutions

obtained during the run because elitism is important for the convergence of MOEAs [26,27]. In every generation, the selection set  $S$ , that contains solutions with the best non-domination ranks selected from the current population  $P$ , is clustered into  $k$  clusters of equal size. This clustering procedure has been shown to enhance the probability of obtaining an evenly-spread non-dominated front [19]. Variation proceeds by processing each solution in the population and restricting mating to the cluster that the solution pertains to. One offspring is created for each population member. Specifically, for each cluster, sets of dependent variables are identified. For each set, new values are generated jointly, for each solution, by sampling a normal distribution that was estimated for these variables from the solutions in the same cluster. Each of these partial modifications is only accepted if they result in an improvement over the previous state; otherwise, the changes are undone. The problem structure of HDR-BT planning allows the impact of such local changes to be efficiently computed to update the objective values of the solutions. Such *partial evaluations* can be straightforwardly employed by MO-RV-GOMEA due to the genetic-local-search-like nature of the GOM operator.

The Interleaved Multi-start Scheme (IMS [28,29]) is employed to operate multiple populations of increasing sizes asynchronously. For every  $b = 8$  generations of population  $P_i$ , the subsequent population  $P_{i+1}$  (with  $|P_{i+1}| = 2 \times |P_i|$ ) is run for one generation. All populations are kept running until the computing budget is over or when satisfying results are obtained. Practitioners thus do not need to determine the population size, which is an important EA control parameter that is notoriously difficult to set, especially for non-trivial real-world applications.

#### 3.2. Multi-resolution scheme

If we consider the number of dose calculation points as a control parameter of the algorithm, its setting can be automated by incorporating it in the IMS as follows. The first population  $P_1$  starts with a small set of dose calculation points  $D$ . In this article, we employ  $n_D = |D| = 5000$ , i.e.,  $n_D^o = 1000$  for each organ  $o \in \{\text{prostate, vesicles, bladder, rectum, urethra}\}$ . When each subsequent population  $P_i$  is initialized, the set of dose calculation points is re-generated and doubled in size, i.e.,  $n_D \leftarrow 2 \times n_D$ . Note that, while each population operates independently, the same set of dose calculation points is used to compute the objective values of all individuals of all populations and the elitist archive members. Therefore, each time the set of dose calculation points is re-generated, elitist archive members and all individuals in (non-terminated) populations need to be re-evaluated with respect to the new set. Fig. 3 shows the pseudocode for the multi-resolution scheme together with the IMS.

#### 3.3. Multi-core CPU parallelization

Modern desktop computers are often equipped with multi-core CPUs, which are beneficial to accelerate evolutionary algorithms, especially when solution evaluations are non-trivial. When the evaluation of each individual is independent from each other, solution evaluations can be straightforwardly parallelized, i.e., multiple individuals can be evaluated at the same time, depending on the number of cores available. In MO-RV-GOMEA, every population member is transformed by GOM into an offspring solution through a series of  $2l - 1$  (partial) modifications, where each corresponds to the variables indicated by a linkage set (see Appendix A4). Note that we do not apply parallelization in this *horizontal* manner because each step in GOM depends on the result of the previous step. Instead, the multi-core parallel computing can be implemented in a *vertical* manner as follows. Each LT of every cluster has exactly  $2l - 1$

MULTI-RESOLUTION SCHEME WITH IMS	
$n_D = 5000$	
1	$n_D \leftarrow 5000$
2	$D \leftarrow \text{GENERATESETODECALCULATIONPOINTS}(n_D)$
3	$\mathbf{P}_1 \leftarrow \text{INITIALIZEPOPULATION}(N_1, D)$
4	$\text{generations}[1] \leftarrow 0$
5	$\text{max\_population\_index} \leftarrow 1$
6	$i \leftarrow 1$
7	<b>while</b> $\text{WITHINALLOWEDCOMPUTINGBUDGET}()$ <b>do</b>
8	<b>EXECUTEONEGENERATION</b> ( $\mathbf{P}_i, D$ )
9	$\text{generations}[i] \leftarrow \text{generations}[i] + 1$
10	<b>if</b> $\text{generations}[i] \bmod b = 0$ <b>then</b>
11	$i \leftarrow i + 1$
12	<b>if</b> $i > \text{max\_population\_index}$ <b>then</b>
13	$n_D \leftarrow 2 \times n_D$
14	$D \leftarrow \text{GENERATESETODECALCULATIONPOINTS}(n_D)$
15	$\text{REEVALUATEALLPOPULATIONSANDELITISTARCHIVE}(D)$
16	$N_i \leftarrow 2 \times N_{i-1}$
17	$\mathbf{P}_i \leftarrow \text{INITIALIZEPOPULATION}(N_i, D)$
18	$\text{generations}[i] \leftarrow 0$
19	$\text{max\_population\_index} \leftarrow i$
20	<b>else</b>
21	$i \leftarrow 1$

**Fig. 3.** Multi-resolution scheme with the interleaved multi-start scheme.

linkage sets; thus, each linkage set can be indexed from the 1st to the  $(2l - 1)$ th linkage set. We traverse from the 1st index to the  $(2l - 1)$ th index. At every index  $i$ th, each population member  $\mathbf{x}$  is modified at the variables indicated by the  $i$ th linkage set of the LT associated with the cluster that  $\mathbf{x}$  belongs to. After all population members are partially modified, solution evaluations are performed for all these partially-altered individuals. Multi-core parallelization can be straightforwardly applied to accelerate these solution evaluations.

#### 4. Experiment settings

In this work, we perform experiments using a desktop computer with the configuration Intel(R) Core(TM) i7-3770S CPU 3.10 GHz. Our data set contains 18 anonymized prostate HDR-BT patient cases from the Amsterdam UMC. For each patient case, the BT treatment planning software Oncentra Brachy (research version 4.6, Elekta AB, Stockholm, Sweden) is used to acquire the organ contours and the catheter delineations, and to activate dwell positions that should be considered for treatment planning. In each inserted catheter, the first dwell position starts at the 5.0 mm offset from the tip of the catheter, and the succeeding dwell positions are activated with a step size of 2.5 mm. Dwell positions within the margin of 5.0 mm extended from the surfaces of the targets (i.e., prostate and seminal vesicles) and outside the margin of 1.0 mm extended from the surface of the urethra are activated. Only these active dwell positions are considered in the planning phase. We randomly uniformly generate 100,000 dose calculation points per organ (i.e., prostate, seminal vesicles, rectum, bladder, and urethra). In total, we have a full set of 500,000 dose calculation points for each patient, offering the same accuracy as in the BT treatment planning software. For each patient case, a full dose rate matrix  $\mathbf{R}$  that contains the dose rates from active dwell positions to these 500,000 points can be created according to the TG-43 dose calculation formalism (the American Association of Physicists in Medicine AAPM Task Group No. 43 Report) [30–32].

As explained in Section 1.3.3, available optimization tools do not use such full sets of 500,000 dose calculation points, but employ much smaller sets instead (e.g., about 5000 points in the case of prostate HDR-BT) for the sake of speed. Therefore, we first

perform experiments with several fixed numbers of dose calculation points: 2500, 5000, 10,000, 20,000, and 40,000, respectively. For each patient case and each number of dose calculation points, we carry out 30 independent runs of MO-RV-GOMEA solving the bi-objective optimization model. The set of dose calculation points is randomly uniformly sampled at the beginning of each run, and the running time is set at 1 h. The status of the elitist archive is logged at every five minutes during the run, which is used to compute the hypervolume development graph. All obtained results are then re-evaluated based on the full set of 500,000 dose calculation points.

Second, we perform experiments with the multi-resolution scheme. For each patient case, 30 independent runs of MO-RV-GOMEA are carried out, and each run is assigned a running time of 1 h. In each run, the number of dose calculation points starts at 5000 points, which is generally similar to the literature. Each time a new population is initialized (as in the IMS), the dose calculation point set doubles its size and is re-generated. The objective values of all the individuals of preceding populations and the elitist archive are re-evaluated with respect to the new dose calculation point set.

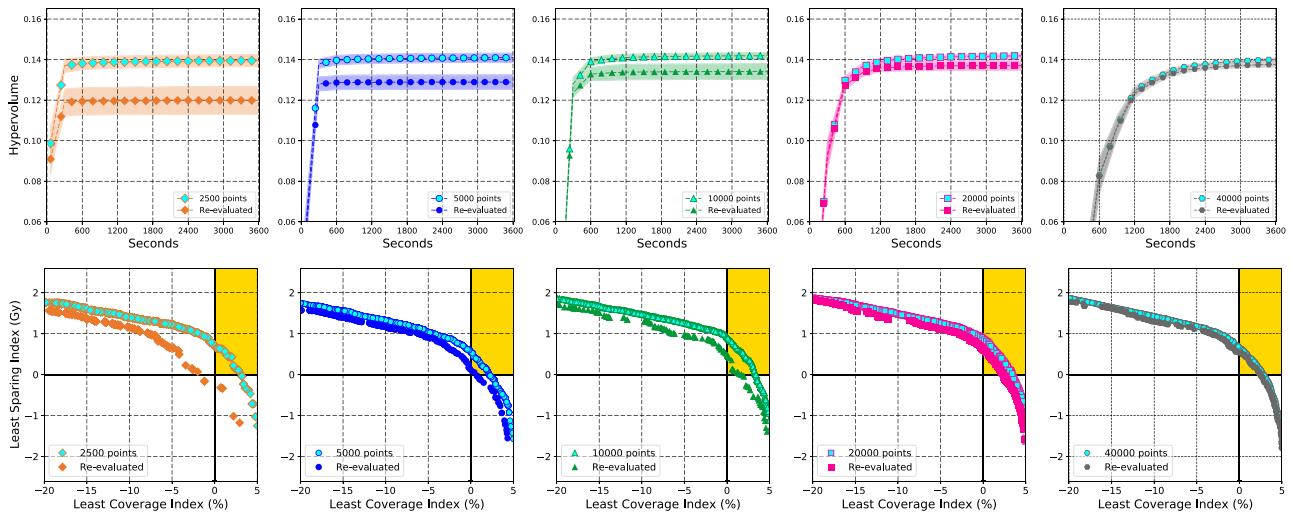
The Mann–Whitney–Wilcoxon statistical hypothesis test was employed with  $p < \alpha = 0.05$  to test whether the hypervolume result obtained by one setting of the dose calculation point set is statistically significantly different from that of another setting.

## 5. Results and discussion

### 5.1. Fixed sets of dose calculation points

**Fig. 4** shows the hypervolume development graphs and the final approximation fronts (that have the median hypervolume values) for Patient 1 from our data set. The results of other cases can be found in the Supplementary Material. For each patient, optimization runs are performed based on five settings of the dose calculation point numbers: 2500, 5000, 10,000, 20,000, and 40,000 points, respectively, and the states of the elitist archives (i.e., non-dominated solutions) at different time points are used to compute the hypervolume values. Optimization runs using small sets (i.e., 2500 or 5000 points) appear to obtain better hypervolume results in the first 5–10 min of the runs compared to when larger sets are used. After 10–20 min (for 2500–20,000 points) or 40–60 min (for 40,000 points), the hypervolume results appear to be the same and, after 1 h, the obtained approximation fronts are quite similar to each other regardless of the numbers of dose calculation points. However, these results are computed based on sets of dose calculation points with different sizes; therefore, they should not be directly compared with each other.

In order to put all results on an equal footing, we re-evaluate the objective values of all (non-dominated) solutions found during the runs using the same full set of 500,000 dose calculation points and then compute the hypervolume again. The more accurate re-evaluations show that using small sets of dose calculation points considerably compromises on the accuracy of the results. After re-evaluation, only the approximation fronts from the runs using 10,000–40,000 points still contain treatment plans in the golden corner while the fronts from the runs using 2500 points do not reach the golden corner anymore. The runs using 5000 points contains very few treatment plans in the golden corner after re-evaluation. Considering that golden-corner solutions indicate treatment plans that satisfy all DV criteria in the clinical protocol, optimization using small sets of dose calculation points fails to truly find good treatment plans (if they exist) even after being run for 1 h. Similarly, the re-evaluated hypervolume values of using small dose calculation point sets are significantly lower. On the other hand, the decreases are not so considerable if larger



**Fig. 4.** Patient 1. Top row: the average hypervolume values of the trade-off fronts over time found by MO-RV-GOMEA with different settings of the number of dose calculation points. Shaded areas represent the standard deviations. Bottom row: the trade-off front that has the median hypervolume value.

**Table 2**

Average re-evaluated hypervolume values and standard deviations (in brackets) of the trade-off fronts of treatment plans obtained after 1 h with four fixed sets of dose calculation points: 2500, 5000, 10,000, and 20,000 points. The best values are bold-faced and statistical significance (with Bonferroni correction) is marked with the asterisk(\*) symbol.

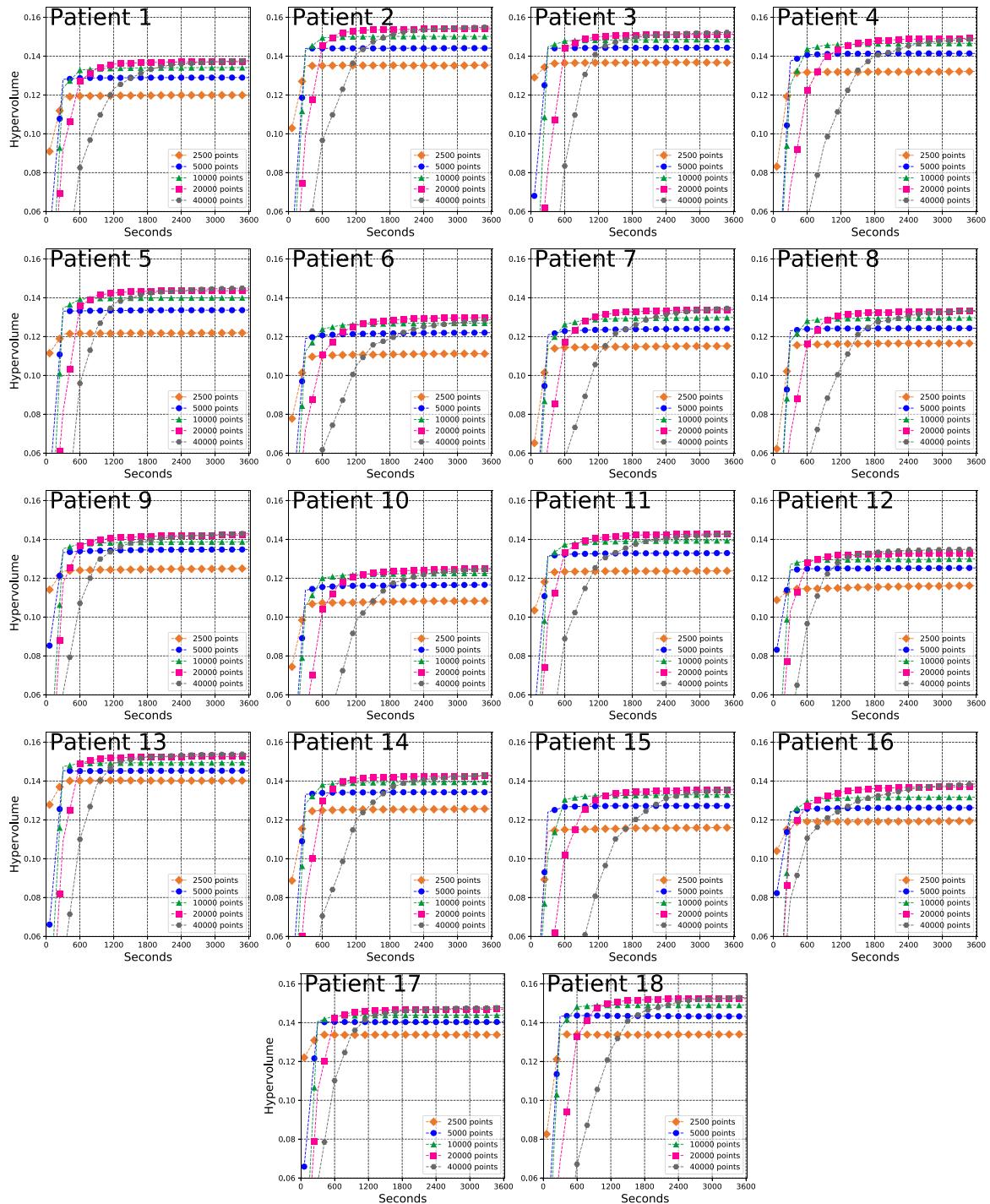
ID	2500	5000	10,000	20,000
1	0.119860 (0.006958)	0.128910 (0.003588)	0.134097 (0.003934)	<b>0.137120</b> (0.002470)*
2	0.135320 (0.005863)	0.144067 (0.005224)	0.150107 (0.004470)	<b>0.154087</b> (0.002394)*
3	0.136653 (0.003600)	0.144323 (0.002174)	0.148513 (0.001758)	<b>0.151270</b> (0.001386)*
4	0.132067 (0.004817)	0.141347 (0.003642)	0.146610 (0.002259)	<b>0.149187</b> (0.001357)*
5	0.121887 (0.008496)	0.133600 (0.006079)	0.140003 (0.003994)	<b>0.143653</b> (0.003063)*
6	0.111123 (0.003745)	0.121993 (0.002338)	0.126993 (0.001456)	<b>0.129807</b> (0.001035)*
7	0.115131 (0.007622)	0.124077 (0.005963)	0.129817 (0.002672)	<b>0.133747</b> (0.002217)*
8	0.116520 (0.004766)	0.124313 (0.004234)	0.129713 (0.002127)	<b>0.133047</b> (0.001608)*
9	0.125023 (0.004535)	0.134780 (0.004195)	0.138753 (0.003547)	<b>0.142210</b> (0.002685)*
10	0.108299 (0.004711)	0.116587 (0.003450)	0.122593 (0.001995)	<b>0.124940</b> (0.001414)*
11	0.123730 (0.005262)	0.132930 (0.003162)	0.139370 (0.001971)	<b>0.142807</b> (0.001309)*
12	0.116152 (0.006804)	0.125283 (0.004020)	0.129857 (0.003569)	<b>0.132870</b> (0.003007)*
13	0.140207 (0.003818)	0.145240 (0.004029)	0.149423 (0.003418)	<b>0.152597</b> (0.002218)*
14	0.125650 (0.005386)	0.134153 (0.004168)	0.139463 (0.003068)	<b>0.142573</b> (0.002013)*
15	0.115981 (0.004569)	0.127167 (0.003091)	0.132930 (0.001970)	<b>0.135407</b> (0.001816)*
16	0.119275 (0.006984)	0.126210 (0.006947)	0.131617 (0.005863)	<b>0.137070</b> (0.003019)*
17	0.133703 (0.003768)	0.140320 (0.002511)	0.143843 (0.002810)	<b>0.146880</b> (0.001705)*
18	0.133973 (0.006031)	0.143200 (0.003301)	0.148967 (0.002584)	<b>0.152377</b> (0.001633)*

sets of dose calculation points are employed. The small differences of both hypervolume graphs and approximation fronts before and after re-evaluation when 20,000 points (or 40,000 points) are used during optimization suggest that employing a dose calculation point set of such a size (or larger sizes) is necessary to ensure acceptable reliability and accuracy of the results obtained for solving our bi-objective optimization model of the prostate HDR-BT planning problem. Determining a proper number of dose calculation points for optimization is important because directly optimizing over the full set of 500,000 points would incur prohibitive computing time (since the evaluation time for each candidate treatment plan increases proportionally with the number of points employed to calculate DV indices).

Fig. 5 shows the comparison results of hypervolume development graphs for all patient cases in our data set when optimizing over 2500, 5000, 10,000, 20,000, and 40,000 dose calculation points, and the presented results are all re-evaluated based on the full set of 500,000 points. The results here agree with our observation mentioned above. In the first 5–10 min, optimization runs that use small sets of dose calculation points (i.e., 2500 or 5000 points) obtain better hypervolume values (and thus better non-dominated fronts). After 10–20 min, optimization runs that use 10,000–20,000 dose calculation points catch up and exceed

the runs that use smaller sets. It takes about 40–60 min for the runs with 40,000 dose calculation points to reach similar results. Table 2 compares the average hypervolume values of the re-evaluated trade-off fronts of treatment plans obtained after 1 h when 2500–20,000 dose calculation points are used. For all patient cases, the optimization runs with 20,000 dose calculation points achieve better hypervolume results than the runs with fewer points, and the differences are found to be statistically significant. It can be seen that after the first few minutes, the runs with smaller sets seem to converge quickly while the runs with larger sets continue to improve their non-dominated fronts. We can draw an analogy between this phenomenon and the overfitting problem in machine learning, where overfitting is more likely to happen when the number of data points in the training set is too few [33]. Similarly, optimization runs might quickly converge to some setting of dwell times such that its resulting dose distribution is (near-)optimal for the DV criteria as considered for small sets of 2500 or 5000 dose calculation points. However, such dwell-time settings are less likely (than the dwell-time settings obtained by the runs with 10,000 or 20,000 points) to satisfy the DV criteria over the full set of 500,000 points (i.e., similar to the validation set or test set in machine learning).

Table 3 compares the average hypervolume values of the re-evaluated trade-off fronts of treatment plans obtained after 1 h



**Fig. 5.** The average hypervolume values of the trade-off fronts over time found by MO-RV-GOMEA with different settings of the number of dose calculation points.

when 20,000 and 40,000 dose calculation points are used. The differences between the results obtained from the two settings are statistically insignificant or negligible. However, Fig. 5 shows that it takes MO-RV-GOMEA with 40,000 points about 40–60 min to obtain similar results which can be achieved in about 20 min by MO-RV-GOMEA with 20,000 points. For 20–30 min of computing time on a single-core CPU, 20,000 points appear to be the proper number of dose calculation points. Therefore, in all remaining experiments, we employ 20,000 points.

There exist several limitations with this rigid setting of a fixed number of points for multi-objective HDR-BT planning. First, Fig. 5 shows that, in our experiments, the hypervolume values

of optimization runs with 20,000 points do not significantly increase after 20 min, indicating that the search starts to converge. However, such an amount of running time pertains to the specific configuration of our computer system. Considering that the total HDR-BT planning time should not exceed 1 h, the optimization running time should be kept as short as possible without compromising on the accuracy because treatment planners need time to investigate solutions from the resulting trade-off front and select a plan that is preferred. Second, the number 20,000 is well-tuned for solving our bi-objective model derived from a specific clinical protocol for prostate HDR-BT. A different protocol at a different clinic, that contains different kinds of clinical criteria, might need

**Table 3**

Average re-evaluated hypervolume values and standard deviations(in brackets) of the trade-off fronts of treatment plans obtained after 1 h with two fixed sets of dose calculation points: 20,000 and 40,000 points. The better values are bold-faced and statistical significance is marked with the asterisk(\*) symbol.

Patient ID	20,000	40,000
1	0.137120 (0.002470)	<b>0.137653</b> (0.001371)
2	0.154087 (0.002394)	<b>0.154960</b> (0.001313)
3	0.151270 (0.001386)	<b>0.152283</b> (0.000816)*
4	<b>0.149187</b> (0.001357)*	0.148967 (0.000887)
5	0.143653 (0.003063)	<b>0.144987</b> (0.001596)
6	<b>0.129807</b> (0.001035)*	0.128693 (0.001689)
7	0.133747 (0.002217)	<b>0.134757</b> (0.001419)*
8	0.133047 (0.001608)	<b>0.133453</b> (0.001128)
9	0.142210 (0.002685)	<b>0.143123</b> (0.001909)
10	<b>0.124940</b> (0.001414)	0.124860 (0.001580)
11	0.142807 (0.001309)	<b>0.142927</b> (0.001387)
12	0.132870 (0.003007)	<b>0.134883</b> (0.002152)*
13	0.152597 (0.002218)	<b>0.153803</b> (0.001097)*
14	0.142573 (0.002013)	<b>0.143073</b> (0.001704)
15	0.135407 (0.001816)	<b>0.135720</b> (0.000822)
16	0.137070 (0.003019)	<b>0.138470</b> (0.001861)*
17	0.146880 (0.001705)	<b>0.147683</b> (0.001050)*
18	0.152377 (0.001633)	<b>0.152983</b> (0.000764)

**Table 4**

Average re-evaluated hypervolume values and standard deviations (in brackets) of the trade-off fronts of treatment plans obtained after 1 h with a fixed set of 20,000 dose calculation points versus with the multi-resolution scheme. The best values are bold-faced and statistical significance (with Bonferroni correction) is marked with the asterisk(\*) symbol.

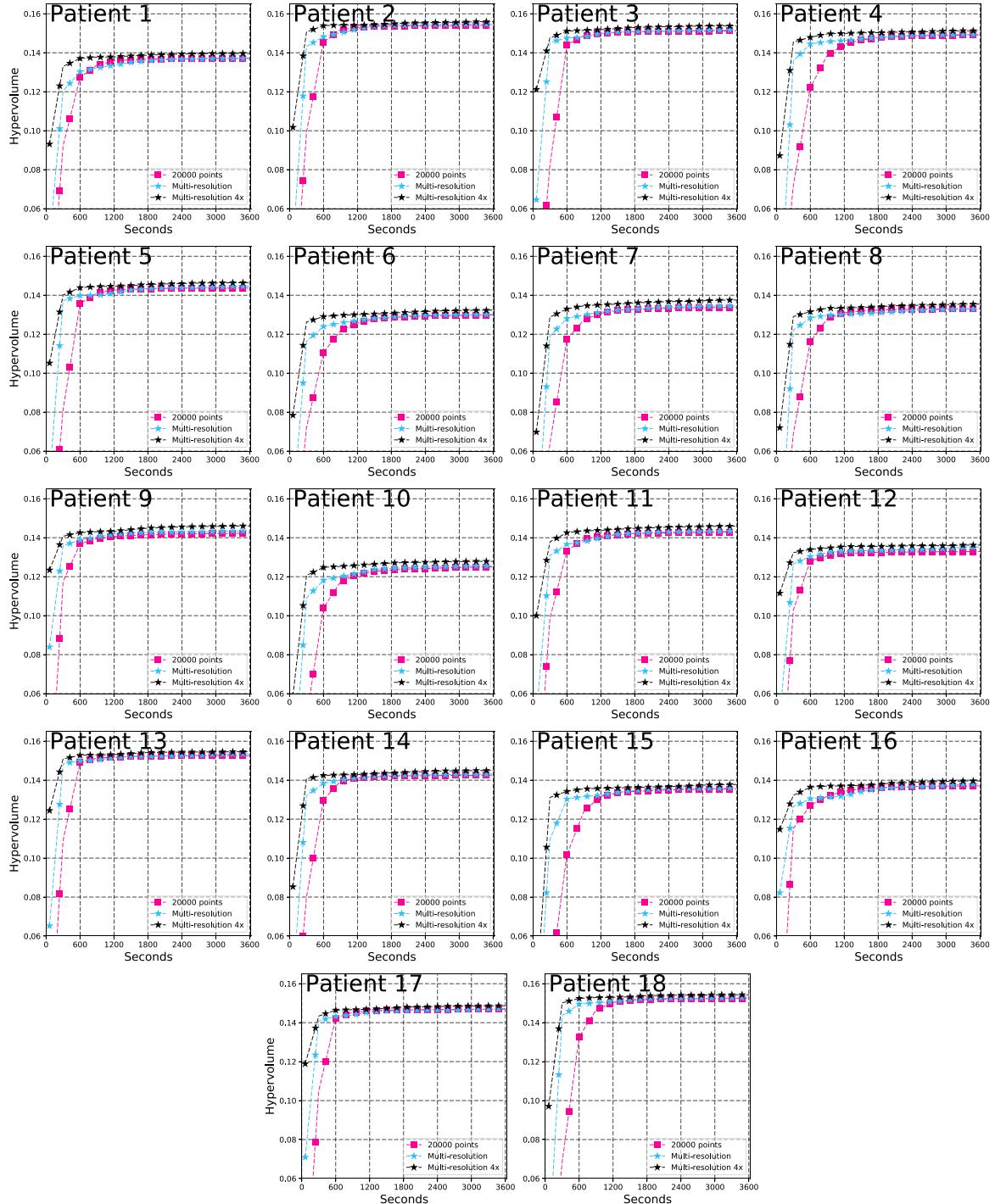
ID	20,000	Multi-resolution	Multi-resolution 4-core
1	0.137120 (0.002470)	0.137100 (0.001987)	<b>0.139637</b> (0.001396)*
2	0.154087 (0.002394)	0.154700 (0.001460)	<b>0.155910</b> (0.001766)*
3	0.151270 (0.001386)	0.151710 (0.001226)	<b>0.153757</b> (0.000481)*
4	0.149187 (0.001357)	0.149397 (0.001466)	<b>0.151310</b> (0.000815)*
5	0.143653 (0.003063)	0.144610 (0.002224)	<b>0.146393</b> (0.001568)*
6	0.129807 (0.001035)	0.130153 (0.001147)	<b>0.132303</b> (0.000570)*
7	0.133747 (0.002217)	0.134480 (0.001961)	<b>0.137520</b> (0.001047)*
8	0.133047 (0.001608)	0.132843 (0.001853)	<b>0.135430</b> (0.001103)*
9	0.142210 (0.002685)	0.143420 (0.001955)	<b>0.146063</b> (0.001322)*
10	0.124940 (0.001414)	0.125563 (0.001726)	<b>0.127843</b> (0.001124)*
11	0.142807 (0.001309)	0.143393 (0.001600)	<b>0.145900</b> (0.000952)*
12	0.132870 (0.003007)	0.135170 (0.001750)	<b>0.136223</b> (0.001728)*
13	0.152597 (0.002218)	0.153510 (0.000767)	<b>0.154613</b> (0.000907)*
14	0.142573 (0.002013)	0.143107 (0.002136)	<b>0.144933</b> (0.001240)*
15	0.135407 (0.001816)	0.136073 (0.001201)	<b>0.137807</b> (0.000954)*
16	0.137070 (0.003019)	0.137980 (0.003402)	<b>0.139610</b> (0.002494)*
17	0.146880 (0.001705)	0.147153 (0.001346)	<b>0.148690</b> (0.000811)*
18	0.152377 (0.001633)	0.152913 (0.001418)	<b>0.154290</b> (0.000989)*

a different number of dose calculation points. Similarly, HDR-BT planning optimization for different treatment targets (e.g., head-and-neck, cervix, or breast) might also require different sizes for the dose calculation point sets. The number of dose calculation points can thus be seen as a control parameter of an HDR-BT planning optimization run. Determining a proper value for an EA control parameter beforehand is difficult in real-world optimization, and it would instead be more practical to employ the IMS to adapt the number of dose calculation points during the run.

## 5.2. Multi-resolution scheme

Fig. 6 shows the hypervolume development graphs for solving the bi-objective model of the prostate HDR-BT planning problem based on a fixed set of 20,000 dose calculation points versus the multi-resolution scheme, in which the number of dose calculation points is adapted along the run. All presented results are re-evaluated based on the full set of 500,000 points. The first two columns of Table 4 compare the hypervolume values of the trade-off fronts of treatment plans obtained after 1 h. Both methods converge to trade-off fronts of similar hypervolume values at the end of the 1-hour run (no statistically significant difference is found), indicating that their results are of equal quality.

The multi-resolution scheme, however, yields faster convergence speed. Because the multi-resolution scheme starts with small sets of dose calculation points, it has the same initial performance as when fixed sets of 2500 or 5000 points are used. This is beneficial since at the beginning of the run, the search is mainly explorative and directly optimizing over large sets of dose calculation points would then be inefficient. If the result is satisfactory, BT planners can terminate the optimization run and investigate the so-far-obtained trade-off front. Otherwise, if the optimization is run further, new populations of larger sizes are initialized and the dose calculation point set is enlarged accordingly, providing better resolution (i.e., accuracy) to the search. Consequently, the longer optimization is allowed to run, the better results (if they exist) could be achieved. This is not the case when small dose calculation point sets (e.g., 2500 or 5000 points) are employed, as Fig. 5 shows that their hypervolume graphs quickly increase but then soon converge to sub-optimal values. In the multi-resolution scheme, early results based on small dose calculation point sets provide effective guidance for later search, which is based on larger dose calculation point sets, while not overly biasing the search for a fixed set of points, because the set of dose calculation points is enlarged before overfitting can occur.



**Fig. 6.** The average hypervolume values of the trade-off fronts over time found by MO-RV-GOMEA with different settings of the number of dose calculation points.

**Fig. 6** and **Table 4** also show that the performance of the multi-resolution scheme can be further improved by simply enabling parallel computation on multi-core CPUs. In this work, we run MO-RV-GOMEA with the multi-resolution scheme on a quad-core CPU, which is a common configuration for desktop computers at the time of writing. The parallelized MO-RV-GOMEA yields better convergence speed and achieves similar or better hypervolume values than the single-core version. Within 5–10 min, the parallelized MO-RV-GOMEA starts to converge, indicating that the optimization run can be terminated because substantial improvements are not expected to be found anymore.

**Fig. 7** compares the resulting trade-off fronts of treatment plans obtained by the runs that have the median hypervolume values when a fixed set of 20,000 dose calculation points is used versus the fronts that could be obtained if the multi-resolution scheme is employed instead. All the presented results are re-evaluated based on the full set of 500,000 dose calculation points. When a fixed set of 20,000 dose calculation points is employed, after being run for 1 h, MO-RV-GOMEA obtained trade-off fronts that contain treatment plans in the golden corner for most patient cases. Since the total planning time should be kept within 1 h, it is unlikely that the whole time budget is used for optimization and the optimization running time should be reduced as much as

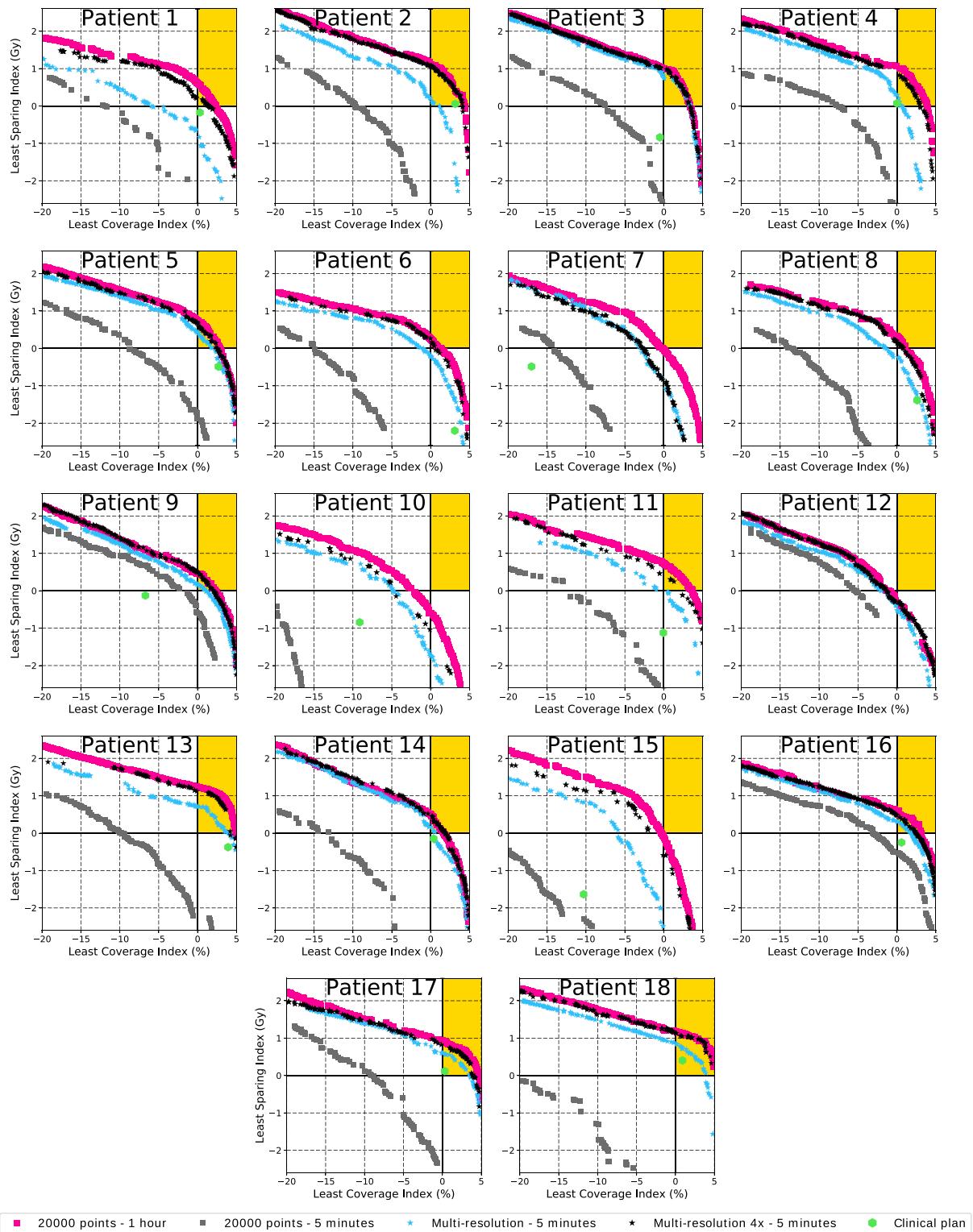


Fig. 7. Pareto fronts comparison.

possible. However, at the time point of 5 min, the trade-off fronts obtained with fixed dose calculation point sets are still far from the golden corner as well as from the 1-hour fronts, indicating that treatment plans of much better quality can still be achieved. Contrastingly, the 5 min trade-off fronts obtained with the multi-resolution scheme are considerably better, closer to the golden corner, and dominate the 5 min fronts obtained with the fixed

dose calculation point sets. Especially, when parallelized computing of solution evaluations is enabled, the multi-resolution scheme achieves trade-off fronts at 5 min runtime that are close to the 1-hour trade-off fronts obtained with the fixed sets of 20,000 dose calculation points. The average hypervolume results in Table 5 further support our claim. This indicates that the parallelized runs could be terminated after 5 min and their results could be presented to BT planners for selecting treatment plans.

**Table 5**

Average re-evaluated hypervolume values and standard deviations (in brackets) of the trade-off fronts of treatment plans obtained after five minutes with a fixed set of 20,000 dose calculation points versus with the multi-resolution scheme. The best values are bold-faced and statistical significance (with Bonferroni correction) is marked with the asterisk(\*) symbol. The 1-hour hypervolume results of using 20,000 points are also included for comparison.

ID	20,000 (5 min)	Multi-resolution (5 min)	Multi-resolution 4 cores (5 min)	20,000 (1 h)
1	0.092210 (0.007147)	0.120333 (0.004767)	<b>0.132980</b> (0.003975)*	0.137120 (0.002470)
2	0.099105 (0.005807)	0.143197 (0.006592)	<b>0.150653</b> (0.003503)*	0.154087 (0.002394)
3	0.082324 (0.013354)	0.145260 (0.002590)	<b>0.147557</b> (0.002453)*	0.151270 (0.001386)
4	0.071905 (0.012951)	0.135883 (0.002846)	<b>0.145497</b> (0.001208)*	0.149187 (0.001357)
5	0.081201 (0.008974)	0.137207 (0.003394)	<b>0.140130</b> (0.003490)*	0.143653 (0.003063)
6	0.072148 (0.007551)	0.116361 (0.005101)	<b>0.126243</b> (0.002020)*	0.129807 (0.001035)
7	0.063972 (0.007704)	0.119053 (0.004933)	<b>0.128750</b> (0.004020)*	0.133747 (0.002217)
8	0.068848 (0.007717)	0.122127 (0.002516)	<b>0.129047</b> (0.001848)*	0.133047 (0.001608)
9	0.117427 (0.006150)	0.135947 (0.001895)	<b>0.140753</b> (0.002352)*	0.142210 (0.002685)
10	0.047569 (0.011441)	0.109066 (0.007975)	<b>0.120660</b> (0.002630)*	0.124940 (0.001414)
11	0.098454 (0.009159)	0.130887 (0.003990)	<b>0.137880</b> (0.003368)*	0.142807 (0.001309)
12	0.102818 (0.010952)	0.126530 (0.005737)	<b>0.132353</b> (0.001962)*	0.132870 (0.003007)
13	0.109106 (0.007456)	0.148427 (0.002415)	<b>0.150877</b> (0.001414)*	0.152597 (0.002218)
14	0.080000 (0.007776)	0.132203 (0.007095)	<b>0.140743</b> (0.001159)*	0.142573 (0.002013)
15	0.035456 (0.009783)	0.109627 (0.005418)	<b>0.131137</b> (0.002119)*	0.135407 (0.001816)
16	0.115160 (0.002799)	0.126453 (0.004493)	<b>0.132170</b> (0.004858)*	0.137070 (0.003019)
17	0.105128 (0.007904)	0.140923 (0.002899)	<b>0.143373</b> (0.003322)*	0.146880 (0.001705)
18	0.068436 (0.011074)	0.143503 (0.001657)	<b>0.150163</b> (0.001326)*	0.152377 (0.001633)

In Fig. 7, we also plot the quality of the clinical plans in terms of the LCI-LSI objective space. These clinical plans were made within 30–60 min by manual optimization for the treatment of the corresponding patient cases. For many cases, the clinical plans lie outside the golden corner, which shows the difficulty of manually designing a BT treatment plan that satisfies all DV criteria in the protocol based on the software currently clinically used. Fig. 7 exhibits that, after being run for 5 min, parallelized MO-RV-GOMEA with the multi-resolution scheme obtains trade-off fronts that dominate the clinical plans in all cases. Such performance supports the usability of our method for being employed in clinical practice. Note that, however, the comparisons here are based solely on Pareto dominance with respect to the (LCI, LSI) objective value vectors of treatment plans. To approve a candidate plan to be used for treatment, BT planners might need to consult additional information (e.g., the patient age and health situation, or the dose distribution outside both target volumes and OARs), which is difficult to be formulated in a meaningful and intuitive way into our current bi-objective optimization model. The inclusion of such information is left for future work.

### 5.3. Discussion

Overall, the multi-resolution scheme addresses two key issues with dose calculation point sets of fixed sizes mentioned in Section 5.1, i.e., (1) reducing the optimization running time as much as possible while still assuring reliability of the obtained results, and (2) setting a proper number of dose calculation points as a control parameter of the optimization algorithm. Note that simply parallelizing the run with a fixed set of 20,000 points would only address the former issue because the latter depends on the specific bi-objective optimization model for HDR-BT planning being solved at hand. A different clinical protocol that contains different DV criteria might require a different number of dose calculation points. Furthermore, other treatment targets (e.g., cervix, breast, head-and-neck) also require different sizes of the dose calculation point set since each target has different characteristics (e.g., volume, shape, surrounding OARs). If the size of the dose calculation point set is fixed, the proper size needs to be determined specifically for each specific case. The multi-resolution scheme thus eliminates such non-trivial parameter tuning when solving HDR-BT planning. In general, our approach is quite flexible and capable of handling various types of problem difficulties and planning aims, thus allowing other kinds of treatment criteria to

be considered during optimization, e.g., hot-spot restriction [34], tumor control probability (TCP) [35], or normal tissue complication probability (NTCP) [35]. Such criteria are important to be used in combination with DV indices, because it has been shown that in certain cases typical DV indices alone are not enough to describe what makes a treatment plan really good [36].

While our approach can be considered sufficiently fast for dwell-time optimization tasks on computer desktops with only CPU cores in current clinical practice, real-time HDR-BT planning problems in the upcoming future, e.g., MR-guided catheter placement [37], would require a faster response time. A potential solution is to further accelerate MO-RV-GOMEA with the use of GPU parallelization, which we recently successfully achieved [38]. To further speed up both the CPU and GPU versions of our approach, we will study alternative ways of sampling dose calculation points that, for the specific DV indices of interest, result in the same precision of the DV indices, but using fewer dose calculation points (e.g., by exploiting shapes and surfaces of organs) [39,40]. Furthermore, our approach has been tested so far with only one protocol for prostate HDR-BT at our clinic (Amsterdam UMC, University of Amsterdam). The future plan is to show and improve its generalizability to other protocols at other clinics and for other tumor sites as well.

### 6. Conclusions

We successfully addressed the purpose of this work, i.e., to speed up MO-RV-GOMEA in solving the bi-objective prostate HDR-BT planning problem. We first discussed the computational challenges associated with automating HDR-BT planning for prostate cancer through computational optimization, especially the importance of properly determining the number of dose calculation points, which are used to evaluate the quality of candidate treatment plans. The more dose calculation points are used, the more accurate the obtained results are, but the more computing time is required. Based on our preliminary research, we recommended the use of MO-RV-GOMEA as the optimization algorithm for solving a bi-objective model for HDR-BT planning. Here, we proposed a multi-resolution scheme that can be seamlessly integrated with an interleaved multi-start scheme for MO-RV-GOMEA, in which the number of dose calculation points is gradually increased during an optimization run. We further described a straightforward approach to multi-core parallelization of MO-RV-GOMEA. Experiments on our data set

of 18 patient cases confirmed the effectiveness and efficiency of the multi-resolution scheme, especially in combination with (4-core) parallel processing, in which the optimization running time can be reduced to 5 min, whereas 30–60 min are required in current practice in BT planning. This result and the recent positive clinical observer study outcomes [23] effectively open the door to using MO-RV-GOMEA together with the bi-objective problem formulation in clinical practice.

### Declaration of competing interest

Dr. T. Alderliesten, Dr. B.R. Pieters, Dr. A. Bel, and Prof. Dr. P.A.N. Bosman are involved in projects supported by Elekta. Dr. Y. Niatsetski is an employee of Elekta.

### Acknowledgments

This work is part of the research program IPPSI-TA with project number 628.006.003, which is financed by the Netherlands Organisation for Scientific Research (NWO) and Elekta. The authors thank Dr. Niek van Wieringen (Department of Radiation Oncology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands) for his support in providing the patient data.

## Appendix A. MO-Rv-GOMEA

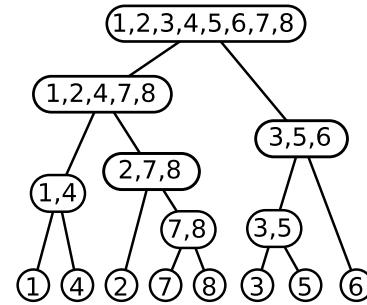
### A.1. Elitist archive

MO-RV-GOMEA employs a secondary population to keep track of the non-dominated solutions obtained during the run because it has been suggested that having such an elitist archive is better for the convergence of MOEAs [26]. The desired archive target size can be set by a user with respect to their available computing resources. It is also shown in [12] that an archive size of 1000 members is suitable for bi-objective prostate HDR-BT planning. Each time a new candidate solution is evaluated, it is added to the archive if it is not dominated by any archive member. Archive members that are dominated by the new solution are removed from the archive. When the target size is reached, a discretization procedure is invoked to divide the objective space into equal hypercubes, where each hypercube can contain only one non-dominated solution. A new non-dominated solution is added into the archive if it resides in an empty hypercube or if it dominates the current occupant of that hypercube. The size of hypercubes is adapted along the optimization run to ensure the archive maintains approximately its desired size. Details about the implementation can be found in [25].

### A.2. Population clustering

Recombination might not be effective when performed on solutions that belong to different regions on the non-dominated front, deteriorating the efficiency of the search. For example, it is generally not beneficial to recombine a treatment plan that maximizes target coverage and a plan that minimizes radiation to OARs because such two plans are at different extremes of the front and typically differ a lot from each other. Population clustering in the objective space is an implementation of the niching concept, such that solution variation can be restricted to solutions that belong to the same cluster and probably bear certain similarities.

MO-RV-GOMEA maintains a working population  $P$  that has  $n$  solutions (i.e., candidate treatment plans), which can be randomly uniformly initialized. In every generation, a selection set



**Fig. A.1.** An example linkage tree with  $l = 8$  problem variables.

$S$  is formed by selecting  $\lfloor \tau n \rfloor$  solutions that have the best non-domination rankings from  $P$ , where  $\tau = 0.35$  [17]. The selection set  $S$  is then clustered into  $k$  clusters of equal size  $c = \frac{2}{k}|\mathcal{S}|$ , where  $k > m$  and  $m$  is the number of objectives. This clustering procedure was shown to enhance the probability of obtaining an evenly-spread non-dominated front [19].

### A.3. Linkage learning

MO-RV-GOMEA bears much resemblance to estimation-of-distribution algorithms (EDAs), in which solution variation is done by sampling from the distribution model that is built over the solutions of the selection set  $S$  in every generation. The distribution model encodes certain linkages (i.e., dependencies) among decision variables (i.e., dwell times) that should be considered when generating offspring solutions. For HDR-BT planning, such linkage-based optimization was shown to result in better results compared to when variable dependencies are disregarded as in classic variation operators (i.e., crossover and mutation) or univariate factorization [12].

MO-RV-GOMEA often employs the so-called *Linkage Tree* (LT) as the linkage model structure. An LT  $F$  over a set of decision variable indices  $L = \{1, 2, \dots, l\}$  is a set of linkage sets  $F = \{\mathbf{F}^1, \mathbf{F}^2, \dots, \mathbf{F}^{|F|}\}$ , where each linkage set  $\mathbf{F}^i$  is a subset of  $L$ , i.e.,  $\mathbf{F}^i \subseteq L$ ,  $i \in \{1, 2, \dots, |F|\}$ . The LT contains  $l$  univariate linkage sets, i.e.,  $\mathbf{F}^i = \{i\}$ ,  $i \in L$ , indicating independence among decision variables. The LT also contains  $l - 1$  multivariate linkage sets, i.e.,  $\mathbf{F}^i \in F$ ,  $|\mathbf{F}^i| > 1$ , where each linkage set as such indicates a group of variables that depend on each other to some extent. A special multivariate linkage set that contains all variable indices is assigned as the root note of the LT, indicating the total-dependent state among variables. The LT has in total  $2l - 1$  linkage sets. An example LT is shown in Fig. A.1.

The LT can be built by the hierarchical clustering algorithm UPGMA (Unweighted Pair Group Method with Arithmetic Mean) that has  $O(nl^2)$  complexity [41]. The input of UPGMA is an  $l \times l$  matrix describing similarity (e.g., a measure of distance) between all pairs of decision variables. The Mutual Information (MI) of variables, which can be inferred from the selected solutions in each cluster, can be used as a similarity metric in the black-box optimization context. However, it has been shown in [12] that, for the HDR-BT planning problem, the Euclidean distance between dwell positions is a better metric for building the LT. Dwell times at dwell positions that are close to each other should exhibit stronger dependence than dwell times at far away dwell positions. Therefore, instead of learning the LT structure in every generation, we employ such geometry knowledge to build an LT beforehand, and this LT is then fixed to be used throughout the optimization run.

In every generation, each cluster of MO-RV-GOMEA is assigned its own LT. While their linkage structure is the same,

the associated distributions of each LT are estimated over the solutions in each cluster. Based on the solutions in each cluster, a univariate normal distribution is estimated for each univariate linkage set while a multivariate normal distribution is estimated for each multivariate linkage set. Each normal distribution involves only the variables that are specified by the linkage set that it is associated with. In total,  $2l - 1$  normal distributions are estimated for each cluster. These distributions will be used to generate offspring solutions when performing variation. Details about estimating and adapting these distributions can be found in [17].

#### A.4. Solution variation by gene-pool optimal mixing

For each cluster, the LT and its normal distributions are used to transform each population member  $\mathbf{x}$  associated with that cluster into an offspring solution  $\mathbf{o}$  by the Gene-pool Optimal Mixing (GOM) operator as follows. All linkage sets in the LT are traversed in a random order. For each linkage set, we sample the associated normal distribution to modify  $\mathbf{x}$  at the variables specified by that linkage set. The partially-altered solution is evaluated to compute the resulting difference in objective values, and the changes are accepted only if they result in an improvement over the previous state; otherwise, the changes are undone. After all linkage sets are traversed, the original population member  $\mathbf{x}$  is evolved into an offspring  $\mathbf{o}$ .

Each cluster is assigned an LT, and the parameters of normal distributions corresponding with each LT are estimated from cluster solutions. A cluster that contains the best solutions for a certain objective (more than solutions in other clusters) is called *extreme* cluster. Clusters that are not extreme clusters are called *middle* clusters. The LT of each cluster is then used to improve population solutions associated with that cluster to create offspring solutions. The improvement check at every GOM step depends on which cluster the population solution belongs to. If the solution belongs to a middle cluster, improvement checks are based on the Pareto dominance relation. If the solution belongs to an extreme cluster, improvement checks are based solely on the objective corresponding to that cluster. Further details about GOM can be found in [17].

#### A.5. Partial evaluations

Due to the genetic-local-search-like nature of the GOM operator, MO-RV-GOMEA performs more evaluations per generation than a typical MOEA (e.g., NSGA-II, MOEA/D). Most evaluations are spent on evaluating partially-altered solutions (i.e., treatment plans), which differ from their previous states at only a few variables (i.e., dwell times). However, the problem structure of HDR-BT planning allows *partial evaluations*, i.e., the impact of such local changes can be efficiently computed to update the solutions' objective values. The radiation dose distribution calculation in Eq. (1) is a multiplication of the dose-rate matrix  $\mathbf{R}$  and a dwell time vector  $\mathbf{t}$  (i.e., a treatment plan). If the values of only a few dwell times are altered, the new treatment plan  $\mathbf{t}'$  will be  $\mathbf{t}' = \mathbf{t} + \Delta\mathbf{t}$ , where  $\Delta\mathbf{t}$  consists of only a few non-zero elements at the corresponding positions. The impact can be computed by performing multiplications between only the columns of  $\mathbf{R}$  that correspond to the changed dwell times and the non-zero elements in  $\Delta\mathbf{t}$ . Details about partial evaluations can be found in [12,17].

Note that partial evaluations are straightforwardly available for MO-RV-GOMEA because each step of GOM only modifies the current solution at a few variables as indicated by a linkage set. Typical MOEAs for real-valued optimization (e.g., NSGA-II, MOEA/D, or MAMaLGaM), however, create a complete offspring

Generation	$P_1$	$P_2$	$P_3$	$P_4$	...
1	✓				
2	✓	✓			
3	✓				
4	✓	✓	✓		
5	✓				
6	✓	✓			
7	✓				
8	✓	✓	✓	✓	
...	...	...	...	...	...

Fig. A.2. Interleaved Multi-start Scheme (IMS) with  $b = 2$ .

solution each time, which differs totally from the parent solutions and requires a full evaluation. Properly implementing a similar LT-based variation procedure for these MOEAs is non-trivial, substantially changing their working principles as well. Partial evaluations were shown to enable obtaining much better performance with MO-RV-GOMEA than many other state-of-the-art MOEAs when solving the HDR-BT planning problem [12].

#### A.6. Interleaved multi-start scheme (IMS)

The IMS is a automatic population-sizing scheme, which was firstly introduced in [28] to eliminate the requirement of population size setting for the Genetic Algorithm and was later adapted for MO-RV-GOMEA [17]. The IMS operates multiple populations of increasing sizes in an asynchronously parallel fashion. The first population  $P_1$  is initialized with some small size  $N_1$ . Each subsequent population  $P_{i+1}$  is then initialized with a population size twice as large as the preceding population, i.e.,  $N_{i+1} = 2 \times N_i$ ,  $i > 0$ . For every  $b$  generations of population  $P_i$ , the subsequent population  $P_{i+1}$  is run for one generation. In principle, there is no upper bound of the population size; all populations are kept running until some termination criteria are satisfied. Fig. A.2 shows an example of the IMS.

The generation bases of IMS  $b = 2$ , and 4 have been suggested for the discrete optimization cases [29] while  $b = 8$  was found to give good results for MO-RV-GOMEA (i.e., for real-valued optimization) [17]. Note that while each population is run separately, they contribute to the same elitist archive. Because new populations of larger sizes are kept being added, smaller populations should be terminated if they are inefficient to solve the current problem instance. For MO-RV-GOMEA with IMS, a population is deemed to be inefficient if it contributes less than 10% of the non-dominated front that is combined from all active populations while a larger population contributes more than 10% of this combined front [17].

## Appendix B. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.asoc.2019.105681>.

## References

- [1] J.N. Lukens, E. McMenamin, What radiation oncology wants medical oncology to know, *J. Adv. Pract. Oncol.* 7 (3) (2016) 307–309.
- [2] J.L.M. Venselaar, D. Baltas, P. Hoskin, A.S. Meigooni, Introduction and innovations in brachytherapy, in: Comprehensive Brachytherapy, Taylor & Francis, 2012, pp. 3–8, <http://dx.doi.org/10.1201/b13075-3>.
- [3] A.M. Dinkla, R. van der Laarse, E. Kaljouw, B.R. Pieters, K. Koedooder, N. van Wieringen, A. Bel, A comparison of inverse optimization algorithms for HDR/PDR prostate brachytherapy treatment planning, *Brachytherapy* 14 (2) (2015) 279–288, <http://dx.doi.org/10.1016/j.brachy.2014.09.006>.

- [4] T. Siauw, A. Cunha, A. Atamtürk, I.-C. Hsu, J. Pouliot, K. Goldberg, IPIP: A new approach to inverse planning for HDR brachytherapy by directly optimizing dosimetric indices, *Med. Phys.* 38 (7) (2011) 4045–4051, <http://dx.doi.org/10.1118/1.3598437>.
- [5] E. Lessard, J. Pouliot, Inverse planning anatomy-based dose optimization for HDR-brachytherapy of the prostate using fast simulated annealing algorithm and dedicated objective function, *Med. Phys.* 28 (5) (2001) 773–779, <http://dx.doi.org/10.1118/1.1368127>.
- [6] A. Karabis, S. Giannouli, D. Baltas, HIPPO: A hybrid inverse treatment planning optimization algorithm in HDR brachytherapy, *Radiat. Oncol.* 76 (2005) S29, [http://dx.doi.org/10.1016/S0167-8140\(05\)81018-7](http://dx.doi.org/10.1016/S0167-8140(05)81018-7).
- [7] R. Alterovitz, E. Lessard, J. Pouliot, I.-C.J. Hsu, J.F. O'Brien, K. Goldberg, Optimization of HDR brachytherapy dose distributions using linear programming with penalty costs, *Med. Phys.* 33 (11) (2006) 4012–4019.
- [8] A. Holm, *Dose Plan Optimization in HDR Brachytherapy using Penalties: Properties and Extensions* (Ph.D. thesis), Linköping University, Linköping, Sweden, 2011.
- [9] S. Cui, P. Després, L. Beaulieu, A multi-criteria optimization approach for HDR prostate brachytherapy: I. Pareto surface approximation, *Phys. Med. Biol.* 63 (20) (2018) 205004, <http://dx.doi.org/10.1088/1361-6560/aae24c>.
- [10] S. Cui, P. Després, L. Beaulieu, A multi-criteria optimization approach for HDR prostate brachytherapy: II. Benchmark against clinical plans, *Phys. Med. Biol.* 63 (20) (2018) 205005, <http://dx.doi.org/10.1088/1361-6560/aae24f>.
- [11] C. Bélanger, S. Cui, Y. Ma, P. Després, J. Adam M Cunha, L. Beaulieu, A GPU-based multi-criteria optimization algorithm for HDR brachytherapy, *Phys. Med. Biol.* 64 (10) (2019) 105005, <http://dx.doi.org/10.1088/1361-6560/ab1817>.
- [12] N.H. Luong, T. Alderliesten, A. Bel, Y. Niatsetski, P.A.N. Bosman, Application and benchmarking of multi-objective evolutionary algorithms on high-dose-rate brachytherapy planning for prostate cancer treatment, *Swarm Evol. Comput.* 40 (2018) 37–52, <http://dx.doi.org/10.1016/j.swevo.2017.12.003>.
- [13] K. Deb, H. Jain, An evolutionary many-objective optimization algorithm using reference-point-based non-dominated sorting approach, part I: Solving problems with box constraints, *IEEE Trans. Evol. Comput.* 18 (4) (2014) 577–601, <http://dx.doi.org/10.1109/TEVC.2013.2281535>.
- [14] M. Lahanas, D. Baltas, N. Zamboglou, A hybrid evolutionary algorithm for multi-objective anatomy-based dose optimization in high-dose-rate brachytherapy, *Phys. Med. Biol.* 48 (3) (2003) 399.
- [15] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.* 6 (2) (2002) 182–197, <http://dx.doi.org/10.1109/4235.996017>.
- [16] L. De Boeck, J. Beliën, W. Egyed, Dose optimization in high-dose-rate brachytherapy: A literature review of quantitative models from 1990 to 2010, *Oper. Res. Health Care* 3 (2) (2014) 80–90, <http://dx.doi.org/10.1016/j.orhc.2013.12.004>.
- [17] A. Bouter, N.H. Luong, C. Witteveen, T. Alderliesten, P.A.N. Bosman, The multi-objective real-valued gene-pool optimal mixing evolutionary algorithm, in: GECCO '17: Proceedings of the Genetic and Evolutionary Computation Conference, ACM, New York, NY, USA, 2017, pp. 537–544, <http://doi.acm.org/10.1145/3071178.3071274>.
- [18] Q. Zhang, H. Li, MOEA/D: A multiobjective evolutionary algorithm based on decomposition, *IEEE Trans. Evol. Comput.* 11 (6) (2007) 712–731, <http://dx.doi.org/10.1109/TEVC.2007.892759>.
- [19] P.A.N. Bosman, The anticipated mean shift and cluster registration in mixture-based edas for multi-objective optimization, in: GECCO '10: Proceedings of the Genetic and Evolutionary Computation Conference, ACM, New York, NY, USA, 2010, pp. 351–358, <http://dx.doi.org/10.1145/1830483.1830549>.
- [20] A. Bouter, T. Alderliesten, C. Witteveen, P.A.N. Bosman, Exploiting linkage information in real-valued optimization with the real-valued gene-pool optimal mixing evolutionary algorithm, in: Proceedings of the Genetic and Evolutionary Computation Conference on - GECCO '17, ACM Press, New York, New York, USA, 2017, pp. 705–712, <http://dx.doi.org/10.1145/3071178.3071272>.
- [21] D. Wierstra, T. Schaul, T. Glasmachers, Y. Sun, J. Peters, J. Schmidhuber, Natural evolution strategies, *J. Mach. Learn. Res.* 15 (2014) 949–980.
- [22] N. Hansen, A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.* 9 (2) (2001) 159–195, <http://dx.doi.org/10.1162/106365601750190398>.
- [23] S.C. Maree, N. Luong, E.S. Kooreman, N. van Wieringen, A. Bel, K.A. Hinnen, H. Westerveld, B.R. Pieters, P.A.N. Bosman, T. Alderliesten, Evaluation of bi-objective treatment planning for high-dose-rate prostate brachytherapy – A retrospective observer study, *Brachytherapy* 18 (3) (2019) 396–403, <http://dx.doi.org/10.1016/j.brachy.2018.12.010>.
- [24] N.H. Luong, A. Bouter, M.C. van der Meer, Y. Niatsetski, C. Witteveen, A. Bel, T. Alderliesten, P.A.N. Bosman, Efficient, effective, and insightful tackling of the high-dose-rate brachytherapy treatment planning problem for prostate cancer using evolutionary multi-objective optimization algorithms, in: Genetic and Evolutionary Computation Conference, GECCO Companion '17, Berlin, Germany, July 15–19, 2017, ACM, New York, NY, USA, 2017, pp. 1372–1379, <http://dx.doi.org/10.1145/3067695.3082491>.
- [25] N.H. Luong, P.A.N. Bosman, Elitist archiving for multi-objective evolutionary algorithms: To adapt or not to adapt, in: Parallel Problem Solving from Nature - PPSN XII - 12th International Conference, Taormina, Italy, September 1–5, 2012, Proceedings, Part II, 2012, pp. 72–81, [http://dx.doi.org/10.1007/978-3-642-32964-7\\_8](http://dx.doi.org/10.1007/978-3-642-32964-7_8).
- [26] J. Knowles, D. Corne, Properties of an adaptive archiving algorithm for storing nondominated vectors, *IEEE Trans. Evol. Comput.* 7 (2) (2003) 100–116, <http://dx.doi.org/10.1109/TEVC.2003.810755>.
- [27] C.A. Coello Coello, S. González Brambila, J. Figueira Gamboa, M.G. Castillo Tapia, R. Hernández Gómez, Evolutionary multiobjective optimization: open research areas and some challenges lying ahead, *Complex Intell. Syst.* (2019) 1–16, <http://dx.doi.org/10.1007/s40747-019-0113-4>.
- [28] G.R. Harik, F.G. Lobo, A parameter-less genetic algorithm, in: Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation-Volume 1, Morgan Kaufmann Publishers Inc., 1999, pp. 258–265.
- [29] N.H. Luong, H. La Poutré, P.A. Bosman, Multi-objective gene-pool optimal mixing evolutionary algorithm with the interleaved multi-start scheme, *Swarm Evol. Comput.* 40 (2018) 238–254, <http://dx.doi.org/10.1016/j.swevo.2018.02.005>.
- [30] R. Nath, L.L. Anderson, G. Luxton, K.A. Weaver, J.F. Williamson, A.S. Meigooni, Dosimetry of interstitial brachytherapy sources: Recommendations of the AAPM radiation therapy committee task group no. 43, *Med. Phys.* 22 (2) (1995) 209–234, <http://dx.doi.org/10.1118/1.597458>.
- [31] M.J. Rivard, B.M. Coursey, L.A. DeWerd, W.F. Hanson, M. Saiful Huq, G.S. Ibbott, M.G. Mitch, R. Nath, J.F. Williamson, Update of AAPM task group no. 43 report: A revised AAPM protocol for brachytherapy dose calculations, *Med. Phys.* 31 (3) (2004) 633–674, <http://dx.doi.org/10.1118/1.1646040>.
- [32] M.J. Rivard, W.M. Butler, L.A. DeWerd, M.S. Huq, G.S. Ibbott, A.S. Meigooni, C.S. Melhus, M.G. Mitch, R. Nath, J.F. Williamson, Supplement to the 2004 update of the AAPM task group no. 43 report, *Med. Phys.* 34 (6 Part1) (2007) 2187–2205, <http://dx.doi.org/10.1118/1.2736790>.
- [33] G. James, D. Witten, T. Hastie, R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, Springer-Verlag New York, 2014.
- [34] R.Y. Kim, A.F. Dragovic, A.C. Whitley, S. Shen, Image-guided brachytherapy for cervical cancer: Analysis of D2 cc hot spot in three-dimensional and anatomic factors affecting D2 cc hot spot in organs at risk, *Brachytherapy* 13 (2) (2014) 203–209, <http://dx.doi.org/10.1016/j.brachy.2013.06.005>.
- [35] S. Anbumani, N.A. Nambiaraj, S. Dayalan, K. Ganesh, P. Anchineyan, R.S. Bilimaggala, A brachytherapy plan evaluation tool for interstitial applications, *Adv. Bioinform.* 2014 (2014) 376207, <http://dx.doi.org/10.1155/2014/376207>.
- [36] D. Todor, P. Lin, I. Barani, M. Axente, M. Hagan, M. Anscher, OC-37 the prostate tumorlet project. critical analysis of D90: tilting at windmills? *Radiat. Oncol.* 103 (2012) S15–S16, [http://dx.doi.org/10.1016/S0167-8140\(12\)72004-2](http://dx.doi.org/10.1016/S0167-8140(12)72004-2).
- [37] W. Wang, A.N. Viswanathan, A.L. Damato, Y. Chen, Z. Tse, L. Pan, J. Tokuda, R.T. Seethamraju, C.L. Dumoulin, E.J. Schmidt, R.A. Cormack, Evaluation of an active magnetic resonance tracking system for interstitial brachytherapy, *Med. Phys.* 42 (12) (2015) 7114–7121, <http://dx.doi.org/10.1118/1.4935535>.
- [38] A. Bouter, T. Alderliesten, B.R. Pieters, A. Bel, Y. Niatsetski, P.A. Bosman, GPU-Accelerated bi-objective treatment planning for prostate high-dose-rate brachytherapy, *Med. Phys.* (2019) <http://dx.doi.org/10.1002/mp.13681>.
- [39] K. Karouzakis, M. Lahanas, N. Milickovic, S. Giannouli, D. Baltas, N. Zamboglou, Brachytherapy dose-volume histogram computations using optimized stratified sampling methods, *Med. Phys.* 29 (3) (2002) 424–432, <http://dx.doi.org/10.1118/1.1450131>.
- [40] M.C. Meer, P.A.N. Bosman, B.R. Pieters, Y. Niatsetski, N. Wieringen, T. Alderliesten, A. Bel, Sensitivity of dose-volume indices to computation settings in high-dose-rate prostate brachytherapy treatment plan evaluation, *J. Appl. Clin. Med. Phys.* 20 (4) (2019) 66–74, <http://dx.doi.org/10.1002/acm2.12563>.
- [41] I. Gronau, S. Moran, Optimal implementations of UPGMA and other common clustering algorithms, *Inform. Process. Lett.* 104 (6) (2007) 205–210, <http://dx.doi.org/10.1016/j.ipl.2007.07.002>.