Clinically Meaningful Dose Distance for Radiotherapy

Paul Dubois[1,2,3][0009−0003−3856−8048], Nikos Paragios[1,2][0000−0002−9668−4763], Paul-Henry Cournède[1][0000−0001−7679−6197]

[1] CentraleSupélec
{p.dubois,nikos.paragios,paul-henry.cournede}@centralesupelec.fr
https://www.centralesupelec.fr/
[2] TheraPanacea {p.dubois,n.paragios}@therapanacea.eu
https://www.therapanacea.eu/
[3] Institut du Cancer de Montpellier paul.dubois@icm.unicancer.fr
https://www.icm.unicancer.fr/fr

Abstract. **Keywords:** Oncology · Radiotherapy · Dose comparison · IMRT · Clinical meaning · Dose volume histogram
Radiotherapy is an essential component of cancer treatment and plays a crucial role in improving patient outcomes and quality of life. With the advancements in technology and the development of sophisticated treatment planning tools, radiation therapy has become more precise and tailored to each patient's individual needs. It is now widely used for a variety of cancer types, either alone or in combination with surgery and chemotherapy.
One of the key challenges in radiation therapy is to assess the quality of treatment plans and verify the accuracy of the radiation delivery. This is particularly important when comparing different treatment plans or treatment techniques, or when evaluating the impact of setup or anatomical changes during the course of treatment.
In this article, we propose a method for comparing radiation doses based on the Dose Volume Histogram (DVH), which captures the importance of different areas of the patient's volume. Our goal is to provide a reliable and clinically relevant tool for dose comparison and evaluation, which can ultimately improve the quality and safety of radiation therapy for cancer patients.

1 Introduction

Radiotherapy is a widely used treatment for cancer, which involves the use of ionizing radiation to kill cancer cells. Intensity-modulated radiation therapy (IMRT) is a type of radiotherapy that delivers high doses of radiation to the tumor while minimizing the radiation dose to the surrounding healthy tissues [5]. Classical IMRT techniques involve the use of several beams (usually 5,7 or 9) of radiation that are directed at the tumor from different angles (usually equispaced angles around the patient) [1]. The intensity of each beam is modulated, or adjusted, to

deliver a higher dose of radiation to the tumor and a lower dose to the surrounding healthy tissues; it has been shown that this technique is much more efficient than 3D-CRT [7] [10] [13] This is achieved by using a computer-controlled device called a multi-leaf collimator (MLC), which shapes the radiation beam to conform to the shape of the tumor.

The success of a radiotherapy treatment plan depends on the optimization process, which involves several steps to ensure the optimal delivery of radiation (that is, delivery that best matches doctor's prescriptions). The optimization process is typically performed using computer software that takes into account the patient's anatomy, the location and size of the tumor/organs, and the radiation objectives defined by doctors. Radiotherapy dose optimization is an inverse problem[14], where the goal is to determine the radiation dose distribution that best matches the desired treatment outcome.

Pre-dose-optimization The first step in the optimization process is to create a virtual model of the patient's anatomy using medical imaging techniques such as CT or MRI scans. This model is then used to identify the location and size of the tumor and to delineate the surrounding healthy tissues that need to be spared from radiation exposure. The next step is to define the radiation dose required for effective treatment. Most of the time, doctors give dose-volume objectives (e.g.: 95% of the PTV should receive at least 75Gy). This is based on the size, location, and type of tumor, as well as the patient's medical history and overall health. The radiation dose is typically expressed in units of Gray (Gy), which is a measure of the amount of radiation energy absorbed by the tissue. These last two steps (structures contouring and dose objectives definition) are performed by doctors.

Radiotherapy doses Once the radiation dose is defined, the optimization process begins. By simulating one MLC parametrization on the patient's body (using the density of the tissues from the scan), we can compute the dose on the patient's volume. This dose is a map from the volume of the patient's body (a region of $\mathbb{R}^3$) to a number of Grays (a number in $\mathbb{R}^+$, since it is impossible to have a negative number of Grays). In practice, we use a discrete version of the dose, that is, for each voxel of the patient's body, we calculate the dose (in Grays).

Dose Optimization To solve the inverse problem of radiotherapy dose optimization, mathematical optimization algorithms are used. These algorithms search for the optimal solution that minimizes a predefined objective function, which represents the treatment goals and constraints.

In order to efficiently solve this optimization problem, the objective function is often designed to be convex, thereby ensuring a well-defined target for the optimization process. Gradient-based, Newtonian, or quasi-Newtonian algorithms are typically employed for this purpose. Despite recent advances in optimization algorithms, determining how closely the best solution found by the optimizer approximates the exact minimum of the objective function remains a challenging task.

Dose-Volume Histograms Doctors have contoured the different structures of interest (that is, associate to which voxels correspond each structure). Thus, it is possible, given a dose, to calculate the dose-volume histogram (DVH) for each structure of interest. The dose-volume objectives then turn to points on the dose volume histogram that the DVH must stay above (in the case of a minimal dose constraint), or under (in the case of a maximal dose constraint).

Formally, for a given dose d, $V(d)$ is defined as the volume of the structure that receives a dose equal to or higher than d. It is mathematically expressed as:

$$V(d) = \frac{Vol(\{p \in \mathcal{V} | D(p) \leq d\})}{Vol(\mathcal{V})}$$

where $\mathcal{V}$ is the set of points of the structure of interest, and $D(p)$ is the dose at particular point p. Then, the DVH curve is the plot of $V(d)$ on the range of d (0 to maximal dose).

Dose Evaluation To evaluate the quality of a dose, doctors look at the 3D distribution of the dose on the patient body (the inter-structure distribution, the existence/number/localization of hot points). They also look at the DVH, and check if the DVH curves matches the DVH objectives.

Stop Criterion Defining a meaningful stopping criterion for the optimization process is a critical challenge in the context of radiotherapy dose optimization. While optimization in clinical practice is often guided by dosimetrists, who may stop optimizing when they are satisfied with the result, the need for full automation of the optimization process necessitates the establishment of a systematic and objective stopping criterion. One possible approach to defining a stopping criterion is to compare the clinical effect of two doses. This would provide a useful tool for comparing different solutions and determining the point at which the optimization process can be terminated.

2 Method

It is not obvious to measure how far apart two doses are, especially in terms of clinical effect: some areas of the patient's volume are more important than others, and dose deposition in some area may, or not, be compensated by dose deposition elsewhere.

2.1 Naive Doses Comparison

The most naive way to compare (and by defining a distance) between two doses is to compare voxel-wise values of the two doses. However, this naive way of comparing doses does not take into account the fact that some extra dose in a region of the patient's body may be compensated by an extra dose in another region. Thus, the voxel-wise distance between two doses may be far while the clinical effect of the two doses is very similar.

We have built a basic example, illustrated in figure 1. This pathological example is derived from a phantom fake patient made of pure water, with one fake cubic PTV and one fake cubic Organ. While acknowledging the lack of realism, the showcased scenario effectively underscores the necessity for employing advanced techniques in dose comparison that surpass mere voxel-wise comparisons.

2.2 Doses Samples

In the rest of this study, we evaluated our proposed method for comparing radiation doses using the TG-119 [8] Prostate case, a widely used benchmark for evaluating radiation therapy plans. TG-119 are provided with dose goals; we used these dose goals to construct our cost function. The cost is a weighted sum of multi-objectives function with one objective for each dose goal, it is constructed as follows:

$$f(\mathbf{d}) = \sum_{o \in \mathcal{O}} w_o f_o(\mathbf{d})$$

$$f_o(\mathbf{d}) = \sum_{d \in \mathbf{d}[o_s]} (d - o_d)_+^2 \text{ if } o \text{ is maximum dose volume constraint}$$

$$f_o(\mathbf{d}) = \sum_{d \in \mathbf{d}[o_s]} (o_d - d)_+^2 \text{ if } o \text{ is minimum dose volume constraint}$$

with:

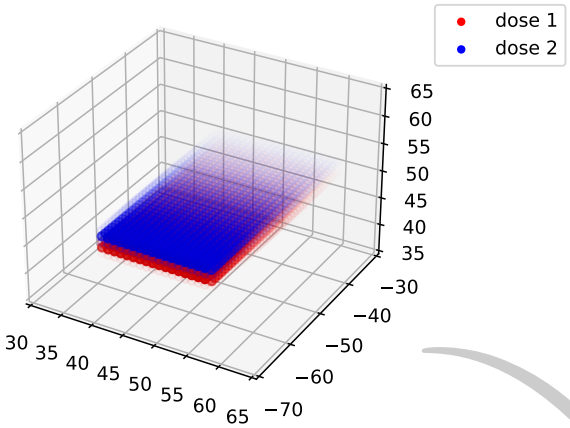
- $\mathbf{d}$ the dose (voxel-wise); $\mathbf{d}[s]$ the dose on voxels of the structure s.
- $\mathcal{O}$ the set of objectives (dose volume goals)
- w_o is the weight of the objective $o \in \mathcal{O}$ (typically ranges between 0 and 100)
- o_s, o_d & o_v the structure, dose and volume goals of $o \in \mathcal{O}$ (e.g.: o_s: PTV; o_d: 80Gy; o_v: 95%)

i.e. a square over/under-dose penalty function. Additionally, we added a regularization term on the bixels[4]: we penalized the variation of bixel values with their neighbors, also with a square penalty. We optimize the bixel values $\mathbf{b}$ with $\mathbf{d} = \mathbf{L}\mathbf{b}$ where $\mathbf{L}$ is a precomputed dose-influence matrix from bixels to voxels. Note that since it is impossible to end negative energy rays, the value of each bixel should be positive ($b \geq 0 \quad \forall b \in \mathbf{b}$)[5].

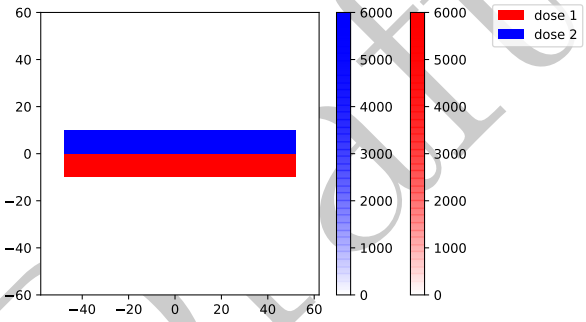
This function is, by construction, convex. Therefore, for a given set of weights, minimizing it should always converge to the same radiotherapy plan. To generate different treatment doses for the same patient case & constraints, we optimized with different weights assignment to each constraint.

[4] A **bixel** (**be**am-**el**ement) represents a discrete element of the radiation beam that can be individually controlled to achieve spatially varying intensity levels; ours are 5mm width, which correspond to the width length of the multi-leaf collimator commonly used.

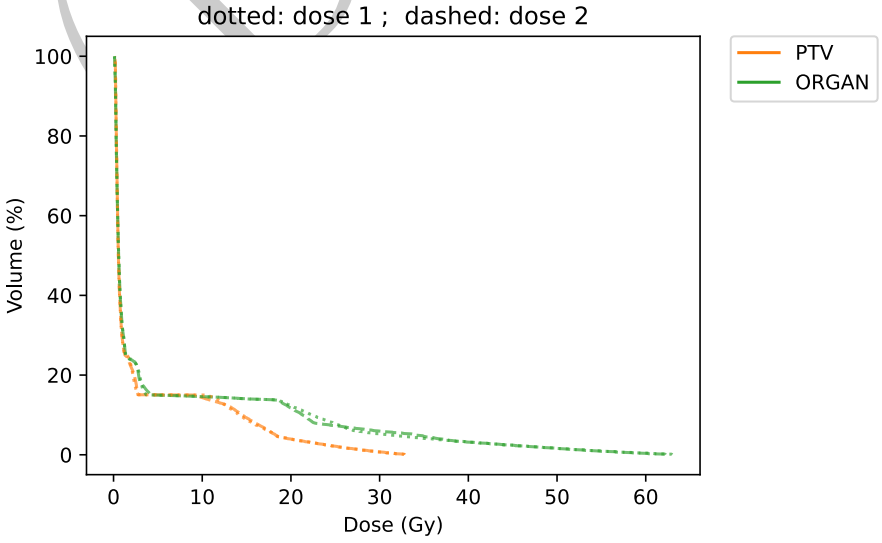
[5] In practice, to ensure positive bixel values, we just compute $\mathbf{d} = \mathbf{L}|\mathbf{b}|$, where $|\mathbf{b}|$ is absolute value of $\mathbf{b}$ element-wise.



(a) 3D visualization of the two doses on the phantom body.



(b) 2D view of the bixels activation of the two doses.



(c) Dose-Volume Histograms of the two doses on the phantom body.

Fig. 1: Example of two doses that have the same clinical effect (measured from the DVHs), but very different voxel-wise dose values.

The optimization of this main objective function was done with the open source optimizer L-BFGS. This optimizer was shown to have the best performances among the open-source optimizers [3].

2.3 Distances Between Doses

Comparing Doses Voxel-wise If two doses are very close together, than voxel-wise comparison is a good measure: we can suppose that the global repartition is similar, and voxel-wise comparison is a good tool to compare local distribution of the dose. Mathematically, we take $\mathbf{d}$, the dose voxel-wise, and define the distance between $\mathbf{d}_1$ and $\mathbf{d}_2$ as the norm of the difference: $\sum_{v \in \mathcal{V}} |d_1(v) - d_2(v)|$, v being voxels in the set of voxels of interest $\mathcal{V}$, and $d_i(v)$ the value of the dose d_i on voxel v (we usually write $|\mathbf{d}_1 - \mathbf{d}_2|$, implicitly summing over voxels).

However, if two regions of equal volume, located in the same structure have their dose values swapped, the voxel-wise difference would be high, although clinically, the two doses are equivalent. Thus, voxel-wise distance is not a good measure between doses for comparing global distribution.

Comparing Dose Volume Histogram Curves We propose to compare the curves on the Dose Volume Histogram (DVH). We have one curve for each structure; we define the distances between doses for each structure, and in the end, we sum up for all structures to end up with a single scalar distance between two doses.

Discrete DVH Approximation The DVH is obtained after sorting the voxel-wise dose of the structure: Let $\mathbf{d}[s]$ bet the voxel-wise dose of the structure s (therefore, a list, of length $n(s)$, the number of voxels that belong to the structure). Let $\mathbf{d}\dot{[}s]$ be the list above in descending order (i.e. $\mathbf{d}\dot{[}s]_i > \mathbf{d}\dot{[}s]_j$ if $0 < i < j \leq n(s)$). Then, the DVH of s can be approximated by the continuous line composed of the segments linking the following points: $\left(\mathbf{d}\dot{[}s]_i, i/n(s)\right) \quad 0 < i \leq n(s)$. Since we compute the dose voxel-wise, we may only have an approximation of the DVH. However, in practice, most structures of interest have more than a hundred voxels, which makes the DVH approximation very precise.

Since we draw one curve per structure of interest, this capture some of the importance of voxel over others. In fact, when analyzing a dose, doctors look at the dose volume (voxel-wise), but they also take a close look at the DVHs; this is an incentive that DVHs should contain meaningful information.

To measure how different two DVH curves, we can imagine several techniques:

- Fréchet distance (treating DVHs as curves in a 2D space)
- Hausdorff distance (treating DVHs as 1D manifolds in a 2D space)
- Wasserstein distance (treating DVHs as probability distributions)
- Kolmogorov-Smirnov test (treating DVHs as probability distributions)
- Total variation between curves (treating DVHs as functions)

We tried all the distances above, but propose to retain only the one giving the best results.

Fréchet Distance We can look at DVH curves as lines in $\mathbb{R}^2$ [6]. In this case, the Fréchet distance is a well known measure of similarity between two curves, and is particularly used for poly-lines [4]. It measures the minimum distance that must be covered by a particle moving along two curves simultaneously. In our case, these curves are the DVH curves of two radiation doses. The Fréchet distance is therefore, a metric of similarity between two curves.

Formally, if P and Q are the curves being compared, γ being a parametrization defined on the interval $[0,1]$, $P(\gamma(t))$ and $Q(\gamma(t))$ the positions of a particle along curves P and Q, respectively, at time t, the Fréchet distance is then defined as:

$$d_{\text{Fréchet}}(P,Q) = \inf_{\gamma} \max_{t\in[0,1]} d(P(\gamma(t)), Q(\gamma(t)))$$

Now, adapting this to DVH curves: Let $\mathcal{C}_A$ and $\mathcal{C}_B$ represent the (discrete) DVH curves of two doses. Let them respectively be consisting of line segments joining the series of points $\{\mathcal{C}_A(i) = (d_i, v_i),\ 1 \le i \le n_A\}$ and $\left\{\mathcal{C}_B(j) = (\tilde{d}_j, \tilde{v}_j),\ 1 \le j \le n_B\right\}$; where d_i and $\tilde{d}_j$ denote the dose levels, v_i and $\tilde{v}_j$ represent the corresponding volumes, and n_A and n_B are the number of points forming $\mathcal{C}_A$ and $\mathcal{C}_B$[7].

The Fréchet distance is the infimum of all possible traversal times. The curves being discrete line segments, we can express the Fréchet distance as:

$$d_{\text{Fréchet}}(\mathcal{C}_A, \mathcal{C}_B) = \min_{\substack{j:[\![1,n_A]\!]\to[\![1,n_B]\!] \\ j\nearrow \ (j \text{ increasing})}} \sum_{i=1}^{n_A} dist\left(\mathcal{C}_A(i), \mathcal{C}_B(j(i))\right)$$

$$\text{where } dist\left(\mathcal{C}_A(i), \mathcal{C}_B(j(i))\right) = \sqrt{(d_i - \tilde{d}_{j(i)})^2 + (v_i - \tilde{v}_{j(i)})^2}$$

Here, j represents a (discrete) a parametrization, and $dist\left(\mathcal{C}_A(i), \mathcal{C}_B(j(i))\right)$ is the distance between points $\mathcal{C}_A(i)$ and $\mathcal{C}_B(j(i))$.

The main drawback of this distance measure is that it is computationally expensive, especially for structures with many voxels. Thus, we had to couple it with an algorithm for curve simplification: the Ramer–Douglas–Peucker algorithm [11]. We used the Ramer–Douglas–Peucker algorithm with $\varepsilon = 0.05$. After testing on a few curves, using this algorithm made calculations 3-5 times faster, while affecting the value of the Fréchet distance by less than 0.5%. It is therefore used in the results presented thereafter.

Hausdorff Distance The Hausdorff distance is another metric of similarity between two curves [6]. It is defined as the maximum distance between any point on one curve to its closest point on the other curve. Formally, let X and Y be two non-empty sets; the Hausdorff distance between X and Y, denoted as $d_{\text{Hausdorff}}(X,Y)$, is defined as:

$$d_{\text{Hausdorff}}(X,Y) = \sup_{x\in X} \inf_{y\in Y} dist(x,y)$$

[6] (Or, in the case of our voxel-wise dose approximation, poly-lines in $\mathbb{R}^2$.)

[7] Since we are always comparing two DVH curves of the same structure, we always have $n_A = n_B$.

where $dist(x, y)$ is the distance between points x and y.

In our analysis, we treat the DVH curves as sets of points in a 2D space ($\mathbb{R}^2$) and use the Hausdorff distance to measure the difference between them. Using the same notation for DVH curves $\mathcal{C}_A$ and $\mathcal{C}_B$ as before, the discrete Hausdorff distance can be computes as follows:

$$d_{\text{Hausdorff}}(\mathcal{C}_A, \mathcal{C}_B) = \max_{i \in [\![1, n_A]\!]} \min_{y \in \mathcal{C}_B} dist(\mathcal{C}_A(i), y)$$

with $\mathcal{C}_B = \left\{ \left((1-\lambda)\tilde{d}_j + \lambda \tilde{d}_{j+1}, (1-\lambda)\tilde{v}_j + \lambda \tilde{v}_{j+1} \right) \mid \lambda \in [0,1], j \in [\![1, n_B - 1]\!] \right\}$, and $dist$ the 2D distance (as before).

Wasserstein Distance The Wasserstein distance, also known as the Earth Mover's Distance, is a metric that quantifies the difference between two probability distributions [9].

Formally, letting μ and ν be two probability distributions defined on a metric space X; the Wasserstein distance, denoted as $d_{\text{Wasserstein}}(\mu, \nu)$, is defined as the infimum cost of transporting the mass from the distribution μ to the distribution ν, where the cost is determined by the distance metric $dist$ on X. That is:

$$d_{\text{Wasserstein}}(P, Q) = \inf_{\gamma \in \Gamma(\mu,\nu)} E_{(x,y)\sim\gamma} \left[dist(x, y) \right]$$

where $\Gamma(\mu, \nu)$ is all possible joint distributions $\gamma(x, y)$ with marginals μ and ν.

In our analysis, we treat the DVH curves as probability distributions and use the Wasserstein distance to measure the difference between them. The main advantage of this distance measure is its ability to capture both local and global differences between the curves. However, it can be computationally expensive, especially when dealing with DVH of large structures.

Kolmogorov-Smirnov Distance Another distance metric that can be used to compare two dose volume histogram (DVH) curves is the Kolmogorov-Smirnov (KS) distance [12]. The KS distance measures the maximum vertical distance between the two curves and is particularly useful when comparing two non-parametric distributions, such as the DVH curves.

Mathematically, let the two DVH curves be described by f and g (from the dose domain to the volume ratio domain), we then have the KS distance d_{KS} given by:

$$d_{KS} = \sup_{x \in \mathbb{R}^+} |f(x) - g(x)|.$$

In our case of discrete DVH, f and g will be linear by segment continuous functions from $\mathbb{R}^+$ to $[0, 1]$ (with value zero above the maximal dose).

Total Variation Distance Finally, we propose a distance metric that computes the integral of the absolute difference between the two DVH curves. This distance measure is simple to compute and provides a good balance between capturing

local and global differences between the curves [2]. Moreover, it is computationally efficient and can handle large structures with many voxels. It is the one that gave the best results, hence, we chose this distance for our analysis.

Initially, the total variation distance measures the integral of the absolute difference between two dose volume histogram (DVH) curves. The dose is (theoretically) unbounded, while the relative volume is bounded (0-100%). We preferred to integrate over a bounded volume instead of an unbounded dose. Hence, we flipped the x and y axes before integrating the absolute dose difference on the range $[0, 1]$; that is, we have the dose on the y axis, and the volume on the x axis.

Mathematically, standard DVHs are given by $V : \mathbb{R}^+ \to [0, 1]$. With two DVHs $V(d)$ and $\tilde{V}(d)$, the total variation will be

$$d_{\text{TotalVariation}} = \int_0^{+\infty} |V(x) - \tilde{V}(d)|.$$

However, we choose to express DVHs with dose as a function of volume, that is $D : [0, 1] \to \mathbb{R}^+$. With two DVHs $D(v)$ and $\tilde{D}(v)$, the total variation will be

$$d_{\text{TotalVariation}} = \int_0^{1} |V(x) - \tilde{V}(d)|.$$

The value of the integral should not change (in theory); we prefer to integrate over a finite domain ($[0, 1]$) than over an unbounded one ($\mathbb{R}^+ = [0, +\infty[$).

An illustration of the differences between classical DVH and after swapping x and y-axes is available on figure 2. The two doses compared were optimized on the TG-119 fake prostate case with different weights (1 and 3) on the PTV objective.

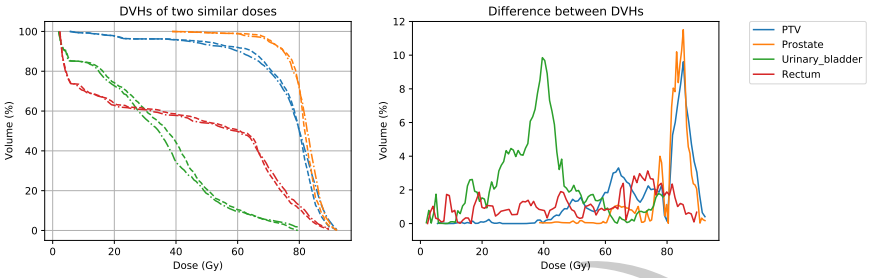
We observe (on fig. 2) that the difference between DVHs is less noisy (it fluctuates less) when we have the dose on the x-axis. This suggest lower numerical error, and is therefore another incentive to have the volume as the x-axis.

The calculation of the total variation distance is computationally efficient, requiring only $\mathcal{O}(n_s)$ operations for each structure, where n_s is the number of voxels in the structure of interest s. Overall, it provides a good balance between capturing local and global differences of DVH curves.

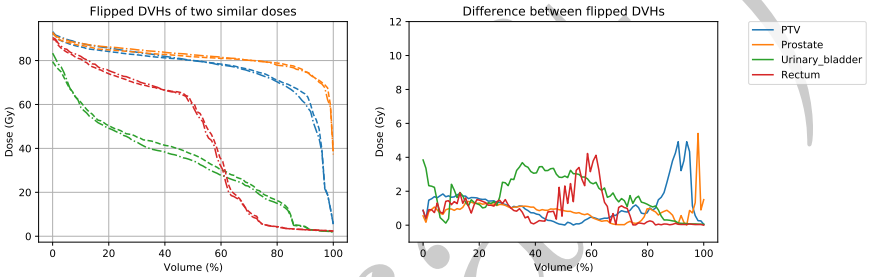
3 Results

3.1 Dose Distances Comparison

For each constraint, we perform an optimization with all weights set to 1, but one, that we sequentially set to 3, 10, 100. This gives us 18 doses that we can compare. We can calculate the pairwise distance between each pair of doses, for each of the distance techniques described above. Considering each optimized dose as a node, we can create a graph for each of the defined distance; and calculating the pairwise distance is equivalent to calculating the adjacency matrix if the



(a) Classical DVH (dose on the x-axis)



(b) Flipped axes DVH (volume on the x-axis)

Fig. 2: DVHs: Comparison of classical and flipped axes styles.

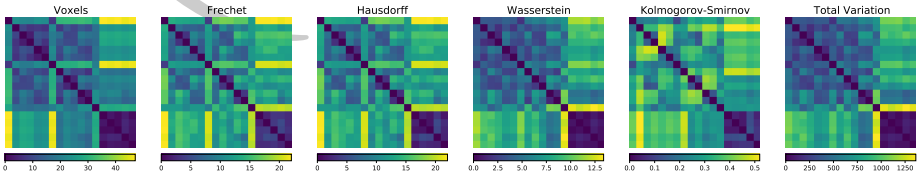


Fig. 3: Pairwise distances between doses
(with different distances calculation method)

associated dense graph. See figure 3 for a comparison of the adjacency matrices.

Ideally, we want a new distance that:

- matches the voxel distance when voxel distance is low
- remains small in some cases of a high voxel distance: when the clinical meaning og the two doses is the same, despite being voxel-wise (very) different

Observing pairwise distances on fig. 3, we have that:

- Frechet and Haussdorff are very similar to voxel-wise; we interpret them as being too sensitive.
  They are therefore not suitable for our purpose.
- Kolmogorov-Smirnov degenerates; it probably captured some noise due to approximation in the DVH calculation (small numerical error).
  It is therefore not suitable for our purpose.
- Wasserstein and Total variation seem to produce acceptable results.
  We therefore chose to continue studying these two distances.

3.2 Link between Total Variation and Wasserstein

One may remark that the adjacency matrices for Wasserstein and Total Variation are very similar. And in fact, since we used Earth moving distance (Wasserstein with $p = 1$), there is an equivalence between the two measures. More precisely, the total variation distance can be seen as a special case of the Wasserstein distance.

The Wasserstein distance, also known as the earth mover's distance, is a metric that quantifies the distance between two probability distributions. Given two probability distributions X and Y with CDFs F and G, the Wasserstein distance between them is defined as:

$$W_p(F, G) = \inf_{\pi \in \Pi(F,G)} \left(\iint_{x,y \in \mathbb{R}^2} |x - y|^p d\pi(x, y) \right)^{1/p}$$

On the other hand, the total variation distance between two curves F and G is defined as:

$$\text{TotalVariation}(F, G) = \int_{x \in \mathbb{R}} |F(x) - G(x)| dx$$

The Wasserstein distance is equivalent to the total variation distance for $p = 1$:

$$W_1(F, G) \equiv \text{TotalVariation}(F, G).$$

In the end, the only difference should be numerical error.

3.3 Bounding of Total Variation and Voxel Distance

It is possible to bound the total variation distance with the Voxel Distance. The other way around, however, is impossible, as shown with the example int the introduction (where two doses have close to identical DVHs, but very different Voxel distances).

We will bound the total variation of one DVH, hence being generalizable to the sum of all DVHs total variation distances.

Definitions Let $\mathcal{S}$ be the structure of interest, and $v_{\mathcal{S}}$ the set of voxels associated to the structure, with $n_{\mathcal{S}} = |v_{\mathcal{S}}|$ the number of voxels. Let d and $\tilde{d}$ be the two doses to be compared.

Sorting lists

Lemma 1. *Let $\dot{l}, l^* \in \mathbb{R}^n$. Let $\dot{l}$ be sorted and $\dot{l^*} \in \mathbb{R}^n$ be sorted version of l^*. Then, we have:*

$$|\dot{l} - l^*| \geq |\dot{l} - \dot{l^*}|$$

Proof. Suppose $a < b$ and $c < d$, and WLOG, $a \leq c$.
We have $|a - d| = |a - c| + |c - d|$ so $|a - d| + |b - c| = |a - c| + |c - d| + |b - c|$ using triangle inequality ($|c - d| + |b - c| \geq |b - c|$): $|a - d| + |b - c| \leq |a - c| + |b - d|$.
Thus, with $\dot{l}$ sorted, swapping elements l_i and l_j ($i < j$) of l^* decreases $|\dot{l} - l^*|$ if $l_i \geq l_j$. Now, bubble sort on l^*, we obtain $\dot{l^*}$ doing only permutations satisfying the condition just stated.

Hence, we obtain $|\dot{l} - l^*| \geq |\dot{l} - \dot{l^*}|$ at the end of the bubble sort. □

Corollary 1. *Let $l, l^* \in \mathbb{R}^n$. Let $\dot{l}, \dot{l^*} \in \mathbb{R}^n$ be sorted version of l, l^*. Then:*

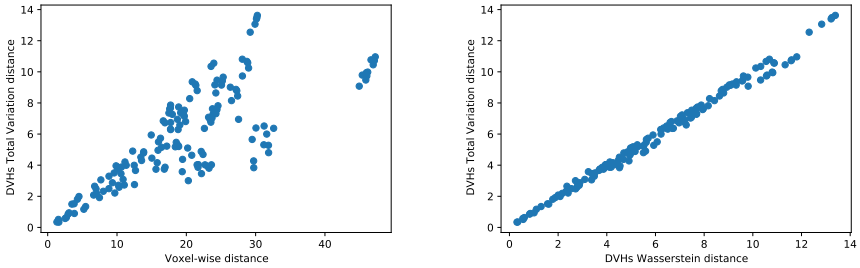
$$|l - l^*| \geq |\dot{l} - \dot{l^*}|$$

Proof. The order in which we perform $|l - l^*| = \sum_{k=1}^{n} |l_k - l_k^*|$ can be chosen, so $|l - l^*| = \sum_{k=1}^{n} |l_{\sigma(k)} - l_{\sigma(k)}^*|$ (with σ a permutation of $[\![1, n]\!]$). Taking σ such that $l_{\sigma(i)} \leq l_{\sigma(j)}$ for $i < j$ and using lemma finishes the proof. □

Proof Outline Suppose the voxel-wise difference is ε-small (i.e. $|d_i - \tilde{d}_i| < \varepsilon$). Then, the total variation of the unsorted vector doses is $|d - \tilde{d}| < n_{\mathcal{S}}\varepsilon$. Let $\dot{d}$ be sorted d and $\dot{\tilde{d}}$ be sorted $\tilde{d}$. Then, by Corollary 1, we have: $|\dot{d} - \dot{\tilde{d}}| \leq |d - \tilde{d}| < n_{\mathcal{S}}\varepsilon$.

Therefore, if d and $\tilde{d}$ are sufficiently close, $\varepsilon \to 0$ and $|\dot{d} - \dot{\tilde{d}}| \to 0$.

Conclusion Thus, voxel-wise very close doses distributions will also have close DVHs distances, which ensure DVHs distances are non-degenerative.



(a) DVHs Total Variation vs Voxel-wise (b) Total Variation vs Wasserstein

Fig. 4: Comparing Distances

3.4 Distances Distribution Comparison

Comparing Total Variation and Voxel-wise The bounding of the total variation DVH distance in terms of voxel-wise distance is clear on figure 4a (there is a linear upper bound to the scatter plot). However, we can remark that some pair of distances are closer in terms of DVH than we imagined using just the voxel-wise distance; hence proving the necessity to analyze further than just the voxel-wise comparison.

Comparing Total Variation and Wasserstein We observe (on figure 4b) that the two DVH distances are almost perfectly proportional. This was expected as they are equivalent mathematically, we only swapped axis of integration in the total variation (hence, the small fluctuations).

4 Discussion

In this article, a novel metric for comparing radiation doses is introduced. This metric has the advantage of being insensitive to dose changes in some regions if being compensated in another region, which was the intended goal. This metric can be useful in several applications, such as dose mimicking and early stopping time for fluence map optimization optimization.

However, there are some limitations to this distance metric. One drawback is that it does not capture spatial distribution, which can be a problem in some cases. There may be some pathological examples where the DVHs are similar, but the clinical interpretation is different. Other factors, such as the distribution of the dose within the target area or the surrounding tissues, can also play a crucial role in determining the effectiveness of the treatment. For example, if two doses have the same amount of high dose, one spread in several small high dose regions, and the other having one large high dose region, then the DVH could look alike, while doctors will interpret the two doses differently. These

types of edge cases are, however, extremely rare in practice. We still advise, for critical cases, to use the voxel invariant metric in conjunction with other techniques to obtain a comprehensive evaluation of radiation doses.

When comparing two different doses, a large distance between them could indicate a significant difference in the intensity or frequency of the treatment, but this does not necessarily mean that one dose is better than the other. The effectiveness of a dose also depends on several other factors such as the individual patient's characteristics, medical history, and response to the treatment. Thus, simply looking at the distance between doses may not provide an accurate picture of which dose is superior or more effective in a particular case. It is essential to consider all relevant factors when evaluating the efficacy of a treatment dose.

Overall, our dose comparison technique is a promising tool. While it has some limitations, it can be a valuable addition to the arsenal of techniques used by radiation oncologists and medical physicists to optimize treatment plans and improve patient outcomes.

Acknowledgment

I thank Antonin Della Noce for his assistance in writing the manuscript.

References

1. Bortfeld, T.: Imrt: a review and preview. Physics in Medicine and Biology **51**(13), R363 (jun 2006). https://doi.org/10.1088/0031-9155/51/13/R21, https://dx.doi.org/10.1088/0031-9155/51/13/R21
2. Chatterjee, S.: Distances between probability measures. https://web.archive.org/web/20080708205758/http://www.stat.berkeley.edu/%7Esourav/Lecture2.pdf
3. Dubois, P.: Radiotherapy dosimetry: A review on open-source optimizer (2023)
4. Efrat, Guibas, H.P.S.M., Murali: New similarity measures between polylines with applications to morphing and polygon sweeping. Reports of Practical Oncology and Radiotherapy **28**(4), 535 – 569 (2002). https://doi.org/10.1007/s00454-002-2886-1, https://doi.org/10.1007/s00454-002-2886-1
5. Ezzell, G.A., Galvin, J.M., Low, D., Palta, J.R., Rosen, I., Sharpe, M.B., Xia, P., Xiao, Y., Xing, L., Yu, C.X.: Guidance document on delivery, treatment planning, and clinical implementation of imrt: Report of the imrt subcommittee of the aapm radiation therapy committee. Medical Physics **30**(8), 2089–2115 (2003). https://doi.org/https://doi.org/10.1118/1.1591194, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.1591194
6. Henrikson, J.T.: Completeness and total boundedness of the hausdorff metric (1999)
7. Kole, T.P., Aghayere, O., Kwah, J., Yorke, E.D., Goodman, K.A.: Comparison of heart and coronary artery doses associated with intensity-modulated radiotherapy versus three-dimensional conformal radiotherapy for distal esophageal cancer. International Journal of Radiation Oncology*Biology*Physics **83**(5), 1580–1586 (2012). https://doi.org/https://doi.org/10.1016/j.ijrobp.2011.10.053, https://www.sciencedirect.com/science/article/pii/S0360301611034614

8. Mynampati, D.K., Yaparpalvi, R., Hong, L., Kuo, H.C., Mah, D.: Application of aapm tg 119 to volumetric arc therapy (vmat). Journal of Applied Clinical Medical Physics **13**(5), 108–116 (2012). https://doi.org/https://doi.org/10.1120/jacmp.v13i5.3382, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1120/jacmp.v13i5.3382
9. Olkin, I., Pukelsheim, F.: The distance between two random vectors with given dispersion matrices. Linear Algebra and its Applications **48**, 257–263 (1982). https://doi.org/https://doi.org/10.1016/0024-3795(82)90112-4, https://www.sciencedirect.com/science/article/pii/0024379582901124
10. Palma, D., Vollans, E., James, K., Nakano, S., Moiseenko, V., Shaffer, R., McKenzie, M., Morris, J., Otto, K.: Volumetric modulated arc therapy for delivery of prostate radiotherapy: Comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy. International Journal of Radiation Oncology*Biology*Physics **72**(4), 996–1001 (2008). https://doi.org/https://doi.org/10.1016/j.ijrobp.2008.02.047, https://www.sciencedirect.com/science/article/pii/S0360301608003428
11. Prasad, D.K., Leung, M.K., Quek, C., Cho, S.Y.: A novel framework for making dominant point detection methods non-parametric. Image and Vision Computing **30**(11), 843–859 (2012). https://doi.org/https://doi.org/10.1016/j.imavis.2012.06.010, https://www.sciencedirect.com/science/article/pii/S0262885612000984
12. Stephens, M.A.: Edf statistics for goodness of fit and some comparisons. Journal of the American Statistical Association **69**(347), 730–737 (1974). https://doi.org/10.1080/01621459.1974.10480196, https://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10480196
13. Verhey, L.J.: Comparison of three-dimensional conformal radiation therapy and intensity-modulated radiation therapy systems. Seminars in Radiation Oncology **9**(1), 78–98 (1999). https://doi.org/https://doi.org/10.1016/S1053-4296(99)80056-3, https://www.sciencedirect.com/science/article/pii/S1053429699800563, radiation Therapy Treatment Optimization
14. Webb, S.: The physical basis of imrt and inverse planning. The British Journal of Radiology **76**(910), 678–689 (2003). https://doi.org/10.1259/bjr/65676879, https://doi.org/10.1259/bjr/65676879, pMID: 14512327