# PREVIS

Furmanová, Katarína; Muren, Ludvig P.; Casares-Magaz, Oscar; Moiseenko, Vitali; Einck, John P.; Pilskog, Sara; Raidou, Renata G.

*Published in:*
Computers and Graphics (Pergamon)

*DOI:*
10.1016/j.cag.2021.04.010



*Publication date:*
2021

Computers & Graphics 97 (2021) 126–138

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on VCBM 2020

# PREVIS: Predictive visual analytics of anatomical variability for radiotherapy decision support

Katarína Furmanová[a,b,*], Ludvig P. Muren[b], Oscar Casares-Magaz[b], Vitali Moiseenko[c], John P. Einck[c], Sara Pilskog[d], Renata G. Raidou[e,f]

[a] Masaryk University, Czechia
[b] Danish Centre for Particle Therapy, Aarhus University Hospital, Denmark
[c] Department of Radiation Medicine and Applied Sciences, UC San Diego, United States
[d] Department of Physics and Technology, University of Bergen, Norway
[e] TU Wien, Austria
[f] University of Groningen, the Netherlands

ARTICLE INFO

*Article history:*
Received 17 January 2021
Revised 10 March 2021
Accepted 13 April 2021
Available online 24 April 2021

*Keywords:*
Medical Visualization
Visual Analytics
Comparative Visualization
Ensemble Visualization
Radiotherapy Planning
Cohort Study

ABSTRACT

Radiotherapy (RT) requires meticulous planning prior to treatment, where the RT plan is optimized with organ delineations on a pre-treatment Computed Tomography (CT) scan of the patient. The conventionally fractionated treatment usually lasts several weeks. Random changes (e.g., rectal and bladder filling in prostate cancer patients) and systematic changes (e.g., weight loss) occur while the patient is being treated. Therefore, the delivered dose distribution may deviate from the planned. Modern technology, in particular image guidance, allows to minimize these deviations, but risks for the patient remain. We present *PREVIS*: a visual analytics tool for (i) the exploration and prediction of changes in patient anatomy during the upcoming treatment, and (ii) the assessment of treatment strategies, with respect to the anticipated changes. Records of during-treatment changes from a retrospective imaging cohort with complete data are employed in *PREVIS*, to infer expected anatomical changes of new incoming patients with incomplete data, using a generative model. Abstracted representations of the retrospective cohort partitioning provide insight into an underlying automated clustering, showing main modes of variation for past patients. Interactive similarity representations support an informed selection of matching between new incoming patients and past patients. A Principal Component Analysis (PCA)-based generative model describes the predicted spatial probability distributions of the incoming patient's organs in the upcoming weeks of treatment, based on observations of past patients. The generative model is interactively linked to treatment plan evaluation, supporting the selection of the optimal treatment strategy. We present a usage scenario, demonstrating the applicability of *PREVIS* in a clinical research setting, and we evaluate our visual analytics tool with eight clinical researchers.



## 1. Introduction

Radiation therapy (RT) is one of the most common cancer treatment methods. It requires careful pre-treatment planning and follows a sophisticated pipeline, which includes the acquisition and registration of anatomical images, delineation of target volumes and organs at risk, and the optimization of the treatment plan [1]. Commonly used fractionation requires that the prescribed dose is delivered over several weeks [2]. During this time, multiple uncertainties may occur and accumulate. Uncertainties due to organ motion lead to deviations in the planned trade-offs between tumor control and healthy tissue complications. In recent years, image-guided RT has been used to monitor the anatomy of the patient prior to treatment administration, and to realign the patient. Still, simple realignment may not suffice, depending on the internal organ changes in size and shape. Adaptive RT, where plan optimization accounts for changing anatomy, has been brought forward to ensure that tumor sites receive the desired dose, while organs at risk are spared [1]. Yet, this process remains resource-intensive.

RT plans are optimized with tumor and organ delineations on a pre-treatment CT scan of the patient. To account for random and systematic uncertainties in patient positioning, internal anatomy and organ motion, safety margins are added to the delineations [1]. These margins are based on population studies; therefore, they may not be adequate for all patients. Some patients may present

* Corresponding author at: Masaryk University, Czechia.
*E-mail address:* katarina.furmanova@gmail.com (K. Furmanová).

https://doi.org/10.1016/j.cag.2021.04.010
0097-8493/

large anatomical changes during treatment, while others are more stable. Accounting for patient-specific motion could be beneficial for all patients. Patients with stable anatomy would require smaller safety margins—resulting in less exposure of normal tissue. In patients with more mobile organs, unnecessary irradiation of healthy tissues can be avoided, if we can anticipate (and accommodate through adaptation) the changes into the dose delivery. Previous work suggests that patterns in anatomical changes of the patients can be identified and anticipated, if few daily observations of the patient's anatomy prior to treatment are available [3,4]. Identifying the type and magnitude of anatomical changes can also indicate which type of treatment modality is the most suitable for a given patient. For example, proton-based RT can offer better localization of the radiation dose, but is more sensitive to anatomical changes than photon-based RT.

We present *PREVIS*, a visual analytics system to support clinical researchers in the exploration, prediction, and assessment of the impact of anatomical changes, occurring during RT. The main contributions of our work are: **(1)** an intuitive representation of the distribution of past patients *within a cohort*, with respect to their anatomy; **(2)** a predictive method for *modeling anatomical changes* of incoming individual patients prior to treatment, by matching them to existing, retrospective cohort data; and **(3)** a predictive strategy for the *pre-treatment assessment and comparison of RT plans*, based on observed and/or modeled anatomical changes of an incoming patient. To the best of our knowledge, a predictive visual analysis of anatomical variability was not possible before *PREVIS*.

This paper is structured as follows. In Section 2, we describe the current clinical practice and research in prostate cancer RT. This is followed by a user—data—tasks analysis (Section 3) and a discussion of related work (Section 4). In Section 5, we describe in detail our system and employed methods. We, then, demonstrate the usability of our system in an exploratory scenario (Section 6), and provide the results of our user evaluation with eight clinical experts (Section 7). Finally, we provide an outlook on future challenges that have not been yet addressed in our approach (Section 8).

## 2. Clinical background

Patients with prostate cancer are commonly treated with external beam radiotherapy (EBRT), where radiation doses are delivered externally using multiple beams or arcs, aimed at the tumor location [1]. When superimposed, these beams accumulate to a high dose for the targeted tumor area, and a lower dose for the surrounding organs. The planned dose is not administered at once, but is delivered in multiple fractions. This means that it is split over several weeks, to allow the recovery of healthy tissue, while minimizing tumor growth. Recent techniques effectively spare healthy tissue, while delivering the desired high dose to the tumor volume. However, parts of healthy organs of the pelvis are still unavoidably irradiated, leading to side-effects and affecting the patient's quality of life.

Recent studies suggest a link between pelvic organ motion/deformation and increased toxicity risks [2]. Anatomical variations may occur naturally across individuals, or may be caused by pathological factors. For example, day-to-day anatomical changes occur, because organs consist of flexible and deformable soft tissues; thus, they are affected by filling changes [5]. Pelvic organs (e.g., the bladder and the rectum) are especially prone to filling changes, and their positions and shape vary significantly [6]. As the inherent complexity of the RT workflow makes it challenging to adapt the treatment plan before every fraction, tumor coverage is prioritized. For this, target volume adjustment on a per-treatment basis has been suggested [7].

Current clinical research aims at choosing the most effective radiotherapy strategy. Treatment plan evaluation and approval is based on the visual examination of the dose distribution with regard to target volumes and organs at risk (on CT slices, or rendered in 3D) and additional plots [1]. First, anatomical views (Fig. 1 (a)) allow experts to see the amount of dose administered to the target volume and surrounding organs for particular instance of patient's anatomy. Second, Dose Volume Histograms (DVHs, in Fig. 1 (b)) provide a 2D representation of the dose distribution for organs, to show how much radiation is administered to the fractional volume of each organ. This supports quick evaluation of compliance with planning objectives and identification of organs at risk of toxicity. Third, Tumor Control Probability (TCP) and Normal Tissue Complication Probability (NTCP) models are statistical models (Fig. 1 (c)). TCP models quantify the probability that a tumor is effectively controlled (i.e., a tumor will not recur), given a specific radiation dose and the sensitivity of the tumor cells to it. NTCP models quantify the probability of normal tissue complications, given a specific radiation dose distribution and its volume effect. In the present work, only the NTCP modeling part will be addressed, as we study only the effect of radiation on the healthy surrounding organs. The anatomical views have scalability limitations in their current form, i.e., they can show only one patient at a time, which is not sufficient to judge the robustness of treatment strategies. The other plots (DVHs, TCP, and NTCP) scale well for the representation of many timesteps and/or patients at the same time, but they do not provide any kind of anatomical information. To facilitate the evaluation of different dose plan alternatives for new incoming patients, our proposed solution integrates anatomical views, DVH-based views and NTCP representations.

Recently, adaptive visual analytics approaches have been investigated, with the purpose of demonstrating the influence of pelvic organ shape and position variability on the dose administration, as well as potential RT-induced toxicity [3,8]. The current paper aims to advance these approaches towards a predictive direction. Our strategy utilizes the information of organ shape and position variability from a retrospective cohort study with complete data, to support clinical researchers in predicting the effect of treatment on new incoming patients with incomplete data.

## 3. User—data—tasks analysis

*Users*. The radiation therapy workflow involves many different clinical experts [1]. This paper is primarily targeting medical physicists and clinical researchers. These experts participate on the design and evaluation of treatment plans. Our system can aid this process by providing insight into the patient's anatomy and its possible changes, and by relating these to radiation dose.

*Data*. We worked with two datasets of prostate cancer patients. For all patients, prostate, bladder, and rectum delineations were available. The first dataset contained eight patients. For each, we had 7–9 timesteps, with anatomical contours based on repeated CTs, acquired twice a week over the treatment period, in addition to the planning CT. For these patients, we had two treatment plans based on two different treatment modalities (one plan for proton and one for photon treatment). The second dataset contained 25 patients. For each, we had 12 timesteps from repeated Cone Beam CTs (CBCTs), in addition to the planning CT. The first five of the repeated CBCTs come from the first five daily sessions of the treatment, and are followed by CBCTs acquired on a weekly basis. For these patients only photon plans were available. All patient data came from clinical protocols, where relevant ethical and institutional review board approvals have been granted.

*Tasks*. In the past, large anatomical changes have been observed during treatment for certain patients, while other patients had a more stable anatomy [3,9]. Also, patterns in the anatomical

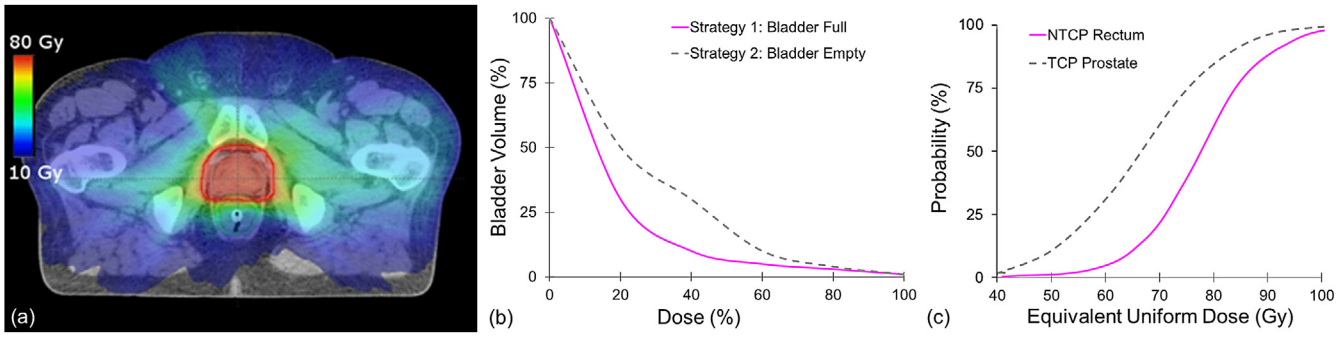


**Fig. 1.** Dose assessment methods. (a) Assessment on an axial CT slice. (b) Dose Volume Histograms (DVHs). (c) Tumor Control Probability (TCP) and Normal Tissue Complication Probability (NTCP).

changes of patients could be identified in the first few fractions of the treatment [4]. Population-based safety margins are often added to the tumor volumes [1], but the current workflow, as discussed in Section 2, does not integrate anatomical variability to the treatment plan. To provide more robust treatment plans with smaller margins for stable patients, and to avoid unnecessary irradiation of healthy tissues in more mobile organs, we determined the main tasks for RT decision support through a predictive visual analytics solution. These tasks have been identified in a long iterative design process. They are based on discussions between the authors of this paper (visualization and clinical experts) and on a detailed analysis of current practices in clinical research, regarding organ motion modeling and treatment evaluation. The tasks are the following:

- **T1 Cohort Analysis:** Analyze a "past" cohort, with respect to anatomical similarities and intra-treatment changes.
- **T2 Patient Matching:** Analyze the anatomical similarities of incoming patients to "past" patients, with complete treatment.
- **T3 Generative Anatomical Modeling:** Predict and analyze the anatomical changes of incoming patients, prior to treatment.
- **T4 Treatment Assessment:** Based on the anticipated anatomical changes, quantify and assess the most suitable treatment.
  - **(a)** How do anatomical changes affect the delivered **dose**?
  - **(b)** How do anatomical changes affect the **NTCP**?
  - **(c)** How can we **compare** multiple treatment plans?

## 4. Related work

In *shape space analysis*, Hermann et al. [10,11] investigate anatomic covariances in ensembles of data, providing also a state of the art report with prospects on the visual analysis of shapes. Busking et al. [12] propose a 2D scatter plot to represent the distribution of elements inside a cohort and to synthesize additional arbitrary objects in the shape space. For comparing objects, they later deal with visualizing intersecting 3D surface meshes [13]. Reiter et al. [14] explore and analyze the variability in multiple pelvic organs, based on spherical harmonics, coupled to dimensionality reduction strategies, such as t-distributed Stochastic Neighborhood Embedding (t-SNE) and Principal Component Analysis (PCA). Generally, former works support the efficient differentiation of shapes, but lack the ability to synthesize and model shapes and motion thereof.

*Shape and motion modeling* has been investigated in clinical studies to evaluate margins for multiple targets in radiotherapy for high-risk prostate cancer using statistical motion models [15–17]. Rios et al. [18] developed a population-based model for prostate cancer radiotherapy that models the bladder motion and deformation based on dominant eigenmodes and mixed-effects models. Similar eigenmode-based approaches have been used for geometric modeling also by other works [19,20]. Yet, none of these approaches goes beyond modeling and does not provide any interactive tools for the visual analysis and exploration of the underlying cohort, coupled to inter-fractional variability and its relation to the evaluation of the treatment plan, which we include in our work.

Previous work in the domain of *uncertainty visualization for medicine* [21] has, among other topics, investigated anatomical variability. Our application might involve additional sources of uncertainty [22], such as from imaging [23,24], but—in the context of this work—we address solely anatomical variability. Von Landesberger et al. proposed tools to support the global analysis of anatomical variability in an examined population [25,26]. Smit et al. presented a novel, educational way of visualizing anatomical variation in complex branching structures, such as arteries or nerves [27], while an older approach within the VOXEL-MAN project investigated morphological organ variations.

In *cohort analysis*, Klemm et al. [28] focus on the extraction of spine-canal variability and the exploration of clusters of similarly shaped spines. This work has been extended to incorporate additional patient information [29], demonstrating how to effectively reduce and visualize image cohort data and to facilitate their understanding. Steenwijk et al. [30] also go beyond shape analysis by proposing a framework for the interactive and structured visual analysis of cohort data. Cohort analysis has also been tackled by Preim et al. [31], Bernard et al. [32] and Alemzadeh et al. [33], for various purposes—also, within *ensemble visualization* [34].

Specifically for *RT*, Wentzel et al. [35] presented a visual computing approach for the estimation of RT plans in head and neck cancer patients, where anatomical similarity based on topology and measures of image fidelity were considered. With this approach, it is still not possible to derive any information about potential RT-induced toxicity. The *Bladder Runner* is a visual analytics system that provides information about the amount of radiation delivered to the bladder across the treatment for a cohort of patients; therefore, also about toxicity risks [8]. The entire approach is based on a 14-D shape descriptor vector for the bladder cohort, which is fed into dimensionality reduction and clustering to detect cohort partitions with similar bladder shapes and evolutions through the treatment period. Later, this was extended into the *Pelvis Runner* to support multiple pelvic organs, by changing the shape description method to support also non-spherical shapes (i.e., the rectum) through linearization [36]. However, the *Pelvis Runner* still did not support the correlation to dose administration, the analysis of its variability and the investigation of potential RT-induced toxicity, which was added in *VAPOR* [3]. Methods from *VAPOR* have also been used to estimate the magnitude of organ variability for individual patients [4]. However, this approach does not support modeling or prediction of anatomical changes.

Regarding the *treatment plan evaluation*, Raidou et al. [37] proposed an approach for the exploration of TCP models, by incorpo-

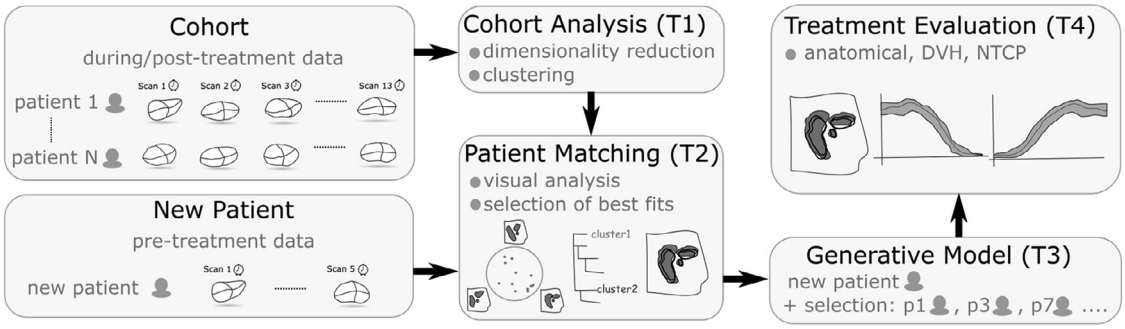


**Fig. 2.** Overview of the analytical workflow in our system. We start by analyzing and preprocessing the existing cohort data. We continue with matching the new patient (with incomplete data) to existing cohort partitions (with complete data). Then, we employ generative modeling to predict upcoming anatomical changes in the new patient. Finally, we assess available treatment strategies, based on our generative model.

rating uncertainty and parameter sensitivity, while enabling inter-patient response variability analysis. This visual analysis tool has been later used in a clinical study [38]. For NTCP models, El Naqa et al. [39] investigated methods for the visualization of the high-dimensional space composed of the interaction between toxicities and treatment, anatomical, and patient-related variables of NTCP in head and neck patients, on the basis of a support vector machine (SVM). Kupchak et al. [40] based, instead, their method for mapping NTCPs onto dose-volumetric regions on a Monte Carlo simulation. There are no other visual analytics approaches to support the integration of the predictive (through the modeling of the organ shape and motion variability) and preventive (through the treatment assessment) aspects, which we integrate into our work.

All previous work focuses solely on retrospective data analysis. For a prospective analysis of new incoming patients with incomplete data, a new approach for inferring their anatomical changes is required. Additionally, previously employed shape analysis methods do not support interactive incorporation of new patients. As we discuss in the upcoming sections, *PREVIS* uses different methods for shape analysis and matching, to enable flexible data exploration and to support the addition of new patients. This also requires the appropriate use of visual representations that provide good insight into patient partitioning, and new strategies for matching new patients to the existing ones. Finally, previous work also lacks support for treatment plan comparison and evaluation, which we now tackle.

## 5. Workflow and functionality of *PREVIS*

We support the previously described tasks **(T1–4)** through an analytical workflow consisting of several steps, as depicted in Fig. 2. Our workflow starts with the analysis and clustering of the retrospective cohort data. When a new patient (with incomplete data) is loaded, he is analyzed and matched to existing cohort partitions (with complete data). Generative modeling is then employed to predict anticipated anatomical changes in the new patient. Finally, different treatment strategies (e.g., proton vs. photon treatment) can be assessed based on the predictive outcomes of the modeling. The steps of our workflow are discussed in detail in the upcoming sections.

### 5.1. Cohort analysis (T1)

Our predictive approach relates new patients to existing patient cohorts, where observations of anatomical changes are done retrospectively. Therefore, our workflow starts with cohort analysis and stratification of past patients, based on their anatomical similarities. Our approach is based on previous work for the analysis of pelvic organ variability [3].

*Data Preprocessing.* Our input data consists of contour delineations for each patient organ at each treatment timestep, and a respective volumetric representation of the administered radiation dose. The contours have been transformed into unified volumetric representations, where the organs are represented by binary coverage masks. The data are aligned based on the center of mass of the prostate. This is a common approach in prostate cancer treatment, as the radiation dose is also centered around the prostate [3]. For a more robust result and to preserve all positional variations, registration based on, e.g., the femoral bones, would be required.

*Shape Description.* We employ a probability density-based 3D shape descriptor, which estimates a probability density of geometric features at selected target points. Our shape descriptor is an extension of the one proposed by Akgül et al. [41,42], to serve better our reconstruction and visualization purposes. This shape descriptor has been demonstrated to be robust for arbitrary shapes and highly efficient, outperforming many other well-known shape descriptors for classification [43]. For many other application scenarios this descriptor might not be suitable, given its lack of rotation and scale invariance. In our case, these are desirable properties, as distinguishing between organs of different sizes and relative positions, is required. The original descriptor proposed by Akgül et al. consists of three features, describing local features at individual points $p$; each, consisting of a direction and a distance component:

- *Radial feature*, consisting of a radial direction $\vec{d_r} = p - m$, where $m$ is the object's center of mass, and a radial distance $|\vec{d_r}|$.
- *Tangent feature*, consisting of a normal direction $\vec{n}$ at point $p$ and a distance of tangent plane at $p$ from object's center of mass $m$.
- *Cross-product feature*, consisting of a cross-product direction $\vec{d_c} = \vec{n} \times \vec{d_r}$ and a radial distance $|\vec{d_r}|$.

While the tangent and cross-product features describe the local shape features of the object at a point $p$, the radial feature directly describes the position of a point $p$. As such, it encodes the coverage probabilities of the organ volumes at the target points (see paragraph *Coverage Probability Volume*, below).

In the original formulation of the descriptor, the target points of individual features are selected as Cartesian products of direction samples (obtained by uniformly sampling a unit sphere), and distance samples (acquired by splitting the range of distance components of a given feature into uniform intervals). With this approach, the reconstruction and visualization of the probability coverage levels (details of which are described below) through radial sampling would be insufficient. Considering the placement of the resulting targets in the 3D space, the targets are very sparse at the upper bound of the radial distance range, and very dense at the lower bound (i.e., close to the center of mass of the object). Therefore, we adjust the descriptor by placing the target points in

a *regular grid*, spanning the bounding box of the evaluated organ shapes. A second adjustment is made for the radial feature, where we evaluate the probability values at the targets based on all volume voxels, while for tangent and cross-product features only surface points are evaluated. We call this modified radial feature *positional feature*. These adjustments allow us to use the positional feature part of the descriptor to represent the volume coverage probability, while the descriptor still fulfills the role of embedding the organ shapes in a low dimensional space that allows for statistical data analysis.

*Clustering*. To create meaningful partitions of our cohort, based on the similarity of the organ shape evolution across patients, we employ a *hierarchical clustering with complete linkage* [44], similarly to the work of Furmanová et al. [3]. Our choice was motivated by the algorithm flexibility, which offers intuitive insight into cluster composition. For this reason, it is also commonly used and well-understood within the medical domain. Given the flexibility of hierarchical clustering, the number of clusters can be adjusted by the user—although an initial value is suggested by the elbow method [45]. The users can select which organs to include in the clustering. In some analytical scenarios, users may focus on a single organ, while in others, the relative positions of the organs also play a role. We chose a complete linkage proximity measure, to avoid the chaining effect of single and average linkage, and bias towards balanced clusters of Ward's method, as shown in the work of Klemm et al. [28], for similar tasks. This method is also less susceptible to noise and outliers, which can be often anticipated in medical data.

In previous related works [3,8] clustering was performed based on only one organ, and the patients were grouped based on the prevailing organ clusters. This is not feasible, when multiple organs need to be considered. Therefore, we employ the following aggregation scheme. We start with individual organ descriptors at individual timesteps. For the three organs—bladder, prostate, rectum—at timestep $i$ we, thus, have three distinct descriptors $b_i$, $o_i$, and $r_i$, respectively. The descriptor of timestep $t_i$ results from chaining the distinct organ descriptors into a single vector $t_i = \{b_i, o_i, r_i\}$.

To identify patients with similar anatomical features throughout treatment, clustering needs to be performed on the patients (not on timesteps). Chaining the available multi-timestep descriptors would lead to very large vectors of inconsistent size. This would be unsuitable input for clustering. It would also introduce artificial time-correspondence among data points. Thus, to obtain a descriptor for a given patient, we compute the mean $\mu$ and standard deviation $\sigma$ for different organs across timesteps and chain them into a patient descriptor $s = \{\mu_b, \mu_o, \mu_r, \sigma_b, \sigma_o, \sigma_r\}$. In this way, we capture for each organ, the average shape and variation. This descriptor is used as input to the clustering.

*Coverage Probability Volume*. For each cluster, we can reconstruct the coverage probability volume. This volume represents the per-voxel probability of encountering a given organ from a given cluster in a 3D space. It inherently encodes uncertainties related to changes of organ shapes. We reconstruct these from the positional feature part of the cluster descriptor. A cluster descriptor $c = \{\mu_b, \mu_o, \mu_r, \sigma_b, \sigma_o, \sigma_r\}$ is computed in the same way as the patient descriptor, but the mean $\mu$ and standard deviation $\sigma$ for different organs are computed from all timestep records belonging to a given cluster, i.e., across different patients. As we previously described, the positional feature is evaluated at a set of target points placed on a regular grid. This is a low-resolution representation of the probability volume, so we resample it using trilinear interpolation.

*Overview Table*. To provide an initial overview of the cohort structure **(T1)**, we adapt the tabular view from VAPOR [3] (Fig. 3). The rows of the table correspond to patients from the cohort and the columns represent timesteps with a recorded state of anatomy.

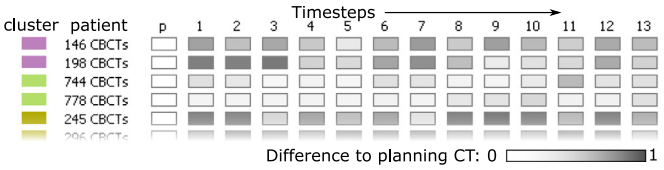


**Fig. 3.** Portion of an Overview Table encoding anatomical differences at individual treatment timesteps (1–13) from planning CT (p).

The cells of the table encode the difference of the anatomy at a given timestep, with respect to the planning CT (first column). These are computed as the Euclidean distance between corresponding timestep descriptors and are normalized according to the maximum distance detected in the cohort (i.e., if a patient with new maximum is added, the encoding is re-scaled). This supports the identification of patients that might benefit from re-planning, due to large anatomical deviations. We place a small color bar at the beginning of each row to indicate cluster adherence of the patients. The table also serves for navigation, as it is interactively linked to the other views.

### 5.2. Patient matching (T2)

After obtaining an overview of the retrospective cohort, we proceed with relating the new patient to the existing complete cohort data. Our system expects as input the volumetric coverage masks of the organs of the incoming patient, which have been acquired and delineated, so far. Ideally, the new patient information would include anatomical data from multiple days, as previous research suggests that this can capture a significant portion of patient-specific organ shape variability [4]. However, our system is designed flexibly, to accommodate patient matching even with one timestep (i.e., only the planning CT).

*Matching calculation*. The data of a new patient are processed in the same way as the cohort data (Section 5.1), to create a low dimensional representation of the organ shape. This enables us to compute the fit of a new patient to the previously identified clusters. The fit is computed as the mean Euclidean distance of the new patient to the patients in the clusters identified during **(T1)**, following the same concepts regarding the considered anatomy as the clustering algorithm (i.e., if the clustering is computed only based on bladder shape, the fit calculation will also consider only the bladder).

*Matching visualization*. To support a detailed view on the matching of a new incoming patient **(T2)**, we employ three linked representations. *A RadViz plot* [46] provides an abstracted overview of the patient distribution within the cohort clusters (Fig. 4, left). A *hierarchy view* [47] supports drilling down to different levels of the hierarchical clustering (Fig. 4, right). Finally, *2D/3D anatomical views* provide an anatomical view on the variability of individual organs (Fig. 5).

In the *RadViz plot*, the clusters, as resulting from the previous step, are placed equidistantly on the circumference of the circle. Each cohort patient is represented as a point within the circle, colored according to the cluster it belongs to—further supporting **(T1)**. The new patient is highlighted with a larger empty circle, and anchor strings connect it to individual clusters. The positions of the points are assigned by a function of attraction to individual clusters (i.e., by how well the patients fit the individual clusters). The fit is computed in the same manner as the fit of the new patient to the clusters, as described above. This enables the users to see how well the clusters are separated. For example, in Fig. 4, patient 198 (see tooltip) belongs to cluster 2, but he is close to the circle's center, indicating that the fit to this cluster is not very strong. The clusters at the circumference of the circle are accompanied by

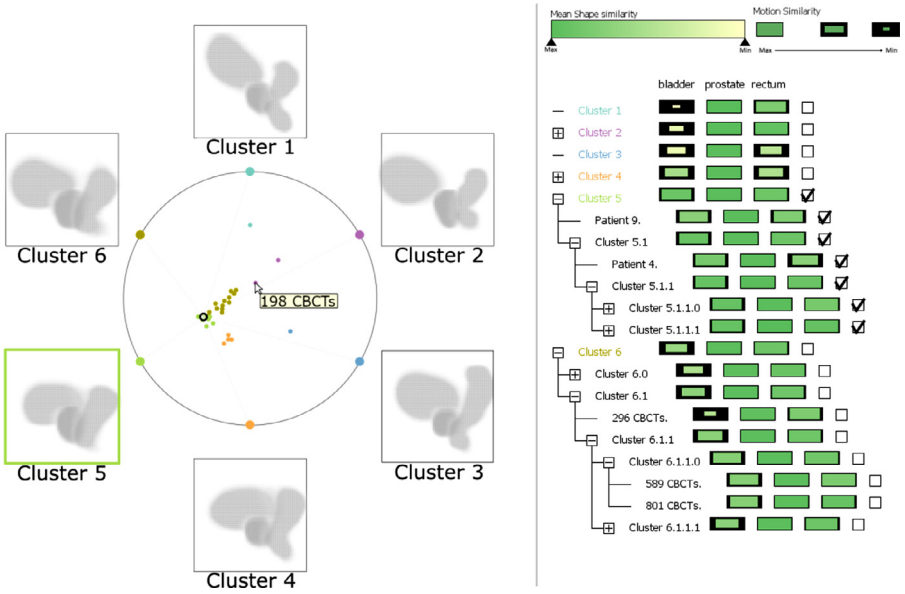


**Fig. 4.** Left: A RadViz plot, showing the distribution of patients into clusters, and anatomical thumbnails of the clusters. Right: A hierarchy plot, showing the data hierarchy, as produced by the hierarchical clustering. The cells encode similarity of individual data points to the new patient. It can be seen that the bladder is the least similar organ in the majority of cases.

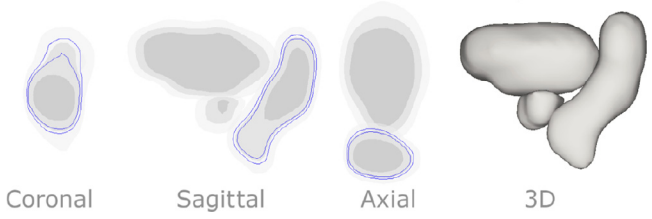


**Fig. 5.** Anatomical views showing three probability coverage levels (25% dark gray, 50% middle gray, and 75% light gray) on the anatomical 2D planes, and a 50% coverage probability level in the 3D view. For the 50% probability level, blue contours indicate standard deviations shown on mouse hover.

anatomical thumbnails. These show the sagittal plane through the center of the coverage probability volume, as computed from the cluster descriptor.

The *hierarchy view* complements the *RadViz plot*, by encoding the hierarchy of the clustering outcome in a foldable tree structure. The sub-clusters (e.g., sub-clusters 6.0 and 6.1, forming cluster 6 in Fig. 4) can also be explored. The hierarchy view is interactively linked to the RadViz plot (i.e, hovering over a hierarchy level, highlights the patients belonging to the respective (sub-)cluster in the RadViz plot). In the hierarchy view, the rows correspond to clusters/patients and the columns to organs. The hierarchy view also encodes the similarity of the new patient to the clusters and patients across individual organs. The similarity is computed as the Euclidean distance between the descriptors corresponding to the given organ. We differentiate between similarity in the mean shapes (obtained by comparing the mean organ descriptors) and similarity in organ variabilities (obtained by comparing the standard deviations of the organ descriptors). The shape similarity is color-coded in the cell of the hierarchy table, and the variability similarity is encoded in the cell size, similarly to previous work [47,48]. Together, the RadViz plot and the hierarchy view support patient matching **(T2)**. While there are numerous alternatives for the representation of hierarchical data (e.g., treemaps or sunburst diagrams), they are not well-suited for encoding complex information about all levels of the hierarchy. The tree structure is a common representation that is easy to understand—also for users outside of the visualization domain.

Additional *anatomical 2D/3D views* complement the anatomical thumbnails of the RadViz plot. The thumbnails provide an overview of the anatomical features of each cluster, and the user can select clusters to explore them in greater detail in larger anatomical views. These include the three anatomical 2D planes (sagittal, axial, and coronal) to show crosscuts across the coverage probability volume and an accompanying 3D view. As discussed above, the coverage probability volume is reconstructed from the mean shape descriptor. In the 2D anatomical views, we show the 25%, 50%, and 75% coverage probability levels, similarly to the work of Raidou et al. [8]. These are color-coded with a luminance scale (lighter equals to higher coverage probability). However, these isovalues can be adjusted to show, e.g., 90%-confidence interval [3]. For the shape variations, we use contours, shown upon hovering the individual coverage levels to indicate the standard deviation of the respective coverage level range. This is shown with additional contours, e.g., in blue in Fig. 5, similar to topographic contours [49]. In the 3D view, an isosurface representation of the volume is shown at a selected iso-level (by default, this is set to the 50% probability level).

### 5.3. Generative anatomical modeling (T3)

The previously discussed techniques facilitate the analysis of cohort clusters and the selection of patients that are most similar to the new patient. Given that for the new patient we have incomplete data (from only one or few timesteps of treatment), this information can serve as an input to a statistical modeling that predicts organ changes for the new patient. Our system automatically recommends as input patients from the most similar cluster. However, users can adjust this selection based on their own assessment of similarity, depicted in hierarchy view (e.g., they can prioritize similarities in one organ over differences in other). Patients can be added in or removed using the RadViz plot, or using a checkbox in the hierarchy view.

To model organ deformations, we use a PCA-based model similar to Budiarto et al. [19]. For each of the included patients $s$ from

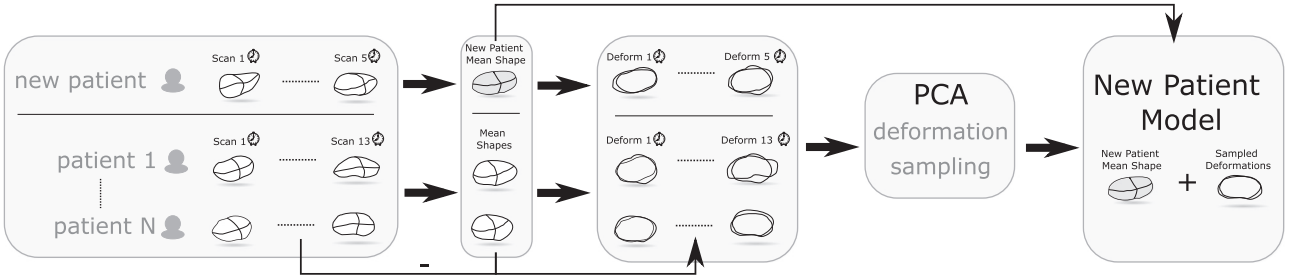


**Fig. 6.** Schematic depiction of our generative PCA-based approach for modeling the organ deformations of a new incoming patient with incomplete data.

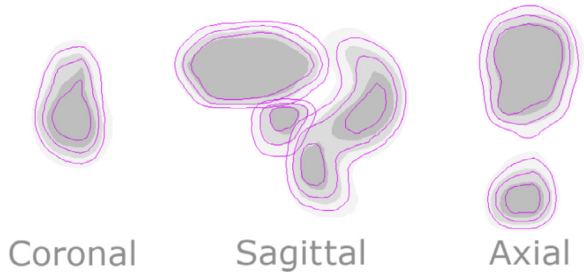


**Fig. 7.** Anatomical view showing the generated patient model. The model that was generated with deformation data from multiple patients is shown with gray probability coverage levels (25%, 50%, and 75%) on the three anatomical 2D planes. Superimposed magenta contours show the model generated from the new patient data (containing 5 timesteps).

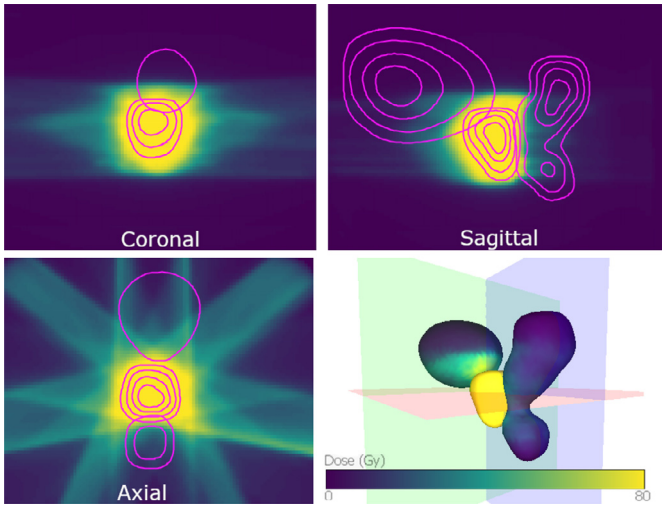


**Fig. 8.** Anatomical view of the new patient, showing the radiation dose according to the treatment plan (viridis color map), and the generated patient model (magenta contours).

the existing cohort, we compute the mean shape descriptor $\bar{s}$ and the corresponding timestep-specific deformations from this mean shape $s_i - \bar{s}$. From these, a covariance matrix $C$ is computed. The eigenmodes of $C$ give us the independent modes of variations. We model the changes of the new patient by deforming the mean shape of the new patient by the weighted sum of $L$ dominant eigenmodes. In our case, the first 5 eigenmodes are selected by default, as these tend to capture the majority of deformations [19,20]. Still, the number of requested eigenmodes, and also the minimal input size can be adjusted. We sample the weights for the $l$-th eigenmode from the normal distribution $\mathcal{N}(0, \sigma^2)$, where the variance is given by eigenvalue $\lambda = \sigma^2$. Fig. 6 schematically depicts our model.

We repeat the sampling 1000 times and proceed to show the results in the same manner as for clusters (i.e., the mean coverage probability volume of the generated results along with standard deviation on hover) in the anatomical view **(T3)**. This number of samples has been chosen experimentally by observing a convergence to a stable coverage probability map. Contour superimposition can be used to highlight differences in the coverage probability of the new patient, as computed from the few available data and as compared to similar patients (Fig. 7, magenta contours).

### 5.4. Treatment assessment (T4)

The treatment plan assessment is supported through *spatial representations* and through the analysis of the *DVHs* and *NTCP models*, similar to common clinical practice **(T4)**.

*Spatial representation*. To support spatial treatment plan assessment, we use the anatomical views of Fig. 8 **(T4 a)**. For the 3D view, we color-code directly on the isosurface generated from the contours the amount of radiation dose corresponding to the given position, according to the treatment plan for the new patient. For the 2D planes, we use the background of the plane and color-code it, according to the treatment plan. In RT, the rainbow color map is often employed to encode the scalar values of the dose [1], which we exchange for the viridis color map. This is familiar to medical experts, while overcoming several limitations of the rainbow color map. For a comparison between two treatments **(T4 c)**, we calculate their difference and employ a diverging color scheme (orange-to-mint) to visually emphasize the differences between the plans [50]. As for the color of the contours, we preserve the same color per treatment plan through the system. The contour colors were chosen such that they can work well with light backgrounds (in the DVH and NTCP plots, as discussed below) and dark backgrounds (in the anatomical views of the treatment plans).

*DVH analysis*. DVHs are widely used in RT to evaluate treatment plans with respect to numerical guidelines for the acceptable/required dose limits for healthy tissues/tumors, respectively. We, therefore, integrate DVH analysis in our approach, as it can support also cluster or individual patients comparison **(T4 a,c)**. To communicate uncertainties stemming from anatomical variations in DVHs, we calculate the respective DVHs at the four levels of coverage probability (i.e., the levels in Fig. 5). For each level, we extract a separate volume and compute the DVH (Fig. 9, gray). The DVH graph is interactively connected to the anatomical views, so the iso-contours and corresponding DVH curves are highlighted upon hovering. We originally considered showing the results in a contour box plot representation [51]. However, this led to confusion, as domain experts tended to interpret it as one particular observation of anatomy with population-based uncertainty margins [1]. It should also be noted that coverage probability levels indicate likelihood of encountering an organ within the given vol-

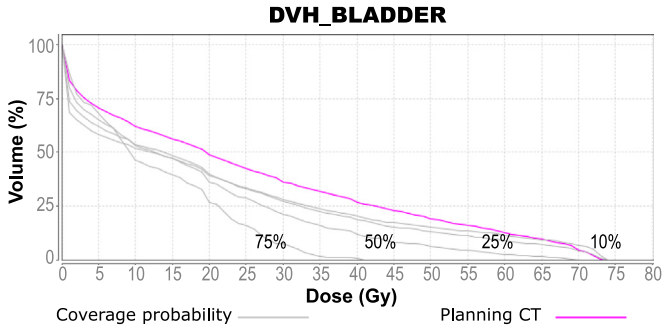


**Fig. 9.** Dose Volume Histograms (DVHs) for the modeled coverage probability map (10%, 25%, 50%, 75%) and the contour from the planning CT for the bladder.

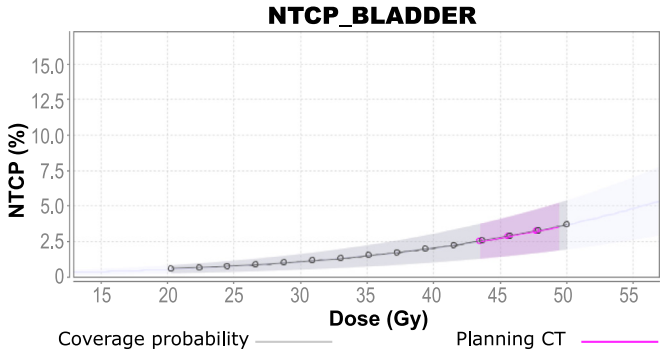


**Fig. 10.** NTCP model enhanced with a mapping of parameter-induced uncertainties (horizontal bands along the curve) and anatomical uncertainties (vertical confidence intervals) for the modeled coverage probability map and the contour from the planning CT.

ume rather than real organ shapes. Thus, a direct comparison of DVH values computed at coverage probability iso-contours should be further looked into. These should not be interpreted as best- or worst-case scenarios. The worst-case scenario would be an organ in the sub-space of probability volume, falling entirely within a high-dose area. This can be seen in Fig. 9.

*NTCP model analysis*. TCP/NTCP modeling is another widely used pre-treatment dose-response evaluation strategy, and TCP/NTCP models are statistical models with parameters describing the dose-response in a population. As discussed before, in this work we are interested only in the dose effect on normal tissues (NTCP models). In particular, we look into the Lyman Kutcher Burman (LKB) NTCP model [52], being the most popular. The LKB model has three main parameters: the volume effect $n$ (a range of values with $n \rightarrow 0$ indicating that normal tissue toxicity is sensitive to hot spots, and $n \rightarrow 1$ indicating that toxicity is dependent on mean dose and hot spots can be tolerated), the dose to a whole or reference volume for 50% risk of endpoint $d50$ (the dose value at which there is a 50% probability of toxicity) and the curve steepness $m$ (how quickly the risk increases as a function of dose).

In this work, we include a strategy for uncertainty analysis of the predicted modeling outcome with regard to these three parameters (i.e., parameter sensitivity analysis). For the representation of this uncertainty, we enhance the traditional representation of Fig. 1(c) with confidence intervals—both horizontal and vertical ones, as depicted in Fig. 10. The *horizontal confidence intervals* relate to the influence of $d50$ and $m$ on the NTCP model curve. This is calculated by sampling the model curve with extreme values of these parameters and displaying the model range around the curve, as computed with the mean of the parameter intervals. The *vertical confidence intervals* are related to the value $n$ of the volume effect, and the existence of multiple coverage probability levels for a given organ. Each coverage level yields one DVH curve (i.e., the dose depends on the coverage level volume). This dose is then used as input to the NTCP, yielding a range instead of a value. As all uncertainties need to be combined, the confidence intervals are combined into a single range of extreme values from all intervals, and the plotted surface indicates the area of normal tissue complications probability. The involved parameters can be interactively modified, and models with different parameter sets can also be compared. This is particularly useful, if treatment plans with different treatment modalities are available (e.g., photon or proton), where the NTCP curve is influenced differently **(T4 b,c)**.

## 6. Exploration scenario

In this section, we present an exploration scenario that demonstrates how our system supports tasks **(T1–4)**. The exploration scenario was initiated by one of the co-designing and co-authoring medical physicists, and it was afterwards employed in a user evaluation, as we will discuss later in this paper. Our scenario can be split into two parts: a *retrospective* and a *predictive* data analysis.

*Retrospective Analysis*. Here, we focus on the analysis of retrospective patient data. For this, we employed both datasets, described in Section 3. Fig. 11 shows the initial overview of the cohort **(T1)** in the overview table (a) and in the RadViz plot (b). The patients are partitioned into six clusters, according to their bladder, rectum, and prostate similarity. From Fig. 11 (a) we see that clusters 1, 2, and 3 containing only one or two patients, i.e., they could be outliers. Looking at the anatomical thumbnails in the RadViz plot (b), we immediately see features that set these patients apart from the rest of the cohort. Cluster 1 (turquoise) demonstrates a unique bladder orientation and size, cluster 2 has a unique separation of the bladder from the prostate, and cluster 3 (blue) significantly differs because of the rectum shape. In other clusters, the differences are more subtle. The overview table shows that patients in cluster 5 (green) have the most stable anatomy, as the color cells do not vary much, but otherwise they are quite similar to cluster 4 (orange). To compare these clusters in greater detail, we select them and compare them in the anatomical views (Fig. 11 (c)). The sagittal plane (middle panel) indicates differences in positioning, where the bladders of cluster 4 (gray) are positioned closer to the rectum than for cluster 5 (magenta contours). We also observe a different orientation of the rectum between the two clusters.

To further explore a single patient, we select patient 8 from cluster 4. For this patient, both photon and proton treatment plans are available and a careful choice must be made, as proton treatment is more costly and not adequate for everyone. We first look at the individual photon and proton treatment plans, and their difference (Fig. 12 (a–c)) **(T4 a,c)**. Different colors of the contours are used to indicate the displayed treatment plans (magenta for photon, orange for proton plan) and link them to the graph-based views (Fig. 13). The most apparent difference is the presence of the so-called *low-dose bath* in the photon plan, a low radiation dose that is delivered to the wide area surrounding the radiation target (Fig. 12 (a,c)). This is expected and unavoidable in photon treatment. In the difference map (c), we also see the areas that receive higher dose in proton treatment. These correspond to the radiation targets (i.e., prostate and lymph nodes in mint color), but overlap with high coverage probability areas of the bladder and rectum. The overview table (Fig. 11 (a)) indicates that patient 8 has a less stable anatomy, which can also be followed through the treatment (Fig. 12 (d-f)). Although in the planning CT, the high dose region of proton plans avoids the bladder and rectum (Fig. 12 (b)), in the subsequent steps the bladder changes its position significantly, overlapping with high radiation dose.

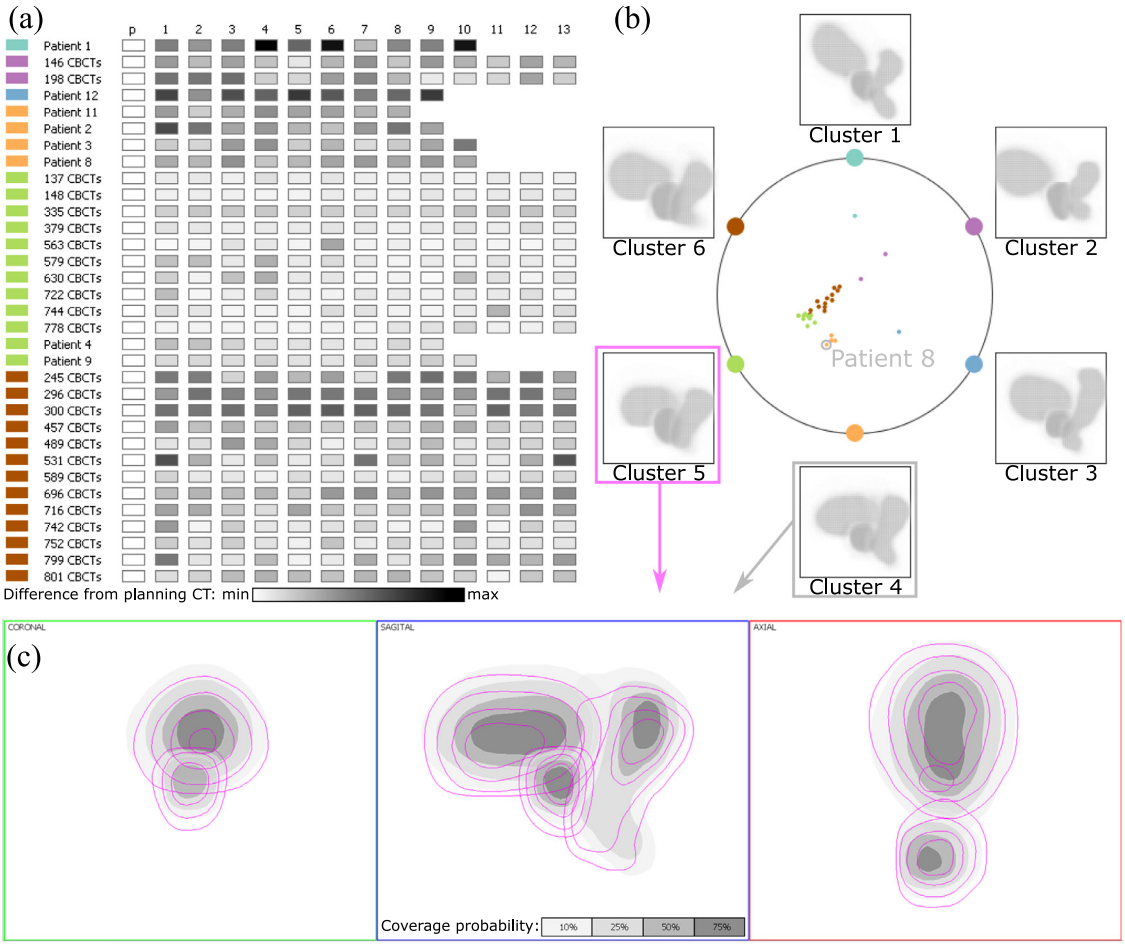


**Fig. 11.** Overview of the initial cohort of 33 patients: (a) Overview table. (b) RadViz plot. (c) Anatomical views comparing cluster 4 (gray values for coverage levels) and cluster 5 (magenta contours).

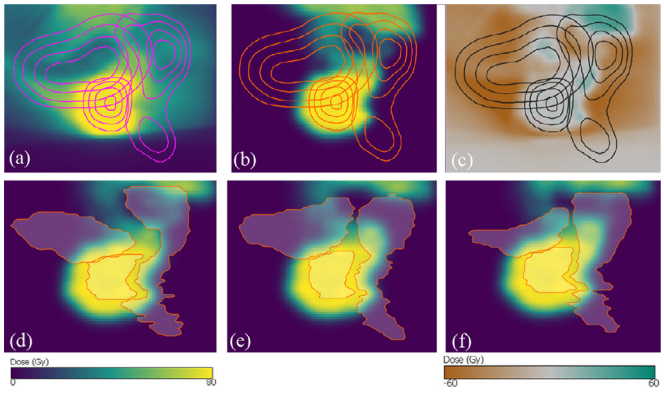


**Fig. 12.** Comparison of treatment plans for a single patient, shown on the sagittal view: (a) Photon plan. (b) Proton plan. (c) Difference between (a) and (b). (d–f) Anatomical changes occurring over treatment. Organ contours from CT acquired (d) prior to treatment, (e) in the first, and (f) in the second treatment week.

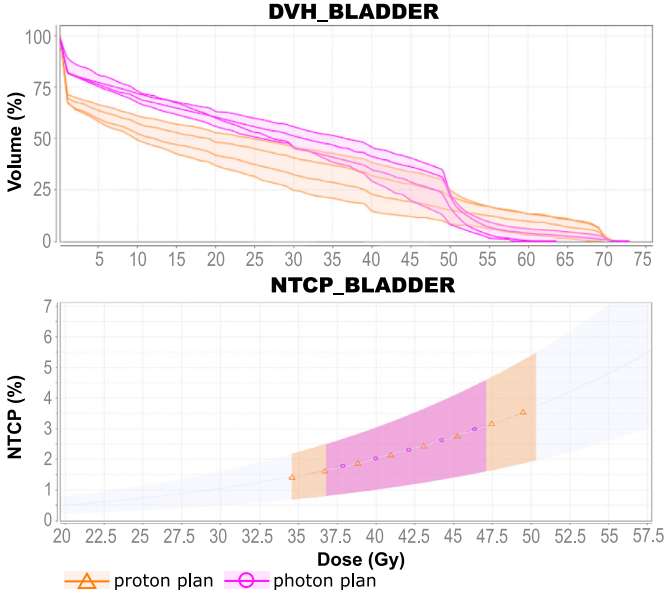


**Fig. 13.** DVH and NTCP model comparison for photon and proton treatment plans, based on the coverage probability levels.

To compare the treatment plans in terms of complication probability for healthy tissues **(T4 b,c)**, we start with the DVH analysis. Fig. 13 (top) shows the comparison of the bladder coverage probability volume DVH for photon (magenta) and proton (orange) treatment plan. The proton plan administers overall a lower dose than photon—only, 25% of the volume receives higher dose with proton plan. The difference map of the plans (Fig. 12 (c)) shows these regions. The NTCP models of these treatment plans (Fig. 13, bottom) indicate a higher NTCP uncertainty range, if treated with protons instead of photons (vertical orange band wider than magenta). Based on this analysis, there might not be any clear advan-

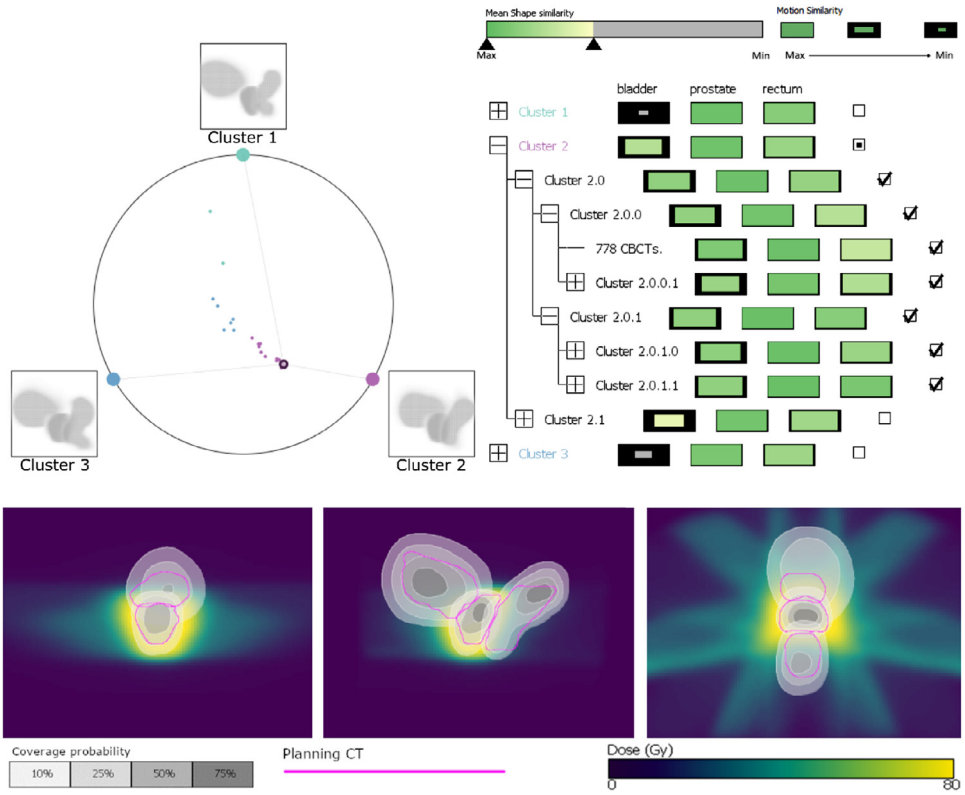


**Fig. 14.** Top: Matching a new patient to an existing cohort. The hierarchy view shows a best fit to cluster 2.0, while the glyph size and color indicate bladder differences with cluster 2.1. Bottom: Comparison of modeled coverage probability map (gray scale) with the contours from planning CT (magenta) over the dose plan (viridis colormap).

tage for proton treatment for this patient. Given his large anatomical changes, proton treatment is more risky.

*Predictive analysis*. Here, we employ only the cohort of 25 patients with photon treatment plans (Section 3). We follow a leave-one-out approach, using 24 patients as a cohort with retrospective data and one patient as the "new" patient with only 5 repeated CBCTs. The 24 patients were grouped into three clusters by our system (Fig. 14, left) **(T1)**. The new patient was then added, and fitted best with cluster 2, as seen in the RadViz plot. The patients of cluster 2 were automatically selected for the predictive modeling, but upon closer inspection of the hierarchy view (Fig. 14, right), we observed that a portion of cluster 2 exhibited significant differences in bladder shape (cluster 2.1). These patients were removed from the generative model, leaving us with 10 similar patients (cluster 2.0) **(T2)**. The coverage probability map predicted by the generative model (Fig. 14, bottom in magenta) for the new incoming patient is investigated with the planning CT (background) **(T3)**. The most significant anatomical differences can be observed in low-dose areas. The healthy organs should, thus, not be affected by unexpected high radiation doses as a result of organ motion. The generated model can be also evaluated in the DVH and NTCP plots (Figs. 9 and 10) **(T4)**.

As we mentioned before, we use the leave-one-out method, so that we can evaluate the precision and convergence of the predictive model. To this end, we compare its results against the coverage probability map computed from the available complete patient data (ground truth), by combining and reconstructing the descriptors of all available timesteps. For each patient, we computed three generative models using the same 10 closest patients, but including incrementally more data: first, we include just the contours from the planning CT (i.e., first timestep), then we added the contours from two repeated CBCTs (i.e., first three timesteps), and finally, we added two more CBCTs (i.e., first five timesteps).

To quantify the convergence of the model, we employ the *Dice Coefficient* metric, defined as $\frac{2|A \bigcap B|}{|A|+|B|}$, where $A$ and $B$ are the probability volumes of the generated model and the ground truth data (complete patient data), respectively. We compute the Dice Coefficient at the four levels of coverage probability. With five timesteps, we can approximate the real coverage probability with an accuracy of $94.69 \pm 3.33\%$ for the bladder, $97.19 \pm 3.54\%$ for the rectum, and $98.44 \pm 1.85\%$ for the prostate, which is in line with findings from previous research [4]. Fig. 15 shows the convergence of our model, when adding more CBCTs. As expected, by adding more timesteps, the differences between the generative model and the actual anatomical variability of the patient are significantly reduced for all three organs. Fig. 16 shows in an anatomical view the convergence of the model for two patients—one with stable anatomy (patient 137) and one with large anatomical deformations (patient 531)—when adding more CBCTs. The numerical convergence for these two patients can be seen in Fig. 17.

## 7. User evaluation

The exploration scenario was then presented to eight medical physicists and clinical researchers. Seven of the participants had 2–6 years of experience in medical physics and treatment planning, and one participant had over 20 years of experience in this field. The evaluation was carried out in small groups of 1–3 participants, similar to common clinical settings. Each evaluation session took approximately 90 min. At the beginning of each session, we introduced the main concepts and components of *PREVIS*. This was then followed by a live demonstration of the exploration scenario. The system was operated by the first author of this paper, while the experts observed the demonstration. Participants were encouraged to think aloud and comment openly regarding the demonstrated functionality of *PREVIS*. This resulted in an active discussion about

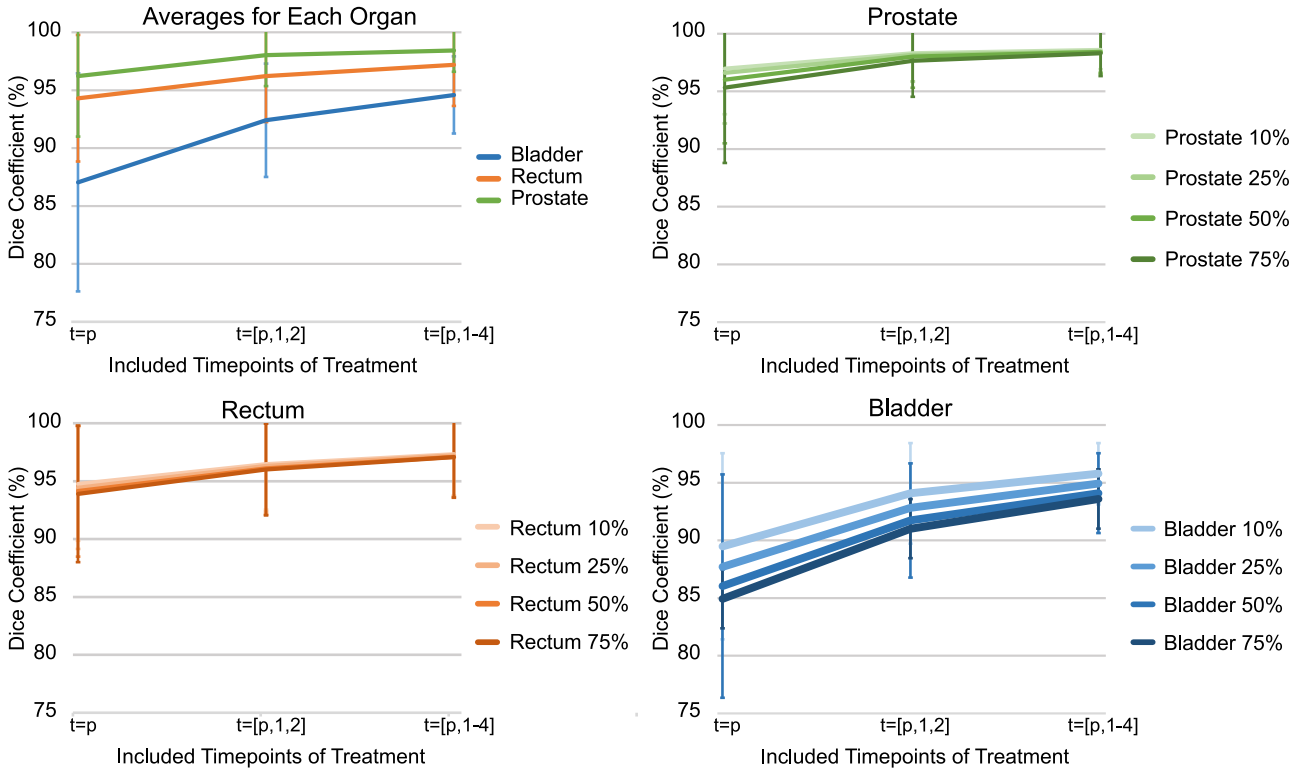


**Fig. 15.** Evolution of the Dice Coefficient for the modeled coverage probability maps, as obtained from our leave-one-out predictive analysis in a cohort of 25 patients, when including incrementally more timepoints to the model. Top left: Average results and standard deviations for three organs. Remaining plots: Results of the model at different coverage probability levels for the three individual organs.

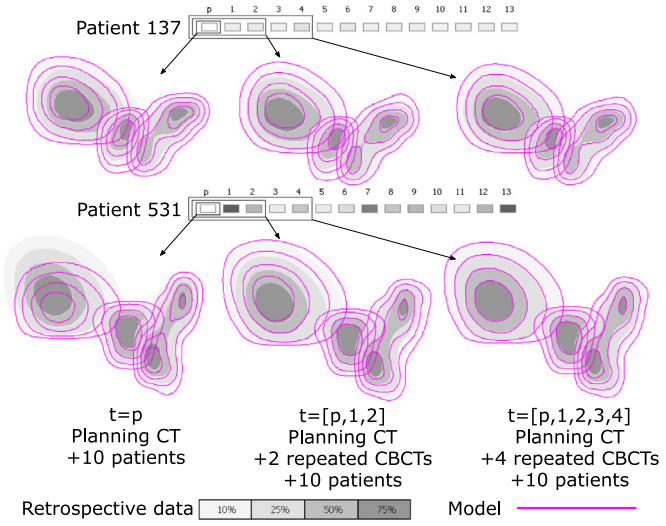


**Fig. 16.** Comparison of the modeled coverage probability maps (magenta; computed when adding more timepoints) against the actual coverage probability (grey scale; computed from all available retrospective data).

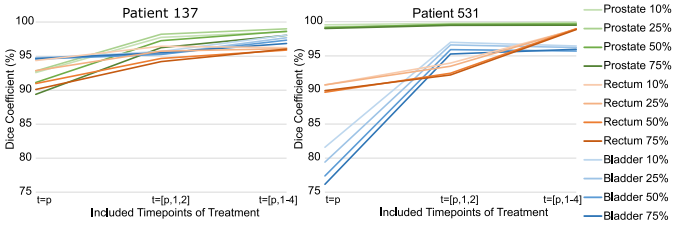


**Fig. 17.** Evolution of the Dice Coefficient for the four modeled coverage probability levels of the two patients presented in Fig. 16 (patient 137: stable, patient 531: mobile), when adding more timepoints to the model.

the employed methods and visualizations, several observed findings, as well as additional exploratory steps in the system, steered by the evaluation participants. Besides the verbal feedback provided during the evaluation sessions, participants filled out also a questionnaire, where the functionality that supports each of the four tasks **(T1–4)** has been rated for *usability* (i.e., effectiveness, efficiency and satisfaction). Additional questions regarding *user experience* (i.e., strengths, limitations, and potential improvements) were also included [53].

All participants were positive about the overview table (Fig. 3) and the RadViz plot (Fig. 4). One participant noted that the overview table indicating difference of repeated CBCT scans to plan is helpful, as physicists often wonder how good the planning CT/contours are when designing the plan. Other participants claimed that the views are *"very nice and intuitive"* and that *"great insight can be gained from the [RadViz plot]"*; for instance, what kinds of groups can be found in a cohort. The views have *"great potential for uncovering patterns we have not been aware of"*. The selection of patients closely matching the new patient was also perceived as intuitive, although one participant commented that *"[it can be] a bit cumbersome to go into the tree of different patients to find good fits to include"*.

The coverage probability representation of anatomical variability (Figs. 5 and 7) was received more critically. While the participants had no problem interpreting it and one participant stated that *"they are very descriptive"*, multiple participants noted that they would need to get used to it. Another participant noted that a drawback of this representation was that *"coverage probability does not communicate the size and mobility of the organ"*. On the

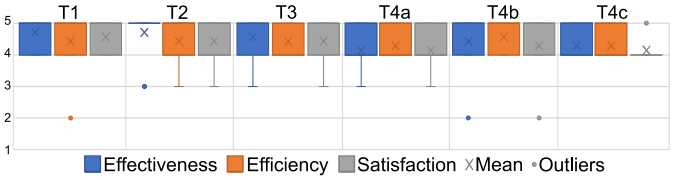


**Fig. 18.** Results of the system usability evaluation (1: low, 5: high).

other hand, multiple participants commented that this representation could be used in treatment planning software where *"it could be useful for exploring the possibility of decreasing the margins, where the probability intervals could be used as margins"*. Also, *"It would be great (given clinically verified results) to include the modeling also in the treatment planning step, not only as 'post-planning' evaluation to avoid re-planning if the results show the treatment plans are not suitable"*. With respect to the treatment plan evaluation features (Figs. 8 and 12), the possibility to compare different plans was appreciated by several participants, stating that *"[the system] may be useful to see which of two treatment modalities, e.g., photon and proton, is best suited for the given patient"*. Several participants suggested that the DVH and NTCP representations (Figs. 9 and 10) would benefit from additional features that are typically present in treatment planning software, such as querying thresholds (e.g., how much of the volume receives 20 Gy of radiation dose). Regarding the NTCP representation one participant noted that *"interactive settings of NTCP parameters with this representation could probably be used also for experimentation and evaluation of parameters."*

Overall, the participants claimed that the system seems intuitive and easy to use. They agreed that the system has great potential for future use and would be most useful, if integrated into a treatment planning software, as it *"offers information that is currently not available on e.g., the extent of anatomical changes"*. They suggested multiple steps of the RT workflow that could benefit from our approach, such as treatment plan design based on predicted coverage probability map, selection of treatment modality based on plan comparison, and adaptive RT by pre-selecting patients for re-planning. However, all participants noted that a robust evaluation of the precision of the generative model would be needed before clinical integration. Apart from the open questions, we asked the participants to rate how well the system supports the four outlined tasks **(T1–4)** in terms of perceived effectiveness, efficiency, and satisfaction on a five-point Likert scale (Fig. 18).

## 8. Conclusion and future work

*PREVIS* is a novel visual analytics approach for the evaluation of RT treatment plans. It provides novel insights into retrospective cohort analysis and enables the predictive analysis of organ motion for new incoming patients, for whom we do not have yet complete data. Also, it supports the assessment of treatment plans with respect to this anticipated organ motion. The predictive aspect of *PREVIS* was previously not possible in any other framework. *PREVIS* was positively evaluated by clinical researchers with experience in treatment planning, who confirmed the applicability of *PREVIS* in multiple steps of the RT workflow.

While the methods have been evaluated positively and our initial assessment indicates that our generative model converges to real data, a more thorough and, preferably, long-term evaluation with a larger number of cases is needed, before our methods can be integrated into clinical practice. Therefore, our immediate priority for future work is the clinical testing of *PREVIS*. Also, an evaluation of the clustering stability, when new patients are incrementally added to cohort, would be required. While we did not observe this to be an issue with our data, it should be investigated further in the future. Another limitation of our approach is that the modeling and the treatment assessment in our system is based mostly on spatial data. However, the organ variability, as well as the risks and potential outcomes for the patient, can be influenced by many other factors, such as co-morbidity, medication, diet, or age. These factors should also be considered in a thorough treatment assessment.

In the future, we also plan to address the limited comparability of coverage probability levels produced by our methods to the real organ shapes in terms of representations familiar to clinicians (DVH, TCP, NTCP). Another possible direction for future work is the adaptation of our methods for different clinical users. Currently, *PREVIS* is more targeted towards clinical researchers, but for users, such as radiation oncologists, the methods would need to be simplified or automated, while new (abstracted) visual representations providing quick insight into results would need to be designed. Finally, many types of uncertainty are present in our application (e.g., uncertainties in the imaging, uncertainties due to contouring or registration errors, uncertainties from our generative model, or due to our human-in-the-loop approach) [22]. Investigating the effect of different uncertainties on top of anatomical variability, and integrating them within the same framework would be another interesting and challenging future direction.

All in all, *PREVIS* is the first visual analytics tool to support the prediction, exploration and assessment of anticipated anatomical changes during RT treatment. Despite having been designed for very specific goals and tasks in the context of RT, several of its components (such as the generative model, or the representations supporting the predictive exploration and analysis of the cohort data) could be employed within other comparative, ensemble, or predictive visual analysis scenarios.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Katarína Furmanová:** Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization. **Ludvig P. Muren:** Conceptualization, Methodology, Investigation, Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition. **Oscar Casares-Magaz:** Conceptualization, Validation, Investigation, Data Curation, Writing - Review & Editing. **Vitali Moiseenko:** Conceptualization, Investigation, Resources, Data Curation, Writing - Review & Editing. **John P. Einck:** Conceptualization, Investigation, Resources, Data Curation, Writing - Review & Editing. **Sara Pilskog:** Conceptualization, Investigation, Resources, Data Curation, Writing - Review & Editing. **Renata G. Raidou:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

## Acknowledgments

This work was supported by Varian Medical Systems of Palo Alto, California, USA in the frame of a research project entitled "*A machine learning centered visualization system for model-based decision making in image-guided and adaptive radiotherapy of cancer*" (principal investigator L.P. Muren).

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.cag.2021.04.010

## References


[1] Schlachter M, Raidou RG, Muren LP, Preim B, Putora PM, Bühler K. State-of-the-art report: Visual computing in radiation therapy planning. Comput Graph Forum 2019;38(3):753–79.
[2] Casares-Magaz O, Moiseenko V, Hopper A, Pettersson NJ, Thor M, Knopp R, Deasy JO, Muren LP, Einck J. Associations between volume changes and spatial dose metrics for the urinary bladder during local versus pelvic irradiation for prostate cancer. Acta Oncol 2017;56(6):884–90.
[3] Furmanová K, Grossmann N, Muren LP, Casares-Magaz O, Moiseenko V, Einck JP, Gröller ME, Raidou RG. VAPOR: visual analytics for the exploration of pelvic organ variability in radiotherapy. Comput Graph 2020a;91:25–38.
[4] Furmanová K, Raidou RG, Grossmann N, Casares-Magaz O, Moiseenko V, Einck JP, Muren LP. Poster: using multiple planning scans to predict organ shape variability during RT for prostate cancer. In: ESTRO 2020; 2020b.
[5] Muren LP, Smaaland R, Dahl O. Organ motion, set-up variation and treatment margins in radical radiotherapy of urinary bladder cancer. Radiotherapy Oncol 2003;69(3):291–304.
[6] Viswanathan AN, Yorke ED, Marks LB, Eifel PJ, Shipley WU. Radiation dose–volume effects of the urinary bladder. Int J Radiat Oncol* Biol* Phys 2010;76(3):S116–22.
[7] Chai X, van Herk M, Hulshof MC, Bel A. A voxel-based finite element model for the prediction of bladder deformation. Med Phys 2012;39(1):55–65.
[8] Raidou RG, Casares-Magaz O, Amirkhanov A, Moiseenko V, Muren LP, Einck JP, Vilanova A, Gröller ME. Bladder runner: visual analytics for the exploration of RT-induced bladder toxicity in a cohort study. Comput Graph Forum 2018;37(3):205–16.
[9] Raidou RG, Furmanová K, Grossmann N, Casares-Magaz O, Moiseenko V, Einck JP, Gröller ME, Muren LP. Lessons learnt from developing visual analytics applications for adaptive prostate cancer radiotherapy. In: VisGap - the gap between visualization research and visualization software; 2020. p. 51–8.
[10] Hermann M, Schunke AC, Schultz T, Klein R. A visual analytics approach to study anatomic covariation. In: 2014 IEEE Pacific visualization symposium; 2014. p. 161–8.
[11] Hermann M, Schunke AC, Schultz T, Klein R. Accurate interactive visualization of large deformations and variability in biomedical image ensembles. IEEE Trans Vis Comput Graph 2016;22(1):708–17.
[12] Busking S, Botha CP, Post FH. Dynamic multi-view exploration of shape spaces. Comput Graph Forum 2010;29(3):973–82.
[13] Busking S, Botha CP, Ferrarini L, Milles J, Post FH. Image-based rendering of intersecting surfaces for dynamic comparative visualization. Vis Comput 2011;27(5):347–63.
[14] Reiter O, Breeuwer M, Gröller ME, Raidou RG. Comparative visual analysis of pelvic organ segmentations. In: Proceedings of the eurographics/IEEE VGTC conference on visualization: short papers; 2018. p. 37–41.
[15] Thörnqvist S, Hysing LB, Zolnay AG, Söhn M, Hoogeman MS, Muren LP, Bentzen L, Heijmen BJM. Treatment simulations with a statistical deformable motion model to evaluate margins for multiple targets in radiotherapy for high-risk prostate cancer. Radiotherapy Oncol 2013a;109(3):344–9.
[16] Thörnqvist S, Hysing LB, Zolnay AG, Söhn M, Hoogeman MS, Muren LP, Heijmen BJM. Adaptive radiotherapy in locally advanced prostate cancer using a statistical deformable motion model. Acta Oncol 2013b;52(7):1423–9.
[17] Bondar L, Intven M, Burbach JPM, Budiarto E, Kleijnen J-P, Philippens M, van Asselen B, Seravalli E, Reerink O, Raaymakers B. Statistical modeling of CTV motion and deformation for IMRT of early-stage rectal cancer. Int J Radiat Oncol* Biol* Phys 2014;90(3):664–72.
[18] Rios R, De Crevoisier R, Ospina JD, Commandeur F, Lafond C, Simon A, Haigron P, Espinosa J, Acosta O. Population model of bladder motion and deformation based on dominant eigenmodes and mixed-effects models in prostate cancer radiotherapy. Med Image Anal 2017;38:133–49.
[19] Budiarto E, Keijzer M, Storchi PR, Hoogeman MS, Bondar L, Mutanga TF, de Boer H, Heemink AW. A population-based model to describe geometrical uncertainties in radiotherapy: applied to prostate cases. Phys Med Biol 2011;56(4):1045.
[20] Söhn M, Birkner M, Yan D, Alber M. Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. Phys Med Biol 2005;50(24):5893.
[21] Ristovski G, Preusser T, Hahn HK, Linsen L. Uncertainty in medical visualization: towards a taxonomy. Comput Graph 2014;39:60–73.
[22] Raidou RG. Uncertainty visualization: Recent developments and future challenges in prostate cancer radiotherapy planning. In: EuroVis workshop on reproducibility, verification, and validation in visualization (EuroRV3); 2018. p. 013–17.
[23] Gillmann C, Arbelez P, Hernndez JT, Hagen H, Wischgoll T. Intuitive error space exploration of medical image data in clinical daily routine. In: EuroVis 2017 - short papers; 2017.
[24] Gillmann C, Wischgoll T, Hamann B, Ahrens J. Modeling and visualization of uncertainty-aware geometry using multi-variate normal distributions. In: Pacific visualization symposium (PacificVis). IEEE; 2018. p. 106–10.
[25] von Landesberger T, Bremm S, Kirschner M, Wesarg S, Kuijper A. Visual analytics for model-based medical image segmentation: opportunities and challenges. Expert Syst Appl 2013;40(12):4934–43.
[26] von Landesberger T, Basgier D, Becker M. Comparative local quality assessment for 3D medical image segmentation with focus on statistical shape model-based algorithms. IEEE Trans Vis Comput Graph 2015.
[27] Smit NN, Kraima A, Jansma D, DeRuiter M, Eisemann E, Vilanova A. Varvis: visualizing anatomical variation in branching structures. In: EuroVis 2016 - short papers; 2016. p. 49–53.
[28] Klemm P, Lawonn K, Rak M, Preim B, Tönnies KD, Hegenscheid K, Völzke H, Oeltze S. Visualization and analysis of lumbar spine canal variability in cohort study data. In: Vision, modeling and visualization; 2013. p. 121–8.
[29] Klemm P, Oeltze-Jafra S, Lawonn K, Hegenscheid K, Völzke H, Preim B. Interactive visual analysis of image-centric cohort study data. IEEE Trans Vis Comput Graph 2014;20(12):1673–82.
[30] Steenwijk MD, Milles J, Buchem M, Reiber J, Botha CP. Integrated visual analysis for heterogeneous datasets in cohort studies. In: IEEE VisWeek workshop on visual analytics in health care, vol. 3; 2010. p. 3.
[31] Preim B, Klemm P, Hauser H, Hegenscheid K, Oeltze S, Toennies K, Völzke H. Visual analytics of image-centric cohort studies in epidemiology. In: Visualization in medicine and life sciences III. Springer; 2016. p. 221–48.
[32] Bernard J, Sessler D, May T, Schlomm T, Pehrke D, Kohlhammer J. A visual-interactive system for prostate cancer cohort analysis. Comput Graph Appl (CG&A) 2015;35(3):44–55.
[33] Alemzadeh S, Hielscher T, Niemann U, Cibulski L, Ittermann T, Vlzke H, Spiliopoulou M, Preim B. Subpopulation discovery and validation in epidemiological data. In: EuroVis workshop on visual analytics (EuroVA); 2017. p. 43–7.
[34] Wang J, Hazarika S, Li C, Shen H-W. Visualization and visual analysis of ensemble data: a survey. IEEE Trans Vis Comput Graph 2018;25(9):2853–72.
[35] Wentzel A-P, Hanula P, Luciani T, Elgohari B, Elhalawani H, Canahuate G, Vock DM, Fuller CD, Marai GE. Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration. IEEE Trans Vis Comput Graph 2019;26:949–59.
[36] Grossmann N, Casares-Magaz O, Muren LP, Moiseenko V, Einck JP, Gröller E, Raidou RG. Pelvis runner: visualizing pelvic organ variability in a cohort of radiotherapy patients. In: Eurographics workshop on visual computing for biology and medicine; 2019. p. 69–78.
[37] Raidou RG, Casares-Magaz O, Muren LP, van der Heide UA, Rørvik J, Breeuwer M, Vilanova A. Visual analysis of tumor control models for prediction of radiotherapy response. Comput Graph Forum 2016a;35(3):231–40.
[38] Casares-Magaz O, Raidou RG, Rorvik J, Vilanova A, Muren LP. Uncertainty evaluation of image-based tumour control probability models in radiotherapy of prostate cancer using a visual analytic tool. Phys Imaging Radiat Oncol 2018;5:5–8.
[39] El Naqa I, Bradley JD, Deasy JO. Nonlinear kernel-based approaches for predicting normal tissue toxicities. In: International conference on machine learning and applications; 2008. p. 539–44.
[40] Kupchak C, Battista J, Van Dyk J. Experience-driven dose-volume histogram maps of NTCP risk as an aid for radiation treatment plan selection. Med Phys 2008;35(1):333–43.
[41] Akgül CB, Sankur B, Yemez Y, Schmitt F. A framework for histogram-induced 3D descriptors. In: 2006 14th European signal processing conference; 2006. p. 1–5.
[42] Akgül CB, Sankur B, Yemez Y, Schmitt F. 3D model retrieval using probability density-based shape descriptors. IEEE Trans Pattern Anal Mach Intell 2009;31(6):1117–33.
[43] Zhang L, da Fonseca MJa, Ferreira A. Survey on 3D shape descriptors; 2007. Technical Report, DecorAR
[44] Everitt BS, Landau S, Leese M, Stahl D. Cluster analysis. 5. John Wiley; 2011.
[45] Thorndike RL. Who belongs in the family? Psychometrika 1953;18(4):267–76.
[46] Hoffman P, Grinstein G, Pinkney D. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In: Proceedings of the 1999 workshop on new paradigms in information visualization; 1999. p. 9–16.
[47] Raidou R, Marcelis F, Breeuwer M, Gröller ME, Vilanova A, van de Wetering H. Visual analytics for the exploration and assessment of segmentation errors. In: Eurographics workshop on visual computing for biology and medicine; 2016b. p. 193–202.
[48] Höllt T, Pezzotti N, van Unen V, Koning F, Eisemann E, Lelieveldt B, Vilanova A. Cytosplore: interactive immune cell phenotyping for large single-cell datasets. Comput Graph Forum 2016;35(3):171–80.
[49] Carter JR. Digital representations of topographic surfaces. Photogramm Eng Remote Sens 1988;54(11):1577–80.
[50] Harrower M, Brewer CA. ColorBrewer.org: An online tool for selecting colour schemes for maps. Cartogr J 2003;40(1):27–37.
[51] Whitaker RT, Mirzargar M, Kirby RM. Contour boxplots: a method for characterizing uncertainty in feature sets from simulation ensembles. IEEE Trans Vis Comput Graph 2013;19(12):2713–22.
[52] Lyman JT. Complication probability as assessed from dose-volume histograms. Radiat Res 1985;104(2s):S13–19.
[53] Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S. Empirical studies in information visualization: Seven scenarios. IEEE Trans Vis Comput Graph 2011;18(9):1520–36.