

Visual Analysis of Tumor Control Models for Prediction of Radiotherapy Response

R.G. Raidou¹, O. Casares-Magaz², L.P. Muren², U.A. van der Heide³, J. Rørvik⁴, M. Breeuwer^{5,1}, A. Vilanova^{6,1}

¹Eindhoven University of Technology, The Netherlands, ²Department of Medical Physics, Århus University Hospital, Denmark

³Department of Radiotherapy, the Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, The Netherlands

⁴Department of Clinical Medicine, University of Bergen, Norway, ⁵Philips Healthcare Best, The Netherlands

⁶Delft University of Technology, The Netherlands

Abstract

In radiotherapy, tumors are irradiated with a high dose, while surrounding healthy tissues are spared. To quantify the probability that a tumor is effectively treated with a given dose, statistical models were built and employed in clinical research. These are called tumor control probability (TCP) models. Recently, TCP models started incorporating additional information from imaging modalities. In this way, patient-specific properties of tumor tissues are included, improving the radiobiological accuracy of models. Yet, the employed imaging modalities are subject to uncertainties with significant impact on the modeling outcome, while the models are sensitive to a number of parameter assumptions. Currently, uncertainty and parameter sensitivity are not incorporated in the analysis, due to time and resource constraints. To this end, we propose a visual tool that enables clinical researchers working on TCP modeling, to explore the information provided by their models, to discover new knowledge and to confirm or generate hypotheses within their data. Our approach incorporates the following four main components: (1) It supports the exploration of uncertainty and its effect on TCP models; (2) It facilitates parameter sensitivity analysis to common assumptions; (3) It enables the identification of inter-patient response variability; (4) It allows starting the analysis from the desired treatment outcome, to identify treatment strategies that achieve it. We conducted an evaluation with nine clinical researchers. All participants agreed that the proposed visual tool provides better understanding and new opportunities for the exploration and analysis of TCP modeling.

Categories and Subject Descriptors (according to ACM CCS): I.3.8 [Computer Graphics]: Applications—Applications; J.3 [Computer Applications]: Life and Medical Sciences—Life and Medical Sciences

1. Introduction

At some stage, 60% of all men diagnosed with prostate cancer receive radiotherapy (RT) treatment [DJFB05]. RT aims at irradiating tumors with a sufficiently high radiation dose, while sparing the surrounding healthy tissues. Before treatment, a plan is performed, during which, different treatment strategies can be followed. These alternatives consider several points, such as dose escalation, uniform or non-uniform tumor irradiation, the amount of the received dose and eventual fractionation of the treatment, i.e., the division of the total radiation into smaller doses per session over a period of time. For the treatment plan, a specific strategy is chosen among these alternatives, based on clinical experience and guidelines. Then, the plan is performed in a planning system, based on information from imaging acquisitions of the patient, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), by radiation oncologists and dosimetrists.

Although clinical practice aims at choosing the most effective RT strategy based on clinical knowledge and guidelines, clinical

research aims at thoroughly evaluating all possibilities. In this way, more targeted treatments can be designed and provided to clinical practice. To simulate and evaluate the effects of a specific strategy, clinical researchers need Tumor Control Probability (TCP) models [WN93].

Conventional TCP models are statistical models that quantify the probability that a tumor is effectively controlled, i.e., treated, given a specific radiation dose. In few words, TCP models answer the question: *What is the probability Y that a tumor is treated with this strategy, given a dose X ?* For example, Figure 1 depicts three different outcomes of a specific TCP model, each from a specific RT treatment strategy. In this case, by providing a total treatment dose of 77 Gy, the first strategy (black) results in 88% probability of treating the whole tumor, while the other two have a lower treatment probability response of 63% (red) and 31% (blue).

In the last years, radiation dose can be delivered in a more localized way and clinical research focuses on providing better tailored tumor treatment. Therefore, patient-specific tissue characteristics

from imaging modalities started being included in planning. This has influenced also TCP modeling, where additional per-voxel information, i.e., properties indicative of tumor characteristics, are being incorporated [TOG06]. In this way, clinical researchers can predict more accurately the tumor treatment probability at a voxel level, by adding in their statistical models radiobiological information, e.g., from Diffusion Weighted (DW) MRI.

So far, several interesting aspects of TCP modeling are not incorporated in clinical research, due to complexity, lack of resources and time constraints. First of all, imaging modalities are subject to *uncertainties* with significant impact on the model outcome and the simulated treatment response [KTH*10]. Additionally, there are many different TCP models and different parameter assumptions in each one of them [WN93,SMB*07]. Usually, these assumptions are educated guesses, and awareness on the *sensitivity* of the models is important. Moreover, TCP modeling is often applied to entire patient cohorts, to investigate the *inter-patient response variability*. This knowledge can help designing more robust treatment strategies. Finally, clinical researchers are interested in exploring and analyzing their data in a *reverse* manner: given a target treatment outcome for a tumor, identify the RT strategy(-ies) to achieve it. In this paper, we introduce a visual analytics approach to extend the exploration of TCP modeling, to cover also these topics that are currently not possible to incorporate in the analysis.

Our *contribution* is the design and implementation of a visual tool that enables clinical researchers to explore the information provided by their TCP models, to discover knowledge and to confirm or generate hypotheses within their data. As far as we know, there is no other tool to serve this purpose. Our approach incorporates the following four main components:

(C1) It supports the quantification and exploration of imaging-induced uncertainty and its propagation to TCP modeling.

(C2) It facilitates exploring and analyzing the sensitivity of TCP models to different assumptions and parameter variations.

(C3) It enables identifying and exploring inter-patient response variability within cohorts.

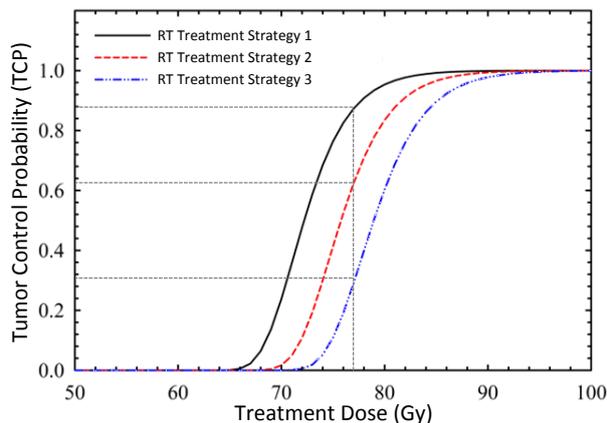


Figure 1: An example of the resulting TCP curves for a given dose, with three different RT treatment strategies.

(C4) It allows, given a targeted treatment outcome, to identify the treatment strategies or parameters that would achieve it.

2. Clinical Background

In the last years, RT research aims at designing more effective and better targeted treatments, to be applied in clinical practice. For the simulation and evaluation of all different treatment strategies, TCP models are being built [WN93]. Conventional TCP models are usually regression models that summarize empirical knowledge about the effect of radiation to tumors and represent the probability that a tumor is effectively treated with a specific dose [WN93,SMB*07].

To achieve a more targeted treatment, tailored to the patient-specific tumor tissue characteristics, information from imaging modalities was recently incorporated to TCP modeling [TOG06]. In this way, properties indicative of tissue characteristics were included to improve the radiobiological accuracy of modeling, at a voxel-level [TOG06]. In this work, we employ a novel TCP model that involves DW-MRI [CMvdHR*16]. This in-vivo imaging technique measures quantitatively the per-voxel water diffusion, from apparent diffusion coefficient (ADC) maps, and is employed to identify high-density tissue like tumors [BJEK*11].

This ADC-based TCP model is subject to uncertainties [KTH*10], often due to calculation restrictions in the clinical setting or due to magnetic field inhomogeneities [BJEK*11]. Although we are considering a specific TCP model, uncertainties are present in all modalities and our proposed approach could be extended also to them. These uncertainties need to be quantified and propagated into modeling to identify their effect on the prediction outcome. More details about the source, quantification and propagation of uncertainty in the employed TCP model are discussed in Section 4.

Additionally, all TCP models, including ADC-based ones, incorporate a number of different parameter assumptions. For example, in the explored TCP model, researchers make assumptions for the amount of dose or fractionation, or when quantifying the per-voxel cell density from ADC maps, or when selecting values for parameters that model the survival or death of tumor cells after irradiation [CMvdHR*16]. Still, it is not known which choices lead to better results, or the effect of different alternatives. Thus, the parameter sensitivity of the model needs to be incorporated in the analysis. In state-of-the-art clinical research, the ADC-derived uncertainty and the model sensitivity to parameter assumptions are not considered yet, as they cannot be explored with the existing tools. This is further obstructed by the fact that the evaluation and analysis of TCP models is usually applied to cohorts of patients, to account for inter-patient response variability.

Finally, the current TCP modeling workflow is based on the question: *What is the probability that a tumor is controlled, given a specific dose?* Yet, clinical researchers have not managed to find an easy and insightful way to answer the inverse: *Which RT delivery strategy can achieve a specific target treatment?*

After an extensive discussion with clinical researchers working on TCP modeling, we defined together the most relevant open tasks for their research, which are also summarized in Figure 2:

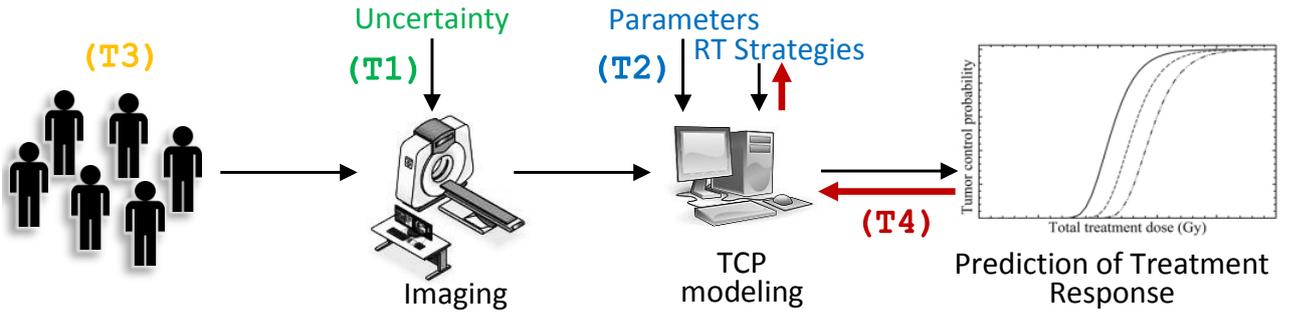


Figure 2: The proposed Visual Analytics approach for the Prediction of RT Treatment Response in TCP Modeling. With colors, we denote the four requirements (T1–T4) described in Section 2, which are our contributions to the workflow employed in clinical research.

(T1) Quantification and interactive exploration of the ADC-induced uncertainty and its propagation to TCP modeling.

(T2) Exploration and analysis of the assumption-induced TCP model sensitivity.

(T3) Identification of inter-patient variability to radiotherapy treatment response.

(T4) Bi-directional TCP modeling workflow (RT strategy ↔ Predicted/desired outcome).

3. Related Work

There are several frameworks that cover topics similar to ours. To the best of our knowledge, there is none for the exploration of TCP modeling. In this section, we review the literature, related to the tasks mentioned in Section 2.

Visualizing Uncertainty. Uncertainty visualization literature is vast [JS03, BHJ*14]. It can be roughly divided into the following main categories: visualizations using visual variables, such as color [Hen03, GR04], brightness [DKLP02], fuzziness [LV02], or texture [BWE05]; visualizations that adapt the basic geometry to represent uncertainty [GR04, ZWK10] or surrounding volume [PH11, PRW11, PMW13]; visualizations with additional graphical variables, such as glyphs [SZB*09, PRJ12, SSSW13]; and visualizations employing animations [LLPY07]. The selection or combination of these approaches is not limitless and must be done in regard to the data, avoiding clutter. In our case, we need to visualize not only the inherent uncertainty of the imaging data itself, but also how it propagates and affects the outcome of the TCP model. Therefore, several of the previously mentioned approaches need to be carefully adapted to suit our application.

Analyzing Parameter Sensitivity. Parameter sensitivity is often connected to forecasting or prediction models. A conceptual framework for parameter sensitivity analysis was presented by Sedlmair et al. [SHB*14]. Other examples of systems for exploring multi-dimensional parameter spaces are the Ensemble-Vis [PWB*09], Noodles [SZD*10], OVis [HMC*13, HMZ*14] and the approach of Berger et al. [BPF11]. Visualizations for parameter sensitivity analysis were also proposed for medical applications [BVPR09, PCR*11, TWSM*11]. Most of them employ multiple views in an

interactive environment, where linking and brushing enables exploration and analysis. Yet, none of these frameworks can be used directly for our purposes.

Studying Cohorts. In many cases, patients are not analyzed individually. Previous work in cohort visualization mainly focuses on the comparative analysis of shape variability [BBP10, SPA*14, HSSK14]. Recently, Steenwijk et al. [SMB*10], Zhang et al. [ZGP14] and Klemm et al. [KLR*13, KOJL*14] proposed interactive visual analysis of cohorts that goes also beyond shape analysis. However, these methods assume that the structure of interest has spatial correspondence between patients and can be compared after matching. This is not valid for tumors. In our case, we need to treat each tumor in the cohort as an entity that we can compare to the rest - still, considering and visualizing the within-cohort heterogeneity.

Redesigning the workflow. Several visualizations for redesigning the usual workflow in a specific application field have been proposed [BM10, CLEK13, DPD*15]. Inspired by these strategies, we adapted their approaches to fit our requirements.

4. Visual Analytics for the Exploration of TCP Models

The proposed visual tool aims at satisfying the specific exploratory needs of clinical researchers working on TCP modeling, as described in Section 2. Our visual tool consists of the four main components (T1–T4) of Figure 2.

Quantification and interactive exploration of uncertainty and its propagation to TCP modeling (T1). The ADC-based TCP model [CMvdHR*16] aims at incorporating cell density (CD) information. This is a common measure in tumor tissue characterization, referring to the number of tumor cells within a volume. The first step in the ADC-based TCP model requires the calculation of CD from ADC values [CMvdHR*16]. There are two main approaches for this: (i) the *sigmoid* approach, where a sigmoid relationship between ADC and CD is established, or (ii) the *Gibbs* approach [GPT07], where histopathological information are used to establish a linear relationship between ADC and CD. Both of these approaches are affected by uncertainty. In the first case, only the uncertainty of the ADC is present, while in the second case, there is an additional uncertainty in the experimental setup that was used to determine the relationship between ADC and CD [GPT07].

Uncertainty in ADC maps has been quantified in previous clinical work [GPT07, KTH*10]. However, each institution has a specific set-up and the uncertainty of quantitative measurements in scanners is usually obtained through experimentation. Our clinical partners modeled experimentally this uncertainty, as the probability that an ADC value m is measured in imaging, given a quantitative real value, r . This is given by a probability density function $p(m|r)$, which is a Gaussian distribution with a standard deviation dependent on the real value r . We are interested, though, in quantifying the probability that the real ADC value r has occurred, given m , i.e., $p(r|m)$. From Bayes' rule, we obtain:

$$p(r|m) = \frac{p(r) \cdot p(m|r)}{p(m)} = \frac{p(r) \cdot p(m|r)}{\int_r p(r) \cdot p(m|r) dr} \quad (1)$$

where $p(r)$ is the prior probability of the value r . This is assumed to be uniform: $p(r) = \frac{1}{R}$, where R is the range of possible values. Since the standard deviation of $p(m|r)$ depends on the value r , the calculation of $p(r|m)$ is not trivial and was approximated analytically, using Taylor expansion. It results to be a skewed Gaussian, dependent on the measured value m . Therefore, in the first case of CD calculation, where it is modeled as a sigmoid function of the ADC, the CD uncertainty is given as a function of $p(r|m)$.

In the second approach, Gibbs et al. [GPT07] conducted an experiment where they associated measured ADC values to CD values, obtained from histopathological slices. These are the data point samples in Figure 3. By calculating the linear fit (red dotted line in Figure 3), Gibbs et al. obtained a relationship between ADC and CD: $CD = \frac{2.1 \cdot 10^{-3} - ADC}{3 \cdot 10^{-5}}$. In this empirical approach, the CD values are obtained from histopathology and, hence, have no uncertainty, but ADC uncertainty is not included. When ADC uncertainty, i.e., $p(r|m)$, is incorporated, the relationship between ADC and CD is affected. To quantify this, we randomly sample the ADC uncertainty distribution of each data point of Figure 3 and calculate the resulting fits for 2 million sets of samples. All generated fits can be seen in Figure 3: with the grayscale colormap, we denote the probability of each one of the fitted lines, which is calculated by the product of the probability functions of the data point samples. White is the least probable and black the most probable. From the generated fits, we calculate the CD probability function $CD(r|m)$.

The remaining steps of the TCP model are mathematical equations [CMvdHR*16], which do not include additional uncertainties and use the CD value as input. Therefore, for the *sigmoid* approach, the uncertainty in the TCP model will depend directly on the ADC uncertainty $p(r|m)$, while for *Gibbs*, it will depend on the CD uncertainty $CD(r|m)$.

For the interactive exploration of the uncertainty, our users are initially interested in having a global overview on the regions of the prostate that are most subject to ADC and CD uncertainty. To simultaneously explore the two uncertainties, we employ a 2D colormap [Bre94]. In this visual representation, we encoded the per-voxel difference between the most probable real ADC value r and the measured value m from the acquisition: $ADC_{diff} = \text{argmax}\{p(r|m)\} - m$ (Figure 4). This difference is always positive and its magnitude depends on the measured ADC values. Therefore, we decided to map it to the luminance dimension of the col-

ormap (Figure 4). Also, we encoded the per-voxel difference between the most probable real CD value - after the propagation of the ADC uncertainty - and the value measured as proposed by the experiment in the literature: $CD_{diff} = \text{argmax}\{CD(r|m)\} - CD_{Gibbs}$ (Figure 4). This is mapped to a divergent hue dimension of the colormap, as both positive and negative values are possible. No transparency is employed in the colormap. Other approaches, such as height fields, were considered, but a discussion with the users showed that the colormap was easier to understand and use.

In addition to color-encoding, we enable users to probe the prostate and interactively explore the entire probability density distributions for the ADC and CD values per voxel (Figure 4). With this dual visualization, the user has an overview on the uncertainty - at prostate level - and locally - at voxel level. Finally, when the user performs TCP modeling, the uncertainty is propagated also to the model outcome, as described before, and visualized on the resulting TCP curve as a density band (Figure 4, zoomed view).

Exploration and analysis of the assumption-induced TCP model sensitivity (T2). In this part, we use two main components. First, the clinical researcher adds a finite number of combinations of TCP parameter sets, for the calculation of the respective TCP models. This is consistent with the traditional way of exploration of TCP modeling, where one or more TCP models are compared to each other. In this case, each combination of TCP parameter sets is encoded to the visualizations depicted in Figure 5-left, which we call pianola plots, inspired by the scrolls used by the musical instrument. In this example, the clinical user has added three parameter sets, depicted by the three white planes. Each row of a pianola plot is a parameter. The first parameter cd is the cell density calculation approach, which is a categorical variable that defines which approach is used for the calculation of CD. It can either take the value *Gibbs* or *sigmoid* - or more, if available. This is encoded with the

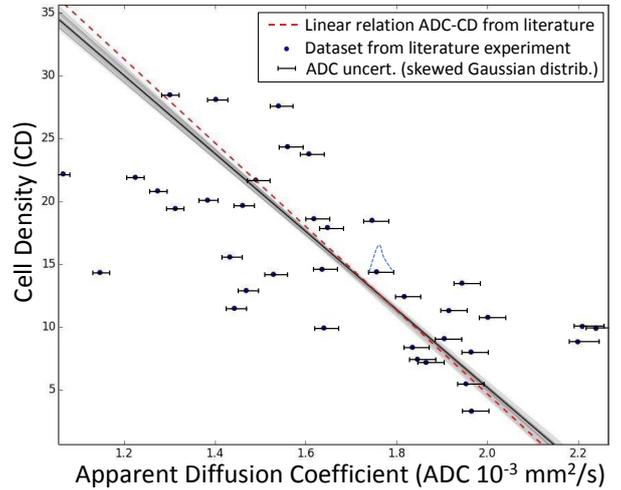


Figure 3: Dataset extracted from the experiment of Gibbs et al. (blue points), and linear relation between ADC and CD (red line) [GPT07], without uncertainty. Incorporating ADC uncertainty results in a set of linear fits (shown with grayscale: dark denotes higher probability).

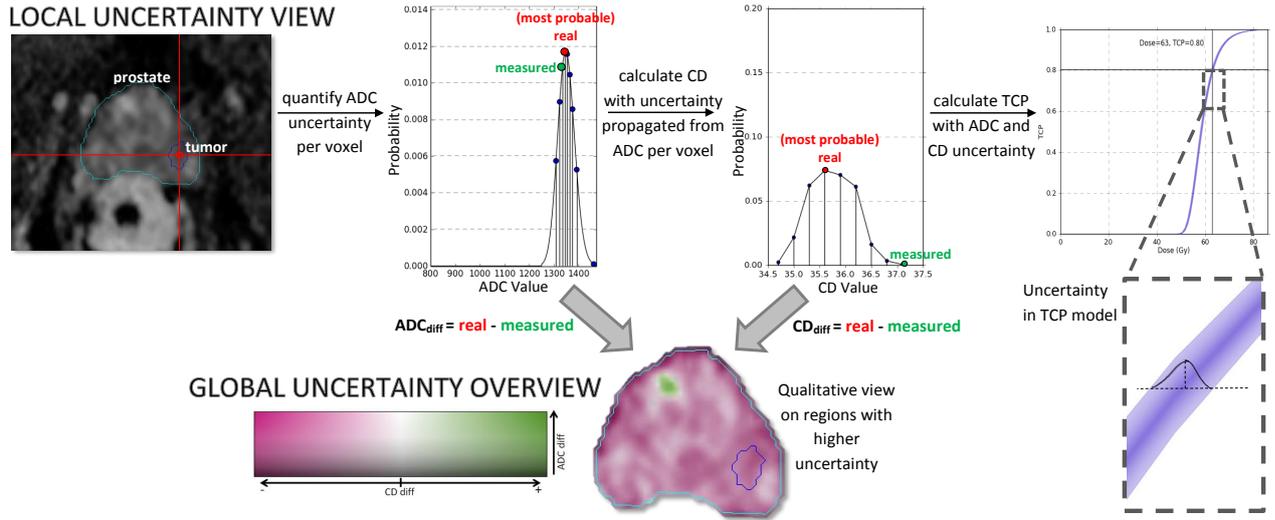


Figure 4: Quantification and interactive exploration of the ADC-induced uncertainty and its effect on TCP modeling (T1).

dot that is located respectively, either in the middle or the end of the first row. The rest of the parameters are continuous variables, which can take values between known and pre-defined ranges. For these, the selected value for each parameter is encoded to the location of each one of the scribbled lines. The user can explore also the effect of varying one or more parameters continuously through a range. This is denoted with a box instead of a line, the width of which depicts the range of values. An example for this is depicted in the second parameter set, for parameter α (Figure 5-left). To intuitively link these sets to the respective TCP curves, the scribbles of each pianola plot are assigned to a different hue. In this way, the clinical user can easily detect how much TCP curves are affected by different parameter choices (Figure 5-right).

With the current workflow, TCP models can be explored only globally. Although TCP curves can be extracted per voxel, clinical researchers currently calculate the expected average response of the whole tumor to a given dose and they only analyze the whole tumor TCP curve. They are not able to perform a voxel-based exploration, to detect whether there are specific parts of the tumor that behave differently than the rest, or to analyze why this happens. To enable this, we provide a functionality to probe the TCP curve: either (i) for a TCP value and see the linked required dose per-voxel (Figure 6), or (ii) for a specific dose and see the linked achieved TCP per-voxel. The latter relates also to task (T4). The linked variable is encoded with a heated-body colormap on the imaging slices of the patient, for direct anatomical reference (Figure 6). When the user has employed several TCP modeling approaches, we visualize also the variability in the respective dose or TCP value, due to the effect of these alternatives, using a circular glyph encoding (Figure 6) [BKC*13]. The size of the glyphs denotes the per-voxel variability and the blue color is chosen to be complementary to the underlying colormap. The circular glyphs were chosen, as they preserve visibility on the underlying color-encoded values. This design also helps identifying the relation between values and variability.

Identification of inter-patient variability to treatment response (T3). The exploration and analysis of the performance of a specific RT strategy is usually evaluated on a cohort of patients. For example, it is interesting for clinical researchers to know how much the per-voxel achieved TCP or the required dose of their patients varies within a cohort. Within-cohort variability is important as it can determine whether a treatment strategy is robust or not with different patients and can aid the design of better treatment strategies. This step is linked to (T2): the user probes the TCP curve for either the TCP response or the dose (Figure 6) and, respectively, the required dose or achieved TCP per-voxel is calculated for the whole cohort, also for multiple TCP modeling approaches, as described in (T2). We provide functionality, with which the users can explore the distributions of the calculated dose or TCP - or multiple sets of these from multiple TCP modeling approaches. Subsequently, they can partition the cohort of patients to identify patients that behave similarly throughout different TCP modeling approaches.

To illustrate our approach for the partitioning, we employ the example depicted in Figure 7. Here, distributions of dose have been calculated for four patients through three different param-

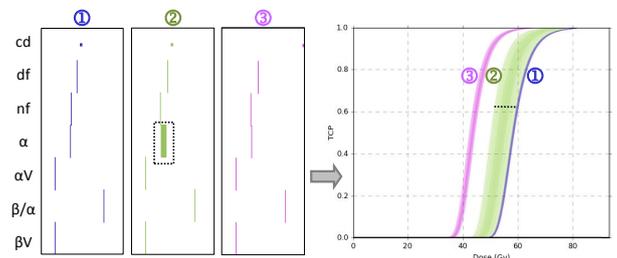


Figure 5: Exploration of the parameter-induced TCP model sensitivity (T2), for three different approaches (1-3).

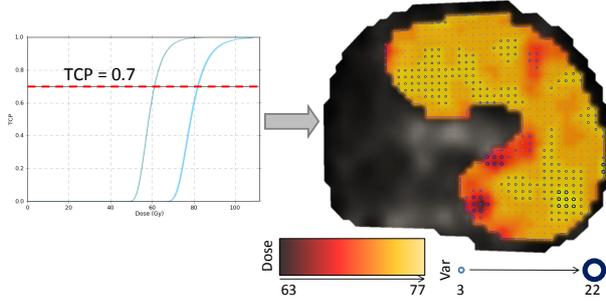


Figure 6: Probing the model curve at a specific TCP value, to inspect the required dose per-voxel (heated-body colormap) and the respective variability (circular glyphs) (T2).

ter settings in TCP modeling. We want to form groups of patients that have similar response patterns along different parameter settings. For this, we need to cluster the patients based on the spreads, i.e., the dispersions in each set of distributions. We quantify the dispersion of each distribution, using the median absolute deviation (*MAD*), which is a robust measure of dispersion [Rup11]: $MAD = \text{median}_i (|X_i - \text{median}_i(X_i)|)$. It is described as the median of the absolute deviations of the distribution data X_i from their median. After calculating all *MAD* measures of the distributions, we employ a *k*-means clustering algorithm on the per patient vectors of calculated *MADs* (Figure 7), which was chosen due to simplicity and computational efficiency.

In our approach, the user interactively selects the number of clusters, *k*. To aid adequate selection of the number of clusters, we employ an additional cluster analysis view. For this, we use a visualization employed in our previous work [RvdHD*15], where the goal of the visual cluster analysis view was to help the users decide whether the visual clusters are well-defined. This is similar to our present goal and we decided to adopt their strategy in our system. In this view, every cluster is mapped to a sphere. For each cluster, we provide internal validity information, i.e., cohesion and separation [RvdHD*15], but also inter-patient and inter-assumption variability. Cohesion is a measure of intra-cluster similarity, while separation is a measure of inter-cluster dissimilarity. For these two measures, we employ the same encoding as in our previous paper [RvdHD*15] (Figure 7-legend): small and opaque spheres depict high cohesion within a cluster, while large and transparent spheres depict low cohesion; also, thin arrows depict well-separated clusters giving the illusion of distance, while thick arrows depict less separated clusters. The inter-patient and inter-assumption variability are encoded to the size of the two dimensions of a box, located at the core of each sphere (Figure 7-legend). With the cluster analysis view, the users interactively change the number of clusters, while following the graphical changes on the glyphs, and decide the most satisfactory result based on the visual optimization of the cluster view, depending on the goal. The users interactively partition the patient cohort inspecting the achieved TCP response, while at the same time, they can identify how much these sub-cohorts of patient responses vary. An automatized initial selection of a good cluster size or number of clusters would be an interesting future extension.

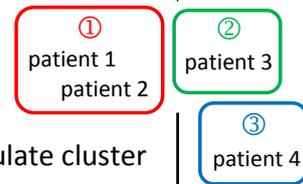
	patient 1	patient 2	patient 3	patient 4
parameter set 1				
parameter set 2				
parameter set 3				

calculate MADs

	patient 1	patient 2	patient 3	patient 4
parameter set 1	MAD ₁₁	MAD ₁₂	MAD ₁₃	MAD ₁₄
parameter set 2	MAD ₂₁	MAD ₂₂	MAD ₂₃	MAD ₂₄
parameter set 3	MAD ₃₁	MAD ₃₂	MAD ₃₃	MAD ₃₄

patient_j = {MAD_{ij}}, i: # sets, j: # patients

k-means clustering



calculate cluster validity measures

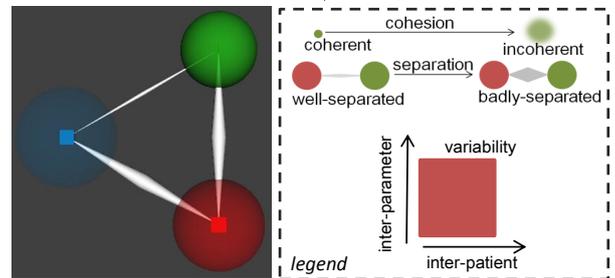


Figure 7: Partitioning a patient cohort based on TCP treatment response (T3). The cluster analysis view adapted from [RvdHD*15] is used for the visual optimization of clustering.

Bi-directional design of TCP modeling workflow (T4). With the introduction of (T4), we enable clinical researchers to start their workflow from the desired outcome, to identify and compare the strategies that achieve it. For this, the user defines an acceptability range for the desired TCP outcome, by sketching it on a canvas (Figure 8). Then, all the acceptable parameter combinations that can achieve this are computed, using a brute-force search. The user is presented with these combinations, using a heatmap matrix (Figure 8). In this matrix, every column corresponds to an acceptable combination and every row to one of the parameters. In the last

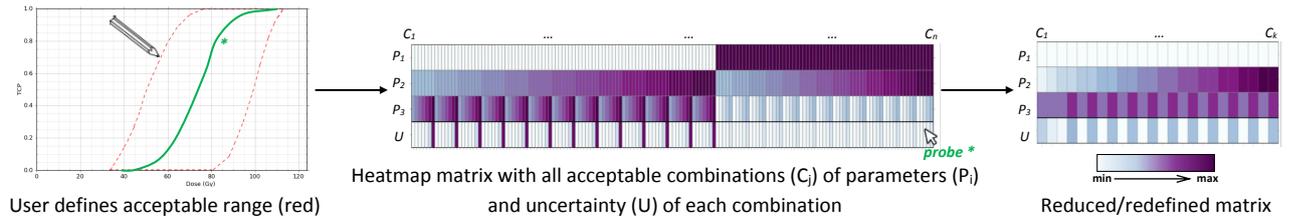


Figure 8: Reversing the workflow in TCP modeling (**T4**).

row, we present also the quantified uncertainty that is introduced by each one of these combinations, calculated in the same way, as in (**T1**). The colormap denotes the range of values for each parameter. The user can interact with the matrix and threshold values that are not plausible or interesting for the analysis, or even select combinations based on their uncertainty (Figure 8). Probing and linking is employed for the inspection of the TCP curve of each combination. The functionality of (**T4**) is expected to open new ways of exploration and analysis for clinical researchers, as up to now the workflow was done in one direction. Now, also the inverse is possible.

Implementation. We implemented the visual tool in Python as a DeVIDE module [BP08], using the Visualization Toolkit (VTK), numpy, scipy, matplotlib and scikit-learn.

5. Results

To assess the value of our visual tool, we performed an evaluation, inspired by the paper of Lam et al. [LBI*12]. The evaluation was performed with nine domain experts from two clinical institutions. The group of participants included three physicists, five medical physicists and one biomedical engineer. Their field experience varies from medium (<5 yrs) to very high (>10 yrs). Two of the participants were actively involved in the design of our tool; both with a very high level of experience in the field of TCP modeling.

In the *first part* of the evaluation, all participants were involved. We demonstrated the visual tool, where we showed the main components, simulating the visual environment for the exploration and analysis of a TCP modeling workflow. The evaluation participants observed the demonstration and were involved in an active discussion about the various visualizations. Then, they completed a questionnaire.

The *second part* was conducted only with the two participants involved in the design of the tool and the analysis was performed with data already familiar to them. For a deeper understanding on the insights that the tool provides, we performed a case study with hands-on exploration. Each of the four tasks of Section 2 was performed with the thinking-out-loud method, as the clinical researchers explained and reasoned on findings in the data. At the end, we asked them to complete again the same questionnaire as previously, to see whether their opinion changed or not, after interacting with the tool.

5.1. Evaluation: Interviews

During the interviews, the participants completed a questionnaire. The first questions were related to the main tasks of Section 2. Each question required an open answer, but also grading using Likert scales (1-5) for the perceived effectiveness, efficiency and satisfaction. To avoid compromising the results, we separated in our analysis the two people involved in the design from the other seven. Half of the tasks were graded higher by the first group and the other half by the second, but it results that the two groups had comparable results without significant difference. Also, we separated our analysis based on level of experience, as it possibly indicates different user categories, performing different tasks. Again, the results were comparable among the different groups. All measured variables received high scores (Figure 9), with a minimum average grade of 4. Uncertainty (**T1**) and Sensitivity (**T2**) received high grades. Partitioning (**T3**) received lower grades, but not lower than 3 (Figure 9). This was explained by the fact that participants wanted to see additional information on the data, when partitioning their cohorts. After the case study of the following section, the two participants involved in the design recompiled the questionnaire. For the efficiency of (**T1**) and effectiveness of (**T3**), the grades improved (Figure 9). This is interesting, as these two participants consist half of our group with very high experience, who had initially graded (**T3**) lower than all the others. This could be an indication that after hands-on exploration, this task became clearer to them. Overall, the results between the two rounds are consistently high (Figure 9).

The nine participants were also asked to compare the visual tool to what they are currently using and to evaluate the overall usefulness of our tool. They commented that they currently, "do not have any other means of analysis, apart from looking at individual graphs". For them, the framework requires training and a level of familiarization, but it "removes a significant overhead from the analysis", giving "important input". More specifically, the uncertainty part (**T1**) provides "directly understandable and quantitative feedback", while with sensitivity analysis (**T2**) they can "perceive directly the influence of the dose prescription". For the cohort partitioning part (**T3**), there were mixed opinions. According to the evaluation participants, it "raises a lot of questions about the subgroups of the cohorts". It could be the "most important clinical application", but it should be done also "based on other variables", or also "for intra-tumor regions". Reversing the workflow (**T4**) can have "great potential". All participants agreed that the visual tool is overall understandable and useful. The strong features of the visual tool are the ability to perform a voxel-based analysis - especially, the probing and linking functionality in the TCP curves and

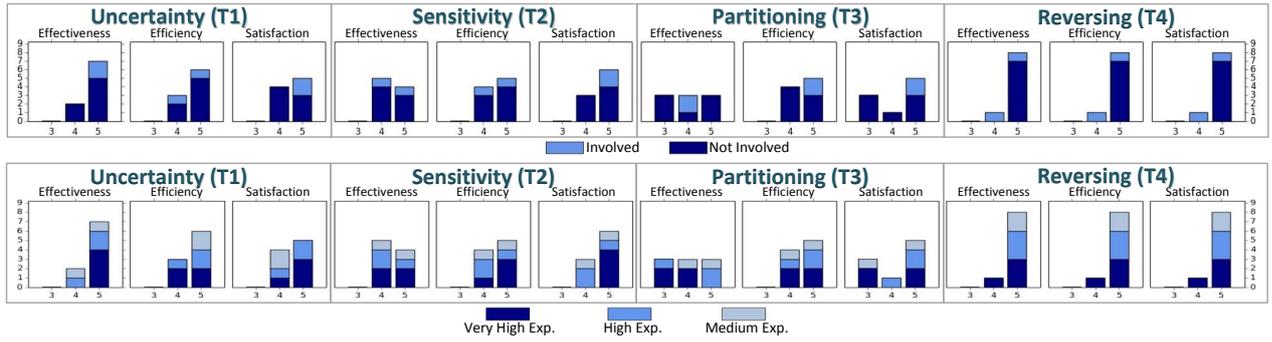


Figure 9: Schematic representation of the evaluation results, for each one of the tasks of Section 2.

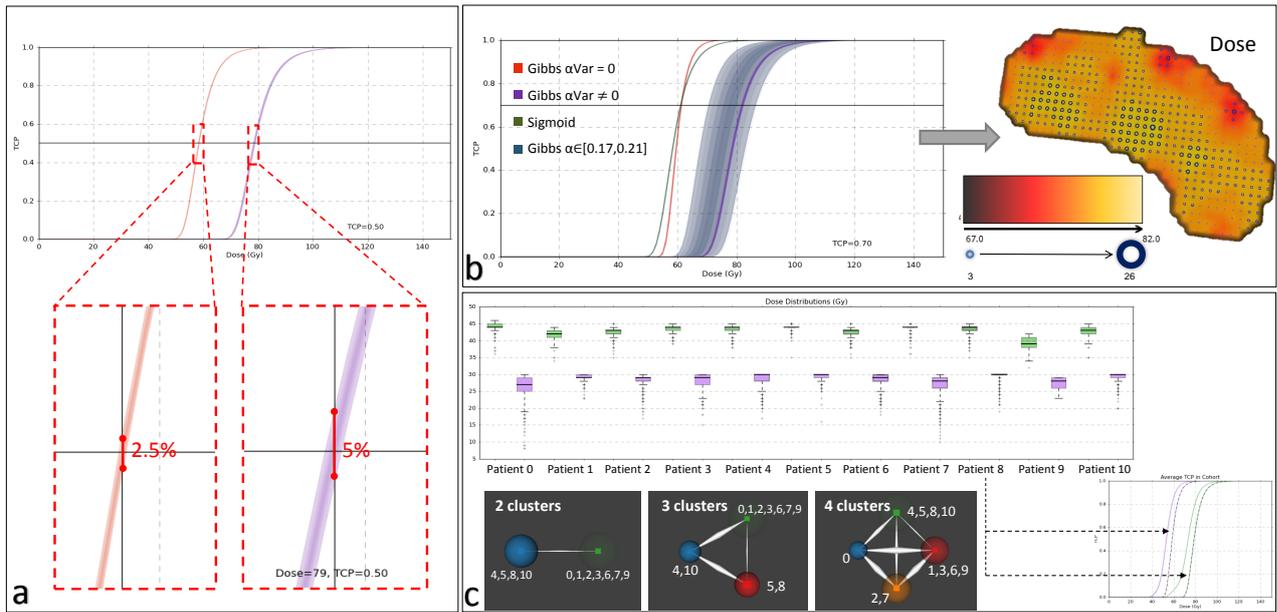


Figure 10: Results from the case study.

the view on the variability from the different modeling approaches (**T2**), as well as the workflow reversing task (**T4**). Improvement proposals were mostly related to cohort partitioning (**T3**).

5.2. Evaluation: Case study

For the case study, ADC data from a cohort of 11 locally advanced prostate tumor patients was used. The ADC maps were derived with a b -value of 1000 and have a size of $256 \times 256 \times 24$ voxels and a resolution of $0.97 \times 0.97 \times 3.6$.

During the task of uncertainty (**T1**), it was noticed that CD might be overestimated in literature, as visualized by the dominant purple color (see Figure 4). Some voxels (green) have been noticed to be sometimes misdelineations of voxels that belong to the bladder or to the urethra. In the rest of cases, like in Figure 4, these are locations in the prostate that should be checked more thoroughly. Less uncertainty is expected within tumors, due to lower ADC val-

ues. The effect of the uncertainty on the TCP was also found to be interesting: it reaches almost 5% of the TCP for the *Gibbs* approach (Figure 10-a, purple curve) and 2.5% for the *sigmoid* approach (Figure 10-a, orange curve), at $D_{50\%}$, i.e., the dose required for achieving 50% control only in the tumor location.

During the exploration of sensitivity (**T2**), four examples of parameter sets were explored. The first two are *Gibbs* approaches with the same RT strategy, but they differ in the αVar parameter. The third is a *sigmoid* approach with the same RT strategy. The last is an additional *Gibbs* approach with a range of α between 0.17 and 0.21. In the resulting TCP curve graph, there are indications that the *sigmoid* model might predict tumor control with a lower dose, i.e., the curve is more to the left, than the respective *Gibbs* (Figure 10-b). Also, when the αVar , which is a parameter that models intra-tumor variability is neglected, then the model suffers less from uncertainty (Figure 10-b). For ranging α , the effect on the TCP is more prominent. In this case, probing the TCP at 70% shows that

a dose ranging between 43 and 82 Gy is required within the tumor only. The variability between the four models, though, is large for the whole tumor (Figure 10-b).

For cohort partitioning (**T3**), after probing the TCP at 70%, we obtain the dose distributions per patient (Figure 10-c). Patient 8 seems to have a different behavior. His TCP curve (Figure 10-c, dotted TCP curve) is on the right side of the average curve, which means that to achieve a 70% TCP, he requires a higher dose, as he is a patient with a much larger tumor. After interactive clustering, the visually optimal cluster analysis view is achieved with two clusters, where patients behave similarly in terms of TCP curves, e.g., the ones in the blue cluster are all on the right side of the average curve.

In reversing the workflow (**T4**), patient 8 was explored to check whether a more satisfying strategy can be identified. A wide range of acceptable TCP is drawn (Figure 8). More than 200 different combinations are identified, as seen by the columns of the heatmap matrix. On first sight, it seems that the *sigmoid* approach (Figure 8 - heatmap, first row, purple section) may be less sensitive to changes in parameters than *Gibbs* (Figure 8 - heatmap, first row, white section), as less combinations are computed for the *sigmoid*. Also, this approach may suffer less from uncertainty (Figure 8 - heatmap, last row), as the range of uncertainties does not go up to the maximum value of uncertainty (no deep purple). By redefining the acceptable limits for the parameters and the uncertainty, only 24 different combinations are preserved (Figure 8). According to the evaluators, (**T4**) functionality could be helpful, to determine the suitability of this patient for a specific therapy. However, for this, no conclusions can be made, as it would also require the involvement of oncologists, and a more extensive study. The examined cases are meant to demonstrate the use of the visual tool; not as an actual analysis with direct clinical inferences.

6. Conclusions and Future Work

In this work, we proposed a visual tool to enable clinical researchers to explore and analyze different aspects of the TCP modeling workflow. We tackled the quantification and interactive exploration of uncertainty and its propagation to TCP modeling, parameter sensitivity analysis of TCP models, cohort partitioning based on treatment response and a novel functionality for enabling also a reverse workflow. Nine clinical researchers evaluated and confirmed the usefulness of the visual tool, as it opens new possibilities and provides access to new insight in the data. We illustrated this also with a case study. A direction for future work includes improving the partitioning of the cohorts to enable clustering also based on other attributes, and also linking to intra-tumor tissue characteristics [RvdHD*15]. In this way, more meaningful inter-patient analysis can also be performed. The proposed visual tool is a promising basis for clinical researchers to gain more knowledge on their complex TCP modeling processes, to explore the data from the models in a more insightful way and to generate and confirm new hypotheses.

7. Acknowledgements

This work was supported by the European Commission's Seventh Framework Programme (Call: FP7-ICT-2011-9, activity ICT-

9-5.2 - Virtual Physiological Human) and is part of the project "DR THERAPAT-Digital Radiation Therapy Patient". We would like to thank the Department of Clinical Medicine, University of Bergen for their data; the evaluation participants from the Netherlands Cancer Institute (NKI-AvL) and Århus University Hospital; and the reviewers for their valuable feedback.

References

- [BBP10] BUSKING S., BOTHA C. P., POST F. H.: Dynamic multi-view exploration of shape spaces. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 973–982. 3
- [BJH*14] BONNEAU G.-P., HEGE H.-C., JOHNSON C. R., OLIVEIRA M. M., POTTER K., RHEINGANS P., SCHULTZ T.: Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization*. Springer, 2014, pp. 3–27. 3
- [BJEK*11] BONEKAMP D., JACOBS M. A., EL-KHOULI R., STOIANOVICI D., MACURA K. J.: Advancements in MR imaging of the prostate: from diagnosis to interventions. *Radiographics* 31, 3 (2011), 677–703. 2
- [BKC*13] BORGIO R., KEHRER J., CHUNG D. H., MAGUIRE E., LARAMEE R. S., HAUSER H., WARD M., CHEN M.: Glyph-based visualization: Foundations, design guidelines, techniques and applications. *Eurographics State of the Art Reports* (May 2013), 39–63. 5
- [BM10] BRUCKNER S., MÖLLER T.: Result-driven exploration of simulation parameter spaces for visual effects design. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (2010), 1468–1476. 3
- [BP08] BOTHA C. P., POST F. H.: Hybrid scheduling in the DeVIDE dataflow visualisation environment. In *SimVis* (2008), pp. 309–322. 7
- [BPF11] BERGER W., PIRINGER H., FILZMOSER P., GRÖLLER E.: Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 911–920. 3
- [Bre94] BREWER C. A.: Guidelines for use of the perceptual dimensions of color for mapping and visualization. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology* (1994), International Society for Optics and Photonics, pp. 54–63. 4
- [BVPR09] BRECHEISEN R., VILANOVA A., PLATEL B., ROMENY B. T. H.: Parameter sensitivity visualization for dti fiber tracking. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 1441–1448. 3
- [BWE05] BOTCHEN R. P., WEISKOPF D., ERTL T.: Texture-based visualization of uncertainty in flow fields. In *Visualization, 2005. VIS 05. IEEE* (2005), IEEE, pp. 647–654. 3
- [CLEK13] COFFEY D., LIN C.-L., ERDMAN A. G., KEEFE D. F.: Design by dragging: An interface for creative forward and inverse design with simulation ensembles. *Visualization and Computer Graphics, IEEE Transactions on* 19, 12 (2013), 2783–2791. 3
- [CMvdHR*16] CASARES-MAGAZ O., VAN DER HEIDE U., RØRVIK J., STEENBERGEN P., MUREN L.: A Tumor Control Probability model for Radiotherapy of prostate cancer using MRI-based Apparent Diffusion Coefficient maps. *Radiotherapy and Oncology (in press)* (2016). 2, 3, 4
- [DJFB05] DELANEY G., JACOB S., FEATHERSTONE C., BARTON M.: The role of radiotherapy in cancer treatment. *Cancer* 104, 6 (2005), 1129–1137. 1
- [DKLP02] DJURCILOV S., KIM K., LERMUSIAUX P., PANG A.: Visualizing scalar volumetric data with uncertainty. *Computers & Graphics* 26, 2 (2002), 239–248. 3
- [DPD*15] DIEHL A., PELOROSSO L., DELRIEU C., SAULO C., RUIZ J., GRÖLLER M. E., BRUCKNER S.: Visual analysis of spatio-temporal data: Applications in weather forecasting. *Computer Graphics Forum* 34, 3 (2015), 381–390. 3

- [GPT07] GIBBS P., PICKLES M. D., TURNBULL L. W.: Repeatability of echo-planar-based diffusion measurements of the human prostate at 3T. *Magnetic resonance imaging* 25, 10 (2007), 1423–1429. 3, 4
- [GR04] GRIGORYAN G., RHEINGANS P.: Point-based probabilistic surfaces to show surface uncertainty. *Visualization and Computer Graphics, IEEE Transactions on* 10, 5 (2004), 564–573. 3
- [Hen03] HENGL T.: Visualisation of uncertainty using the hsi colour model: computations with colours. In *7th International Conference on GeoComputation* (2003), Hrvatska znanstvena bibliografija i MZOS-Svibor. 3
- [HMC*13] HOLLT T., MAGDY A., CHEN G., GOPALAKRISHNAN G., HOTEIT I., HANSEN C. D., HADWIGER M.: Visual analysis of uncertainties in ocean forecasts for planning and operation of off-shore structures. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific* (2013), IEEE, pp. 185–192. 3
- [HMZ*14] HOLLT T., MAGDY A., ZHAN P., CHEN G., GOPALAKRISHNAN G., HOTEIT I., HANSEN C. D., HADWIGER M.: Ovis: A framework for visual analysis of ocean forecast ensembles. *Visualization and Computer Graphics, IEEE Transactions on* 20, 8 (2014), 1114–1126. 3
- [HSSK14] HERMANN M., SCHUNKE A. C., SCHULTZ T., KLEIN R.: A visual analytics approach to study anatomic covariation. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE* (2014), IEEE, pp. 161–168. 3
- [JS03] JOHNSON C. R., SANDERSON A. R.: A next step: Visualizing errors and uncertainty. *Computer Graphics and Applications, IEEE* 23, 5 (2003), 6–10. 3
- [KLR*13] KLEMM P., LAWONN K., RAK M., PREIM B., TOENNIES K. D., HEGENSCHIED K., VÖLZKE H., OELTZE S.: Visualization and analysis of lumbar spine canal variability in cohort study data. In *VMV* (2013), pp. 121–128. 3
- [KOJL*14] KLEMM P., OELTZE-JAFRA S., LAWONN K., HEGENSCHIED K., VOLZKE H., PREIM B.: Interactive visual analysis of image-centric cohort study data. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (2014), 1673–1682. 3
- [KTH*10] KALLEHAUGE J. F., TANDERUP K., HAACK S., NIELSEN T., MUREN L. P., FOKDAL L., LINDEGAARD J. C., PEDERSEN E. M.: Apparent diffusion coefficient (ADC) as a quantitative parameter in diffusion weighted MR imaging in gynecologic cancer: Dependence on b-values used. *Acta Oncologica* 49, 7 (2010), 1017–1022. 2, 4
- [LBI*12] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9 (2012), 1520–1536. 7
- [LLPY07] LUNDSTROM C., LJUNG P., PERSSON A., YNNERMAN A.: Uncertainty visualization in medical volume rendering using probabilistic animation. *Visualization and Computer Graphics, IEEE Transactions on* 13, 6 (2007), 1648–1655. 3
- [LV02] LEE C. H., VARSHNEY A.: Representing thermal vibrations and uncertainty in molecular surfaces. In *Electronic Imaging 2002* (2002), International Society for Optics and Photonics, pp. 80–90. 3
- [PCR*11] PRETORIUS A. J. B. M.-A. P., CARPENTER A. E., RUDDLE R., ET AL.: Visualization of parameter space for image analysis. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (2011), 2402–2411. 3
- [PH11] PÖTHKOW K., HEGE H.-C.: Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *Visualization and Computer Graphics, IEEE Transactions on* 17, 10 (2011), 1393–1406. 3
- [PMW13] PFAFFELMOSER T., MIHAI M., WESTERMANN R.: Visualizing the variability of gradients in uncertain 2d scalar fields. *Visualization and Computer Graphics, IEEE Transactions on* 19, 11 (2013), 1948–1961. 3
- [PRJ12] POTTER K., ROSEN P., JOHNSON C. R.: From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing*. Springer, 2012, pp. 226–249. 3
- [PRW11] PFAFFELMOSER T., REITINGER M., WESTERMANN R.: Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 951–960. 3
- [PWB*09] POTTER K., WILSON A., BREMER P.-T., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C. R.: Ensemble-Vis: A framework for the statistical visualization of ensemble data. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on* (2009), IEEE, pp. 233–240. 3
- [Rup11] RUPPERT D.: *Statistics and Data Analysis for Financial Engineering*. Springer Texts in Statistics, 2011. 6
- [RvdHD*15] RAIDOU R. G., VAN DER HEIDE U. A., DINH C. V., GHOBADI G., KALLEHAUGE J., BREEUWER M., VILANOVA A.: Visual analytics for the exploration of tumor tissue characterization. *Computer Graphics Forum* 34, 3 (2015), 11–20. 6, 9
- [SHB*14] SEDLMAIR M., HEINZL C., BRUCKNER S., PIRINGER H., MOLLER T.: Visual parameter space analysis: A conceptual framework. *Visualization and Computer Graphics, IEEE Transactions on* 20, 12 (2014), 2161–2170. 3
- [SMB*07] SØVIK Å., MALINEN E., BRULAND Ø. S., BENTZEN S. M., OLSEN D. R.: Optimization of tumour control probability in hypoxic tumours by radiation dose redistribution: a modelling study. *Physics in medicine and biology* 52, 2 (2007), 499. 2
- [SMB*10] STEENWIJK M. D., MILLES J., BUCHEM M., REIBER J., BOTHA C. P.: Integrated visual analysis for heterogeneous datasets in cohort studies. In *IEEE VisWeek Workshop on Visual Analytics in Health Care* (2010), vol. 3. 3
- [SPA*14] SCHMIDT J., PREINER R., AUZINGER T., WIMMER M., GRÖLLER M. E., BRUCKNER S.: YMCA-Your mesh comparison application. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on* (2014), IEEE, pp. 153–162. 3
- [SSSSW13] SCHULTZ T., SCHLAFFKE L., SCHÖLKOPF B., SCHMIDT-WILCKE T.: HiFiVE: a Hilbert space embedding of fiber variability estimates for uncertainty modeling and visualization. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 121–130. 3
- [SZB*09] SANYAL J., ZHANG S., BHATTACHARYA G., AMBURN P., MOORHEAD R. J.: A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 1209–1218. 3
- [SZD*10] SANYAL J., ZHANG S., DYER J., MERCER A., AMBURN P., MOORHEAD R. J.: Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *Visualization and Computer Graphics, IEEE Transactions on* 16, 6 (2010), 1421–1430. 3
- [TOG06] TANDERUP K., OLSEN D. R., GRAU C.: Dose painting: art or science? *Radiotherapy and Oncology* 79, 3 (2006), 245–248. 2
- [TWSM*11] TORSNEY-WEIR T., SAAD A., MÖLLER T., HEGE H.-C., WEBER B., VERBAVATZ J.-M., BERGNER S.: Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration. *Visualization and Computer Graphics, IEEE Transactions on* 17, 12 (2011), 1892–1901. 3
- [WN93] WEBB S., NAHUM A.: A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density. *Physics in Medicine and Biology* 38, 6 (1993), 653. 1, 2
- [ZGP14] ZHANG Z., GOTZ D., PERER A.: Iterative cohort analysis and exploration. *Information Visualization* (2014), 1473871614526077. 3
- [ZWK10] ZEHNER B., WATANABE N., KOLDITZ O.: Visualization of gridded scalar data with uncertainty in geosciences. *Computers & Geosciences* 36, 10 (2010), 1268–1275. 3