A new voxel principal component analysis for predicting and spatially characterizing rectal toxicity following prostate cancer radiotherapy

Auréline FARGEAS[1,2], Amar KACHENOURA[1,2], Louis H. LIU[1,2], Frédéric COMMANDEUR[1,2], Gaël DRÉAN[1,2], Caroline LAFOND[3], Laurent ALBERA[1,2], Oscar ACOSTA[1,2], Renaud DE CREVOISIER[1,2,3]

[1]INSERM, U1099, Rennes, F-35000, France.

[2]Laboratoire Traitement du Signal et de l'Image (LTSI) Université de Rennes 1, Rennes, F-35000, France.

[3]Centre Eugène Marquis Département de Radiotherapie, Rennes, F-35000, France.

aureline.fargeas@gmail.com

Résumé – En radiothérapie du cancer de la prostate, il existe une corrélation entre la distribution de dose et l'apparition d'évènements de toxicité suite au traitement. La plupart des méthodes actuelles portant sur la prédiction des événements de toxicité sont basées sur l'histogramme dose-volume. Cependant, ces méthodes ne sont pas en mesure d'établir une corrélation entre l'apparition d'évènements secondaires et de la distribution de dose 3D au niveau voxel. Dans ce travail, basé sur la distribution de dose 3D complète, une analyse en composantes principales a été effectuée pour prédire l'apparition d'effets secondaires et pour mettre en évidence des zones sur-irradiées en corrélation avec l'apparition de saignement rectal suite au traitement. Nous avons cherché à identifier, à partir de la distribution de dose 3D, deux bases (caractérisant respectivement les patients avec et sans saignement rectal) en utilisant l'analyse en composantes principales. Nous avons réalisé la prédiction en évaluant une nouvelle variable prédictive calculée en mesurant la distance entre la distribution de dose d'un nouvel individu et les deux sous-espaces représentés par les bases. La méthode, appliquée à une base de données de 118 patients (traités par radiothérapie externe), a été comparée à une récente approche basée sur une analyse en composantes principales des pas de dose, obtient de bon résultat (AUC=0.87).

Abstract – In prostate cancer radiotherapy, understanding the correlation between the dose distribution and the occurrence of undesirable side-effects is crucial to correlate the treatment outcome with the planning parameters. Most of the current methods addressing the prediction of the toxicity events are based on dose volume histogram. However, these methods are not able to correlate the toxicity and the spatial dose distribution at the voxel level. Using the whole 3D planned dose distribution, a principal component analysis based approach was performed to predict late rectal toxicity and to construct a dose pattern characterizing the difference between patients with rectal bleeding and those without. After a non-rigid registration, the method aimed at identifying, from 3D dose distribution, two basis (characterizing patients with/without rectal bleeding). The prediction was performed by evaluating a new variable computed by measuring the distance of a new individual 3D dose distribution to both subspaces spanned by the bases. The method, applied to a total of 118 patients treated for prostate cancer radiotherapy and compared with a recent principal component analysis approach based only on the DVH, showed good performance (AUC=0.87) and suggested that the method is able to establish the correlation between dose and toxicity outcomes.

1 Introduction

The main challenge in prostate cancer radiotherapy (PCRT) is to deliver the prescribed dose to the clinical target volume (prostate and seminal vesicles) while minimizing the dose to the neighboring organs at risk (the rectum and the bladder) thus avoiding subsequent toxicity-related events. Several studies have been proposed to assess the risk of toxicity according to the dose : the normal tissue complication probability (NTCP) models. On of the most used NTCP model were proposed by Lyman in the context of uniform radiation therapy, and subsequently adapted to 3D-conformal radiation [13]. These models operate in two different steps: 1) dose volume histogram (DVH) reduction; and 2) probability-mapping. In [15], based on principal component analysis (PCA), the most relevant DVH bins are extracted and used to predict late toxicity. These methods do not perform a formal prediction exploiting the spatial characteristics of the dose distributions since they considered the organs as having homogeneous radio-sensitivity. Buettner *et al.* [2, 3] proposed the use of the three-dimensional (3D) planned dose distribution addressing the issue of spatial information loss. In [2], a classification approach based on locally connected neural network using a two-dimensional dose-surface maps was performed. In [3], a parameterized representation of the dose to describe its geometrical properties was proposed. Nevertheless, in these methods, the voxel across the whole population information was not jointly exploited. Performing classification by simultaneously exploiting the 3D signal across a population is challenging because the inter-individual anatomical variability leads to a misalignment of information. To cope with this issue, non-rigid registration methods can be used to map all the data to a common coordinate system where voxel analysis is meaningful in terms of spatial localization [6]. Following this idea, previous classification approaches exploiting the 3D signal across a given population have been proposed. For instance, Principal Component Analysis (PCA) was used by Fripp *et al.* [10] to discriminate Alzheimer's disease and normal elderly control participants. With the same objec-

tive, Higdon *et al.* [11] performed a comparison of different classification methods (logistic regression, Linear Discriminant Analysis (LDA) and quadratic discriminant analysis). In the context of rectal bleeding after PCRT, we proposed in [7] to use PCA to analyze non-rigidly registered dose distributions. We identified one basis of orthogonal vectors from 3D dose distributions of the whole database (patients with and without rectal bleeding) allowing for classification. In the same context, we previously proposed in [4, 8] to identify, from 3D dose distribution, two basis (characterizing patients with and without rectal bleeding) using novel methods based on a deterministic multi-way analysis and a semi-nonnegative independent component analysis, respectively. Thus, the classification was performed by measuring the distance of a new individual 3D dose distribution to both subspaces spanned by the bases.

We proposed, in this paper, a novel voxel-based PCA method to predict late rectal toxicity and to construct a dose pattern that spatially characterizes the difference between patients with and without rectal bleeding. This procedure aimed at computing two bases of vectors from 3D planned dose distribution (3DpDD) of patients with and without rectal bleeding, respectively. A predictive parameter was then calculated using its distances to the subspaces spanned by both bases. The method, tested on real clinical data, outperformed a recent PCA approach based only on the DVH [15].

2 Materials and method

Materials and preprocessing step

A total of 118 patients treated for localized prostate cancer with intensity-modulated radiation therapy (IMRT) were included in the study. The used treatment planning system was Pinnacle V7.4 (Philips Medical System, Madison, WI). The total prescribed dose was 46 Gy to the seminal vesicles delivered in 4.6 weeks, and 80 Gy to the prostate delivered in 8 weeks, with a standard fractionation of 2 Gy per fraction. The patient positioning, CT acquisition, volume delineations and dose constraints complied with GETUG 06 recommendations as described in [1]. For the rectal wall, the constraints were: the maximum dose in 1.8cm^2 had to be lower than 76Gy and a V72 Gy (volume receiving at least 72Gy) lower than 25%. Rectal toxicity events were prospectively collected and scored according to the Common Terminology Criteria for Adverse Events (CTCAE) version 3.0. The events were defined as rectal bleeding ($\geq$ Grade 1), at least one episode occurring between 6 months and 2 years after RT. Patients with a history of hemorrhoids were not allowed to be scored as Grade 1 bleeding. A total of 31 patients presented at least a Grade 1 late rectal bleeding event. For each patient, the planning data were available: the planning CT, the manual delineation of the organs and planned dose distribution.

The rectum of each patient has been described using both the Euclidean distance maps outside the rectum and the Laplacian scalar fields inside the rectum computed for each patient [12, 14]. rectal descriptors map were non-rigidly registered using the demons algorithm, towards a selected template to be used as the common coordinate system. This typical selected individual maximized a similarity criterion which is the sum of squared differences computed after rigid registration.

For each patient, a $(N_1 \times N_2 \times N_3)$ 3DpDD (inside the rectum) is vectorized to obtain an $(N \times M^{(1)})$ matrix $\mathbf{X}^{(1)}$ for patients with rectal bleeding and an $(N \times M^{(2)})$ matrix $\mathbf{X}^{(2)}$ for those without rectal bleeding, where $N = N_1 \times N_2 \times N_3$ represents the number of voxels, and $M^{(1)}$ and $M^{(2)}$ the number of individuals with and without rectal bleeding, respectively.

A PCA-based subspace identification

This section presents the statistical analysis used to estimate the vector bases $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, that characterize patients with and without rectal bleeding, respectively. In the sequel, we present the algorithm in the general case, namely to estimate the base $\mathbf{A}^{(i)}$, whatever $i \in \{1, 2\}$. Let $\mathbf{X}^{(i)}$ be the vertorized 3DpDD of patients of the i^{th} group. Assume that each individual's 3DpDD, $\mathbf{x}^{(i)}[m] = [x_1^{(i)}[m], \ldots, x_N^{(i)}[m]]^T$, is one realization of a N-dimensional random process $\{\mathbf{x}^{(i)}[m]\}$. $\mathbf{x}^{(i)}[m]$ is modeled as a linear combination of a basis vectors $(\mathbf{a}_1^{(i)}, \ldots, \mathbf{a}_{R^{(i)}}^{(i)})$ and its coordinates, $\mathbf{s}^{(i)}[m] = [s_1^{(i)}[m], \ldots, s_{R^{(i)}}^{(i)}[m]]^T$. Thus, the problem we tackle can be formulated as a blind source separation one: given a real random vector process $\mathbf{x}^{(i)}[m]$, find a $(N \times R^{(i)})$ matrix $\mathbf{A}^{(i)}$ and a $R^{(i)}$-dimensional random process $\mathbf{s}^{(i)}[m]$, such that for each index m: $\mathbf{x}^{(i)}[m] = \mathbf{A}^{(i)}\mathbf{s}^{(i)}[m]$. In addition, the coordinate of $\mathbf{x}^{(i)}[m]$ on basis $\mathbf{A}^{(i)}$ are decorrelated. So the vector basis $\mathbf{A}^{(i)}$ can be estimated using a principal component analysis. PCA is one of the oldest and popular technique in data analysis. Typically, the PCA of $\mathbf{x}^{(i)}[m] = [x_1^{(i)}[m], \ldots, x_N^{(i)}[m]]^T$ consists in looking to an overdetermined $(N^{(i)} \times R^{(i)})$ (i.e $R^{(i)} \leq N^{(i)}$) orthonormal linear transform $\mathbf{W}^{(i)}$ such that the $R^{(i)}$ components of the vector $\mathbf{s}^{(i)}[m] = [s_1^{(i)}[m], \ldots, s_{R^{(i)}}^{(i)}[m]]^T$ are mutually uncorrelated. Once the matrices $\mathbf{W}^{(i)}$ are estimated, the vector basis $\mathbf{A}^{(i)}$ are calculated as the pseudo-inverse of $\mathbf{W}^{(i)}$.

Feature selection and predictive parameter identification

Using a training database, the PCA was used to identify two vector bases $\mathbf{A}^{(1)} = [\mathbf{a}_1^{(1)}, \ldots, \mathbf{a}_{R^{(1)}}^{(1)}]$ and $\mathbf{A}^{(2)} = [\mathbf{a}_1^{(2)}, \ldots, \mathbf{a}_{R^{(2)}}^{(2)}]$, $\boldsymbol{A}^{(i)}$, spanning two vector subspaces, $\boldsymbol{\mathcal{A}}^{(1)}$ and $\boldsymbol{\mathcal{A}}^{(2)}$, that characterize patients with and without rectal bleeding, respectively. The key question was: how to select the most informative features of $\boldsymbol{A}^{(i)}$, that better discriminates the two classes, namely the rectal and non-rectal bleeding groups. In this study, to choose of the optimal couple $(\mathbf{a}_{opt}^{(1)}, \mathbf{a}_{opt}^{(2)})$ of features, we proposed to use the correlation ratio criterion. For this criterion, the optimal features are selected by minimizing the ratio between the correlation within each group and the correlation between the two groups. Thus, using the validation database, the vectorized 3D dose distribution $\boldsymbol{g}$ of a new patient is orthogonally projected onto the two selected subspaces $\boldsymbol{\mathcal{A}}_{opt}^{(i)}$.

The orthogonal matrix projector $\boldsymbol{\Delta}_{\boldsymbol{\mathcal{A}}_{opt}^{(i)}}^{\perp}$ on subspace $\boldsymbol{\mathcal{A}}_{opt}^{(i)}$ can then be computed as follows:

$$\forall i \in \{1, 2\}, \qquad \boldsymbol{\Delta}_{\boldsymbol{\mathcal{A}}_{opt}^{(i)}}^{\perp} = \mathbf{a}_{opt}^{(i)} \mathbf{a}_{opt}^{(i)\,\mathsf{T}} \tag{1}$$

Consequently, the euclidean distances between $\boldsymbol{g}$ and its orthogonal projection were computed by:

$$\forall i \in \{1,2\}, \qquad d_{\boldsymbol{g}}^{(i)}(\mathcal{A}_{opt}^{(i)}) = ||\boldsymbol{g} - \boldsymbol{\Delta}_{\mathcal{A}_{opt}^{(i)}}^{\perp} \boldsymbol{g}|| \tag{2}$$

where $||.||$ is the euclidean norm operator. Finally, a new distance-based parameter, denoted $v_{\boldsymbol{g}}$, was computed as follows:

$$v_{\boldsymbol{g}} = \frac{1}{2} - \frac{d_{\boldsymbol{g}}^{(1)}(\mathcal{A}_{opt}^{(1)})}{d_{\boldsymbol{g}}^{(1)}(\mathcal{A}_{opt}^{(1)}) + d_{\boldsymbol{g}}^{(2)}(\mathcal{A}_{opt}^{(2)})} \tag{3}$$

It was interesting to note that this step was performed by the euclidean norm and did not require any evolutive classification method. Because of the reduced number of patients in our database (especially those suffering from rectal bleeding), a leave-one-out cross validation was performed to evaluate the proposed approach. A 3D individual planned dose distribution was extracted (validation step) and our approach was applied to the remaining 3DpDD (training step). This was repeated such that each patient was used as a test sample.

Assessment of the predicted capabilities

In order to assess the predictive capabilities of the distance-based parameter $v_{\boldsymbol{g}}$ for each patient, the receiving operator characteristic (ROC) curve was used on these parameters. The area under the curve (AUC) was used as a measure of the performance of the proposed model, as previously described [9]. We also used standard logistic regression of the following ratio $v_{\boldsymbol{g}}$ for each patient to estimate the risk of toxicity, in order to assert that our parameter is statistically-significant.

Spatial characterization

Eventually, we can wonder if the selected basis vectors $\mathbf{a}_{opt}^{(i)}$ computed by PCA could highlight region correlated with rectal bleeding. To do so, we first reconstructed the 3DpDD of all the patients using only the optimal features ($\mathbf{a}_{opt}^{(1)}$, $\mathbf{a}_{opt}^{(2)}$). Then, the mean normalized differences between the two groups, in the rectum, was calculated.

3 Results and discussions

Figure 1 shows the ROC curves and the respective AUC of the different models used to predict 2-years Grade$\geq$ 1 rectal bleeding, using a leave-one-out cross-validation scheme. Our approach obtained an AUC of 0.87 and a significant p-value, p= 4.141^{-5}, when using an univariate analysis on the following distance-based parameter, $v_{\boldsymbol{g}}$, for each patient. These results are higher than the one obtained using the PCA approach based only on the DVH proposed by Sohn *et al* [15]. For this method, the correlation with toxicity of the first three principal components (PC) computed using DVH as single variables was investigated. According to this analysis, 1st PC and 3rd PC were the only variable significantly associated with rectal bleeding ($p = 0.03$ and $p = 0.02$, respectively). Although, the predictive capabilities of multivariate logistic regression using both 1st and 3rd PCS was not significantly correlated with the toxicity events ($p = 0.72$). Even if the obtained results are promising in terms of performance, they shall be confirmed with larger cohorts. Besides, future work will be focused on the inclusion of clinical factors in our classification procedure in order to increase the efficiency of the proposed method. For a thorough evaluation, our work will also compare the approach with NTCP model. One of the issue for this comparison to be carried out lies in the need of large cohorts to estimate the population specific parameters, witch is not the case on this study.

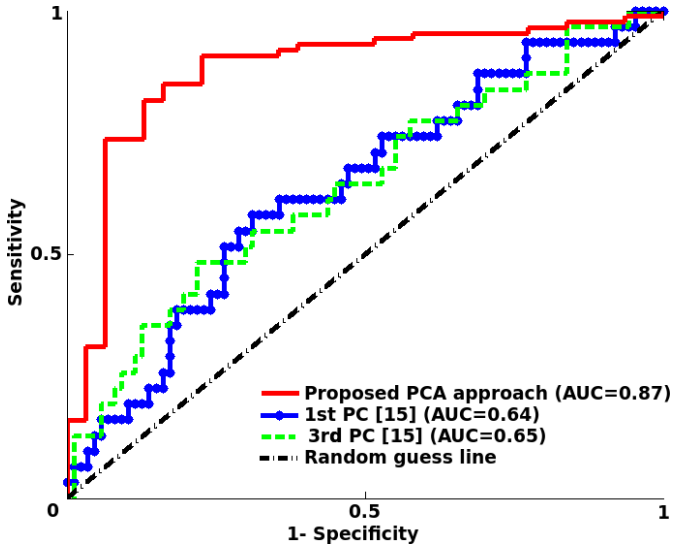

FIGURE 1 – ROC and AUC to predict 2-year grade>1 rectal bleeding using the proposed PCA approach and PCA based on DVH proposed by Sohn *et al.* [15].

Regarding the spatial characterization, figure 2 depicts a sagittal slice of the 3-dimensional mean normalized differences, in the rectum, between the mean reconstructed 3DpDD of the two groups (using the two selected subspaces $\boldsymbol{\mathcal{A}}_{opt}^{(i)}$). For a sake of clarity, the overlaid prostate and rectum of the template patient in the sagittal plane were added. The dominant pattern observed with PCA approach is in agreement with the results given in [16] which revealed associations between bowel quality of life and inferior rectal dose that could significantly influence radiation planning and prognostic models. Obviously, all these observations should be confirmed with a deeper statistical analysis based on a larger database which will be the object of forthcoming work. Identifying such spatial patterns is crucial if we aim at guiding the planning of dose distribution for patients, mainly at the inverse planning systems of IMRT.

4 Conclusion

This study focused on the analysis of 3D registered planned dose distributions across a population to predict the risk for a patient of suffering from rectal bleeding after prostate radiotherapy. The originality of the proposed PCA approach is its jointly exploits the 3D spatial patterns of dose. Representative features are extracted and spatial characterization of region correlated with rectal bleeding by a correlation ratio criteria. A distance-based predictive parameter is derived by comparing two Euclidean distances. We also reconstructed the 3DpDD of all the patients using only the optimal features. The obtained results suggest that spatial patterns representative of the

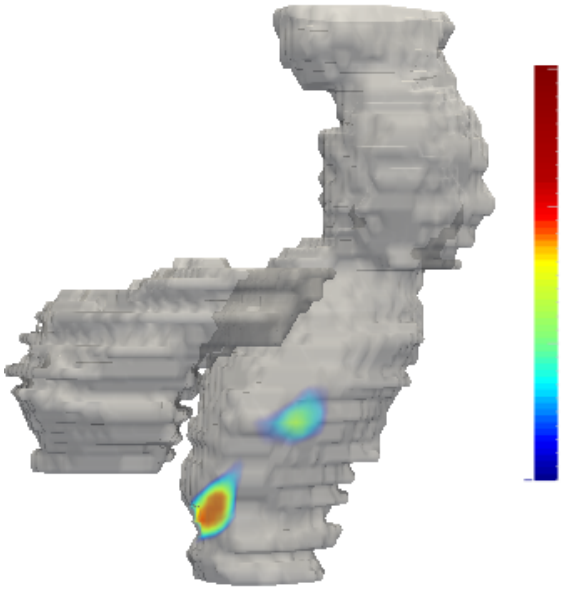

FIGURE 2 – Identification of a region where rectal bleeding would be related to the received dose: normalized mean differences between the two groups.

rectal bleeding group may be identified. The obtained results shed some light on the difficult problem of understanding dose-toxicity relationships. Thus, the PCA parameters emerges as a promising predictive variable for prediction in toxicity studies after prostate cancer radiotherapy.

Acknowledgements

This work was supported by “Region Bretagne” and has received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01 and has been partially funded by the Institut de Recherche en Sant Publique-IreSP (France), under the Plan Cancer 2009-2013 Call (No. C13005N5).

Références

[1] V. Beckendorf, S. Guerif, E. Le Prise, *et al.* ’70 Gy versus 80 Gy in localized prostate cancer: 5-year results of GETUG 06 randomized trial.’ *Int J Radiat Oncol Biol Phys*, vol. 80, pp. 1056–1063, 2011.

[2] F. Buettner, S.L. Gulliford, S. Webb and M. Partridge ’Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach’ *Phys in Med and biol*, vol. 54, no. 17, pp. 5139, 2009.

[3] F. Buettner, S. L. Gulliford, S. Webb and M. Partridge ’Modeling late rectal toxicities based on a parameterized representation of the 3D dose distribution,’ *Phys in Med and Biol*, vol. 56, no. 7, p. 2103, 2011.

[4] J. Coloigner, A. Fargeas, A. Kachenoura *et al.* ’A novel classification method for prediction of rectal bleeding in prostate cancer radiotherapy based on a semi-nonnegative ICA of 3D planned dose distributions.’, *IEEE J of Bio and Health Inf*, DOI 10.1109/JBHI.2014.2328315, 2014.

[5] L.A. Dawson, M. Biersack, G. Lockwood, *et al.* ’Use of principal component analysis to evaluate the partial organ tolerance of normal tissues to radiation.’ *Int J Rad Onco Biol Phys*, vol. 62, pp. 829–837, 2005.

[6] O. Acosta, G. Drean, J -D. Ospina *et al.*, ’Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy ’ *Phys in Med and Biol*, vol. 58, no. 8, pp. 2581–2595, 2013.

[7] A. Fargeas, A. Kachenoura, O. Acosta *et al.*, ’Feature extraction and classification for rectal bleeding in prostate cancer radiotherapy: A PCA based method.,’ *IRBM*, vol. 34, no. 4-5, pp. 296–299, 2013.

[8] A. Fargeas, L. Albera, A. Kachenoura *et al.*, ’On feature extraction and classification in prostate cancer radiotherapy using tensor decompositions’, *Med Eng and Phys*, vol. 37, no. 1, pp. 126–131, 2015.

[9] C. Fiorino, T. Rancati, G. Fellin, *et al.*, ’Late fecal incontinence after high-dose radiotherapy for prostate cancer: better prediction using longitudinal definitions.’ *Int J Rad Oncol Biol Phys*, vol. 83, pp. 38–45, 2012.

[10] J. Fripp, P. Bourgeata, O. Acosta, *et al.*, ’Appearance modeling of 11C PiB PET images: characterizing amyloid deposition in Alzheimer’s disease, mild cognitive impairment and healthy aging.’ *Neuroimage*, vol. 43, no. 3, pp. 430, 2008.

[11] R. Higdon, N. Foster, R. Koeppe, *et al.*, ’A comparison of classification methods for differentiating fronto-temporal dementia from alzheimer’s disease using FDG-PET imaging,’ *Stat in Med*, vol. 23, no. 2, pp. 315–326, 2004.

[12] S. E. Jones, B. R. Buchbinder, and I. Aharon, ’Three-dimensional mapping of cortical thickness using Laplace’s Equation,’ *Hum Brain Mapp*, vol. 11, no. 1, pp. 12–32, 2000.

[13] JT. Lyman, ’Complication probability as assessed from dose-volume histograms,’ *Rad Res Suppl*, vol. 104, pp. S13–S19, 1985.

[14] K. R. Rocha, A. J. Yezzi, and J. L. Prince, ’A hybrid Eulerian-Lagrangian approach for thickness, correspondence, and gridding of annular tissues,’ *Im Process IEEE Trans On*, vol. 16, no. 3, pp. 636–648, 2007.

[15] M. Sohn, M. Alber, and Di. Yan, ’Principal component analysis-based pattern analysis of dose-volume histograms and influence on rectal toxicity.’ *Int J of Rad Onc Biol Phys*, vol. 69, no 1, p. 230–239, 2007.

[16] M. H. Stenmark, A. S. C. Conlon *et al.*, ’Dose to the inferior rectum is strongly associated with patient reported bowel quality of life after radiation therapy for prostate cancer’, *Rad and Onc*, vol. 110, no. 2, pp. 291–297, 2014.