Prediction of high-resolution portal images for treatment verification in radiotherapy by coupling Monte Carlo simulations to non-parametric Bayesian denoising

Delphine Lazaro-Ponthus, Eric Barat, Cindy Le Loirec, Thomas Dautremer, Thierry Montagu, D. Patin, Guérin L., A. Batalla

PREDICTION OF HIGH-RESOLUTION PORTAL IMAGES FOR TREATMENT VERIFICATION IN RADIOTHERAPY BY COUPLING MONTE CARLO SIMULATIONS TO NON PARAMETRIC BAYESIAN DENOISING

D. Lazaro-Ponthus[1,*], E. Barat[1], C. Le Loirec[1], T. Dautremer[1], T. Montagu[1], D. Patin[1], L. Guérin[2], A. Batalla[2]
[1]CEA LIST, F-91191 Gif-sur-Yvette, France; [2]Centre F. Baclesse, F-14076 Caen Cedex, France.

I. INTRODUCTION

In order to ensure safety and efficiency in the delivery of Intensity Modulated Radiotherapy (IMRT) treatments, amorphous silicon (a-Si) electronic portal imaging devices (EPIDs) are now routinely used for dosimetric verifications. A straightforward way to do this is to compare the measured EPID image with a reference image which can be calculated in the treatment planning system (TPS). Among the various prediction models proposed to compute the reference image, Monte Carlo (MC) simulation is highly attractive due to its ability to predict accurately and directly the dose to the detector in a wide range of configurations, without requiring the conversion to dose in water [1]. However, MC remains to date so time consuming that MC computation of portal images with meaningful statistical uncertainty is only feasible for pixel sizes around 2 mm, which could lead to misinterpretations due to the loss of image resolution. To overcome this problem, we developed a new method to compute high resolution reference EPID images in reasonable computing times. This method is based on the denoising of MC calculated images with a non parametric Bayesian algorithm called DPGLM (for Dirichlet Process Generalized Linear Model) [2], particularly suited to very noisy images. In this study, an accurate model of an a-Si EPID was first developed and validated. Then, the performances of the denoising algorithm were assessed and compared to those obtained with IRON (Iterative Reduction Of Noise [3]), a denoising algorithm usually employed in radiotherapy. Finally, MC images of a head-and-neck treatment plan were computed and denoised, and then compared to acquired EPID images.

II. MATERIALS AND METHODS

II.1. EPID model development and validation

Experiments were carried out with a Siemens Optivue1000 EPID mounted on a Siemens ARTISTE linear accelerator (linac). The Optivue1000 is an a-Si flat panel device of 1024 × 1024 pixels of 0.39 × 0.39 mm² each, representing a 41 × 41 cm² active detection area. The PENELOPE MC code was used both to model the linac [4] (including a full description of the Siemens 160 leaf MLC) and the EPID. The EPID model consists of 13 layers described in terms of geometry and materials according to manufacturer's information. To mimic backscattering coming from structures surrounding the EPID, uniform water-equivalent slabs of varying thicknesses (from 1 to 70 mm) were added below this model and corresponding images were simulated for 10 × 10 and 20 × 20 cm² fields. By comparing profiles drawn in the inline and crossline directions on simulated and acquired images, a non-uniform map of water-equivalent slabs can be deduced.

The final model was validated against experimental data for two configurations. First, portal images without phantom in the beam were acquired for different jaw defined field sizes (5 × 5, 10 × 10, 15 × 15, 20 × 20 and 25 × 25 cm²) with the EPID positioned at a 100 cm source to detector distance (SDD). Second, a layered heterogeneous phantom (cf Figure 1) was placed in the beam, its entrance face located at 67.8 cm from the source. This phantom is made of two 30 × 30 × 5 cm^3 slabs of water equivalent material (1.04 g/cm^3), a 30 × 30 × 8 cm^3 slab of CIRS bone equivalent material (1.8 g/cm^3) and two 30 × 16 × 8 cm^3 slabs of CIRS lung equivalent material (0.3 g/cm^3), separated by a 3 cm air gap. For the model validation step, portal images were computed on a 256 × 256 pixel grid (pixel size: 1.6 mm) to decrease the simulation run time. Acquired and simulated images were then compared using a 2 % / 2 mm gamma-index, after normalization of the simulated images with respect to acquired images.

II.2. Denoising of portal images

II.2.1. IRON denoising method

The IRON denoising method relies on the minimization of a criterion combining two terms: one accounting for the data adjustment and the other one encouraging low curvature. But since the curvature penalty in IRON is non-convex, a global minimum solution is not guaranteed. Another difficulty in the IRON method lies in the roughness of the non differentiable penalty. Minimization routines like conjugate gradient or quasi-Newton methods are known to be non optimal for such non smooth functions.

With a pixel size of about 2 mm, these algorithmic difficulties tend to be mitigated since the MC calculated dose images can exhibit a convenient signal-to-noise ratio (SNR). In this situation, the initial point of the optimization, which is taken to be the MC data, is not "so far" from the desired solution. But this is not the case when one wants to respect the EPID's physical pixel size (0.39 mm) in the MC simulations. We are faced here to a much noisier environment for reasonable computation times. Initialization through raw MC data can thus

*Corresponding author: delphine.lazaro@cea.fr

reveal the ill-behavior of the minimization routine and extremely slow convergence to a local minimum. This initialization point's dependency may appear troublesome and this work aims at proposing a method relaxing this constraint.

II.2.2. Principles of the DPGLM denoising method

In the statistical interpretation of the IRON criterion, the curvature penalty can also be seen as a kind of *prior* term characterizing our degree of belief in a smooth dose deposit. This Bayesian rephrasing of the denoising problem forms in that way the framework of the proposed approach. A key point of Bayesian methods is that they give access to the estimation's uncertainty. Namely, we seek for the whole set of solutions, expressed by their *posterior* distribution, instead of looking for just a particular one. We retain for the dose estimate the posterior mean – which minimizes the L2 risk –. As a side-effect, our method is able to propagate the whole information present in the MC data.

Another characteristic of our approach is its nonparametric feature. Since the number of variables is huge in the EPID's MC data denoising problem (n=1024 × 1024 pixels), it turns out that it is convenient to consider the problem as the estimation of a continuous surface in R^2 which amounts to infer over a potentially infinite number of parameters, leading to a so-called *Bayesian nonparametric regression* approach. All statistical material cannot be expressed here and readers may refer to [2, 5] to gain an insight into involved methodologies. We model the n computed EPID's data (x_i,y_i) for i=1,…,n, where $x_i \in R^2$ stands for the pixel coordinates and y_i for the pixel's calculated dose. The method lies in estimating $f(x,y)$, the join distribution of (x,y), from simulated points (x_i,y_i) in a nonparametric way and to take for the denoised dose $d(x)$ for all $x \in R^2$:

$$d\,(x) = E\,(y|x) = \int_R y\,.\,f(y|x)\,dy = \frac{\int_R y\,.f(x,y)\;dy}{\int_R f(x,y)\;dy} \tag{1}$$

Nonparametrics arise from the choice of a Dirichlet Process Mixture (DPM) for prior specification of the joint density $f(x,y)$. Roughly, the DPM structure involves an open-ended number of components which relies only on the dataset and DPM parameters.

From the elicited prior and data (x_i,y_i), we want to compute the posterior distribution $f(x,y|x_1,y_1,\dots,x_n,y_n)$ and conditional expectation $\hat{d}(x) = E(y|x,x_1,y_1,\dots,x_n,y_n)$. The exact computation of the posterior is intractable and we use a Markov Chain Monte-Carlo (MCMC) approximation scheme (Gibbs sampler) to draw samples from the target distribution.

At each iteration (t) of the MCMC procedure, we are thus able to sample a denoised dose surface $d(x)^{(t)}$. For T samples, the posterior distribution is given by the set of $d(x)^{(t)}$ for t = 1,…,T, and the dose estimate (posterior mean) is expressed as:

$$\hat{d}\,(x) \approx \frac{1}{T}\sum_{t=1}^{T} d\,(x)^{(t)} \tag{2}$$

We can as well compute the posterior standard deviation or credible intervals from the collection $\{d(x)^{(t)}\}$.

All parameters of the DPM prior distribution are also sampled at each iteration, assuming an additional degree of hierarchy in the dose data model and putting vague priors on these parameters.

II.2.3. Denoising test case

Denoising effectiveness of the DPGLM algorithm was assessed on 1024 × 1024 images simulated for the heterogeneous phantom irradiated by a 15 ×15 cm² field. MC calculations of phase space files (PSF) storing 50, 100, 500, 1000, 1700, 3000 and 5000 million photons were performed with associated statistical uncertainties of better than 15, 10, 5, 3.5, 2, 1 and 0.7 % of the maximum dose. Portal images were then simulated using these PSF and recycling particles with a splitting factor of 10, and they were then denoised with the DPGLM and IRON algorithms. To assess performances of both algorithms, we calculated in a 600 × 600 pixel central area of the image the fraction of pixels presenting a difference of more than 1 % of the maximum normalized dose, with respect to the reference image. Due to computational time limitations linked to EPIDs MC calculations, the choice of a common reference image remains a tricky issue since the image with the best statistical uncertainty (0.7 %) is still too noisy to be taken as the reference. In order to avoid any bias in the comparison, we then resorted to use a reference for each denoising method, namely the image with 0.7 % statistical uncertainty denoised with the algorithm under test. The reference for raw MC image evaluation is the 0.7 % statistical uncertainty MC image itself.

II.2.4. Application to a head-and-neck treatment plan

The portal image associated to one of the beams used in a head-and-neck IMRT treatment plan was simulated with an associated statistical uncertainty of better than 5 % and was denoised with DPGLM and IRON.

III. RESULTS

III.1. EPID model development and validation

The model for the Optivue1000 EPID that best matches experimental data includes two kinds of non-uniform layers of water-equivalent material beneath the 13 layers model: a 50 mm water layer of 16 × 33 cm², centered in (x=0 cm, y=-4 cm) (the Y axis refers to the inline direction) and anywhere else a 30 mm water layer. Profiles drawn through acquired and simulated images in the crossline direction are shown on Figure 2, for portal images without and with phantom in the beam, respectively. 2D gamma-index values are also given in Table I. These results demonstrate the need to include in the model a correction for backscatter to accurately predict portal images in any configuration, especially for large field sizes.

Table I. Comparison of gamma-index values obtained for the different configurations.

	Images without phantom		Images with phantom
Field size (cm²)	Without backscatter correction	With backscatter correction	With backscatter correction
5 x 5	99.4 %	99.4 %	99.4 %
10 x 10	95.0 %	98.4 %	98.1 %
15 x 15	–	97.4 %	97.5 %
20 x 20	57.2 %	95.4 %	96.1 %
25 x 25	24.6 %	93.1 %	–

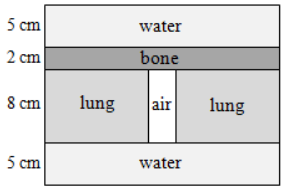

Figure 1. Description of the heterogeneous phantom.

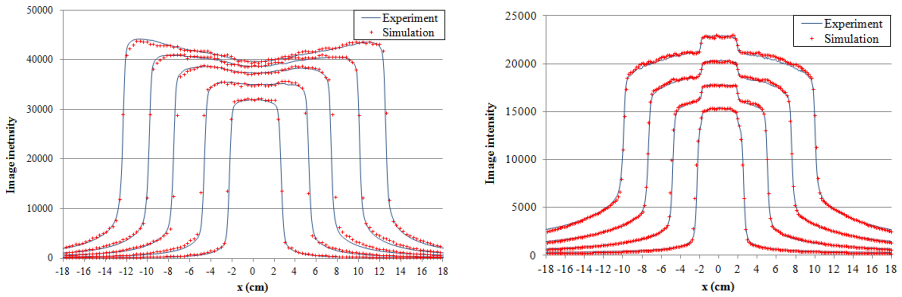

Figure 2. Profiles for acquired (blue) and simulated (red) portal images in the crossline direction, for the configuration without phantom (left) and with phantom (right).

III.2. Denoising of portal images

III.2.1. Denoising test case

As expected, we experienced slow convergence for the IRON algorithm for low SNR images. Note that, due to the large amount of variables, we resorted to use a limited-memory Broyden, Fletcher, Goldfarb, Shanno algorithm (LM-BFGS) [6]. Despite the needs of significant computing requirements, the structure of DPGLM algorithm makes it suitable for parallelization contrary to the LM-BFGS optimization algorithm. As a consequence, the effective computation times are similar for both methods.

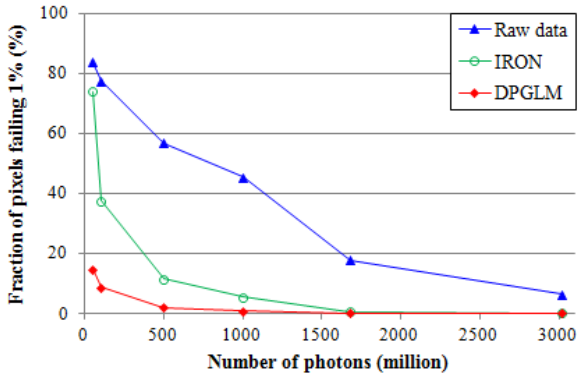

Figure 3. Fraction of pixels failing the 1 % difference test for MC images (raw data), images denoised with IRON and with DPGLM.

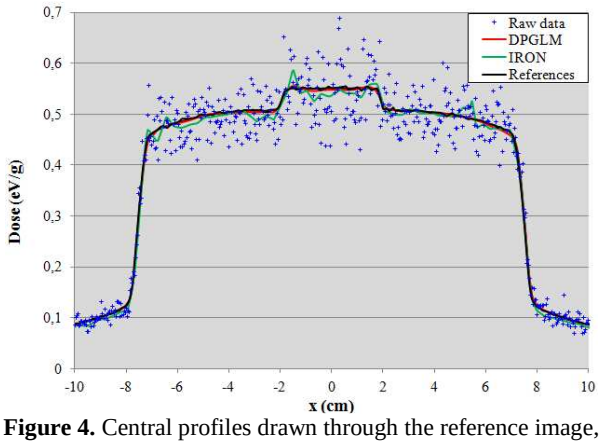

Figure 4. Central profiles drawn through the reference image, the MC image, the image denoised with IRON and the image denoised with DPGLM, for a 100 million photons PSF.

Figure 3 summarizes the fraction of pixels which fail the 1 % criterion. In all cases, the interest of using any of the denoising algorithms is evident, even for MC images with a high SNR. IRON and DPGLM exhibit similar

performances for images with a statistical uncertainty better than 2 %. At lower SNR, the statistical basis of DPGLM offers more robustness with respect to noise. This allows maintaining below 2 % the fraction of pixels failing the chosen dose criterion for images with a statistical uncertainty lower than 5 %. Obtaining the same image quality with IRON would require a PSF about three times larger. We also observe on the profiles shown in Figure 4 that DPGLM produces smoother images than IRON while preserving edges in high-gradient dose regions. These results demonstrate that it is possible to reach image quality compatible with clinical interpretation for PSF storing between 100 (10 % statistical uncertainty) and 500 million photons (5 % statistical uncertainty) with DPGLM. For instance, the computation of the simulated image on 100 processors (2.26 GHz) of our Linux cluster lasts in half an hour when running 100 million photons from the PSF and 2,5 hours when running 500 million photons. In the same configurations, DPGLM denoising on 1024 × 1024 images necessitates 1,5 hour. The complete computation of the portal image takes 2 hours and 4 hours for 100 and 500 million photons, respectively. Shorter MC simulation times were observed for smaller field sizes.

III.2.2. Head-and-neck treatment plan

The acquired image and its reference image calculated by combining MC simulation and DPGLM denoising are shown on Figure 5. Profiles drawn along the white line on the acquired image, the undenoised MC image, the MC image denoised with IRON and DPGLM are compared on Figure 6. Here again, DPGLM produces smoother images than IRON and allows direct comparison with the acquired image at the same resolution.

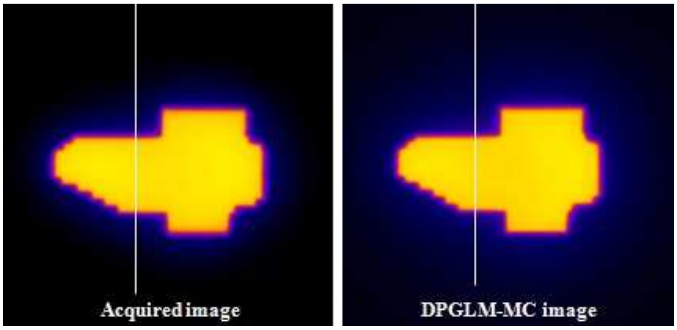

Figure 5. Acquired EPID image (left) and associated simulated image after denoising with DPGLM (right), for the head-and-neck treatment beam.

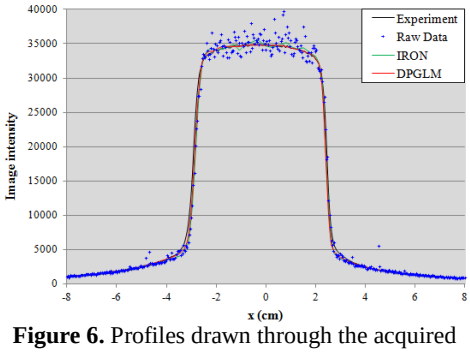

Figure 6. Profiles drawn through the acquired image, the MC image undenoised, the MC image denoised with IRON and with DPGLM.

IV. Discussion and Conclusion

This study shows that the combination of MC simulations with efficient denoising methods enables the accurate computation of high-resolution portal images for computational burden now compatible with clinical settings and acceptable for TPS. Particularly, in a context of low SNR, DPGLM reveals interesting performances. In addition, some features of the proposed method have not been fully investigated. Among them, based on a nonparametric regression approach, DPGLM is able to interpolate the dose deposit at any coordinates of the portal image's plane. This could offer flexibility in the choice of final image resolution. Another key feature of the approach lies in the uncertainty estimation over the whole image. This could bring enhanced information which can help in practical situations to determine credible intervals containing the desired dose image.

V. References

[1] R S Cufflin, E Spezi, A E Millin and D G Lewis, An investigation of the accuracy of Monte Carlo portal dosimetry for verification of IMRT with extended fields, Phys. Med. Biol. **55** (2010): 4589–4600.

[2] L. Hannah, D. M. Blei, and W. B. Powell, Dirichlet process mixtures of generalized linear models, Journal of Machine Learning Research **12** (2011):1923–1953.

[3] M Fippel and F Nüsslin, Smoothing Monte Carlo calculated dose distributions by iterative reduction of noise, Phys. Med. Biol. **48** (2003): 1289–1304.

[4] D. Lazaro-Ponthus, L. Guérin, A. Batalla, T. Frisson and D. Sarrut, Commissioning of PENELOPE and GATE Monte Carlo models for 6 and 18 MV photon beams from the Siemens Artiste linac, 11th Biennal ESTR0, London UK, May 2011.

[5] N. L. Hjort, C. Holmes, P. Müller, S. G. Walker, S. Ghosal, A. Lijoi, I. Prünster, Y. W. Teh, M. I. Jordan, J. Griffin, D. B. Dunson, and F. Quintana, Bayesian Nonparametrics (2010), Cambridge Series in Statistical and Probabilistic Mathematics.

[6] D. C. Liu and J. Nocedal, On the Limited Memory Method for Large Scale Optimization, Mathematical Programming B **45** (3) (1989): 503–528.