Denoising techniques combined to Monte Carlo simulations for the prediction of high-resolution portal images in radiotherapy treatment verification

Delphine Lazaro, Eric Barat, Cindy Le Loirec, Thomas Dautremer, Thierry Montagu, L. Guerin, A. Batalla

Denoising techniques combined to Monte Carlo simulations for the prediction of high-resolution portal images in radiotherapy treatment verification

D Lazaro[1], E Barat[1], C Le Loirec[1], T Dautremer[1], T Montagu[1], L Guérin[2] and A Batalla[2]

[1]CEA, LIST, F-91191 Gif-sur-Yvette, France

[2]Service de Physique Médicale, Centre François Baclesse, F-14076 Caen Cedex, France

E-mail address: delphine.lazaro@cea.fr

Abstract

This work investigates the possibility of combining Monte Carlo (MC) simulations to a denoising algorithm for the accurate prediction of images acquired using amorphous silicon (a-Si) electronic portal imaging devices (EPIDs). An accurate MC model of the Siemens OptiVue1000 EPID was first developed using the PENELOPE code, integrating a non-uniform backscatter modelling. Two already existing denoising algorithms were then applied on simulated portal images, namely the iterative reduction of noise (IRON) method and the locally adaptive Savitzky-Golay (LASG) method. A third denoising method, based on a nonparametric Bayesian framework and called DPGLM (for Dirichlet Process Generalized Linear Model) was also developed. Performances of the IRON, LASG and DPGLM methods, in terms of smoothing capabilities and computation time, were compared for portal images computed for different values of the RMS pixel noise (up to 10%) in three different configurations, a heterogeneous phantom irradiated by a non-conformal 15 × 15 cm² field, a conformal beam from a pelvis treatment plan, and an IMRT beam from a prostate treatment plan. For all configurations, DPGLM outperforms both IRON and LASG by providing better smoothing performances and demonstrating a better robustness with respect to noise. Additionally, no parameter tuning is required by DPGLM, which makes the denoising step very generic and easy to handle for any portal image. Concerning the computation time, the denoising of 1024 × 1024 images takes about 1 h 30, 2 h and 5 min using DPGLM, IRON, and LASG, respectively. This paper shows the feasibility to predict within a few hours and with the same resolution as real images accurate portal images, combining MC simulations with the DPGLM denoising algorithm.

PACS code: see http://www.aip.org/pacs: 87.10.Rt (Monte Carlo simulations), 87.55.Qr (Quality assurance in radiotherapy), 87.56.Fc (Quality assurance equipment), 02.50.-r (Probability theory, stochastic processes, and statistics), 02.60.Ed (Interpolation; curve fitting).

Submitted to: Physics in Medicine and Biology

1. Introduction

Over the past several years, the need for patient-specific quality assurance (QA) has become increasingly important, driven by the complexity of current delivery techniques such as intensity modulated radiation therapy (IMRT) and intensity modulated arc therapy (IMAT). In this context, electronic portal imaging devices (EPIDs) have proved to be valuable tools not only for patient positioning control, but also for dosimetric evaluation of complex treatment plans in IMRT and more recently in IMAT (McCurdy and Greer 2009, Iori *et al* 2010). Many dosimetric verification approaches using EPIDs were investigated until now and were presented in detail in a recent comprehensive review (van Elmpt *et al* 2008). They can be broadly classified into two groups: one is based on the comparison of a predicted fluence or dose at the detector with a portal image acquired before or during the treatment (McCurdy *et al* 2001, van Esch *et al* 2004, Spezi and Lewis 2002, van Elmpt *et al* 2005), while the other relies on the estimation of the dose delivered to the patient from the portal image, using either a backprojection algorithm (Wendling *et al* 2006) or fluence profiles extracted from the acquired portal images (Steciw *et al* 2005).

Prediction accuracy is essential to guarantee highly accurate QA using EPIDs. Indeed, the high resolution of current EPIDs (pixel size down to 0.39 mm) allows recording variations in the energy fluence due to MLC specific effects (inter- and intraleaf leakage, rounded leaf end transmission, tongue-and-groove). The portal image prediction model should hence include all these effects so that the discrepancies between measured and predicted images could be correctly interpreted. Two methods were proposed for the portal dose image prediction; one is based on full forward Monte Carlo (MC) simulation of the image formation process, which offers a high accuracy at the expense of a large computation time. The kernel convolution method was introduced as an alternative to reduce this computation time. In this method, the predicted image is calculated by convolving the incident fluence, either extracted from the treatment planning system (TPS) or calculated by MC simulations, with EPID dose response kernels. Two kinds of kernels were used: analytical or parameterized dose kernels (Van Esch *et al* 2004, Kirkby and Sloboda 2005, Wendling *et al* 2006, Van Elmpt *et al* 2006, Greer *et al* 2007, Vial *et al* 2008, Greer *et al* 2009), and pre-calculated MC-based dose kernels (McCurdy *et al* 2001, Warkentin *et al* 2003, Steciw *et al* 2005, Li *et al* 2006, Chytyk and McCurdy 2006, 2009, Wang *et al* 2009). As already pointed out (Vial *et al* 2008, Greer *et al* 2009), kernel convolution methods may reach a high level of precision, only if they combine both an accurate method to derive the energy fluence and an accurate modelling of the EPID dose response kernels. This latter point is particularly delicate because of the complexity of the EPID dose response. The phosphor scintillation layer contained in current EPIDs causes signal blurring, due to optical glare (Warkentin *et al* 2003) and scatter within the scintillation layer: this effect could be explicitly modelled using an additional kernel or simply by a uniform material layer beneath the imager adding to the model. The high atomic number of the scintillation layer is also responsible for an oversensitivity to low energy radiation. As a consequence, EPIDs exhibit an energy-spectra dependent response (Schach von Wittenau 2002, Kirkby and Sloboda 2005) and are very sensitive to all potential causes of beam energy variations, such as field size and beam hardening in the linac head (Sheikh-Bagheri and Rogers 2002) caused by the beam modifiers such as the MLC (Kim *et al* 2001) and by the patient or the phantom placed in the beam (Jarry and Verhaegen 2007). Backscattered radiation arising from the support arm of the EPID contributes also significantly to the signal and should be explicitly modelled in the kernel. As shown by some authors (Greer *et al* 2009, Ko *et al* 2004, Moore and Siebers 2005, Siebers *et al* 2004, Wang *et al* 2009), this backscatter contribution can be non-uniform and has to be taken into account for extended field modelling (Cufflin *et al* 2010). Even though many improvements were included in the generation of dose kernels these past years to take all these effects into account, kernels are usually generated for a limited number of irradiation configurations and are thus not specific. In the work of Wang *et al* (2009), there was an attempt to generate adaptive MC-based EPID kernels in a more flexible way but they are, for instance, not adapted to the prediction of extended field as they did not model a non-uniform backscatter.

Although computationally intensive, MC simulations offer today the most appropriate and accurate technique to predict reliably both energy fluence maps and dose to the detector, overcoming the limitations of convolution methods in modelling all different irradiation configurations. Several works were devoted to full forward MC prediction of portal images (Spezi and Lewis 2002, Chin *et al* 2003, Siebers *et al* 2004, Jarry and Verhaegen 2005, Parent *et al* 2006, Cufflin *et al* 2010) using the

BEAM/DOSXYZ MC codes. When considering the real pixel size of portal images acquired by current EPIDs, i.e about 0.4 mm, the calculation of a 0.5% Root Mean Square (RMS) pixel noise image would require a computation time estimated to be about one month on 96 CPUs and a phase-space file (PSF) storing about 4×10^{10} photons, which means a storage space of about 3.5 To. In order to balance the trade-off between acceptable computation times and production of portal images with meaningful RMS pixel noise, the pixel size in the MC simulated image was chosen around 2 mm most of the time. As a direct consequence, the predicted image is more blurred than the measured one and effects of rounded leaf ends, tongue-and-grooves, and interleaf transmission could be washed, hence potentially leading to problems in the interpretation of errors.

In this study, we propose to explore a new methodology to predict portal images, based on the denoising of MC pre-calculated images. The aim is to compute reference EPID images at the same resolution as the measured images, while keeping computing times at a reasonable level. The objectives of this work are hence twofold: (1) to develop and commission a MC model for the Siemens Optivue1000 EPID using the PENELOPE code, and (2) to develop a new portal image prediction method combining MC simulations to an adequate 2D denoising technique. To this end, we tested smoothing techniques usually employed in radiotherapy for 3D dose distribution denoising, namely the IRON method (Iterative Reduction Of Noise, Fippel and Nüsslin 2003) and the locally adaptive Savitky-Golay filtering or LASG (Kawrakow 2002), and we assessed their performances in denoising MC predicted images. A new 2D denoising technique, based on non parametric Bayesian framework and called DPGLM (for Dirichlet Process Generalized Linear Model) (Hannah *et al* 2011), was also developed and was demonstrated to be particularly suited to very noisy images. Denoising performances of DPGLM were compared to those of IRON and LASG in terms of accuracy and computation time.

2. Materials and methods

2.1. Materials

Experiments were carried out on a Siemens ARTISTE™ linear accelerator (linac) (Siemens Medical Solutions, Erlangen, Germany), operating at 6 MV and equipped with a 160-MLC™. The 160-MLC features a single focused design that consists of 160 leaves, mounted on two leaf banks. They are arranged in an alternative pattern of upper and lower leaves, so that the source-to-collimator distance is 460 mm and 452 mm for lower and upper leaves, respectively. Each leaf is 95 mm thick, with a projected leaf width of 5 mm at the machine isocenter, and has a rounded edge in combination with an S-shaped middle-part, whereas its sides are completely flat, without any tongue or groove. A new slanted leaf design replaced the tongue and groove system to allow complete interdigitation. More details on the design of the 160-MLC can be found elsewhere (Tacke *et al* 2008, Klüter *et al* 2011, Prah *et al* 2011).

A Siemens OptiVue™ 1000 EPID (Siemens Medical Solutions) is mounted on the linac gantry: it is an amorphous silicon (*a*-Si) flat panel device with an active detection area of 41×41 cm², and a matrix of 1024×1024 pixels, each pixel having a square pitch of 0.39 mm. It was operated by the software coherence Therapist Workspace (Siemens Medical Solutions). The EPID was positioned with the surface of the detector at 100 cm from the source. No additional buildup was used on the EPID since this is the standard acquisition configuration for nonzero gantry angle measurements.

Calibration curves allowing the conversion of MC dose to an acquired portal image intensity were obtained for a 10×10 cm² field at isocenter, by taking portal images of solid water slabs of different thicknesses, with machine dose rate settings of 100 MU min^{-1}.

2.2. Monte Carlo simulations

2.2.1. Linac head and MLC modelling. MC simulations of the 6 MV photon beam of the ARTISTE linac were performed using the 2006 release of the PENELOPE code (Salvat *et al* 2006). A new version of the main program PENMAIN was developed in our institute, in which several conventional variance reduction techniques (selective bremsstrahlung splitting, Russian roulette combined with angular splitting, and rotational splitting) were implemented. This new version is also parallelized based on the standard MPI message passing interface and can be run on a cluster consisting of seven machines, each with $12 \times 2{,}26$ GHz processors, and twelve machines, each with $24 \times 2{,}26$ GHz processors, hence a total of 372 processors. In this work, only the selective

bremsstrahlung splitting and the Russian roulette combined with angular splitting techniques were activated for the linac part. The geometry of the linac head was described in detail, based on information supplied by Siemens. The energy-loss cutoff values W_{cc} and W_{cr} were set equal to 500 keV and 10 keV, respectively, and the multiple-scattering parameters C_1 and C_2 to 0.05, everywhere. Commissioning of the linac model was performed by comparing measured and simulated percentage depth doses (PDD) and lateral dose profiles at d_{max}, 5 and 10 cm for open square fields from 5 × 5 to 30 × 30 cm² at a source-to-surface distance of 100 cm (Lazaro-Ponthus *et al* 2011). The simulations reproduced the measurements within 1% for the PDDs and within 2% / 2 mm for the lateral dose profiles.

The modelling of the MLC is an important issue and should be conducted carefully, since high resolution EPID images will be simulated. The 160-MLC geometry (leaf shape and arrangement) was fully described following the manufacturer's specifications and using the PENGEOM geometry routine distributed with PENELOPE (Le Loirec *et al* 2012). An in-house calibration method based on the work of Boyer and Li (1997) was developed to predict the leaf positions for all the MLC leaves from the information reported in the DICOM file (LeafJawPosition and LeafJawBoundaries). A script was then written to convert automatically leaf information stored in the DICOM file into the geometry file required by the PENELOPE code. The movable skin method developed by Brualla *et al* (2009) was also implemented to speed up calculations within the jaws and the MLC. This method can be seen as a range-rejection technique by means of the geometry definition and presents the advantage of keeping unchanged the physics, thus preserving the accuracy provided by MC simulations. In the MLC structure, the following parameters were set: in the non-skin region, $C_1 = C_2 = 0.1$, $W_{cc} = 200$ keV, $W_{cr} = 50$ keV, the energy absorption for electrons and positrons was set to 200 keV, and the energy absorption for photons was set to 50 keV; in the skin region, the energy absorption of electrons and positrons was increased to 1 MeV, while keeping the other parameters unchanged.

In simulations dedicated to the commissioning of the EPID model, the PSF used to store the information about each particle exiting the linac head was recorded in a plane located 40 cm downstream from the target, just below the jaws. A number of 5×10^8 histories were simulated: this represents about 100 million and 1000 million photons stored in the PSFs, necessary to achieve uncertainties of less than 2% in computing portal images of the 5 × 5 cm² and 20 × 20 cm² fields, respectively.

2.2.2.OptiVue1000™ EPID modelling. The EPID was also modelled using the PENELOPE code. The model consists of a sequence of thirteen layers described in terms of geometry and materials according to information provided by Siemens. It includes the main components of the EPID, such the 1 mm-thick copper plate and the phosphor scintillation screen (Lanex Fast Back screen, Eastman Kodak Comp., Rochester, NY), as well as the printed circuit-board materials and the rear housing. As suggested by Siebers *et al* (2004), the lateral dimensions of the EPID were set larger in the model than the current dimensions (51 × 51 cm² instead of 41 × 41 cm²) to approximate the materials surrounding the cassette. Portal images were computed by scoring the energy deposited in the phosphor layer on a virtual grid.

To mimic backscattered radiation arising from the rear housing of the detector and from structures surrounding the EPID (gantry), and to model as close as possible this non-uniform backscattering, an effective representation of the backscattering components was introduced in the model following the method proposed by Cufflin *et al* (2010). To this end, uniform water-equivalent slabs of varying thicknesses, ranging from 1 to 70 mm, were added downstream from the initial thirteen-layer 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 optical glare effect was not explicitly modelled here but is included in the additional backscatter layer.

2.2.3. Commissioning of the EPID model

The EPID model was validated against experimental data for two irradiation configurations. First, portal images without any 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, described in 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 × 2 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.

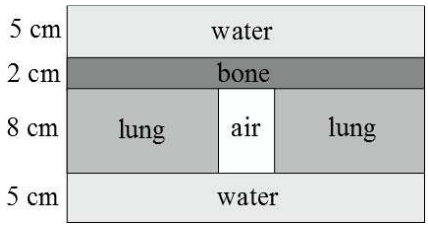

Figure 1. Description of the heterogeneous phantom used in the commissioning of the EPID model.

For this validation step, the denoising techniques were not already available; hence, portal images were computed on a 256 × 256 pixel virtual grid (pixel size: 1.6 mm) to decrease the simulation runtime. For both configurations (in-air and with the phantom in the beam), acquired and simulated images were compared using a 2% / 2 mm gamma-index. For the in-air configuration only, portal images were also computed without including the backscatter correction into the model, and gamma-index values obtained with both EPID models were compared.

2.3. Description of EPID image denoising techniques and implementation

In order to denoise simulated portal images, three denoising techniques were implemented: IRON, LASG and DPGLM. The principles of the IRON and LASG methods are first quickly reminded; the principles of the newly developed technique DPGLM, and the rationale for its development, are also described in detail below.

2.3.1. Iterative Reduction Of Noise (IRON). The IRON denoising method proposed by Fippel and Nüsslin (2003) 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.

Initially designed to perform on 3D dose distributions in radiotherapy, IRON is facing new challenges in the context of portal image denoising. With a pixel size of about 2 mm, the algorithmic difficulties mentioned above 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 “too far” from the desired solution. But this is no longer the case when one wants to maintain the EPID physical pixel size (0.39 mm) in the MC simulations. The IRON technique faced here to a much noisier environment in order to reach reasonable MC computation times. Initialization through raw MC data can thus reveal the ill-behavior of the minimization routine and extremely slow convergence to a local minimum. This initialization point dependency may appear troublesome and this is one of our motivations in proposing a new method which allows relaxing this constraint.

Note that, due to the large amount of variables ($N = 1024 \times 1024 = 10485764$), we resorted to use a limited-memory Broyden, Fletcher, Goldfarb, Shanno algorithm (LM-BFGS) (Liu and Nocedal 1989). The purpose of LM-BFGS is to avoid in computations the approximation of the inverse of the Hessian matrix whose storage size is $N \times N$. Instead, this algorithm maintains only the history of the m most recent updates (points and gradients) which are used in place of explicit matrix-vector products involving the inverse of the Hessian. For EPID images denoising, we fixed m = 100. We also notice that the best performances of the algorithm were reached when tuning the trade-off parameter of IRON original implementation $w = 0.05$.

2.3.2. Locally Adaptative Savitzky-Golay filtering (LASG). The LASG algorithm has gained popularity in 3D dose denoising since its introduction by Kawrakow (2002). The original Savitzky-Golay smoothing algorithm (Savitzky and Golay 1964) relies on a polynomial regression over a neighbourhood around each data point and retains the smoothed value at the central point. Savitzky and Golay showed that, once the window has been defined, the smoothing algorithm amounts to a linear filter. The modification proposed by Kawrakow allows choosing a window size in order to optimize the trade-off between smoothing capacities and bias reduction. The algorithm starts with an initial maximum window size which is reduced, following a specific strategy, until the satisfaction of a $\mathcal{X}^2$- goodness of fit test. This test is based on the assumption that the noise follows a Gaussian distribution entirely characterized by the standard deviation stored in the simulated dose data file. Performances of this approach for 3D MC dose denoising were reported in El Naqa *et al* (2005) and contrasted with alternative methods. When applying LASG to EPID simulated data, detector characteristics have to be considered. First, the small pixel size and the low detection efficiency induce a non-gaussian noise distribution which makes suboptimal the underlying model of the goodness of the fit test. Second, the MC estimated standard deviation may become itself highly noisy. These concerns highlight the particular attention to be paid to the tuning of the $\mathcal{X}^2_{max}$ parameter.

2.3.3 Dirichlet Process Generalized Linear Model (DPGLM). 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 the framework of the proposed approach. A key point of the Bayesian methods is that they give access to the estimation uncertainty. Namely, we seek for the whole set of solutions, expressed by their *posterior* distribution, instead of looking for a particular one only. We retain for the dose estimate the posterior mean – which minimizes the L2 risk. As a side-effect, DPGLM 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 very large in the EPID MC data denoising problem ($n = 1024 \times 1024$ pixels), it turns out that it is convenient to consider the problem as the estimation of a continuous surface in $\mathbb{R}^2$ 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 are invited to refer to other references to get an insight into involved methodologies (Müller *et al* 1996, Hjort *et al* 2010, Hannah *et al* 2011). In this framework the n computed EPID data (x_{1i}, x_{2i}, y_i) for $i = 1, \dots, n$, are modelled, where (x_{1i}, x_{2i}) stands for the pixel coordinates and y_i for the pixel calculated dose. The method lies in the estimation of $f(x_1, x_2, y)$, the joint distribution of (x_1, x_2, y), from simulated points (x_{1i}, x_{2i}, y_i) in a nonparametric way, and to take for the denoised dose $d(x_1, x_2)$ for all $(x_1, x_2) \in \mathbb{R}^2$:

$$d\,(x_1, x_2) = E\,(y|x_1, x_2) = \int_{\mathbb{R}} y\,.\,f(y|x_1, x_2)\,dy = \frac{\int_{\mathbb{R}} y\,.\,f(x_1, x_2, y)\;dy}{\int_{\mathbb{R}} f(x_1, x_2, y)\;dy} \tag{1}$$

Nonparametrics arise from the choice of a Dirichlet Process Mixture (DPM) for prior specification of the joint density $f(x_1, x_2, y)$. This model is based on the random distribution G,

$$G \sim \mathrm{DP}(\alpha, G_0)$$

where the symbol "$\sim$" means "is distributed as" and $\mathrm{DP}(\alpha, G_0)$ stands for the Dirichlet Process distribution (Hjort *et al* 2010) with mean measure G_0 and concentration parameter α. This distribution over random probability distributions plays a central role in the nonparametric modeling. An interpretation of DP by Sethuraman (1994) lets us express the random measure $G(\cdot)$ as

$$G(\cdot) = \sum_{k=1}^{\infty} w_k\,\delta_{\theta_k}(\cdot)$$

with $w_1 = V_1,\ w_k = V_k \prod_{l=1}^{k-1}(1 - V_l)$ such that $V_k \sim \mathrm{Beta}(1, \alpha)$ (see Appendix B for definition) and $\delta_u(\cdot)$ stands for the Dirac delta function located at u. Here, $\theta_k \sim G_0$ represents the parameters

associated to the kth component of G. Nonparametrics come from the (potentially) infinite sum in the DP prior for G. This leads to the following random density $f(x_1, x_2, y)$

$$f(x_1, x_2, y) = \sum_{k=1}^{\infty} w_k \, f_{\theta_k}(x_1, x_2, y)$$

where f_{θ_k} stands for k^{th} component of $f(x_1, x_2, y)$, parameterized by θ_k. We now make explicit f_{θ_k} for the application of DPGLM to EPIDs images denoising. We use the additional notations $\theta_k = \left(\mu_k, \tau_k, \beta_k, \sigma_k^2\right)$ and $\mu_k = (\mu_{1k}, \mu_{2k})$, $\tau_k = (\tau_{1k}, \tau_{2k})$,

$$f_{\theta_k}(x_1, x_2, y) = \mathcal{N}(x_1|\mu_{1k}, \tau_{1k}) \, \mathcal{N}(x_2|\mu_{2k}, \tau_{2k}) \, \mathcal{N}\left(y\middle|\beta_k' \cdot \tilde{X}_k(x_1, x_2), \sigma_k^2\right)$$

where $\mathcal{N}(\cdot\,|\mu, \tau)$ represents the normal (gaussian) distribution with mean μ and variance τ, $\tilde{X}_k(x_1, x_2)$ is a centered regressors vector and β_k the regression coefficient vector. Note that the regression model for the outcome y given covariates (x_1, x_2) may be chosen as linear (i.e. $\tilde{X}_k(x_1, x_2) = (1, x_1 - \mu_{1k}, x_2 - \mu_{2k})'$ in this case) or polynomial in the DPGLM approach (e.g. $\tilde{X}_k(x_1, x_2) = (1, x_1 - \mu_{1k}, x_2 - \mu_{2k}, (x_1 - \mu_{1k})(x_2 - \mu_{2k}), (x_1 - \mu_{1k})^2, (x_2 - \mu_{2k})^2)'$ for quadratic regression).

The prior measure $G_0$is taken as follows. Priors for means μ are normal distributions, priors for variances τ and σ^2 are inverse-gamma, and prior for regression coefficients β is multivariate normal. Using Bayes rule, we deduce from the expression of $f(x_1, x_2, y)$ that

$$f(y|x_1, x_2) = \sum_{k=1}^{\infty} \frac{w_k \, \mathcal{N}(x_1|\mu_{1k}, \tau_{1k}) \, \mathcal{N}(x_2|\mu_{2k}, \tau_{2k})}{\sum_{l=1}^{\infty} w_l \, \mathcal{N}(x_1|\mu_{1l}, \tau_{1l}) \, \mathcal{N}(x_2|\mu_{2l}, \tau_{2l})} \mathcal{N}\left(y\middle|\beta_k' \cdot \tilde{X}_k(x_1, x_2), \sigma_k^2\right)$$

and

$$E(y|x_1, x_2) = \sum_{k=1}^{\infty} \frac{w_k \, \mathcal{N}(x_1|\mu_{1k}, \tau_{1k}) \, \mathcal{N}(x_2|\mu_{2k}, \tau_{2k})}{\sum_{l=1}^{\infty} w_l \, \mathcal{N}(x_1|\mu_{1l}, \tau_{1l}) \, \mathcal{N}(x_2|\mu_{2l}, \tau_{2l})} \, \beta_k' \cdot \tilde{X}_k(x_1, x_2)$$

The latter expression reveals our prior in the Bayesian nonparametric approach for the smoothed EPID image.

Note that heteroscedasticity and non-gaussian behaviour are handled in the model of $f(y|x_1, x_2)$ thanks to the mixture of Gaussian distribution $\mathcal{N}\left(y\middle|\beta_k' \cdot \tilde{X}_k(x_1, x_2), \sigma_k^2\right)$ for observed dose y at any x. Another remark is that we may consider a multivariate normal prior for $x = (x_1, x_2) \sim \mathcal{N}(x|\mu, \Gamma)$ which avoids privileging EPID matrix directions. But, since jaws and MLC geometries share the same axes, this oriented prior reveals appealing.

From the elicited prior defined above and data (x_{1i}, x_{2i}, y_i), we aim at computing the posterior distribution $f(x_1, x_2, y|x_{11}, x_{21}, y_1, \dots, x_{1n}, x_{2n}, y_n)$ and conditional expectation $\hat{d}\,(x_1, x_2) = E(y|x_1, x_2, x_{11}, x_{21}, y_1, \dots, x_{1n}, x_{2n}, y_n)$. The exact computation of the posterior distribution is intractable and we use a Markov Chain Monte-Carlo (MCMC) approximation scheme (Gibbs sampler) to draw samples from the target distribution. In order to make inference feasible over infinite dimensional objects (infinite sum in DP), we follow a slice sampler approach from Kalli *et al*. (2011), where only a random finite number κ of components are involved per iteration. At each iteration (t) of the MCMC procedure, we are thus able to sample a denoised dose surface $d(x_1, x_2)^{(t)}$. For T samples, the posterior distribution is given by the set of $d(x_1, x_2)^{(t)}$ for $t = 1, \dots, T$, and the dose estimate (posterior mean) is expressed as:

$$\hat{d}\,(x_1, x_2) \approx \frac{1}{T} \sum_{t=1}^{T} d\,(x_1, x_2)^{(t)} \tag{2}$$

We can as well compute the posterior standard deviation or credible intervals from the collection $\left\{d(x_1, x_2)^{(t)}\right\}$.

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.

Algorithm details of the Gibbs sampler are given in Appendix A.

2.4. *Assessment of the denoising technique performances*

2.4.1. Description of the denoising test cases. Denoising effectiveness of the three algorithms presented above was assessed on the two following test cases.

In the first test case, portal images corresponding to the irradiation of the heterogeneous phantom with a squared 15 × 15 cm² field were computed using the commissioned EPID model. A virtual grid of 1024 × 1024 pixels, such as the physical detection matrix of the OptiVue1000 EPID, was considered in the simulation. Monte Carlo simulations of 50, 100, 500, 1000, 1700, 3000 and 5000 million photons were performed, resulting in computed portal images with associated statistical uncertainties on the maximum dose of 15, 10, 5, 3.5, 2, 1.8 and 1.6%. A splitting factor of 10 was used to recycle the particles stored in the PSFs for the portal image calculation.

Simulated portal images were then denoised with the IRON, LASG and DPGLM algorithms. Due to computation time limitations linked to EPID MC calculations, the choice of a common reference image remains a tricky issue since the image with the smallest RMS pixel noise (1.6%) 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 image for each denoising method, namely the image with 1.6% RMS pixel noise denoised with the algorithm under test. The reference for the raw MC image evaluation is the 1.6% RMS pixel noise MC image itself.

The goal of the second test case is to challenge the behavior of the denoising algorithms even further. For this, a 'picket fence' field with alternating leaves open and closed was set on the linac, resulting in many steep dose gradient regions. The portal image was acquired by irradiating directly the EPID located at 140 cm from the source. Two portal images were simulated using a virtual grid of 1024 × 1024 pixels: one with a high RMS pixel noise, referred as 'low SNR' image in the following (PSF storing about 4.5×10^7 photons and recycled ten times) and the other with a low RMS pixel noise, referred as 'high SNR' in the following (PSF storing about 3×10^8 photons and recycled ten times). Once the portal images simulated, they were denoised with the IRON, LASG and DPGLM algorithms, and compared to the acquired image.

2.4.2. Assessment criteria. For the first test case described above, the performances of each denoising algorithm were assessed not only in terms of accuracy but also in terms of computation time. Regarding to accuracy, two criteria were considered:

- the visual inspection of profiles drawn through portal images: this qualitative criterion is based on a subjective appreciation but it can help to appreciate the smoothness added by each denoising method,
- the 1% dose difference test proposed by Fippel and Nüsslin (2003), which is a modified and more severe version of the x% / y mm test (Van Dyke et al 1993). It consists here in calculating the fraction of pixels presenting a difference of more than 1% of the maximum normalized dose with respect to the reference image, in a 600 × 600 pixel central area of the image.

The efficiency of the method combining MC simulations and denoising was also assessed in terms of computation time. This issue is crucial because we should know if it is rather spending CPU time for smoothing or for simulating additional particle histories. In other words, this should help the user to determine the optimal trade-off between computation time and accuracy (or statistical uncertainty in the smoothed portal image). To this end, the 1% difference test was also expressed with respect to the total computation time necessary to the portal image calculation, equal to the time spent for the MC simulations plus the time spent for smoothing.

For the 'picket fence' test, profiles were drawn through the acquired portal image and the denoised simulated images, along the direction perpendicular to the leaf motion and on the central axis. They were then compared.

2.5. *Application to clinical treatment plans*

The developed methodology for portal image prediction was tested with two clinical examples that involve different types of fields.

A pelvis treatment plan using conformal beams was first considered. In this case, the irradiation field is large. The EPID was located at 100 cm from the source and it was irradiated directly. Only one of the beams was simulated with PENELOPE. The portal image associated to this beam was then

simulated: the image contained 1024 × 1024 pixels and the number of primary particles was chosen so that the image was simulated with a RMS pixel noise of 3%, 5.7% and 10% of the maximal dose, which represents in the three cases an acceptable SNR from the point of view of denoising. For statistical uncertainties of 3% and 5.7%, the PSF used stored about 2.3×10^9 photons, and it was recycled ten times for the first, and only once for the second. For a RMS pixel noise of 10%, the PSF used stored about 7.1×10^8 photons, and was played once. The quantitative evaluation of the performances of IRON, LASG and DPGLM was done by calculating the gamma index between the experimental image and the denoised image for all statistical uncertainties, with a 2% / 2.5 mm criteria.

A prostate IMRT clinical plan was chosen as a second case. This plan was optimized and calculated on Konrad (Siemens Medical Systems) and contained five fields equiangularly distributed, with step-and-shoot delivery of tiny segments. The EPID was located at 140 cm from the source and it was irradiated directly. Only one field (field 2) made of nine segments was simulated. For each segment, a PSF was computed so that the RMS pixel noise in the resulting simulated portal image was no more than 5%. Each PSF thus contains about 55 millions of photons recycled 20 times. Here again, portal images were simulated on a 1024 × 1024 virtual grid. All simulated portal images were denoised with IRON, LASG and DPGLM, for comparison. The gamma analysis was performed for each individual segment before considering the whole modulated beam. This latter was computed following two ways: first, each simulated image was smoothed then weighted by the corresponding MU retrieved from the TPS and the resulting weighted images for all segments are added together. In a second test, each simulated image was first weighted before adding the images for all segments, and the smoothing is performed at the end on the final modulated image. This test was done in order to check if smoothing must be done on individual segments or only on the final modulated image. Gamma criteria of 2% (of maximum dose) and 2.5 mm in distance were considered for the pixels in the image for which the value is greater than 20% of the maximum dose.

3. Results

3.1. Commissioning of the EPID model

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 layer 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 and inline directions are presented on figure 2, for portal images without phantom in the beam. Profiles obtained in the crossline direction for the portal images with the heterogeneous phantom in the beam are shown in figure 3. 2D gamma-index values for gamma criteria of 2% / 2 mm are also given in table 1. These results demonstrate the need to include a correction for backscatter radiation in the model to accurately predict portal, especially for large field sizes. They also validate the accuracy of our model for a wide range of field sizes and for different irradiation configurations, with and without phantom in the beam.

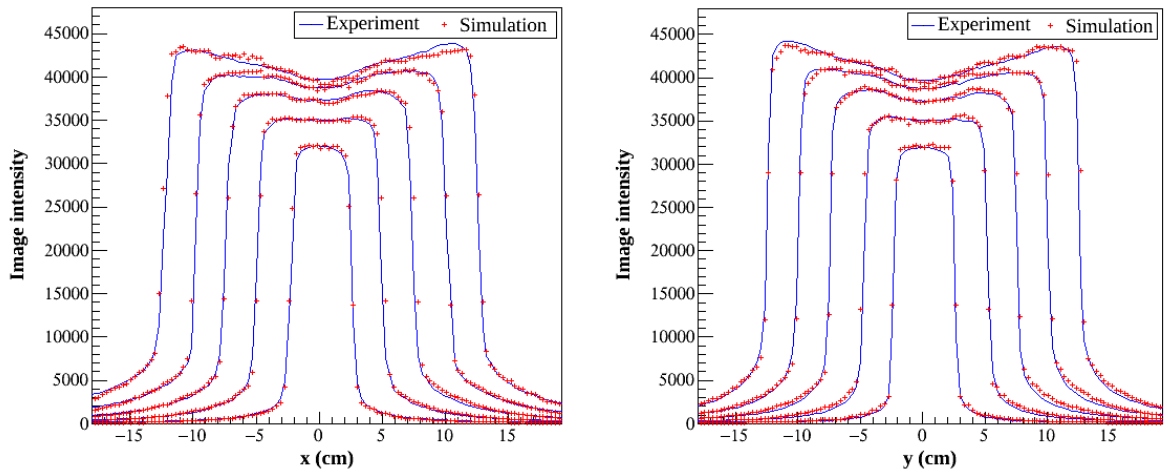

Figure 2. Profiles for acquired (blue line) and simulated (red crosses) portal images in the crossline (left) and inline (right) directions, for the configuration without phantom in the beam.

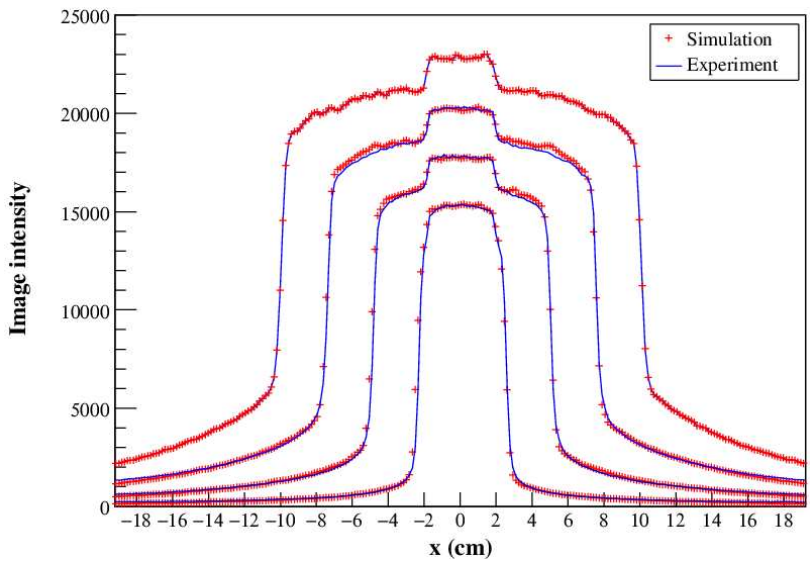

Figure 3. Profiles for acquired (blue line) and simulated (red crosses) portal images in the crossline direction, for the configuration with the heterogeneous phantom in the beam.

Table 1. 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	88.3%	97.4%	97.5%
20 x 20	57.2%	95.4%	96.1%
25 x 25	24.6%	93.1%	95.0%

3.2. Denoising of portal images

3.2.1. Denoising test case: heterogeneous phantom. Figure 4 summarizes the fraction of pixels which fail the 1% difference dose criterion. In all cases, the interest of using any of the tested denoising algorithms is evident, even for MC images with a high SNR. IRON, LASG, and DPGLM exhibit similar performances for images with a RMS pixel noise better than 2%. For images with higher RMS pixel noise, the statistical basis of DPGLM offers more robustness with respect to noise. Indeed, MC images computed with 5% RMS pixel noise could be denoised correctly using DPGLM, as proved by the 2% fraction of pixels failing the 1% difference dose criterion. The LASG and IRON algorithms would require much less noisy MC pre-calculated images, of about 2% and 3.5% for IRON and LASG, respectively in order to reach the same result. We also observe on the profiles shown in figure 5 that DPGLM produces smoother images than IRON and LASG while preserving well the edges in the high-gradient dose regions. These results demonstrate that it is possible to reach an image quality compatible with a clinical interpretation using DPGLM for PSFs that store between 100 million (10% RMS pixel noise) and 500 million photons (5% RMS pixel noise). Obtaining the same image quality using LASG would require a PSF two times larger. With IRON, a PSF three times larger would be necessary. Also let notice that figure 5 exhibits the unstable behavior of LASG algorithm in high noise context. Indeed, LASG interval's size adaptation scheme may lead to a zero length smoothing window. This situation arises as soon as no window size configuration can satisfy the $\mathcal{X}^2$- goodness of fit test. As a consequence, these related data points remain unsmoothed and appear as outliers in the resulting image. This drawback has been pointed out by authors in Kawrakow

(2002). In order to mitigate this effect, they suggest to generate several MC batches and to use a weighted average as smoothed estimate owing to the low probability to observe outliers on the same pixel for different batches. Nevertheless, for noisy EPID images this probability reveals not negligible.

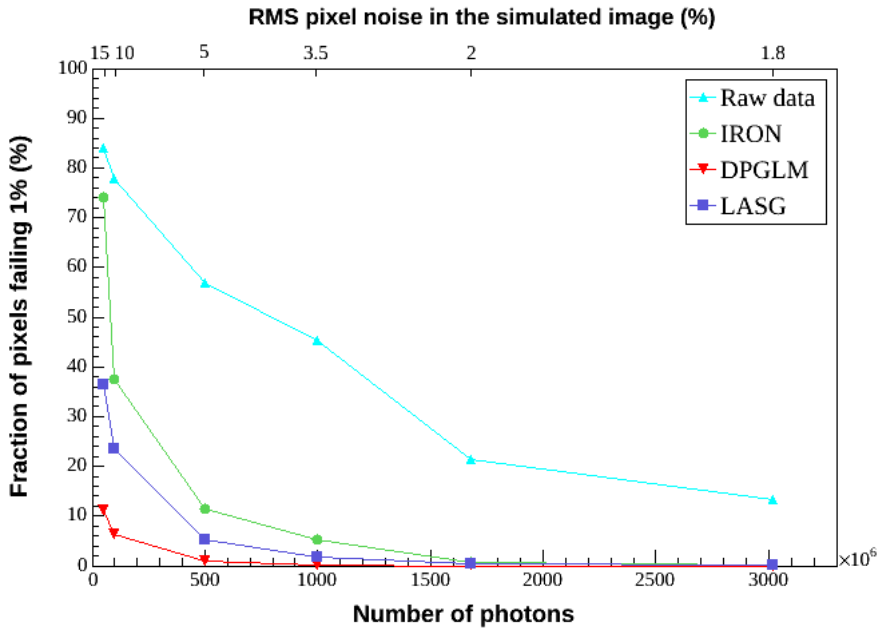

Figure 4. Fraction of pixels failing the 1% difference test for MC raw images (cyan), images denoised with IRON (green), LASG (blue) and DPGLM (red).

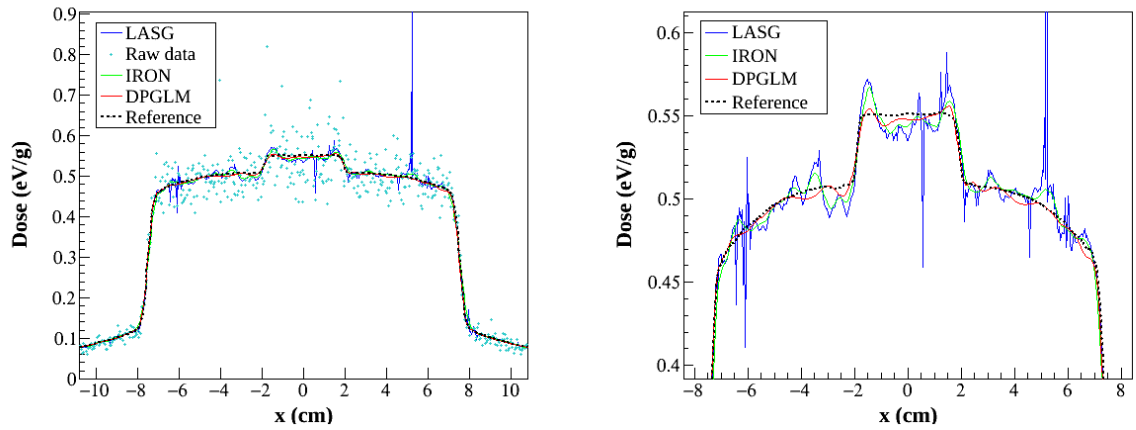

Figure 5. Central profiles drawn through the reference image (black), the MC raw image (cyan), the image denoised using IRON (green), LASG (blue) and DPGLM (red), for the 100 million photon PSF.

Concerning computation times, we experienced slow convergence for the IRON algorithm for low SNR images, as expected. Despite the needs of significant computing requirements, the structure of DPGLM algorithm makes it suitable for parallelization contrary to the LM-BFGS optimization algorithm used for IRON. As a consequence, the effective computation times are similar for DPGLM and IRON. The MC computation of the 1024 × 1024 simulated image on 100 processors (2.26 GHz) of our Linux cluster lasts in 30 minutes when running over 100 million photons from the PSF and 2 h 30 min when running 500 million photons. In the same configurations, denoising on 1024 × 1024 images necessitated about 1 h 30 min, 2 h, and 5 min using DPGLM, IRON, and LASG, respectively. The complete computation of the portal image using IRON took thus 2 h 30 min and 4 h 30 min for 100 and 500 million photons, respectively. Using DPGLM, it took 2 h and 4 h for 100 and 500 million photons, respectively. Using LASG, it took 35 min and 2 h 35 min for 100 and 500 million photons,

respectively. Figure 6 shows the fraction of pixels failing the 1% difference dose criterion test with respect to the global computation time of the portal image.

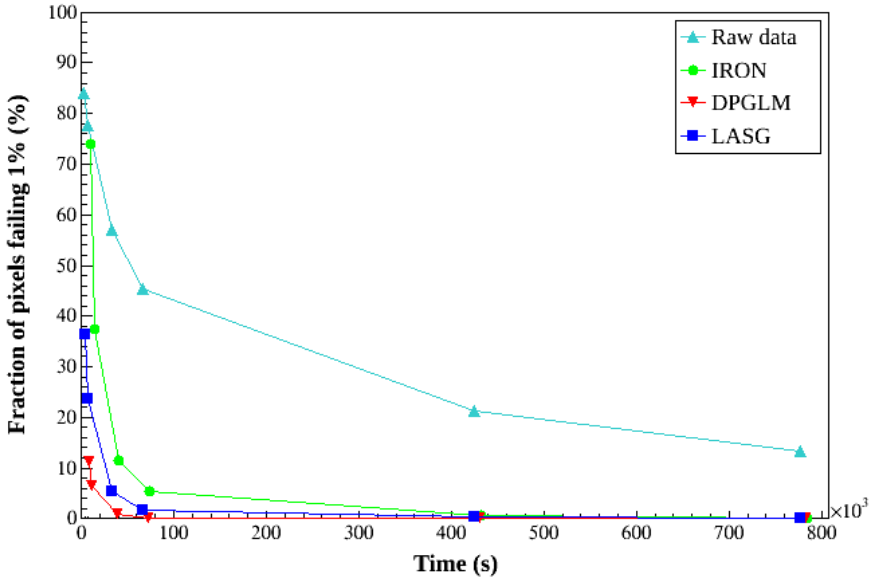

Figure 6. Fraction of pixels failing the 1% difference test for MC raw images (cyan), images denoised with IRON (green), LASG (blue) and DPGLM (red) with respect to the global computation time.

Although LASG was the fastest algorithm for denoising, we nevertheless observed an optimum in terms of accuracy and computing time for DPGLM, on the portal image simulated with a RMS pixel noise of 5%. This denoising test case also helped us to define a range of RMS pixel noise values in the MC simulated image for which the denoising techniques could be applied relevantly. Hence, we noticed that IRON performed well for a RMS pixel noise (σ) smaller than 2 or 3%, LASG for σ smaller than 3.5%, and DPGLM for σ of 10% at most. That is why, in the following, portal images were then simulated with a σ of no more than 10%.

3.2.2. Denoising test case: picket fence field. Figure 7 shows the comparison between the profile drawn through the experimentally acquired image and through the simulated images for the two RMS pixel noise values considered here. These results show that DPGLM performs equally in denoising the high and low RMS pixel noise simulated images. The percentage difference between the experimental profile and the simulated profiles denoised with DPGLM are within 2% or within 2 mm in the region of steep dose gradient. This excellent agreement validates the reliability of the parameters used in our MLC model (leaf-end radius, leaf width and leaf side design), and ensures that the MLC leaf positions are well reproduced. As already noticed above for the first test case, profiles denoised using DPGLM are smoother than those denoised using IRON or LASG (with $\mathcal{X}^2_{\max} = 1.7$), for the high RMS pixel noise image. For the low RMS pixel noise image, both IRON and LASG produce many outliers, giving here again a good example of the effects caused by the unstable behvior of both algorithms in high noise context. By tuning the $\mathcal{X}^2_{\max}$ value to 10 for the LASG algorithm, the simulated images, including the one with high RMS pixel noise, seemed to be better smoothed; but this is at the expense of the signal amplitude, which is no more correctly reproduced.

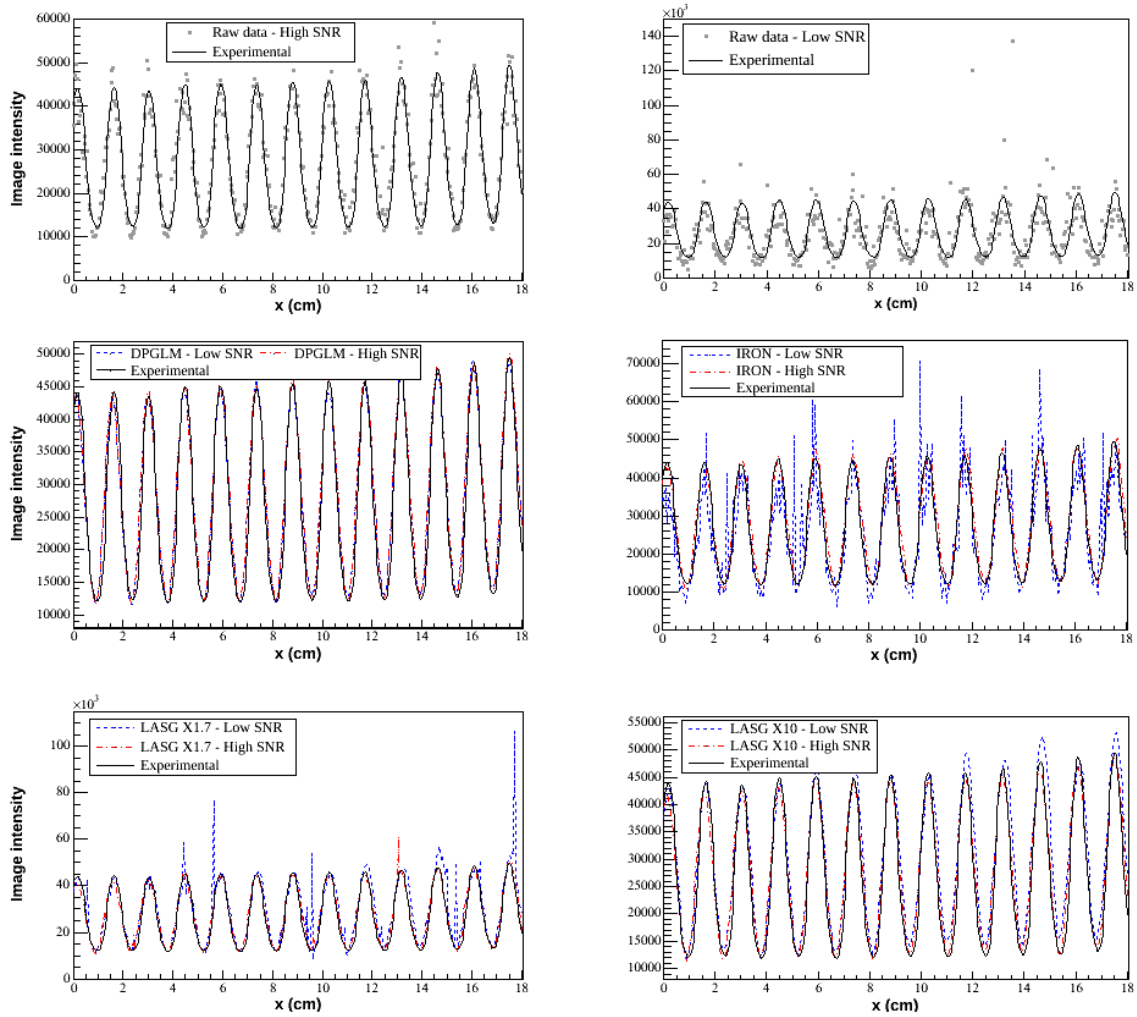

Figure 7. Profiles drawn through the experimental image (black) and the MC raw image (grey squares) simulated for a high SNR value (top left)) and a low SNR value (top right). Profiles through images denoised with DPGLM (middle left), IRON (middle right), LASG with $\mathcal{X}^2_{max} = 1.7$ (bottom left) and LASG with $\mathcal{X}^2_{max} = 10$ (bottom right) are also compared to the experimental profile (in black), for the high SNR value (in red) and the low SNR value (in blue).

3.2.3. Pelvis treatment plan. Figure 8 shows the acquired portal image for the pelvis treatment plan (figure 8a) as well as details in a zoomed area outlined in yellow in image (a). In this outlined area, simulated images obtained for a RMS pixel noise of 3%, 5.7% and 10% before smoothing (first column), after smoothing with DPGLM (second column), IRON (third column), and LASG (last column) are also shown.

Images smoothed with IRON looked particularly grainy in the irradiation field: this effect became more pronounced as the RMS pixel noise in the simulated image increased. LASG gave images with a better quality than IRON, although some spikes could be observed. The amplitude and frequency of these spikes became more important as the simulated image became noisy. For all the statistical uncertainties considered here, DPGLM outperformed IRON and LASG by providing smoother images. However, for the σ = 10% case, it can be seen that details in the high-gradient dose region were significantly altered compared to results obtained for σ = 3% or 5.7% on all smoothed images. In this case, the level of noise was too high to allow a faithful reproduction of the details in the image towards the field edges. These results suggested to model MC portal images with a RMS pixel noise smaller than 5% for conformal and IMRT beams.

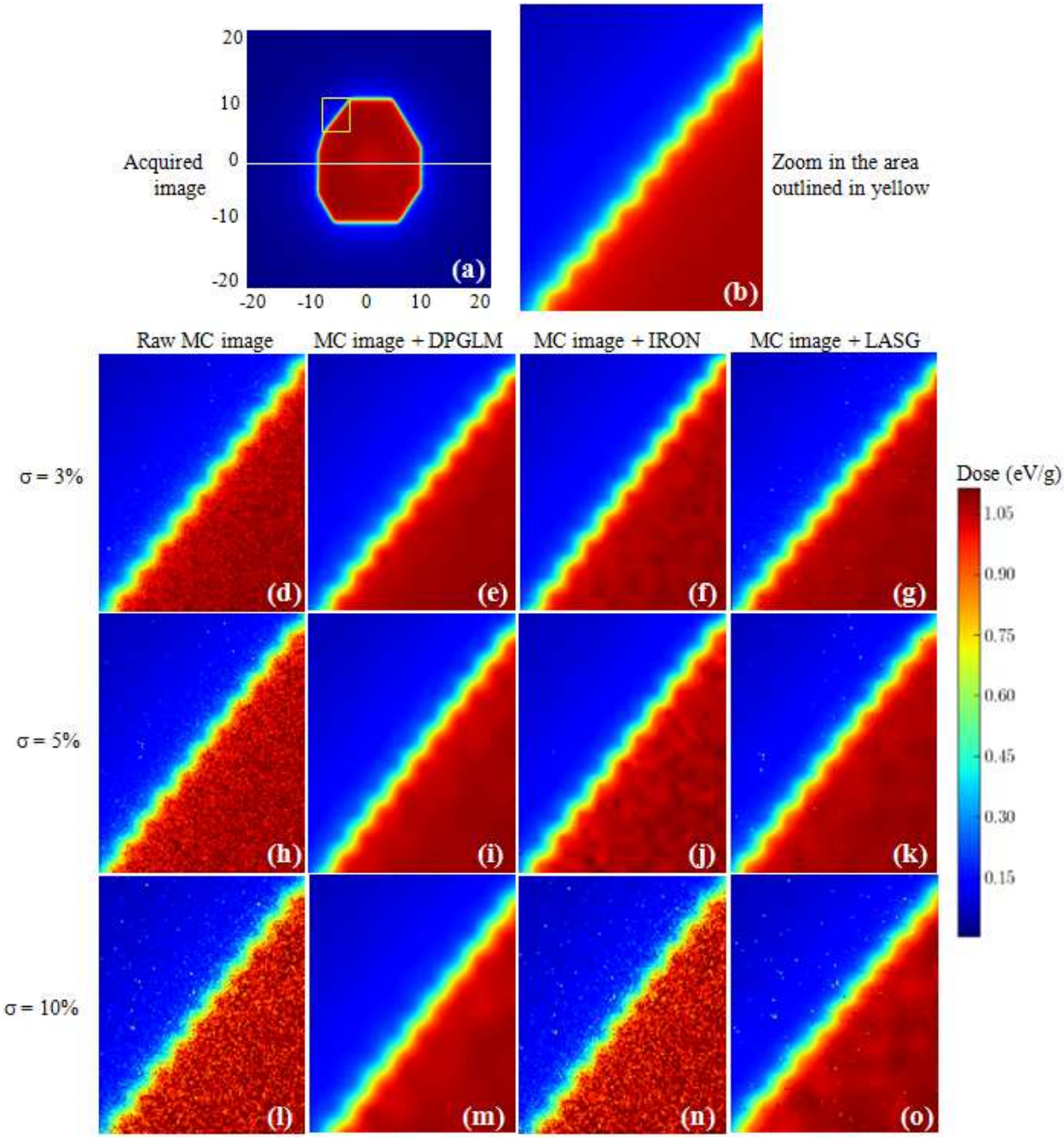

Figure 8. (a) Acquired portal image for the pelvis treatment beam and (b) zoom in the area outlined in yellow in (a). In this outlined area, simulated images obtained without denoising with a RMS pixel noise of 3%, 5.7% and 10% (d, h and l), using the DPGLM method (e, i and m), the IRON method (f, j and n) and the LASG method (g, k and o).

For the different simulated statistics, the profiles were drawn in the unsmoothed MC image and in images smoothed with DPGLM, IRON and LASG, along the white line represented in figure 8a. The comparison of profiles is shown in figure 9. Spikes caused by the LASG algorithm can also be clearly visible on the profiles and they are particularly important for σ = 5.7% or 10%. For all σ values, DPGLM produced smoother profiles than IRON and LASG, and reproduced very well the experimental profile.

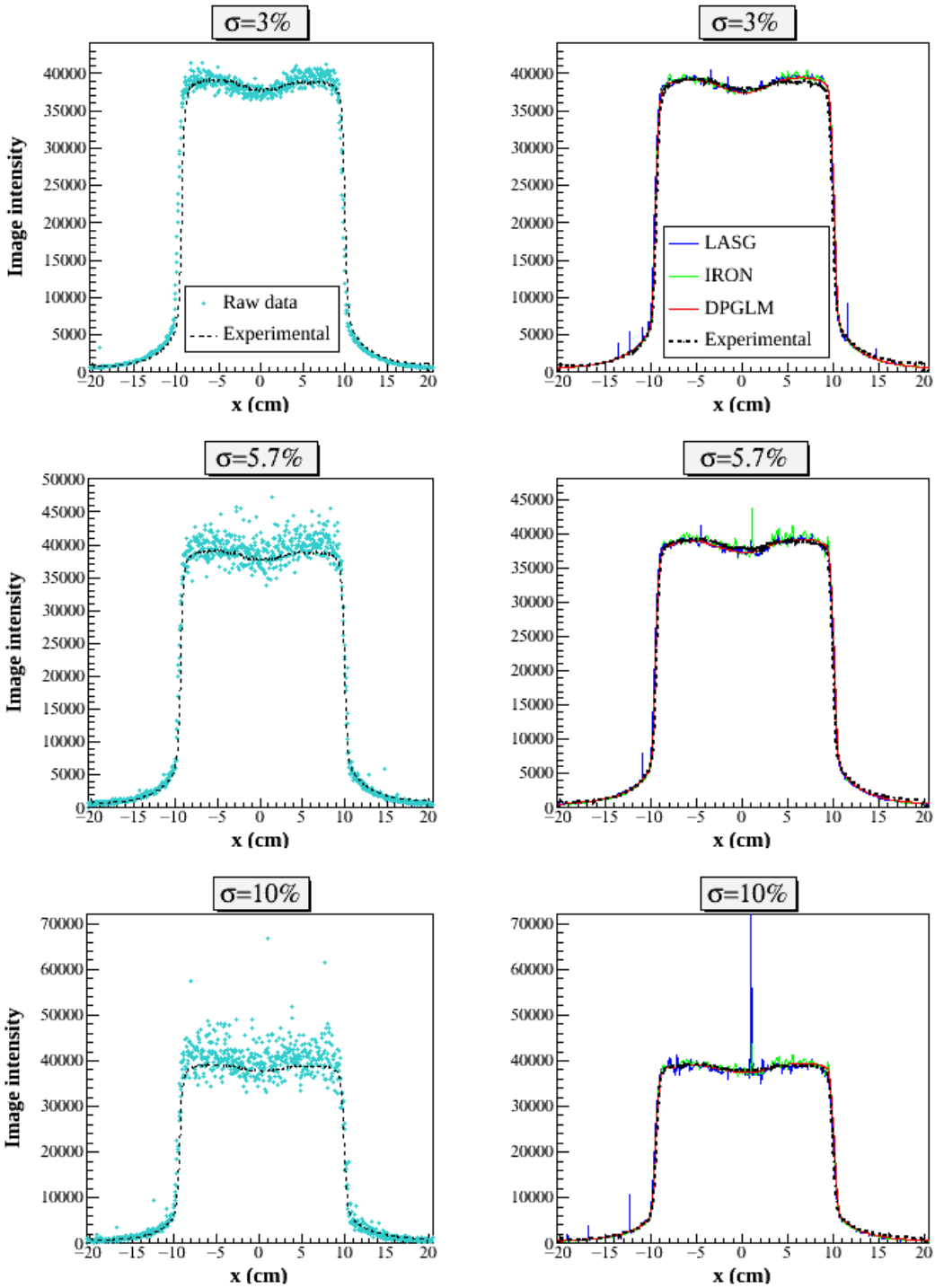

Figure 9. Profiles drawn through the acquired image (black dashed line), the MC image undenoised (cyan crosses), the MC image denoised with IRON (green), LASG (blue) and DPGLM (red), for three different statistical uncertainties in the simulated image.

Table 2 gives the results of the gamma analysis performed on MC images computed for the three different statistical uncertainties and denoised using IRON, LASG and DPGLM. For σ = 3%, LASG and IRON performed equally well, but a little bit worse than DPGLM that gave excellent results. IRON showed limitations when σ becomes 5.7%, to the contrary, LASG and DPGLM maintained their smoothing capabilities very well. At σ = 10%, only DPGLM managed to smooth correctly the image. These results stressed the expected behaviour from IRON, which is at its most successful when images are moderately noisy, i.e for σ smaller than 3%.

Table 2. Percentage of pixels passing gamma criteria of 2% / 2.5 mm for the simulated images calculated for three different statistical uncertainties, and smoothed with DPGLM, IRON and LASG.

RMS pixel noise	DPGLM	IRON	LASG ($\mathcal{X}^2_{max} = 1.7$)
σ = 3%	97.2	94.4	93.5
σ = 5.7%	96.5	88.4	94.0
σ = 10%	95.7	46.8	83.0

Dealing with computation time, the simulation of the portal images for RMS pixel noise of 5.7% and 10% took 40 minutes and 14 minutes on 72 processors, respectively. Denoising times are fixed by the size of the image to be denoised, and are thus equal to those obtained in section 3.2.1, i.e 1 h 30 min, 2 h and 5 min using DPGLM, IRON, and LASG, respectively. The overall time required to predict portal images for σ = 5.7% was thus of 2 h 10 min, 2 h 40 min and 45 min for DPGLM, IRON, and LASG, respectively.

3.2.4. IMRT prostate treatment plan. Results of the gamma analysis are summarized in table 3. The gamma analysis performed on each segment individually showed that 95.2% of points in average passed the chosen gamma criteria for DPGLM, 73.3% for IRON and 89.3% for LASG. Low values obtained for IRON could be explained by the high level of noise still present in the image after smoothing, in particular within the irradiation field. The spikes caused by the use of LASG also penalized the gamma value. Here again, DPGLM proved its ability to smooth simulated images with a RMS pixel noise of about 5%.

Table 3. Percentage of pixels passing gamma criteria of 2% / 2.5 mm for the simulated images smoothed with DPGLM, IRON, and LASG.

Segment	DPGLM	IRON	LASG ($\mathcal{X}^2_{max} = 1.7$)
1	97.1	72.4	91.3
2	97.0	70.8	91.2
3	96.0	67.2	89.0
4	95.0	67.3	87.3
5	94.3	69.6	87.3
6	96.7	68.5	87.9
7	97.1	75.4	90.3
8	90.7	80.4	88.8
9	93.3	88.7	90.4
All segments smoothed individually, then summed	97.8	77.2	92.4
All segments summed, then final image smoothed	98.2	81.0	80.0 ($\mathcal{X}^2_{max} = 1.7$) 94.6 ($\mathcal{X}^2_{max} = 10$)

With DPGLM and IRON, very similar results were obtained when either smoothing first each segment individually or smoothing the final image obtained after summing all segments. We could also notice that the gamma-index value calculated for the whole beam is a little bit better than values taken individually for each segment, showing an averaging effect which tends to decrease dose gradients in the summed image. This effect could be observed also for IRON and LASG. For LASG, we noticed poor results in smoothing after summing all segments if we keep $\mathcal{X}^2_{max} = 1.7$. This is due to the correlations between the MC data of each segment. Indeed, a common PSF recorded after the jaws was used for computing each segment. In this case, the variance of the sum is far greater than the sum of the variances and we had to increase $\mathcal{X}^2_{max}$ value to 10 in order to get better performances. Varying the $\mathcal{X}^2_{max}$ value from 1.7 to 10, the percentage of pixels passing the 2% / 2.5 mm criteria

increased from 80.0% to 94.6%, almost reaching the same performances as DPGLM. It is likely that this inter-segments correlation phenomenon might also affect the IRON performances when working on the sum of segments since computed noise values are also used in this method. These results showed that an accurate prediction of the portal image for a modulated beam is possible by smoothing only the final modulated image rather than each segment individually, which contributes to save time during the smoothing step of the prediction algorithm.

The simulation of the portal image for one segment took about 15 minutes on 72 processors. Added to the computation time devoted to smoothing, the overall time for the portal image prediction of a segment was thus of 1 h 45 min, 2 h 15 min and 20 min for DPGLM, IRON, and LASG, respectively, if the image is smoothed once after summation of all segments.

4. Discussion

The present study is, to our knowledge, the first study that proves that the PENELOPE MC code is suitable for the prediction of accurate portal images. As already shown by other authors on the Varian aS500 EPID (Cufflin *et al* 2010), a non-uniform backscatter model had to be integrated within the MC model in order to take into account for backscattered radiation arising from components that surround the EPID cassette, from the arm that supports the EPID and also from the gantry. An accurate backscatter modelling is mandatory in order to ensure an accurate prediction of portal images, in particular for large field sizes.

Regarding the denoising part of this work, we have shown the feasibility of the approach based on the combination of a smoothing technique with MC simulations to predict portal images with the same resolution, and potentially with the same accuracy as experimental images. Using MC simulations alone, this objective could not have been reached due to the intensive computational resources that are required. Performances of three denoising algorithms were evaluated in this context, in terms of accuracy and computation time. Results obtained for the IRON algorithm confirmed that IRON is not suitable for denoising problems characterized by a huge number of variables, i.e pixels in the image, but much more relevant in denoising 3D dose distributions where less and far bigger voxels are used. Better performances for IRON might have been observed if we had applied it on smaller images, for instance a portion of the 1024 × 1024 portal image. As expected, IRON was also shown to be heavily hindered when very noisy images were considered and it was unable to smooth portal images simulated with a RMS pixel noise higher than 3%. In order to facilitate the convergence of IRON, a proposed alternative could be to initialize IRON with an image already denoised, for instance the image smoothed by LASG since it is the fastest of all the algorithms tested here.

Good results could be obtained with LASG only if the $\mathcal{X}^2_{max}$ parameter was tuned in order to balance the trade-off between introduced bias and smoothing capabilities. If the $\mathcal{X}^2_{max}$ value is too high, steep dose gradients are too smoothed and too much more bias is introduced; otherwise, if the $\mathcal{X}^2_{max}$ value initially proposed in Kawrakow *et al* (2002) was kept, smoothing is not optimal. The tuning of the $\mathcal{X}^2_{max}$ should be done in a trial-and-error process, for each configuration, which could be cumbersome and time consuming, and finally not convenient in practice. As noticed in Kawrakow *et al* (2002), some pixels are not denoised at all by LASG and appear as outliers on the smoothed image. This issue might be mitigated by working on several batches. But the non gaussianity of the noise is increased by working with smaller PSF. This can lead to reinforce the non optimal behaviour of LASG in this situation.

For all the configurations studied in this work, DPGLM outperformed LASG and IRON by producing smoother images and being able to smooth relevantly images characterized by a RMS pixel noise smaller than 5%. Moreover, to the contrary of LASG and IRON, DPGLM requires neither parameter adjustment nor knowledge about the variance of the noise associated to each pixel, making this denoising method more generic than LASG. In the case of the IMRT beam where portal images for each segment are summed, DPGLM is less sensitive than LASG or IRON to errors caused by the inter-segment correlations due to the use of a common PSF to compute each segment. The tuning of both IRON and LASG algorithms also reflects this problem and should be revised specifically in order to take it into account. One possible alternative could be to compute one independent PSF per segment in order to decrease such correlations.

The overall computation time for the prediction of the portal image corresponds in our method to the sum of the time spent for the MC simulation and the time spent for smoothing. It is well known

that the MC simulation time depends highly on the irradiation field size, and it was shown to vary from 2 h 30 min for a 15 × 15 cm² field to about 10 min (on 100 processors) for a small IMRT beamlet (for a RMS pixel noise of 5% in the simulated image). Considering smoothing time alone, LASG is by far the fastest algorithm compared to other techniques, when implemented on the same hardware resources and with identical parallelization efforts, i.e. hybrid parallelization MPI/OpenMP in our case. However, when considering also the smoothing capabilities, DPGLM reveals interesting performances and allows simulating a portal image with a similar quality than using LASG but with a PSF storing twice fewer particles. Working with smaller PSF, DPGLM offers hence the possibility to save disk storage compared to LASG, which might reveal attractive. It should also be noticed that in contrast with LASG or IRON, DPGLM outputs are expressed as functions (as mentioned in section 2.2.3) and not as matrix values taken over a fixed input data grid. It is thus possible with DPGLM to compute the denoised dose at any portal plane coordinates whether belonging to the initial simulated data grid or not. This feature may offer some additional flexibility in tuning the trade-off between image resolution and smoothing time. Indeed, since DPGLM smoothing time is approximately proportional to the number of input points, we can save time by using a coarser input grid for areas with moderate dose gradients (e.g. out-of-field regions), reducing then the number of data for these regions. Furthermore, it seems reasonable to systematically try to use the coarsest input grid which does not entail bias in the denoised image computed at the desired output resolution using DPGLM interpolation capabilities.

To conclude, DPGLM reveals very interesting denoising properties on 2D portal images. Some features of the proposed DPGLM method have not been fully investigated in this work. Among them, a 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. This could also help in the assessment of the denoising method since here, a reference image could not be determined easily due to very large computational resources required in order to do so. An example is given in figure 10 where it is showed that credible intervals may be assessed and that they contain the expected true dose value.

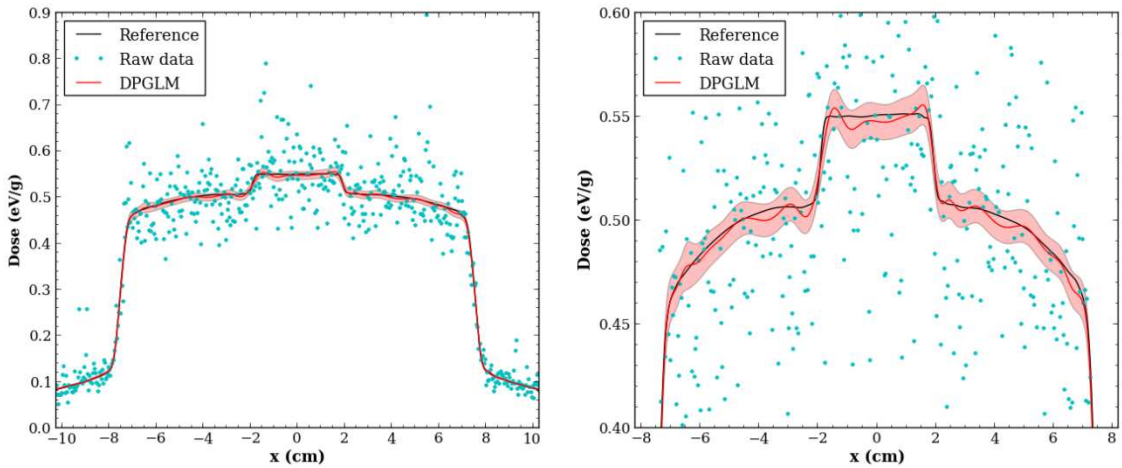

Figure 10. Results obtained for the 100 million photon denoising test case PSF. The 3σ-uncertainty interval computed with DPGLM (red band) is plotted round the DPGLM mean dose estimate (red) and compared to reference dose profiles (black). Raw data are represented by cyan dots.

Additionally, DPGLM performance needs to be studied more thoroughly, for instance the systematic bias introduced by DPGLM in denoised images. The sliding window test proposed by De Smedt *et al* (2006) to evaluate the systematic bias was tested here but images very similar to those obtained with the gamma analysis were obtained, yielding no additional relevant information regarding to the systematic bias evaluation. We plan to perform this study using rather a method based on data replications in the future. Finally, the performance of DPGLM will also be assessed very soon on portal images computed with a patient in the beam, and compared to LASG and IRON.

5. Conclusions

This paper presents a new methodology that combines MC simulations to an efficient denoising method which enables the accurate computation of high-resolution MC portal images for computational burden compatible with clinical settings and acceptable for TPS. The Siemens OptiVue1000 EPID was first accurately simulated using the MC code PENELOPE. One objective of this work was to identify a denoising algorithm adapted to the specific requirements of portal image denoising, i.e able to deal with a huge number of variables (a highly pixelized image), and also able to denoise potentially very noisy images, presenting a non Gaussian noise with a variable amplitude depending on the localization of the pixel. A new denoising method based on a nonparametric Bayesian framework and called DPGLM, was compared in terms of smoothing capabilities and computing time with two standard denoising algorithms, namely IRON and LASG. DPGLM outperformed both IRON and LASG by providing smoother images and allowing the smoothing of portal images with a RMS pixel noise of 5% at most, in all tested configurations. Interesting features of the DPGLM algorithm were thus demonstrated in this work regarding to 2D dose denoising. Future work will involve thorough assessment of the systematic bias introduced by DPGLM, optimization of the algorithm to even further shorten denoising computing time, assessment of its robustness for portal images computed with a patient in the beam, and its generalization for 3D dose denoising.

Acknowledgments

The authors would like to thank Olivier De Dreuille, Sébastien Richard and Carsten Schulze for providing the technical specification of the linac and the EPID imager, and for their assistance with the IMRT beam modelling. The authors also like to acknowledge OSEO for funding the French INSPIRA project and supporting this work.

Appendix A. DPGLM Gibbs sampler

In order to express all steps of the sampling procedure, we first give a generative prior model for the DPM parameters. Definitions of probability density functions are given in Appendix B.

$$
\begin{gathered}
V_1, V_2, \dots \sim \mathrm{Beta}(1, \alpha) \\
\text{let } w_1 = V_1, w_2 = V_2(1 - V_1), \dots \\
\mu_{11}, \mu_{12}, \dots \sim \mathcal{N}\left(m_1, \sigma_{\mu 1}^2\right) \\
\mu_{21}, \mu_{22}, \dots \sim \mathcal{N}\left(m_2, \sigma_{\mu 2}^2\right) \\
\tau_{11}^{-1}, \tau_{12}^{-1}, \dots \sim \mathrm{Gamma}(a_{\tau 1}, b_{\tau 1}) \\
\tau_{21}^{-1}, \tau_{22}^{-1}, \dots \sim \mathrm{Gamma}(a_{\tau 2}, b_{\tau 2}) \\
\beta_1, \beta_2, \dots \sim \mathcal{N}\left(M_\beta, \Gamma_\beta\right) \\
\sigma_1^{-2}, \sigma_2^{-2}, \dots \sim \mathrm{Gamma}(a_{\sigma^2}, b_{\sigma^2}) \\
\text{let } \theta_1 = (\mu_{11}, \mu_{21}, \tau_{11}, \tau_{21}, \beta_1, \sigma_1^2),\ \theta_2 = (\mu_{12}, \mu_{22}, \tau_{12}, \tau_{22}, \beta_2, \sigma_2^2), \dots
\end{gathered}
$$

This generative model is fully equivalent to the prior probability density $f(w_1, w_2, \dots, \theta_1,\ \theta_2, \dots)$, and corresponds to the DP prior for the random measure $G(\cdot) = \sum_{k=1}^{\infty} w_k\ \delta_{\theta_k}(\cdot) \sim \mathrm{DP}(\alpha, G_0)$.

Given the DPM parameters, we complete the description by a generative model for the observed data (x_{1i}, x_{2i}, y_i). We introduce the latent classification variables K_i, defined for $i = 1, \dots, n$, such that $K_i = k$ if (x_{1i}, x_{2i}, y_i) is distributed from the kth component of the mixture $f(x_1, x_2, y)$.

For all $i \leq n$,

$$
\begin{gathered}
K_i |\ w_1, w_2, \dots \sim \sum_{k=1}^{\infty} w_k\ \delta_k(\cdot) \\
x_{1i}, x_{2i} | K_i, \theta_1,\ \theta_2, \dots \sim\ \mathcal{N}\left(x_{1i} |\ \mu_{1K_i}, \tau_{1K_i}\right) \mathcal{N}\left(x_{2i} |\ \mu_{2K_i}, \tau_{2K_i}\right) \\
y_i |\ K_i, x_{1i}, x_{2i},\ \theta_1,\ \theta_2, \dots \sim \mathcal{N}\left(y_i |\ \beta'_{K_i} \cdot \tilde{X}_{K_i}(x_{1i}, x_{2i}), \sigma_{K_i}^2\right)
\end{gathered}
$$

This generative model is fully equivalent to the likelihood

$$
f(K_1, \dots, K_n, x_{11}, x_{21}, \dots, x_{1n}, x_{2n}, y_1, \dots, y_n |\ w_1, w_2, \dots, \theta_1,\ \theta_2, \dots)
$$

From these probability distributions and using Bayes rule, we want to compute the conditional probability density

$$f(w_1, w_2, \dots, \theta_1,\ \theta_2, \dots \mid K_1, \dots, K_n, x_{11}, x_{21}, \dots, x_{1n}, x_{2n}, y_1, \dots, y_n)$$

This is achieved thanks to a Gibbs sampler which, at iteration (t) of the algorithm, successively draws samples from, for all k,

$$\tau_{1k} \mid \mu_{1k}, K_1, \dots, K_n, x_{11}, \dots, x_{1n}$$
$$\tau_{2k} \mid \mu_{2k}, K_1, \dots, K_n, x_{21}, \dots, x_{2n}$$
$$\mu_{1k} \mid \tau_{1k}, K_1, \dots, K_n, x_{11}, \dots, x_{1n}$$
$$\mu_{2k} \mid \tau_{2k}, K_1, \dots, K_n, x_{21}, \dots, x_{2n}$$
$$\beta_k \mid \mu_{1k}, \mu_{2k}, \tau_{1k}, \tau_{2k}, \sigma_k^2, K_1, \dots, K_n, x_{11}, x_{21}, \dots, x_{1n}, x_{2n}, y_1, \dots, y_n$$
$$\sigma_k^2 \mid \mu_{1k}, \mu_{2k}, \tau_{1k}, \tau_{2k}, \beta_k, K_1, \dots, K_n, x_{11}, x_{21}, \dots, x_{1n}, x_{2n}, y_1, \dots, y_n$$
$$w_1, w_2, \dots \mid K_1, \dots, K_n$$

And for all $i \leq n$,

$$K_i \mid x_{1i}, x_{2i}, y_i, \mu_{11}, \mu_{12}, \dots, \mu_{21}, \mu_{22}, \dots, \tau_{11}, \tau_{12}, \dots, \tau_{21}, \tau_{22}, \dots, \beta_1, \beta_2, \dots, \sigma_1^2, \sigma_2^2, \dots$$

The nonparametric behavior of the method is characterized by a potentially infinite collection of parameters w_k and θ_k. This computational difficulty can be tackled by truncation of the DP at a predefined huge number of components (Ishwaran and James 2001). We suggest here another approach (Kalli *et al*. 2011). This latter method introduces auxiliary variables u_i which allow sampling only a finite random number κ of components at iteration (t) while avoiding any hard truncation of the model. We are now ready to describe all steps of the algorithm.

- *Initialize at random.*
  - Sample for $1 \leq i \leq n, u_i \sim \text{Uniform}(0, 1/n)$, and set $u^* = \min_u(\{u_i\})$
  - Sample $V_1 \sim \text{Beta}(1, \alpha)$. Set $w_1 = V_1$ and $r_1 = 1 - V_1$.
  - For $k > 1$, generate $V_k \sim \text{Beta}(1, \alpha)$, set $w_k = V_k\, r_{k-1}$, set $r_k = r_{k-1}(1 - V_k)$, until $r_k < u^*$.
  - Set to κ^* the maximum value of k.
  - Sample θ_k for $1 \leq k \leq \kappa^*$, following prior distributions.

For iteration $t = 1, \dots, T$,

- Sample $\left(K_i \mid x_{1i}, x_{2i}, y_i, \mu_{11}, \dots, \mu_{1\kappa^*}, \mu_{21}, \dots, \mu_{2\kappa^*}, \tau_{1\kappa^*}, \dots, \tau_{1\kappa^*}, \tau_{2\kappa^*}, \dots, \tau_{2\kappa^*}, \beta_1, \dots, \beta_{\kappa^*}, \sigma_1^2, \dots, \sigma_{\kappa^*}^2\right)$ for $i \leq n$.
  - Compute for $1 \leq k \leq \kappa^*$,
  - $\lambda_k =$
    $1(w_k > u_i)\, \max(w_k, 1/n)\, \mathcal{N}(x_{1i} \mid \mu_{1k}, \tau_{1k})\, \mathcal{N}(x_{2i} \mid \mu_{2k}, \tau_{2k})\, \mathcal{N}\left(y_i \middle| \beta_k' \cdot \tilde{X}_k(x_{1i}, x_{2i}), \sigma_k^2\right)$
    where $1(A)$ is the indicator function : $1(A) = 1$ if A is true and $1(A) = 0$ otherwise.
  - Generate $K_i \sim \frac{1}{\sum_{k=1}^{\kappa^*} \lambda_k} \sum_{k=1}^{\kappa^*} \lambda_k\, \delta_k(K_i)$
- Reorder components labels following their *order of appearance* when generating K_i. Set to κ_n the number of distinct K_i. Set, for all $k \leq \kappa_n$, $n_k = \#\{K_i = k\}$, the number of data assigned to component k.
- Sample$(w_1, w_2 \dots, u_1, \dots, u_n \mid K_1, \dots, K_n)$
  - Sample $\left(w_1, w_2, \dots, w_{\kappa_n}, r_{\kappa_n}\right) \sim \text{Dirichlet}\left(n_1, n_2, \dots, n_{\kappa_n}, \alpha\right)$
  - Sample for $i \leq n$, $u_i \sim \text{Uniform}\left(0, \min(w_k, 1/n)\right)$
  - Set $u^* = \min_u(\{u_i\})$
  - For $k > \kappa_n$, generate $V_k \sim \text{Beta}(1, \alpha)$, set $w_k = V_k\, r_{k-1}$, set $r_k = r_{k-1}(1 - V_k)$, until $r_k < u^*$.
  - Set κ^* to the maximum value of k.
- Sample $(\tau_{1k} \mid \mu_{1k}, K_1, \dots, K_n, x_{11}, \dots, x_{1n})$ for $k \leq \kappa^*$,

$$\tau_{1k}^{-1} \sim \text{Gamma}\left(a_{\tau 1} + \frac{n_k}{2}, b_{\tau 1} + \frac{1}{2} \sum_{\{i:\, K_i = k\}} (x_{1i} - \mu_{1k})^2\right)$$

- Sample $(\tau_{2k} \mid \mu_{2k}, K_1, \dots, K_n, x_{21}, \dots, x_{2n})$ for $k \leq \kappa^*$,

$$\tau_{2k}^{-1} \sim \text{Gamma}\left(a_{\tau 2} + \frac{n_k}{2}, b_{\tau 2} + \frac{1}{2} \sum_{\{i:\, K_i = k\}} (x_{2i} - \mu_{2k})^2 \right)$$

- Sample $(\mu_{1k} |\, \tau_{1k}, K_1, \dots, K_n, x_{11}, \dots, x_{1n})$ for $k \leq \kappa^*$,

$$\mu_{1k} \sim \mathcal{N}\left(\frac{\sigma_{\mu 1}^2 \left(\sum_{\{i:\, K_i = k\}} x_{1i} \right) + m_{\mu 1}\, \tau_{1k}}{n_k\, \sigma_{\mu 1}^2 + \tau_{1k}}, \frac{\sigma_{\mu 1}^2\, \tau_{1k}}{n_k\, \sigma_{\mu 1}^2 + \tau_{1k}} \right)$$

- Sample $(\mu_{2k} |\, \tau_{2k}, K_1, \dots, K_n, x_{11}, \dots, x_{1n})$ for $k \leq \kappa^*$,

$$\mu_{2k} \sim \mathcal{N}\left(\frac{\sigma_{\mu 2}^2 \left(\sum_{\{i:\, K_i = k\}} x_{2i} \right) + m_{\mu 2}\, \tau_{2k}}{n_k\, \sigma_{\mu 2}^2 + \tau_{2k}}, \frac{\sigma_{\mu 2}^2\, \tau_{2k}}{n_k\, \sigma_{\mu 2}^2 + \tau_{2k}} \right)$$

- Sample $\left(\beta_k |\, \mu_{1k}, \mu_{2k}, \tau_{1k}, \tau_{2k}, \sigma_k^2, K_1, \dots, K_n, x_{11}, x_{21}, \dots, x_{1n}, x_{2n}, y_1, \dots, y_n\right)$ for $k \leq \kappa^*$,
  - Set

$$\Gamma_{\beta_k} = \left(\frac{\sum_{\{i:\, K_i = k\}} \tilde{X}_k(x_{1i}, x_{2i})' \cdot \tilde{X}_k(x_{1i}, x_{2i})}{\sigma_k^2} + \Gamma_\beta^{-1} \right)^{-1}$$

  - Set

$$\mathrm{M}_{\beta_k} = \Gamma_{\beta_k} \cdot \left(\frac{\sum_{\{i:\, K_i = k\}} \tilde{X}_k(x_{1i}, x_{2i})\, y_i}{\sigma_k^2} + \Gamma_\beta^{-1} \cdot M_\beta \right)$$

  - Sample

$$\beta_k \sim \mathcal{N}\left(\mathrm{M}_{\beta_k}, \Gamma_{\beta_k} \right)$$

- Sample $\left(\sigma_k^2 |\, \mu_{1k}, \mu_{2k}, \tau_{1k}, \tau_{2k}, \beta_k, K_1, \dots, K_n, x_{11}, x_{21}, \dots, x_{1n}, x_{2n}, y_1, \dots, y_n\right)$ for $k \leq \kappa^*$,

$$\sigma_k^{-2} \sim \text{Gamma}\left(a_{\sigma^2} + \frac{n_k}{2}, b_{\sigma^2} + \frac{1}{2} \sum_{\{i:\, K_i = k\}} \left(y_i - \beta_k' \cdot \tilde{X}_k(x_{1i}, x_{2i}) \right)^2 \right)$$

- Compute at iteration (t) for any chosen (x_1, x_2),

$$d\,(x_1, x_2)^{(t)} = E(y|x) = \sum_{k=1}^{\infty} \frac{w_k\, \mathcal{N}(x_1 | \mu_{1k}, \tau_{1k})\, \mathcal{N}(x_2 | \mu_{2k}, \tau_{2k})}{\sum_{l=1}^{\infty} w_l\, \mathcal{N}(x_1 | \mu_{1l}, \tau_{1l})\, \mathcal{N}(x_2 | \mu_{2l}, \tau_{2l})}\, \beta_k' \cdot \tilde{X}_k(x_1, x_2)$$

The smoothed EPID image may or may not share the same grid as initial MC data since one can choose any point (x_1, x_2) of interest. In particular, it is easy to interpolate between initial data points.

Then, from all T iterations of the Gibbs sampler, we can compute the dose estimate

$$\hat{d}\,(x_1, x_2) \approx \frac{1}{T} \sum_{t=1}^{T} d\,(x_1, x_2)^{(t)}$$

In addition, we can also compute the dose posterior standard deviation

$$\sigma_{\hat{d}\,(x_1, x_2)} \approx \left(\frac{1}{T-1} \sum_{t=1}^{T} \left(d\,(x_1, x_2)^{(t)} - \hat{d}\,(x_1, x_2) \right)^2 \right)^{\frac{1}{2}}$$

Hyperparameters.

Note that some hyperparameters can be fixed considering knowledge from MC simulations. Typically, the mean of the prior distribution on σ_k^2 can be fixed to the mean of variances stored in the EPID's dose data file for pixels allocated to the kth component. Hyperparameters may either be estimated through another level of hierarchy and setting to them vague priors. As an example, we can put a $\text{Gamma}(\varphi_b, \psi_b)$ prior on the scale parameters $b_{\tau 1}$ and $b_{\tau 2}$ in the distribution of components excursions. This choice leads to a Gamma posterior distribution. E.g.

$$b_{\tau 1} \sim \text{Gamma}\left(\varphi_b + \kappa_n, \psi_b + \sum_{k=1}^{\kappa_n} \frac{1}{\tau_{1k}}\right)$$

It is also worthwhile to estimate the concentration parameter α of the DP following Escobar and West (1995).

We may also notice that we found efficient to replace the Gaussian prior on regression coefficients β_k by a so-called empirical Bayes prior based on sampling P locations $(\tilde{x}_{11}, \tilde{x}_{21}, \dots, \tilde{x}_{1P}, \tilde{x}_{2P})$ from $\mathcal{N}(\mu_{1k},, \tau_{1k})\mathcal{N}(\mu_{2k}, \tau_{2k})$ and take for $(\tilde{y}_1, \dots, \tilde{y}_P)$ the nearest value of y_i corresponding to each sampled points. P has to be greater than the length of vector β_k. We use then $\beta_k \sim \mathcal{N}(\mathrm{M}_\beta, \Gamma_\beta)$ as prior with

$$\Gamma_\beta = \left(\frac{\sum_{p=1}^{P} \tilde{X}_k(\tilde{x}_{1p}, \tilde{x}_{2p})' \cdot \tilde{X}_k(\tilde{x}_{1p}, \tilde{x}_{2p})}{\sigma_k^2}\right)^{-1}$$

$$\mathrm{M}_\beta = \Gamma_\beta \cdot \left(\frac{\sum_{p=1}^{P} \tilde{X}_k(\tilde{x}_{1p}, \tilde{x}_{2p})\, \tilde{y}_p}{\sigma_k^2}\right)$$

Expression for the posterior of β_k is straightforward.

Appendix B. Reminder on some probability density functions definitions used in DPGLM.

- Uniform(a, b) distribution, for $a \le x \le b$ with $b > a$.

$$f_{\text{Uniform}(a,b)}(x) = \begin{cases} \dfrac{1}{\mathrm{b-a}} & \text{if } a \le x \le b \\ 0 & \text{if } a > x \text{ or } x > b \end{cases}$$

- Beta(a, b) distribution, for $0 \le x \le 1$.

$$f_{\text{Beta}(a,b)}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Where $\Gamma(a)$ represents the Gamma function in a.

- Gamma(a, b) distribution, for $x \ge 0$,

$$f_{\text{Gamma}(a,b)}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

- Multivariate normal (gaussian) distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Omega})$, for $\boldsymbol{x} \in \mathbb{R}^p$,

$$f_{\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Omega})}(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Omega}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \cdot \boldsymbol{\Omega}^{-1} \cdot (\boldsymbol{x}-\boldsymbol{\mu})}$$

Where $|\boldsymbol{A}|$ stands here for the determinant of matrix $\boldsymbol{A}$. Note that this definition covers the univariate case when $p = 1$.

- Dirichlet$(\alpha_1, \alpha_2, \dots, \alpha_K)$ distribution for $\boldsymbol{x} = (x_1, x_2, \dots, x_K)' \in \mathbb{R}^K$, $x_i > 0$ for $1 \le i \le K$, $\sum_{i=0}^{K-1} x_i < 1$, $x_K = 1 - \sum_{i=0}^{K-1} x_i$ and $\alpha_i > 0$ for $1 \le i \le K$,

$$f_{\text{Dirichlet}(\alpha_1,\alpha_2,\dots,\alpha_K)}(\boldsymbol{x}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_K)}{\Gamma(\alpha_1)\Gamma(\alpha_2) \dots \Gamma(\alpha_K)} {x_1}^{\alpha_1 - 1} {x_2}^{\alpha_2 - 1} \dots {x_K}^{K-1}$$

References

Boyer AL and Li S 1997 Geometric analysis of light-field position of a multileaf collimator with curved ends *Med. Phys.* 24 757-62

Brualla L, Salvat F and Palanco-Zamora R 2009 Efficient Monte Carlo simulation of multileaf collimators using geometry-related variance-reduction techniques *Phys Med Biol.* **54** 4131-49

Chin PW, Spezi E and Lewis DG 2003 Monte Carlo simulation of portal dosimetry on a rectilinear voxel geometry: a variable gantry angle solution *Phys. Med. Biol.* **48** N231–8

Chytyk K and McCurdy BM 2006 Investigation of tilted dose kernels for portal dose prediction in a-Si electronic portal imagers *Med. Phys.* **33** 3333–39

Chytyk K and McCurdy BM 2009 Comprehensive fluence model for absolute portal dose image prediction *Med. Phys.* **36** 1389–98Greer PB 2007 Off-axis dose response characteristics of an amorphous silicon electronic portal imaging device *Med. Phys.* **34** 3815–24

Cufflin RS, Spezi E, Millin AE and Lewis DG 2010 An investigation of the accuracy of Monte Carlo portal dosimetry for verification of IMRT with extended fields *Phys. Med. Biol.* **55** 4589–4600

De Smedt B, Fippel M, Reynaert N and Thierens H 2006 Denoising of Monte Carlo dose calculations: Smoothing capabilities versus introduction of systematic bias *Med. Phys.* **33** 1678-87

El Naqa I, Kawrakow I, Fippel M, Siebers J V, Lindsay P E, Wickerhauser M V, Vicic M, Zakarian K, Kauffmann N and Deasy J O 2005 A comparison of Monte Carlo dose calculation denoising techniques *Phys. Med. Biol.* **50** 909-22

Escobar M D and West M 1995 Bayesian Density Estiamtion and Inference Using Mixtures *J. Amer. Stat. Assoc*. **90** 577–88

Fippel M and Nüsslin F 2003 Smoothing Monte Carlo calculated dose distributions by iterative reduction of noise *Phys. Med. Biol.* **48** 1289–1304

Greer P B, Vial P, Oliver L and Baldock C 2007 Experimental investigation of the response of an amorphous silicon EPID to intensity modulated radiotherapy beams *Med. Phys.* **34** 4389-98

Greer P B, Cadman P, Lee C and Bzudsek K 2009 An energy fluence-convolution model for amorphous silicon EPID dose prediction *Med. Phys.* **36** 547-55

Hannah L, Blei D M and Powell W B 2011 Dirichlet process mixtures of generalized linear models *J. Mach. Learn. Res.* **12** 1923–53

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

Iori M, Cagni E, Paiusco M, Munro P and Nahum AE 2010 Dosimetric verification of IMAT delivery with a conventional EPID system and a commercial portal dose image prediction tool *Med Phys.* **37** 377-90

Ishwaran H and James LF 2001 Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.* **96** 161–173

Jarry G and Verhaegen F 2005 Electron beam treatment verification using measured andMonte Carlo predicted portal images *Phys. Med. Biol.* **50** 4977–94

Jarry G and Verhaegen F 2007 Patient-specific dosimetry of conventional and intensity modulated radiation therapy using a novel full Monte Carlo phase space reconstruction method from electronic portal images *Phys. Med. Biol.* **52** 2277–99

Kalli M, Griffin J E and Walker SG 2011 Slice Sampling Mixture Models, *Statistics and Computing*, **21**, 93–105

Kawrakow I 2002 On the denoising of MC calculated dose distribution *Phys. Med. Biol.* **47** 3087–103

Kim JO, Siebers JV, Keall PJ, Arnfield MR and Mohan R 2001 A Monte Carlo study of radiation transport through multileaf collimators *Med. Phys.* **28** 2497–2506

Klüter S, Sroka-Perez G, Schubert K and Debus J 2011 Leakage of the Siemens 160 MLC multileaf collimator on a dual energy linear accelerator *Phys. Med. Biol.* **56** N29-N37

Ko L, Kim JO and Siebers JV 2004 Investigation of the optimal backscatter for an aSi electronic portal imaging device *Phys. Med. Biol.* **49** 1723–38

Kirkby C and Sloboda R 2005 omprehensive Monte Carlo calculation of the point spread function for a commercial a-Si EPID *Med. Phys.* **32** 1115–27

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

Le Loirec C, Lazaro D, Garcia-Hernandez J-C, Guérin L, Batalla A and Poumarède B 2012 PENSSART, a new Monte Carlo system based on PENELOPE for dose calculations in radiotherapy: dosimetric evaluation submitted to *Phys. Med. Biol.*

Li W, Siebers JV and Moore JA 2006 Using fluence separation to account for energy spectra dependence in computing dosimetric a-Si EPID images for IMRT fields *Med. Phys.* **33**, 4468–80

Liu D C and Nocedal J 1989 On the Limited Memory Method for Large Scale Optimization, *Math. Prog. B* **45** (3) 503–28

McCurdy BM, Luchka K and Pistorius S 2001 Dosimetric investigation and portal dose image prediction using an amorphous silicon electronic portal imaging device *Med. Phys.* **28** 911–24

McCurdy BM and Greer PB 2009 Dosimetric properties of an amorphous-silicon EPID used in continuous acquisition mode for application to dynamic and arc IMRT *Med Phys.* **36** 3028-39

Moore JA and Siebers JV 2005 Verification of the optimal backscatter for an aSi electronic portal imaging device *Phys. Med. Biol.* **50** 2341–50

Müller P, Erkanli A and West M I K E 1996 Bayesian curve fitting using multivariate normal mixtures *Biometrika* **83** 67-79

Parent L, Seco J, Evans P M, Fielding A and Dance D R 2006 Monte Carlo modelling of a-Si EPID response: the effect of spectral variations with field size and position *Med. Phys.* **33** 4527–40

Prah DE, Kainz K, Peng C, Li XA 2011 The dosimetric and delivery advantages of a new 160-leaf MLC *Technol Cancer Res Treat.* **10** 219-29

Savitzky A and Golay M J E 1964 Smoothing and differentiation of data by simplified least squares procedures *Anal. Chem.* **36**(8) 1627-39

Salvat F, Fernandez-Varea JM and Sempau J 2006 Penelope 2006, a code system for Monte Carlo simulation of electron and photon transport OECD ISBN 92-64-02301-1 (Barcelona, Spain)

Schach von Wittenau AE, Logan CM, Aufderheide 3rd MB and Slone DM 2002 Blurring artifacts in megavoltage radiography with a flat-panel imaging system: Comparison of Monte Carlo simulations with measurements *Med. Phys.* **29** 2559-70

Sethuraman J 1994 A constructive definition of Dirichlet priors, *Stat. Sinica*, **4**, 639–650

Sheikh-Bagheri D and Rogers DW 2002 Monte Carlo calculation of nine megavoltage photon beam spectra using the BEAM code *Med. Phys.* **29** 391–402

Siebers J V, Kim J O, Ko L, Keall P J and Mohan R 2004 Monte Carlo computation of dosimetric amorphous silicon electronic portal images *Med. Phys.* **31** 2135-46

Spezi E and Lewis DG 2002 Full forward Monte Carlo calculation of portal dose from MLC collimated treatment beams *Phys. Med. Biol.* **47** 377–90

Steciw S, Warkentin B, Rathee S and Fallone BG 2005 Three-dimensional IMRT verification with a flat-panel EPID *Med. Phys.* **32** 600–12

Tacke MB, Nill S, Häring P and Oelfke U 2008 6 MV dosimetric characterization of the 160 MLC™, the new Siemens multileaf collimator *Med. Phys.* **35** 1634-42

Van Dyke J, Barnett R, Cygler J and Schragge P 1993 Commissioning and quality assurance of treatment planning computers *Int. J. Radiat. Oncol. Biol. Phys.* **26** 261-73

Van Elmpt WJ, Nijsten SMJJG, Mijnheer BJ and Minken AWH 2005 Experimental verification of a portal dose prediction model *Med. Phys.* **32** 2805–18

Van Elmpt WJ, Nijsten SMJJG, Schiffeleers RFH, Dekker ALAJ, Mijnheer BJ, Lambin P and Minken AWH 2006 AMonte Carlo based three-dimensional dose reconstruction method derived from portal dose images *Med. Phys.* **33** 2426–34

Van Elmpt WJ, McDermott L, Nijsten S, Wendling M, Lambin P and Mijnheer BJ 2008 A literature review of electronic portal imaging for radiotherapy *Radiother. Oncol.* **88** 289–309

Van Esch A, Depuydt T and Huyskens D P 2004 The use of an aSi-based EPID for routine absolute dosimetric pre-treatment verification of dynamic IMRT fields *Radiother. Oncol.* **71** 223–34

Vial P, Greer PB, Hunt P, Oliver L and Baldock C 2008 The impact of MLC transmitted radiation on EPID dosimetry for dynamic MLC beams *Med. Phys.* **35** 1267-77

Wang S, Gardner JK, Gordon JJ, Li W, Clews L, Greer PB and Siebers JV 2009 Monte Carlo-based adaptive dose kernel accounting for different field size responses of imagers *Med. Phys.* **36** 3582–95

Warkentin B, Steciw S, Rathee S and Fallone BG 2003 Dosimetric IMRT verification with a flat panel EPID *Med. Phys.* **30** 3143–55

Wendling M, Louwe RJW, McDermott LN, Sonke J-J, van Herk M and Mijnheer BJ 2006 Accurate two-dimensional IMRT verification using a back-projection EPID dosimetry method *Med. Phys.* **33** 259–73