Magnetic resonance-based computed tomography metal artifact reduction using Bayesian modelling

Nielsen, Jonathan Scharff; Edmund, Jens Morgenthaler; Van Leemput, Koen

PAPER • OPEN ACCESS

Magnetic resonance-based computed tomography metal artifact reduction using Bayesian modelling

View the article online for updates and enhancements.

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Magnetic resonance-based computed tomography metal artifact reduction using Bayesian modelling

Jonathan Scharff Nielsen[1,2], Jens Morgenthaler Edmund[2,3] and Koen Van Leemput[1,4]

[1] Department of Health Technology, Technical University of Denmark, 2820 Lyngby, Denmark
[2] Radiotherapy Research Unit, Department of Oncology, Gentofte and Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark
[3] Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark
[4] Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, United States of America

E-mail: jonathanscharff2@gmail.com

Abstract

Metal artifact reduction (MAR) algorithms reduce the errors caused by metal implants in x-ray computed tomography (CT) images and are an important part of error management in radiotherapy. A promising MAR approach is to leverage the information in magnetic resonance (MR) images that can be acquired for organ or tumor delineation. This is however complicated by the ambiguous relationship between CT values and conventional-sequence MR intensities as well as potential co-registration issues. In order to address these issues, this paper proposes a self-tuning Bayesian model for MR-based MAR that combines knowledge of the MR image intensities in local spatial neighborhoods with the information in an initial, corrupted CT reconstructed using filtered back projection. We demonstrate the potential of the resulting model in three widely-used MAR scenarios: image inpainting, sinogram inpainting and model-based iterative reconstruction. Compared to conventional alternatives in a retrospective study on nine head-and-neck patients with CT and T1-weighted MR scans, we find improvements in terms of image quality and quantitative CT value accuracy within each scenario. We conclude that the proposed model provides a versatile way to use the anatomical information in a co-acquired MR scan to boost the performance of MAR algorithms.

1. Introduction and purpose

1.1. Background

Medical x-ray computed tomography (CT) images of patients with metal implants often display major corruption from streak artifacts (Bal and Spies 2006, Jäkel and Reiss 2007), which affects both the visual quality of the images and the quantitative CT value accuracy. The latter is a potential hazard in radiotherapy (RT), where the CT values are used in treatment planning to provide electron density and relative stopping power estimates (Andersson *et al* 2014, Giantsoudi *et al* 2017). This is of particular concern in head-and-neck RT, where dental implants and fillings occur frequently and are simultaneously close to both the tumor site and critical organs. In this situation, metal artifact reduction (MAR) plays an important role in error management (Kim and Tomé 2007, Dietlicher *et al* 2014, Kwon *et al* 2015, Maerz *et al* 2015).

MAR is in general a difficult problem, as demonstrated by its approximately 40 year long history that has spawned numerous algorithmic approaches (Koehler *et al* 2011, Andersson *et al* 2014, Gjesteby *et al* 2016). A reason for this difficulty is the origin of the artifacts, which arise from multiple contributions that are amplified in the presence of metal (Beister *et al* 2012, Gjesteby *et al* 2016). Some of these contributions stem from incorrect assumptions in the CT reconstruction model that relates the image coefficients to the x-ray projection data (sinogram) (Goldman 2007, Buzug 2008). In particular, the mono-energetic approximation of the x-ray source spectrum in, e.g. the wide-spread filtered back projection (FBP) algorithm leads to incorrect modelling of the projections that are acquired through metal, and thus beam hardening artifacts (Stonestrom *et al* 1981, Hsieh

et al 2000). Other contributions are more model-independent, such as the photon starvation in the metal projections that leads to noise artifacts (Nuyts *et al* 1998, Gjesteby *et al* 2016).

MAR algorithms may be categorized into three overall approaches. *Image inpainting* algorithms replace corrupted CT values with better estimates by post-processing already reconstructed images (Lell *et al* 2013). *Sinogram inpainting* algorithms, on the other hand, replace metal projections by estimates that fit the reconstruction model to a higher degree, which may be particularly effective in dealing with photon starvation (Kalender *et al* 1987, Meyer *et al* 2009, Zhang *et al* 2011, Lell *et al* 2012, Meyer *et al* 2012). Finally, *model based iterative reconstruction* (MBIR) algorithms change the CT reconstruction model itself to a more complex probabilistic forward model that better accounts for the artifact sources, at the cost of having to optimize a generally non-linear image functional in a slow, iterative algorithm (De Man *et al* 2000, 2001, Elbakri and Fessler 2002, Thibault *et al* 2007, Hamelin *et al* 2008, Stayman *et al* 2012, Van Slambrouck and Nuyts 2012, Nuyts *et al* 2013, Tilley *et al* 2016, Jin *et al* 2015, Stille *et al* 2016, Fu *et al* 2017).

An important part of many MAR algorithms is the inclusion of prior information about the image that is being estimated. While this is especially true for image inpainting algorithms, which impose such information directly in image space, prior information is also used in some of the most successful sinogram inpainting algorithms. An example is the 'normalized MAR' (nMAR) method (Meyer *et al* 2009, 2012), which replaces metal projections by (scaled) projection estimates simulated from a template image. MBIR models also include prior information as they consist of two parts: a sinogram data likelihood that addresses, e.g. noise and beam hardening artifacts by modelling the detector noise, the x-ray source spectrum and the implant material; and an image prior distribution that may be used to guide the reconstruction with statistical knowledge about the image being reconstructed (De Man *et al* 2001, Elbakri and Fessler 2002, Hamelin *et al* 2008, Van Slambrouck and Nuyts 2012, Nuyts *et al* 2013, Tilley *et al* 2016, Fu *et al* 2017).

In a general CT setting, the quality of the available prior information is rather limited. In sinogram inpainting, for example, the required template is typically generated by post-processing a CT image reconstruction that is heavily corrupted by metal artifacts (Meyer *et al* 2009, Lell *et al* 2012, Meyer *et al* 2012). Being generated from corrupted data, the resulting prior information may itself be compromised, thereby introducing new artifacts in complex, highly corrupted regions such as the head and neck near the teeth and oral cavity. Similarly, in MBIR the limited availability of accurate prior information often motivates the use of relatively simple functional priors that merely impose mathematical regularities in the reconstructed image (De Man *et al* 2000, Webster Stayman and Fessler 2001, Pan *et al* 2010, Makeev and Glick 2013).

1.2. Contributions of this paper

In the specific context of RT, a potential source of prior information for improved MAR is the magnetic resonance (MR) image that is commonly co-acquired for tumor- and normal tissue delineation. Since metal artifacts are often more localized in MR compared to CT (Park *et al* 2015, Bell and Bashir 2018), the MR scan can provide superior anatomical prior information in regions that are heavily corrupted in CT (Jonsson *et al* 2010, Bell and Bashir 2018). Furthermore, since co-registration and acquisition of the MR and CT scans in the same patient fixation is already part of the tumor-delineation process, this can be done with minimal interruption to the existing clinical workflow.

Despite the obvious potential, to the best of our knowledge only a few prior attempts have been made to use MR for reducing CT metal artifacts, most notably the image inpainting algorithms described in Anderla *et al* (2013), Delso *et al* (2013), Park *et al* (2015). The principal difficulty faced by such MR-based approaches to CT MAR lies in the image contrast disparities between the two modalities, especially between bone and air, which are easily distinguishable in the CT but not in the MR scan unless dedicated sequences are used (Delso *et al* 2013). Additional difficulties arise from co-registration issues, mainly due to to inter-acquisitional motion, which limit the accuracy of MR-based predictions.

In order to overcome these difficulties, we propose a novel Bayesian algorithm to predict uncorrupted CT values from a combination of corrupted CT measurements and corresponding MR intensities. The contributions of this work can be summarized as follows:

1. Due to the contrast disparity between MR and CT, single-voxel MR intensities are poor CT value predictors. In contrast to prior work, our model therefore relies on MR image *patches* (Andreasen *et al* 2015, 2016) instead, i.e. collections of MR intensities taken from local spatial neighborhoods that encode higher-order anatomical features.

2. In contrast to existing methods that simply replace CT values that are deemed to be corrupt with entirely new values (Anderla *et al* 2013, Delso *et al* 2013, Park *et al* 2015), our method automatically blends MR-based predictions with the original, corrupted CT values. As we demonstrate experimentally, this way of retaining the information of the original CT helps to discern bone from air as well as to address co-registration issues.

3. Rather than directly targeting only one particular MAR strategy such as image inpainting (Anderla *et al* 2013, Delso *et al* 2013, Park *et al* 2015), the predictive model developed here can be used flexibly within several widely-used MAR scenarios. In particular, we demonstrate experimentally the benefit of using the proposed model to directly calculate blended CT value replacements and thus perform image inpainting; to generate a template for sinogram inpainting; and to define an image reconstruction prior for MBIR.

4. In order to facilitate clinical adoption, our model uses data from only the patient under consideration and automatically tunes its parameters to fit this specific patient. We demonstrate how the model thereby accounts for variations in patient-specific features, such as the severity of the artifact corruption as well as variations in the MR sequence parameters. It also removes the need for external training data of well-aligned MR and CT image pairs that is commonly required for MR-based CT prediction in the literature (Dowling *et al* 2012, Andreasen *et al* 2015, Torrado-Carvajal *et al* 2015, Han 2017, Torrado-Carvajal *et al* 2019).

The paper is structured as follows. We first derive and discuss the proposed predictive model, cover the automatic parameter estimation and present three MAR algorithms that are based on this model (section 2 'Methods'). Next, (in section 3 'Experiments') we describe the practical implementation of the three MAR algorithms, and benchmark their performance in comparison to conventional alternatives; the results are afterwards presented in sequence for the three MAR algorithms (section 4 'Results'). We end with a final summary of our findings and main conclusions, followed by a discussion of, in particular, the potential of clinical implementation of our methods (section 5 'Conclusion and Discussion').

An early version of our approach, with only qualitative (visual) results on a small set of patients, was previously presented (Nielsen *et al* 2018). The current manuscript provides substantially more technical derivations and analysis, in particular putting more focus on the self-tuning of the model parameters; proposes a novel sinogram inpainting variant as well as an improved MBIR algorithm; and contains detailed quantitative results on a larger set of patients.

2. Methods

2.1. Predictive model

Let a set of voxels covering a patient volume be assigned indices i from the index set $\mathcal{T}$ ($i \in \mathcal{T}$), and denote the voxel center locations as $\{\mathbf{x}_i\}_{i \in \mathcal{T}}$. To avoid cluttered notation, we will suppress the index set and use $\{\cdot\} \equiv \{\cdot\}_{i \in \mathcal{T}}$ in the remaining. Consider now a CT volume reconstructed using FBP (Buzug 2008), with per-voxel FBP values $\{t_i\}$. Assuming the availability of an MR image that is co-registered with the FBP reconstruction, we extract a set of small cuboidal MR volumes ('patches') centered at $\{\mathbf{x}_i\}$, with intensities stacked in the vectors $\{\mathbf{m}_i\}$. The values in $\{t_i\}$ are potentially corrupted by metal artifacts, and our task is to estimate the underlying true CT values, $\{y_i\}$, given the observation of $\{t_i, \mathbf{m}_i\}$.

To accomplish this, we infer the *posterior predictive distribution* $p(\{y_i\}|\{t_i, \mathbf{m}_i\}, \boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ denotes the set of model parameters, from a generative, probabilistic model of the corrupted CT values contingent on the observed MR scan. We specifically model the two-step process by which the uncorrupted CT values are first sampled from a prior distribution, assumedly in an independent fashion for each voxel:

$$p(\{y_i\}|\{\mathbf{m}_i\}, \boldsymbol{\sigma}) = \prod_{i \in \mathcal{T}} p(y_i|\mathbf{m}_i, \boldsymbol{\sigma}),$$

and then corrupted by sampling from a noise model $p(\{t_i\}|\{y_i\})$, again independently for each voxel:

$$p(\{t_i\}|\{y_i\}, \boldsymbol{\sigma}) = \prod_{i \in \mathcal{T}} p(t_i|y_i, \boldsymbol{\sigma}).$$

Under this generative model, applying Bayes' rule yields:

$$p(\{y_i\}|\{t_i, \mathbf{m}_i\}, \boldsymbol{\sigma}) = \prod_{i \in \mathcal{T}} p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma}), \tag{1}$$

$$\text{where} \quad p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma}) = Z^{-1} p(t_i|y_i, \boldsymbol{\sigma}) \, p(y_i|\mathbf{m}_i, \boldsymbol{\sigma}) \tag{2}$$

$$\text{with} \quad Z = \int_{-\infty}^{\infty} p(t_i|y_i, \boldsymbol{\sigma}) \, p(y_i|\mathbf{m}_i, \boldsymbol{\sigma}) dy_i. \tag{3}$$

The predictive distribution of equation (1) depends on an FBP-dependent likelihood function $p(t_i|y_i, \boldsymbol{\sigma})$ and an MR-determined prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$, which we detail below.
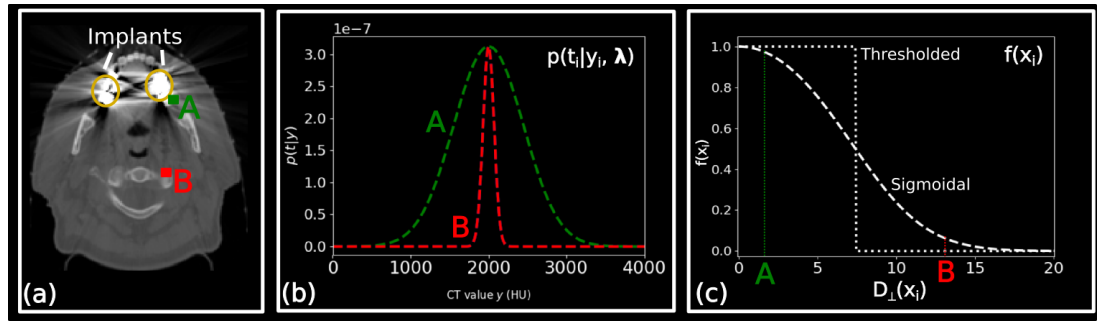
Figure 1. Illustrations of the artifact noise model. (a) Two points, A and B, at different distances to the metal implants. (b) The Gaussian artifact noise models for points A and B. (c) The sigmoidal function used to scale the variance with distance to the metal. Also shown is a thresholding at $f = 0.5$, useful for automatic distinction of corrupted and uncorrupted regions.

2.1.1. Likelihood

The likelihood $p(t_i|y_i, \boldsymbol{\sigma})$ models the distribution of the corrupted CT value given the underlying true CT value. Introducing the maximal noise variance σ_t^{*2}, we model the artifact corruption as additive Gaussian noise with variance σ_t^2 smoothly decreasing with the distance to the implants, such that it is equal to σ_t^{*2} within the implants but 0 far away where the FBP is free of artifacts. In particular:

$$p(t_i|y_i, \boldsymbol{\sigma}) = \mathcal{N}(t_i|y_i, \sigma_t^2) \quad \text{with} \quad \sigma_t^2 = \sigma_t^{*2} f(\mathbf{x}_i) \tag{4}$$

$$\text{and} \quad f(\mathbf{x}_i) = 1 + \tanh(-\frac{D_\perp^2(\mathbf{x}_i)}{\kappa}). \tag{5}$$

Here, $D_\perp(\mathbf{x}_i)$ is the perpendicular spatial distance to a segmentation of the metal implants; $\mathcal{N}(\cdot|\psi, \nu^2)$ denotes a Gaussian with mean ψ and variance ν^2; and κ sets the decrease rate of the sigmoidal function $f(\mathbf{x}_i)$. As an illustration, consider the two voxels at different distances to the metal in figure 1(a), for which the noise models (figure 1(b)) have different widths due to the modulation with $f(\mathbf{x}_i)$ (figure 1(c)).

The metal segmentation and κ parameter must be defined by the user; section 3.2 will present the definitions used for our experiments. The maximal artifact noise variance σ_t^{*2} will however be chosen automatically given the observed data, as described in section 2.3.

2.1.2. Prior

Letting $\mathcal{T}_u \subseteq \mathcal{T}$ denote an assumed *uncorrupted* part of the patient volume (see figure 2(a)), we learn the prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$ from all matched CT/MR pairs $\{y_n, \mathbf{m}_n\}_{n \in \mathcal{T}_u}$ in $\mathcal{T}_u$ using kernel regression (Bishop 2006). In particular, we estimate the joint distribution $p(y_i, \mathbf{m}_i|\boldsymbol{\sigma})$ with kernel density estimation (Bishop 2006) using Gaussian kernels with diagonal covariance matrices:

$$\begin{aligned} &p(y_i, \mathbf{m}_i|\boldsymbol{\sigma}) \\ &= |\mathcal{T}_u|^{-1} \sum_{n \in \mathcal{T}_u} \mathcal{N}(y_i|y_n, \sigma_y^2)\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \sigma_m^2 \mathbf{I}_M), \end{aligned} \tag{6}$$

and obtain the prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$ accordingly:

$$p(y_i|\mathbf{m}_i, \boldsymbol{\sigma}) = \frac{p(y_i, \mathbf{m}_i|\boldsymbol{\sigma})}{p(\mathbf{m}_i|\boldsymbol{\sigma})} = \sum_{n \in \mathcal{T}_u} w_n^i \mathcal{N}(y_i|y_n, \sigma_y^2) \tag{7}$$

$$\text{with} \quad w_n^i = \frac{\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \sigma_m^2 I_M)}{\sum_{n' \in \mathcal{T}_u} \mathcal{N}(\mathbf{m}_i|\mathbf{m}_{n'}, \sigma_m^2 I_M)}. \tag{8}$$

Here, $\mathcal{N}(\cdot|\psi, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian with mean ψ and covariance matrix $\boldsymbol{\Sigma}$, $\mathbf{I}_M$ is the identity matrix of dimension M (the number of voxels in a patch) and $|\cdot|$ denotes set cardinality. σ_y^2 and σ_m^2 are the kernel variances, which, together with the artifact noise variance σ_t^{*2} (see section 2.1), will be automatically estimated given the observed data as described in section 2.3. We obtain the required uncorrupted part of the patient volume $\mathcal{T}_u$ (as well as its complement $\mathcal{T}_c$) by thresholding $f(\mathbf{x}_i)$ as shown in figure 1(c), leading to the following sets:

$$\mathcal{T}_u \equiv \{i \in \mathcal{T}|f(\mathbf{x}_i) \leqslant 0.5 \quad \text{and} \quad \mathcal{T}_c \equiv \mathcal{T} \setminus \mathcal{T}_u, \tag{9}$$

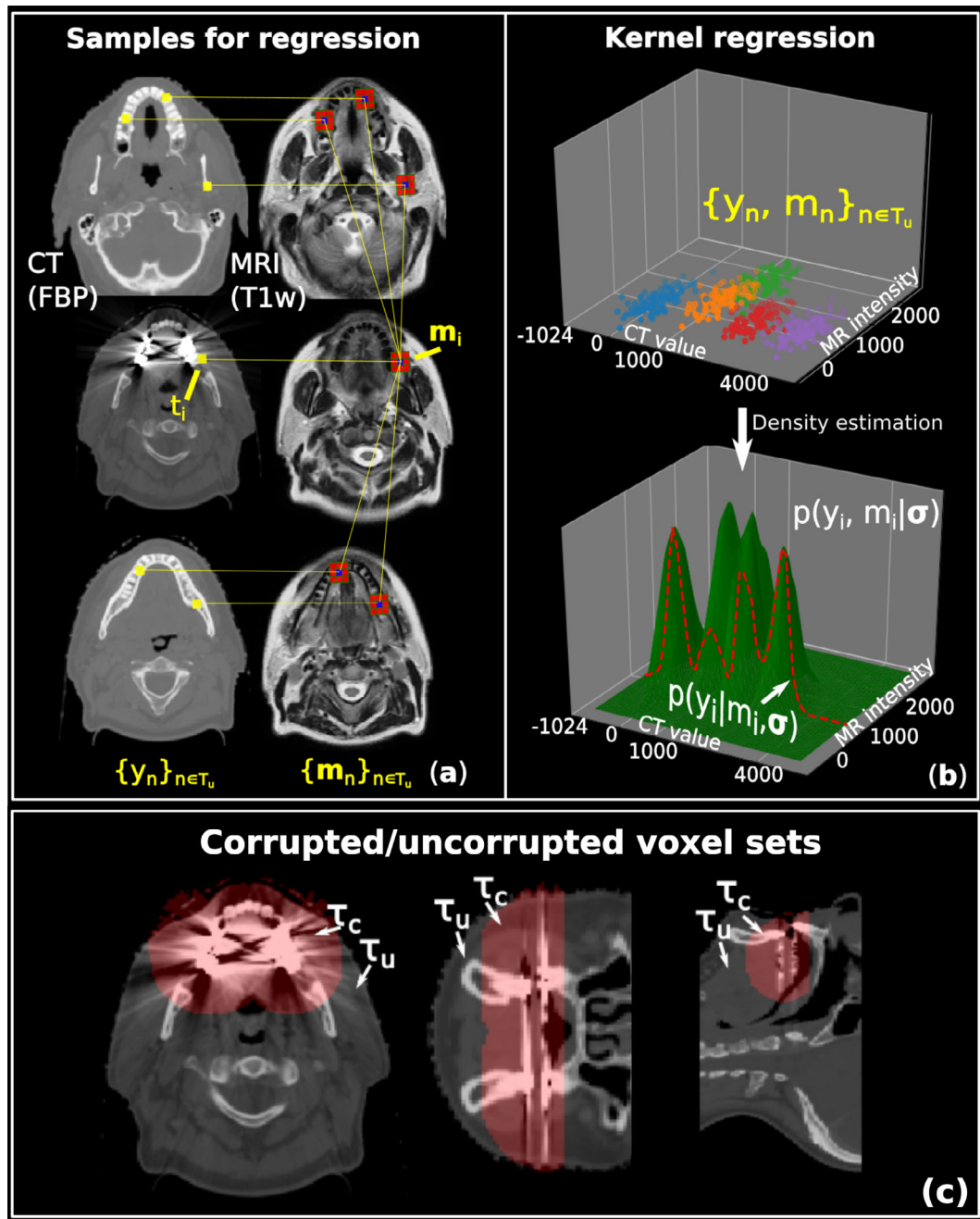which are illustrated in figure 2(c).

Figure 2. Illustration of the kernel regression on an uncorrupted part of the patient volume. (a) Illustration of the regression points $\{y_n, \mathbf{m}_n\}_{n\in\mathcal{T}_u}$; shown are a few particularly good matches that would have a large weight in the prior model. (b) On the regression point set, *kernel density estimation* is used to define the joint distribution $p(y_i, \mathbf{m}_i|\boldsymbol{\sigma})$ (shown as a surface for $1 \times 1 \times 1$ patches). The red curve corresponds to $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$, results from *kernel regression* and is a normalized trace on the kernel density estimate surface at a specific $\mathbf{m}_i$. (c) Illustration of the corrupted voxel set $\mathcal{T}_c$ (equation (9)), with the uncorrupted set $\mathcal{T}_u$ as its complement.

The kernel regression process is illustrated in figure 2(b) for the special case where 1x1x1 patches are used. After estimating $p(y_i, \mathbf{m}_i|\boldsymbol{\sigma})$ using uncorrupted CT/MR samples $\{y_n, \mathbf{m}_n\}$, the prior distribution $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$ corresponds to drawing a trace at the observed $\mathbf{m}_i$ and normalizing, leading to a one-dimensional Gaussian mixture model that displays several peaks along the CT-axis. The mixture contains a large number of Gaussians, i.e. one for every sample in $\{y_n, \mathbf{m}_n\}_{n\in\mathcal{T}_u}$. Since the samples in practice are clustered corresponding to tissue types, as illustrated in figure 2(b), the effective number of modes is however automatically reduced. In this way, kernel regression automatically performs an implicit tissue classification and reflects it in the prior model.

The expectation value of $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$ presents a way to calculate a (Bayesian) prior estimate of the uncorrupted CT:

$$\hat{y}_i^{\text{pCT}} \equiv \int_{-\infty}^{\infty} y_i p(y_i|\mathbf{m}_i, \boldsymbol{\sigma}) dy_i = \sum_{n\in\mathcal{T}_u} w_n^i y_n, \quad \forall i \in \mathcal{T}. \tag{10}$$

This straighforward *pseudo-CT* (Andreasen *et al* 2015, Hofmann *et al* 2008) (pCT) estimate, which is dependent only on the MR scan, will serve as a baseline for our experiments.

2.2. Posterior predictive distribution

Using the prior and likelihood that we have now defined, the posterior in equation (1) may be calculated. Defining the parameter set $\boldsymbol{\sigma} \equiv \{\sigma_t^*, \sigma_y, \sigma_m\}$ and plugging in equations (7) and (4):

$$p(y_i|t_i, \mathbf{m}_n, \boldsymbol{\sigma}) = Z^{-1} \sum_{n \in \mathcal{T}_u} w_n^i \mathcal{N}(y_i|y_n, \sigma_y^2) \mathcal{N}(t_i|y_i, \sigma_t^2). \tag{11}$$

Using that:

$$\mathcal{N}(t_i|y_i, \sigma_t^2) \mathcal{N}(y_i|y_n, \sigma_y^2)$$
$$= \mathcal{N}(y_i|\mu_n^i, (\sigma_t^{-2} + \sigma_y^{-2})^{-1}) \mathcal{N}(t_i|y_n, \sigma_t^2 + \sigma_y^2) \tag{12}$$

with

$$\mu_n^i \equiv \frac{1}{1 + \frac{\sigma_t^2}{\sigma_y^2}} t_i + \frac{1}{1 + \frac{\sigma_y^2}{\sigma_t^2}} y_n, \tag{13}$$

the normalizing factor (equation (3)) becomes:

$$Z = \sum_{n' \in \mathcal{T}_u} w_n^i \mathcal{N}(t_i|y_n, \sigma_t^2 + \sigma_y^2)$$

and we get:

$$p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma}) = \sum_{n \in \mathcal{T}_u} v_n^i \mathcal{N}(y_i|\mu_n^i, (\sigma_y^{-2} + \sigma_t^{-2})^{-1}), \tag{14}$$

with weights:

$$v_n^i \equiv \frac{w_n^i \mathcal{N}(t_i|y_n, \sigma_t^2 + \sigma_y^2)}{\sum_{n' \in \mathcal{T}_u} w_{n'}^i \mathcal{N}(t_i|y_{n'}, \sigma_t^2 + \sigma_y^2)}. \tag{15}$$

Equation (14) is our main result. We will later use it to calculate a Bayesian estimate of $\{y_i\}$ via its expectation:

$$\hat{y}_i \equiv \int_{-\infty}^{\infty} y_i p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma}) dy_i = \sum_{n \in \mathcal{T}_u} v_n^i \mu_n^i, \quad \forall i \in \mathcal{T}, \tag{16}$$

which will be used in the proposed MAR methods.

2.2.1. Interpreting the posterior

The means and weights $\{\mu_n^i, v_n^i\}_{n \in \mathcal{T}_u}$ in the Gaussian mixture of equation (14) are similar to the ones in the prior model of equation (7), but are now determined by both the potentially corrupted FBP value t_i and the regression points $\{y_n, \mathbf{m}_n\}_{n \in \mathcal{T}_u}$. How the two are blended is determined by the exact values of the parameters $\boldsymbol{\sigma}$, which are therefore critical for the shape of the posterior.

Kernel regression parameters

The influence of the kernel regression variances σ_y^2 and σ_m^2 is isolated to the prior $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$, and is as illustrated in figure 3(a).

The σ_y^2 parameter controls the width of the Gaussians in the mixture of equation (7). This affects the modelled CT value variance within the implicit tissue classes defined by the kernel regression, which is reflected by the width of the peaks in the prior.

The σ_m^2 parameter scales the squared MR image patch differences in the Gaussian weights w_n^i, such that decreasing it translates to more weight on a smaller set of Gaussians with smaller $\|\mathbf{m}_i - \mathbf{m}_n\|$. Compared to larger settings (figure 3(a), right), a small value (figure 3(a), left) means that only patches in uncorrupted areas that have very similar intensity profiles contribute to the prediction, which shows as an amplification of the corresponding peaks in the prior.

Artifact noise variance

The influence of σ_t^{*2} is isolated to the likelihood $p(t_i|y_i, \boldsymbol{\sigma})$ and governs the expected level of corruption near the metal voxels (near the metal, $\sigma_t^2 \approx \sigma_t^{*2}$). For a given voxel, σ_t^2 directly determines the width of the likelihood and thus how the prior is modified to yield the posterior. Typical values of σ_t^2 lead to posteriors such as those in figure 3(b), where peaks near the observed FBP value t_i are amplified.
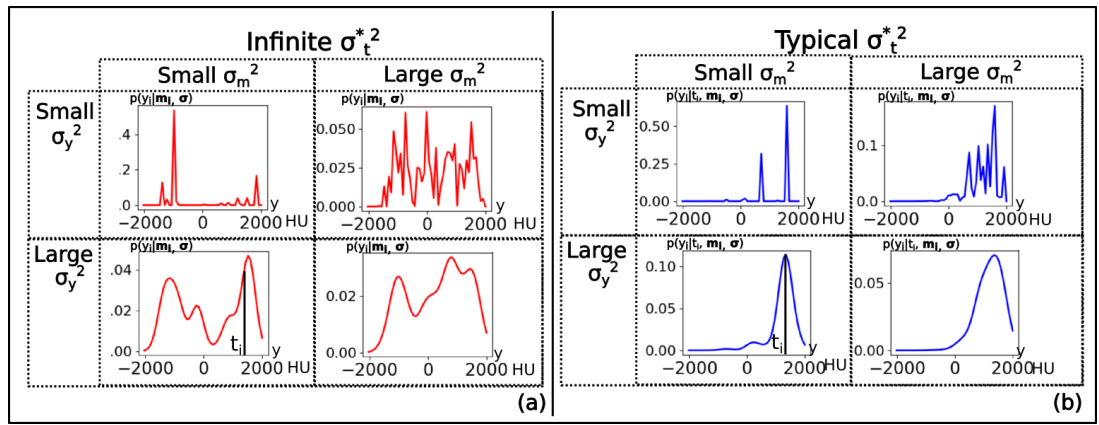
Figure 3. Influence of the parameter settings on the posterior for a single voxel. Each plot is for a different combination of the CT and MR image kernel variances σ_y^2 and σ_m^2, whose values increase respectively from top to bottom and left to right. The FBP value in the voxel, t_i, is indicated. (a) The posterior for $\sigma_t^{*2} \to \infty$, corresponding to the MR-based prior model; (b) the posterior for an appropriately chosen σ_t^{*2} (see section 2.3).

In the special case where σ_t^2 is very large, however, the likelihood becomes effectively a constant, so that the prior dominates the estimate and $\mu_n^i \approx y_n^i$, $v_n^i \approx w_n^i$, and the estimate in equation (16) therefore approaches the pCT in equation (10) ($\hat{y}_i \approx \hat{y}_i^{\text{pCT}}$). The purely MR-based pCT thus arises as a special case of equation (16).

In the other extreme where $\sigma_t^2 \to 0$, the likelihood becomes a sharp function, while $\mu_n^i \to t_i$ and $\hat{y}_i \to t_i$. In that scenario the estimate therefore simply copies the FBP without reference to the MR scan. The FBP is thus another special case of equation (16).

2.3. Automatic parameter tuning

2.3.1. Parameter tuning

As we have seen, the $\boldsymbol{\sigma}$ parameters critically alter the shape of the posterior as well as the predictions derived from it, and so it is important that they are chosen appropriately. The observed data themselves may be used as a guide for this process; for instance, a highly corrupted FBP would suggest a large σ_t^{*2} while a certain degree of CT value variation between voxels containing the same tissue type would suggest a certain setting for σ_y^2. We will now use a Bayesian method to automatically pick the parameters that best explain the observed data given our probabilistic model.

Equations (6) and (4) together define a joint distribution over the observed data $\{t_i, \mathbf{m}_i\}$ together with the unobserved variables $\{y_i\}$:

$$p(\{t_i, \mathbf{m}_i, y_i\}|\boldsymbol{\sigma}) = \prod_{i \in \mathcal{T}} p(t_i, \mathbf{m}_i, y_i|\boldsymbol{\sigma})$$

$$\text{with} \quad p(t_i, \mathbf{m}_i, y_i|\boldsymbol{\sigma}) = p(y_i, \mathbf{m}_i|\boldsymbol{\sigma})\, p(t_i|y_i, \boldsymbol{\sigma}). \tag{17}$$

Marginalizing equation (17) over the unobserved variables $\{y_i\}$, using equations (6) and (4) for the components, yields

$$p(\{t_i, \mathbf{m}_i\}|\boldsymbol{\sigma}) = \prod_{i \in \mathcal{T}} p(t_i, \mathbf{m}_i|\boldsymbol{\sigma}) \quad \text{with}$$

$$p(t_i, \mathbf{m}_i|\boldsymbol{\sigma}) = \int_{-\infty}^{\infty} p(t_i, \mathbf{m}_i, y_i|\boldsymbol{\sigma})\, dy_i$$

$$= |\mathcal{T}_u|^{-1} \sum_{n \in \mathcal{T}_u} \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \sigma_m^2 \mathbf{I}_M)$$

$$\left[\int_{-\infty}^{\infty} \mathcal{N}(t_i|y_i, f(\mathbf{x}_i)\sigma_t^{*2})\mathcal{N}(y_i|y_n, \sigma_y^2)\, dy_i \right]$$

$$= |\mathcal{T}_u|^{-1} \sum_{n \in \mathcal{T}_u} \mathcal{N}(t_i|y_n, f(\mathbf{x}_i)\sigma_t^{*2} + \sigma_y^2)\mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \sigma_m^2 \mathbf{I}_M). \tag{18}$$

In the last step, we used equation (12) to carry out the integral. Given the observed variables $\{t_i, \mathbf{m}_i\}$, appropriate parameter values $\boldsymbol{\sigma}$ may now be obtained by maximizing equation (18), effectively fitting the model to the available data.

We perform the optimization using an iterative expectation-maximization (EM) algorithm (Dempster *et al* 1977) that only requires closed-form updates. Starting from an initial setting of $\boldsymbol{\sigma}$, EM alternately

applies an expectation (E) and a maximization (M) step. In the E-step, a lower bound to the objective function $\log p(\{t_i, \mathbf{m}_i\}|\boldsymbol{\sigma})$ is constructed given the current estimate of $\boldsymbol{\sigma}$, derived directly from Jensen's inequality (Borman 2009):

$$
\begin{aligned}
\log p(\{t_i, \mathbf{m}_i\}|\boldsymbol{\sigma}) &= \sum_{i \in \mathcal{T}} \log \sum_{n \in \mathcal{T}_u} \phi_n^i \\
&= \sum_{i \in \mathcal{T}} \log \sum_{n \in \mathcal{T}_u} v_n^i \left(\frac{\phi_n^i}{v_n^i} \right) \\
&\geqslant \underbrace{\sum_{i \in \mathcal{T}} \sum_{n \in \mathcal{T}_u} v_n^i \log \left(\frac{\phi_n^i}{v_n^i} \right)}_{(*)},
\end{aligned}
$$

where v_n^i are the weights in equation (15) and we defined:

$$
\phi_n^i = |\mathcal{T}_u|^{-1} \mathcal{N}(t_i|y_n, f(\mathbf{x}_i)\sigma_t^{*2} + \sigma_y^2) \mathcal{N}(\mathbf{m}_i|\mathbf{m}_n, \sigma_m^2 \mathbf{I}_M).
$$

The lower bound $(*)$ is, by design, equal to $\log p(\{t_i, \mathbf{m}_i\}|\boldsymbol{\sigma})$ at the current parameter estimate, so when it is maximized during the M-step, the objective is guaranteed to increase (Dempster *et al* 1977). In general, the non-analytic $f(\mathbf{x}_i)$ prevents a closed-form formulation of this maximization, and so we simplify the objective by thresholding at $f = 0.5$, as illustrated in figure 1[5]. This removes $f(\mathbf{x}_i)$ from the equations while splitting $\mathcal{T}$ into the corrupted and uncorrupted sets $\mathcal{T}_c$ and $\mathcal{T}_u$, leading to a closed-form M-step and algorithm 1:

Algorithm 1. Automatic parameter estimation.

1: Choose an initial estimate of the parameters
   $(\boldsymbol{\sigma} \leftarrow \{10^6, 10^6, 10^6\})$, and set $\delta \leftarrow 1$.

2: **while** $\delta > 10^{-3}$ **do**

3:     $\boldsymbol{\sigma}_0 \leftarrow \boldsymbol{\sigma}$

4:     **E-step**: Calculate v_n^i, $\forall i \in \mathcal{T}$ and $\forall n \in \mathcal{T}_u$,
   using equation (15).

5:     **M-step**: Update the parameter estimates:

6:
   $$
   \begin{aligned}
   [\sigma_y]^2 &\leftarrow \frac{1}{|\mathcal{T}_u|} \sum_{i \in \mathcal{T}_u} \sum_{n \in \mathcal{T}_u} v_n^i (t_i - y_n)^2 \\
   [\sigma_t^*]^2 &\leftarrow \frac{1}{|\mathcal{T}_c|} \sum_{i \in \mathcal{T}_c} \sum_{n \in \mathcal{T}_u} v_n^i (t_i - y_n)^2 - [\sigma_y]^2 \\
   [\sigma_m]^2 &\leftarrow \frac{1}{|\mathcal{T}|} \sum_{i \in \mathcal{T}} \sum_{n \in \mathcal{T}_u} v_n^i \frac{\|(\mathbf{m}_i - \mathbf{m}_n)\|^2}{M}.
   \end{aligned}
   $$

7:     $\delta \leftarrow \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_0\| / 3$

8: **end while**

The calculation of v_n^i in the E-step may be viewed as a probabilistic assignment of the data points to the tissue classes that were implicitly defined during kernel regression. Given this classification, the update equations in the M-step estimate the within-class variance, over different parts of the patient volume:

- σ_y^2 is calculated over the uncorrupted set $\mathcal{T}_u$, which reflects the observed noise in the CT values within a tissue class in the absence of the metal artifacts.
- σ_t^{*2} updates to the additional variance in the metal artifact corrupted volume $\mathcal{T}_c$, reflecting the level of artifact corruption.
- σ_m^2 updates to the MR image patch variance over the entire volume.

During the iterations, the update of the parameters in the M-step improves the soft classification in the E-step as the model increasingly fits the data, which in turn improves the parameter estimates. This continues until the objective is maximized and the parameters stop changing.

2.4. Application to MAR: three algorithms

We propose to use the posterior predictive distribution $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma})$, specified in equation (14) and with parameters $\boldsymbol{\sigma}$ automatically tuned using algorithm 1, as the basis for three different MAR algorithms. Specifically, we define an image inpainting method that simply uses the mean of the predictive distribution; a sinogram

---

[5] Since $f(\mathbf{x}_i)$ displays a relatively sharp drop-off, this approximation does not notably change the results, as we also verified experimentally.

inpainting method that uses the image inpainting result as a template to estimate the metal projections; and an MBIR algorithm that uses the full distribution $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma})$ as a reconstruction prior together with a Poisson likelihood model of the x-ray intensity measurements.

2.4.1. Image inpainting

The image inpainting method that we propose, which we will call *kernel regression MAR* (kerMAR) in the remainder, directly calculates a CT estimate as the mean of $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma})$, which is given by equation (16), with one exception: the metal implants, as segmented for the definition of $f(\mathbf{x}_i)$ (see section 2.1), are left untouched. For a comparative evaluation of our model, we also define a similar algorithm that we call pCT, which instead uses the mean of $p(y_i|\mathbf{m}_i, \boldsymbol{\sigma})$, given by equation (10). This latter algorithm is entirely MR-dependent, effectively ignoring the information in the available FBP reconstruction.

2.4.2. Sinogram inpainting

The sinogram inpainting method we propose builds upon the nMAR algorithm (Meyer *et al* 2009, 2012), which uses simulations on a template image to replace the metal projections. Rather than directly replacing the metal projections, nMAR smoothly integrates the simulated metal projections in the sinogram as follows: the ratio of the original sinogram and the simulation is calculated. In the 'sinogram' of ratios, the metal projections, labelled by a similar projection simulation on a binary mask derived from the metal implant segmentation, are then linearly interpolated between the nearest-lying values. This provides a set of weights that reflect the deviations between the simulation and original. These are used to weigh the simulated metal projections to smooth the transition over the metal implants between the original projections and the simulated ones. The resulting, inpainted sinogram is finally used for a reconstruction with FBP, and the image is post-processed: in particular, since the template does not in general reproduce the metal implants accurately, their contribution to the metal projections is not accurately simulated, leading to errors near and in the metal implants. The implant segmentation is therefore used to reintroduce the original metal CT values, upon which a frequency split (Meyer *et al* 2012) is performed to preserve high frequency information (details) near the implants.

Our MR-based method, nMAR-k, simply uses our inpainted kerMAR image as the template within the established nMAR framework. For comparison, we also apply the nMAR algorithm with a conventionally generated FBP-derived template, calculated by thresholding using K-means clustering (Bishop 2006) on an initial linear interpolation MAR (liMAR) (Kalender *et al* 1987) image, followed by bulk CT value assignment (Meyer *et al* 2009, Lell *et al* 2012, Meyer *et al* 2012).

2.4.3. MBIR

MBIR maximizes the posterior distribution of a CT reconstruction given the acquired x-ray intensity data and a reconstruction prior, for which we propose to use our predictive distribution:

$$p(\{y_i\}|\{n_j\}) \propto p(\{n_j\}|\{y_i\}) \prod_{i \in \mathcal{T}} p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma}), \tag{19}$$

where $p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma})$ is given by equation (14). Here $\{n_j\}$ is the set of x-ray intensity measurements for difference paths through the patient volume. Following (Nuyts *et al* 1998), we use a Poisson likelihood:

$$p(\{n_j\}|\{y_i\}) = \prod_j p(n_j|\{y_i\}),$$

$$p(n_j|\{y_i\}) = \mathsf{Poisson}\{\lambda_j\}, \quad \lambda_j = \Gamma_j e^{-\sum_{i \in \mathcal{T}} l_{j,i} y_i},$$

where Γ_j is the emitted x-ray count toward detector j, and $\mathbf{L}$ is the *system matrix* whose entry $l_{j,i}$ defines the intersection between the x-ray path j and voxel i. To maximize the reconstruction posterior in equation (19) we use the iterative MLTR algorithm (Nuyts *et al* 1998, Van Slambrouck and Nuyts 2012) in its extended form De Man *et al* (2000), De Man (2001), which allows for a general reconstruction prior. Starting from an initial reconstruction estimate, the algorithm maximizes the log-posterior by iteratively applying an additive term. In addition to the x-ray data likelihood, this term depends on the first and second derivatives of the log-prior (De Man *et al* 2000, De Man 2001), which in our case become

$$\ln\big(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma})\big)' = (\sigma_t^{-2} + \sigma_y^{-2})\Big[\sum_{n \in \mathcal{T}_u} \tilde{v}_n^i \mu_n^i - y_i\Big] \quad \text{and} \tag{20}$$

$$\ln\big(p(y_i|t_i, \mathbf{m}_i, \boldsymbol{\sigma})\big)''$$
$$= (\sigma_t^{-2} + \sigma_y^{-2})^2 \Big(\sum_{n \in \mathcal{T}_u} \tilde{v}_n^i {\mu_n^i}^2 - \big[\sum_{n \in \mathcal{T}_u} \tilde{v}_n^i \mu_n^i\big]^2\Big) - (\sigma_t^{-2} + \sigma_y^{-2}) \tag{21}$$

$$\text{where} \quad \tilde{v}_n^i = \frac{v_n^i \mathcal{N}(y_i | \mu_n^i, (\sigma_t^{-2} + \sigma_y^{-2})^{-1})}{\sum_{n' \in \mathcal{T}_u} v_{n'}^i \mathcal{N}(y_i | \mu_{n'}^i, (\sigma_t^{-2} + \sigma_y^{-2})^{-1})}. \tag{22}$$

We will refer to the resulting reconstruction algorithm as MLTR-k.

3. Experiments

3.1. Materials

We evaluated the three proposed algorithms on an anonymized retrospective data set of nine head-and-neck radiotherapy patients containing dental implants and/or fillings. The image sets had a resolution of $1.2 \times 1.2 \times 2.0$ mm (CT) and $0.5 \times 0.5 \times 5.5$ mm (MR). The MR images were resampled to the CT resolution after rigid, multi-modal co-registration by mutual information (Wells *et al* 1996, Maes *et al* 1997) using the MatLab (version R2019a) image processing toolbox (Mathworks 2019). The CT scanner was a Philips Brilliance Big Bore with kVp120, the MR scanner a Philips Panorama 1.0T HFO. The patients were MR scanned with a T1-weighting 2D spin-echo sequence at TE $=10$ ms and TR $=520.2$–572.2 ms (varying between scans). The patients were positioned in the same fixation for both MR and CT scans. For MBIR and sinogram inpainting, we exported the sinograms from the CT scanner.

3.2. Practical implementation

A generic procedure for implementing the three MAR algorithms is described in algorithm 2:

Algorithm 2. Summary of implementation.

1: Calculate the FBP.

2: Segment the metal implants.

3: Calculate $f(\mathbf{x}_i)$ according to equation (5).

4: Obtain the uncorrupted set $\mathcal{T}_u$ by thresholding

  $f(\mathbf{x}_i)$ at 0.5 (see equation (9)).

5: Estimate $\boldsymbol{\sigma}$ using algorithm 1, then calculate the

  weights $\{w_n^i\}_{n \in \mathcal{T}_u}$ using equation (8) and $\{\mu_n^i, v_n^i\}_{n \in \mathcal{T}_u}$

  using equations (13) and (15), $\forall i \in \mathcal{T}$.

6: Apply the three MAR algorithms:

  1. For **image inpainting**, use equation (16) and the metal segmentation of step 2.

  2. For **sinogram inpainting**, use the image inpainting result as a template and proceed as outlined in section 2.4.

  3. For **MBIR**, use the derivatives in equations (20) and (21) along with equation (22) in an MLTR implementation.

For our experiments, we implemented the various steps as follows:

Step 1: We acquired the FBPs as reconstructed by the vendor-provided scanner software.

Step 2: We performed the metal segmentation automatically using Otsu's method (Otsu 1979).

Step 3: We found the exact value of κ in the expression for $f(\mathbf{x}_i)$ to be non-critical, and used $\kappa = (10 \text{ mm})^2$ for all our head-and-neck patients.

Step 5: The vast majority of the weights $\{w_n^i\}_{n \in \mathcal{T}_u}$ will in practice attain very small values and therefore not contribute to the model in a meaningful way. In order to speed up computations, we therefore used a fast patch matching algorithm (Ta *et al* 2014) to identify 200 regression points for each voxel i with particularly small patch differences $\|\mathbf{m}_n - \mathbf{m}_i\|$ and therefore large weights w_n^i, effectively clamping the weights of all other regression points to zero. We used $5 \times 5 \times 5$ patches, i.e. $6 \times 6 \times 10$ mm.

Step 6: For sinogram inpainting, we used 3D spiral forward projection to detect the metal projections and calculate the prior projections. For the interpolation of the metal projections, we used 2D barycentric linear interpolation over the cylindrical detector array on a triangular grid. For image reconstruction, since the MBIR implementation in particular required numerous slow but highly parallellizable forward and back projection operations, speed was a priority both for potential clinical implementation and to facilitate the experiments. We therefore used the GPU-accelerated primitives in the ASTRA (van Aarle *et al* 2015, 2016) package, as well as the bundled FBP implementation (with a Shepp–Logan filter). We performed the reconstructions in 2D, after rebinning and interpolating the 3D spiral sinograms from the scanner to sets of 2D sinograms with a linear detector geometry.

We initialized the iterative reconstruction process in MBIR with uniform images with an attenuation coefficient of 10^{-6} mm^{-1}. We stopped iterating once the voxel-averaged change between iterations fell below 10^{-6} mm^{-1}.
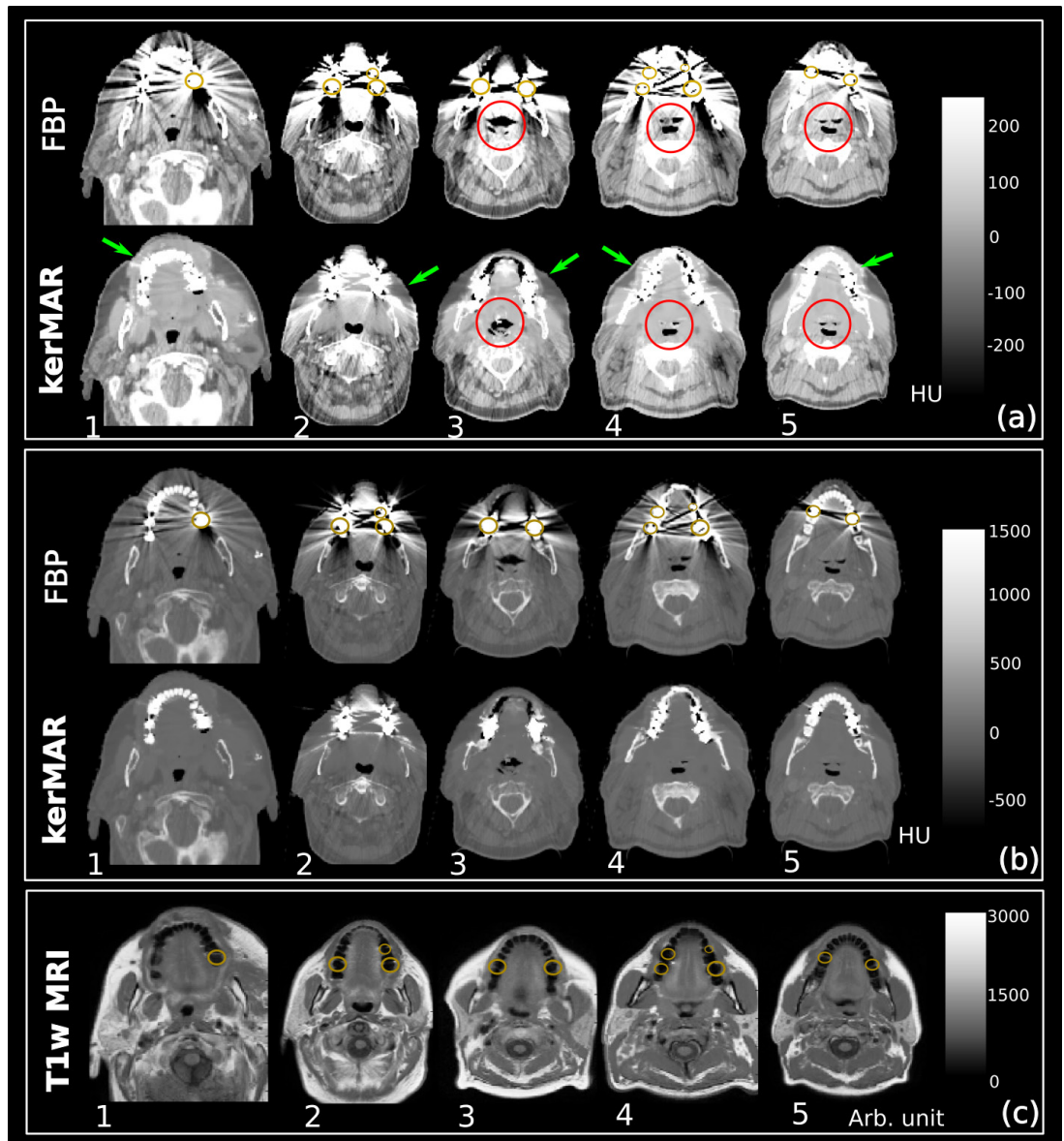
Figure 4. H&N image results. Gold circles indicate the metal implants and patient numbers are shown below the images. (a) and (b) FBP (top) and kerMAR (bottom) at different window levels. (a) A narrow, soft tissue enhancing window level. The arrows indicate regions where the anatomical information in the MR scan led to artifact reduction, while the red circles indicate compromised areas due to poor CT/MR co-registration. (b) A wider window level. (c) The T1w MR images.

Our entire computational framework used Python. The MAR algorithms, parameter estimation algorithm and sinogram rebinning methods in particular used NumPy and SciPy, while the reconstruction algorithms used ASTRA in its Python-wrapped form.

3.3. Quantitative evaluation

For each MAR algorithm, we evaluated the quantitative accuracy of the artifact-reduced CT images as follows: first, we acquired manual delineations of the oral cavity and the teeth for each patient, drawn in the FBP CT images guided by the MR scans. We split these delineations into a corrupted and uncorrupted part using our definitions of $\mathcal{T}_c$ and $\mathcal{T}_u$, and calculated reference mean CT values for each structure and each patient using the CT values in the uncorrupted parts. Around these mean values, we calculated the CT value standard deviations (STDs) over the corrupted parts; this provided an image quality metric for each structure, with each MAR, and for each patient. We performed the calculations in Hounsfield Units (HU) (Buzug 2008). We finally contrasted the STD observations for the different MAR algorithms using a repeat measurements Student's t-test, and report the patient-averaged STD results for all MAR algorithms and structures, as well as the p-values of the tests.

The use of the STD metric as an image quality metric relies on the assumption that different partitions of each ROI do not display large differences in mean HU values in the absence of artifacts as a result of large anatomical variations; such differences would lead to a systematic error in the reference mean, directly affecting the
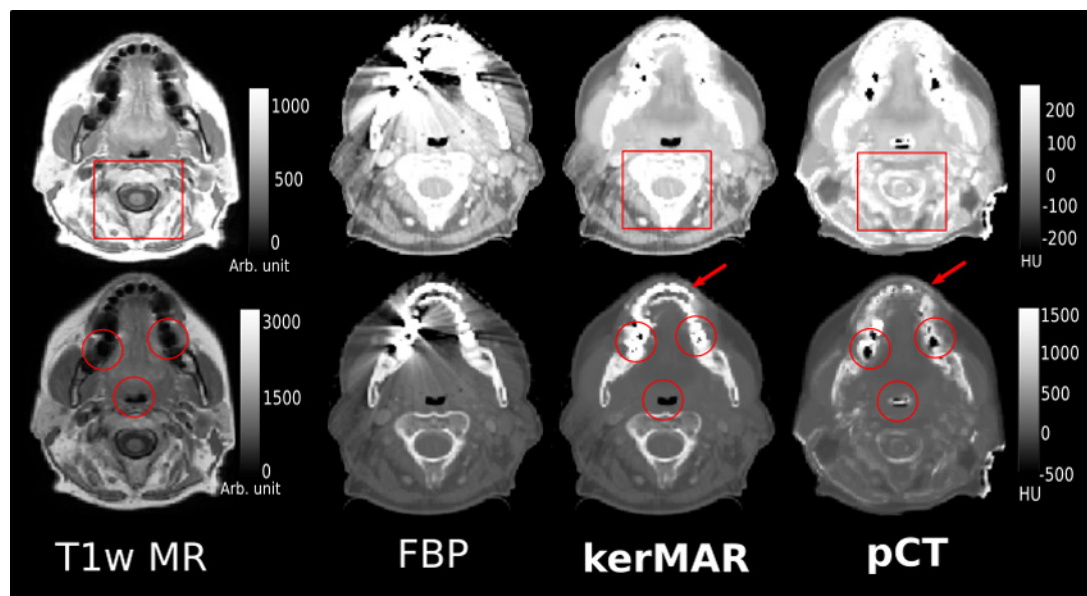
Figure 5. Axial slices of kerMAR and its special cases, i.e. the MR-based pCT (kerMAR with infinite artifact noise variance) and the FBP (kerMAR with 0 noise variance). Results are shown at a wide (top) and narrow, tissue enhancing (bottom) window level. The kerMAR simultaneously displays artifact reduction, improved anatomical fidelity in difficult bone and air regions (rings) and handling of co-registration errors due to inter-scan motion (arrows). The squares focus on a slice of the spinal cord, where the kerMAR noticeably improves upon the pCT by referencing the FBP.

STD results. To estimate the magnitude of this error, we used the uncorrupted parts of our nine-patient cohort to emulate the artifact-free tissue distributions in the oral cavity and teeth. Using that the corrupted and uncorrupted sets in our experiments were, on average, not far from equal-sized, we then created 10^6 random bipartions of the extracted set of CT values, for each calculating the difference in mean between the partitions. These differences between means turned out to be near-Gaussian distributed; for the oral cavity the average was ~ 0 HU and the standard deviation 1.1 HU, while for the teeth the average was ~ 0 HU and standard deviation 6.3 HU. Even after multiplying the standard deviations by a factor 10 for a more conservative estimate, the error on the reference means in our experiments should thus be in the tens of HU, which is an order of magnitude smaller than the variations between MAR algorithms that we present in the following section. Consequently, the STD metric appears to be a valid indicator of artifact suppression.

4. Results

We now report our results in sequence for the proposed image inpainting, sinogram inpainting and MBIR algorithms.

4.1. Image inpainting

Figure 4 shows visual results of the MR-based image inpainting algorithm kerMAR for representative head-and-neck patients, at a narrow, soft tissue enhanced window level (a) and a wider level (b). The T1w MR images are shown in (c). The green arrows in (a) indicate regions where kerMAR provided notable artifact reduction by suppressing both high and low intensity streaks, while at the same time preserving complex structures where the CT values are difficult to predict from the MR scan, such as the teeth.

The region where kerMAR was least successful was the windpipe, indicated by the rings. Here, the sometimes severe misalignments of the CT and MR images, due to inter-scan patient motion, led to a highly inaccurate MR-based prior and thus anatomical errors. However, in similar but less severe cases, such as with smaller misalignments, mechanisms in the kerMAR algorithm allowed it to avoid such anatomical errors. In particular, as we saw in section 2.2.1, the kerMAR image is an intermediate between the extreme special cases of pCT ($\sigma_t^{*2} \to \infty$) and FBP ($\sigma_t^{*2} \to 0$), and thus corresponds to a non-trivial blending of the FBP and MR-based prediction. Figure 5 shows the kerMAR image along with these two special cases for a patient where the MR and CT scans were misaligned in the teeth. In this less severe case of misalignment, while we see poor pCT performance and thus a poor prior-based prediction, the blending with the FBP led to a much improved kerMAR. Similar results occurred in the spinal cord (red square), where similar inaccuracies led to an apparent introduction of MR features in the pCT, which were successfully suppressed in the kerMAR.
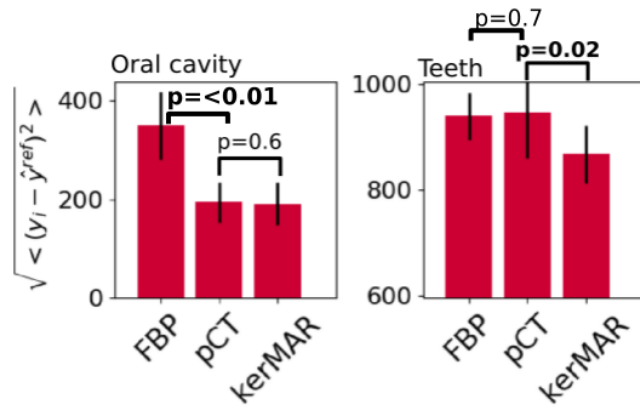
Figure 6. Results of the quantitative analysis in the oral cavity and teeth for FBP, purely MR-based pCT and our kerMAR algorithm. The error bars are standard deviations.
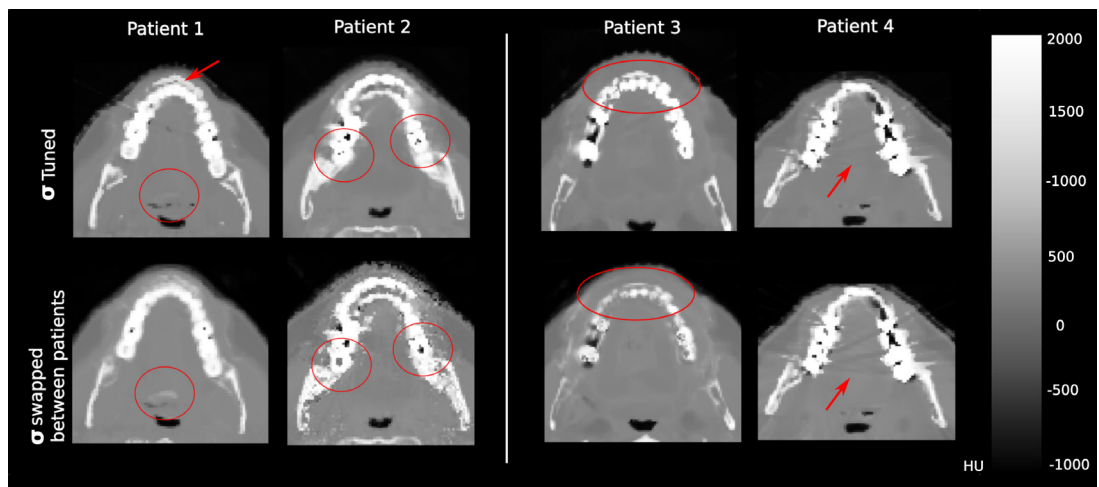

Figure 7. kerMAR calculated for four head-and-neck patients using tuned parameters $\boldsymbol{\sigma}$ (top) and with these parameters intentionally swapped between patients with different tunings; in particular, 1 with 2 and 3 with 4 (bottom). The rings and circles highlight the most substantial differences owed to the parameter swaps.

To summarize the results of the visual inspection, we saw (1) clear visual artifact reduction with both kerMAR and pCT and (2) a clear performance difference in bone and air regions between the two MAR algorithms. For the quantitative analysis of the oral cavity and teeth, these observations are reflected in the STDs shown in figure 6, which measure the CT value standard deviation in the corrupted parts of the structures around the mean intensity in the uncorrupted parts. In the oral cavity, kerMAR and pCT were equally effective, with significant STD reductions relative to the FBP of $\sim$150 HU. In the teeth, however, the pCT did not improve on the FBP, consistent with our observations of poor pCT performance in such thick, bony regions. Here, only kerMAR showed a benefit, providing a significant improvement over the pCT of $\sim$100 HU.

4.1.1. Influence of automatic parameter tuning

We finally considered the influence of the automatic patient-specific parameter tuning (algorithm 1) by intentionally swapping the estimated parameters between patients. Figure 7 shows the results for two representative patient-pairs, where we swapped the parameters for patient 1 and 3 with those of patient 2 and 4, respectively. Patients 1 and 3 displayed relatively smaller settings of σ_y^2 and σ_t^{*2}, and swapping the parameters with those of patient 2 and 4 therefore led to increased σ_y^2 and σ_t^{*2}. This substantially altered the predictive model; as illustrated in figure 3, a σ_y^2 increase widens the prior peaks while a σ_t^{*2} increase widens the likelihood, leading to a prediction that was both less precise and less accurate in areas where the MR-based prediction was compromised. These changes in the predictive model caused errors especially in the teeth (rings and arrow), as well as blur for patient 1, as compared to the automatically tuned parameters.

Patients 2 and 4 correspondingly displayed larger settings of σ_y^2 and σ_t^{*2}. Upon swapping, the ensuing decrease in σ_y^2 and σ_t^{*2}, and thus narrowing of the posterior, led to image noise and less effective artifact reduction; this is clearly visible for patient 2. Along with the larger σ_t^{*2} and σ_y^2, patients 2 and 4 further displayed smaller values of
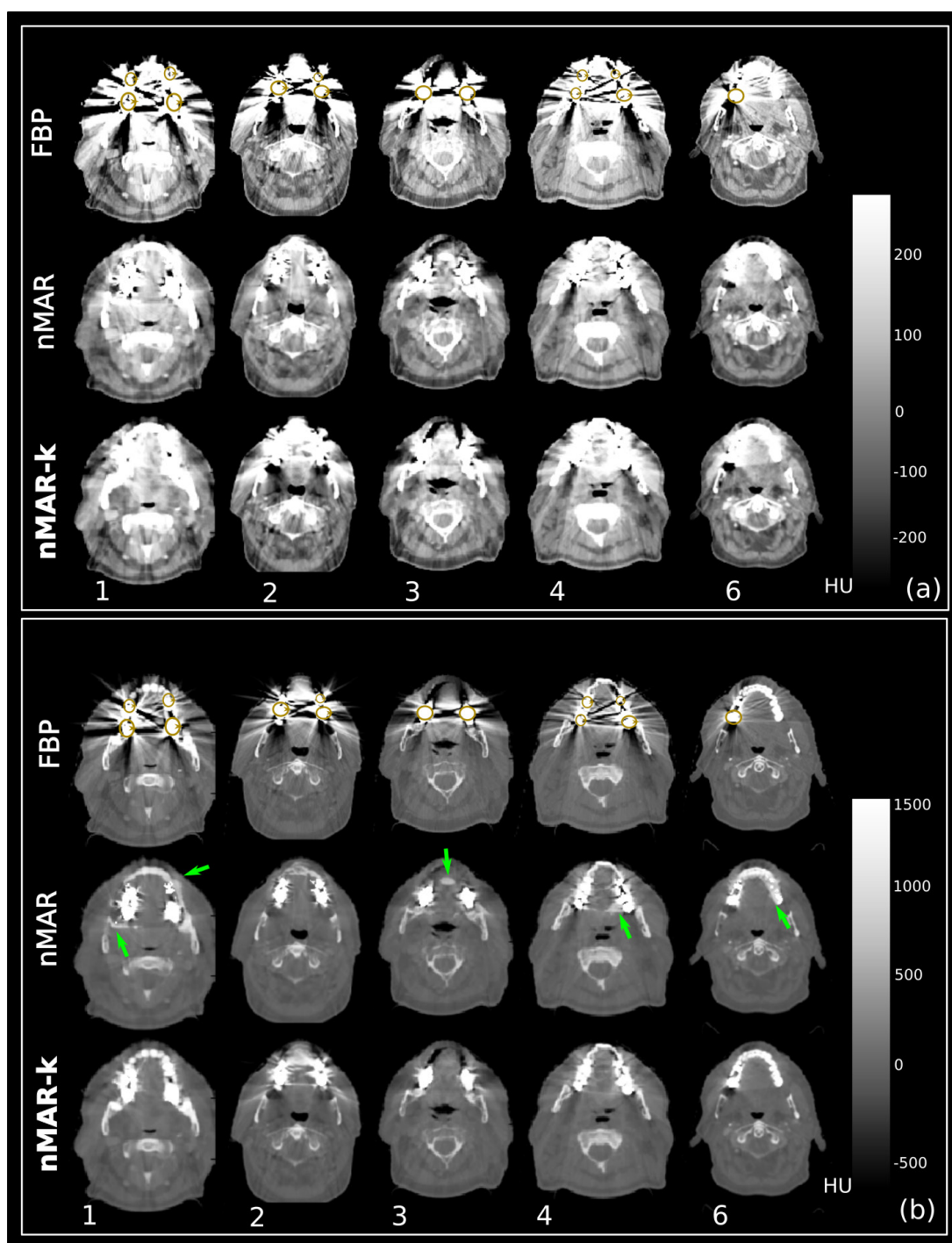
Figure 8. Head-and-neck results at different window levels for (top) standard nMAR, where a conventional FBP-based template is used for projection estimation and (bottom) the proposed nMAR-k where we used the MR-based kerMAR as template. Golden circles indicate the metal implants, and patient numbers are shown beneath the images. (a) A narrow, soft tissue enhancing window level. (b) A wider window level. Arrows point to artifacts introduced in conventional nMAR due to errors in the CT-based template.

σ_m^2. The swap therefore also led to an increase in σ_m^2 and thus a prior model with a less sensitive MR patch-based distinction between its regression points. Compared to using the automatically tuned parameters, this led to errors for patient 2 in the thick molars indicated by the circles, where the relevant MR patch differences are subtle.

4.2. Sinogram inpainting

Figure 8 shows visual results of the proposed MR-based sinogram inpainting method nMAR-k, alongside those produced with a standard nMAR implementation based on a conventional, FBP-based template image. As seen at the narrow, soft tissue enhancing window level in figure 8(a), nMAR and nMAR-k performed similarly in terms of artifact reduction; neither algorithm reconstructed an artifact-free oral cavity, while both reduced artifacts elsewhere to similar degrees (see regions indicated by arrows).
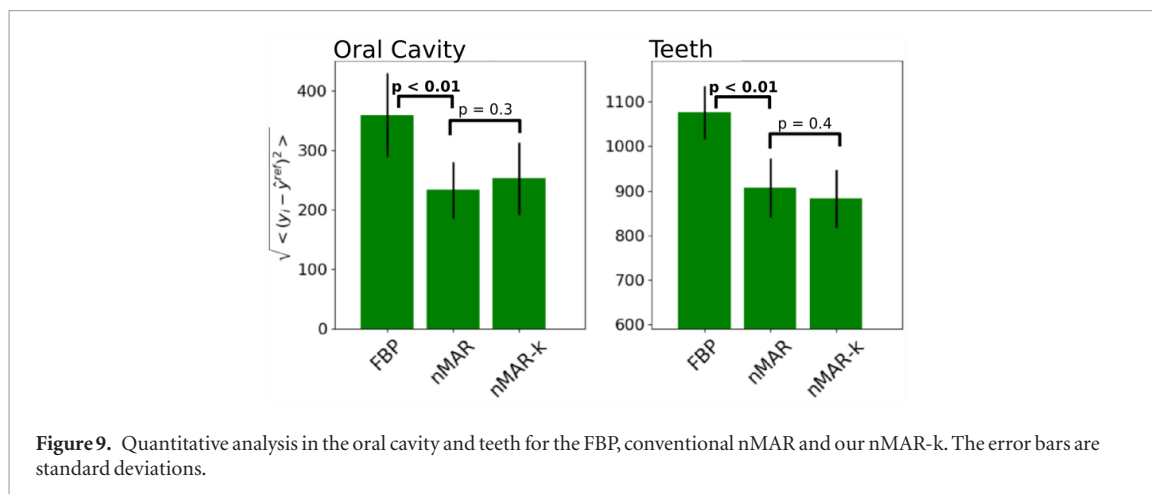
Figure 9. Quantitative analysis in the oral cavity and teeth for the FBP, conventional nMAR and our nMAR-k. The error bars are standard deviations.

At the wider window level in figure 8(b), we see a clearer difference between nMAR and nMAR-k: The sometimes lower quality of the FBP-based template in standard nMAR led to the visible artifacts (arrows), which are absent using the proposed MR-based method. A clear benefit of using the higher quality MR-based template thus appears to be to avoid certain artifacts introduced by the CT-based one. The influence of such improvements on the quantitative STD metric was comparatively minor, as figure 9 shows significant improvement of both methods over the FBP, but little quantitative difference between them.

4.3. Model-based iterative reconstruction

In order to evaluate the benefit of including the predictive distribution $p(\{y_i\}|\{t_i, \mathbf{m}_i\}, \boldsymbol{\sigma})$ as an image reconstruction prior in the MLTR algorithm, we ran the algorithm both with (the proposed MLTR-k algorithm) and without this prior (referred to in the remainder as simply 'MLTR'). We ran each algorithm until the voxel-averaged difference accross two subsequent iterations of the reconstructed image decreased below a predefined threshold (10^{-6}). Figure 10 displays the obtained reconstructions at the specified convergence threshold for both methods in five representative patients. Results are shown at a soft tissue enhancing window level (a) and at a wider level (b). At the narrow window level, the main detectable benefit of MLTR-k compared to MLTR was less blurry reconstructions, which indicate that the algorithm was closer to full convergence; in terms of artifact reduction, MLTR-k more successfully eliminated low intensity streaks but did not wholly eliminate high intensity ones. At the wider window level, MLTR-k more clearly improved the MAR over MLTR (arrows in (b)), although these benefits came, in a few cases, at the expense of some newly introduced artifacts in the teeth (see black rings), that however were not a general trend.

Figure 11(a) further shows the logarithm of the voxel-averaged difference used as convergence criterion for both algorithms and our nine patients across iterations. As we can see, the rate of convergence is universally larger for MLTR-k with the MR-based prior, as compared to the purely likelihood-based MLTR. In particular, the convergence speed was essentially doubled with MLTR-k, decreasing the number of iterations from ~400–600 to ~200–300.

Visually, MLTR-k with its MR-based prior thus provided superior MAR to the prior-free MLTR, as well as an image closer to full convergence at the specified convergence threshold. These results are supported by the quantitative results in figure 11(b) that show significant standard deviation improvements compared to MLTR, in both the oral cavity and teeth, of respectively ~200 HU and ~150 HU.

4.4. Comparison to gold standard

The nMAR sinogram inpainting algorithm (with the CT-based template) may be used as a gold standard for benchmarking and comparing the three proposed MR-based algorithms between themselves. To more easily perform this comparison, we summarize the results of our quantitative evaluations in figure 12 along with p-values of the comparison to nMAR; the results for the MR-based results are highlighted. We see no case of significantly worse performance compared to the gold standard of the three MR-based algorithms, and further see a quantitative improvement in the oral cavity using the image space approach (kerMAR). Additionally, the quantitative results should be taken together with the visual comparison, and when we compared the sinogram inpainting method (nMAR-k) to nMAR in section 4.2, we found the negligible quantitative results to be accompanied by visual improvements. The comparison is more difficult with the MBIR approach (MLTR-k), where we again see insignificant STD results but where, as we saw in section 4.3, the visual quality may have been compromised by additional artifacts.
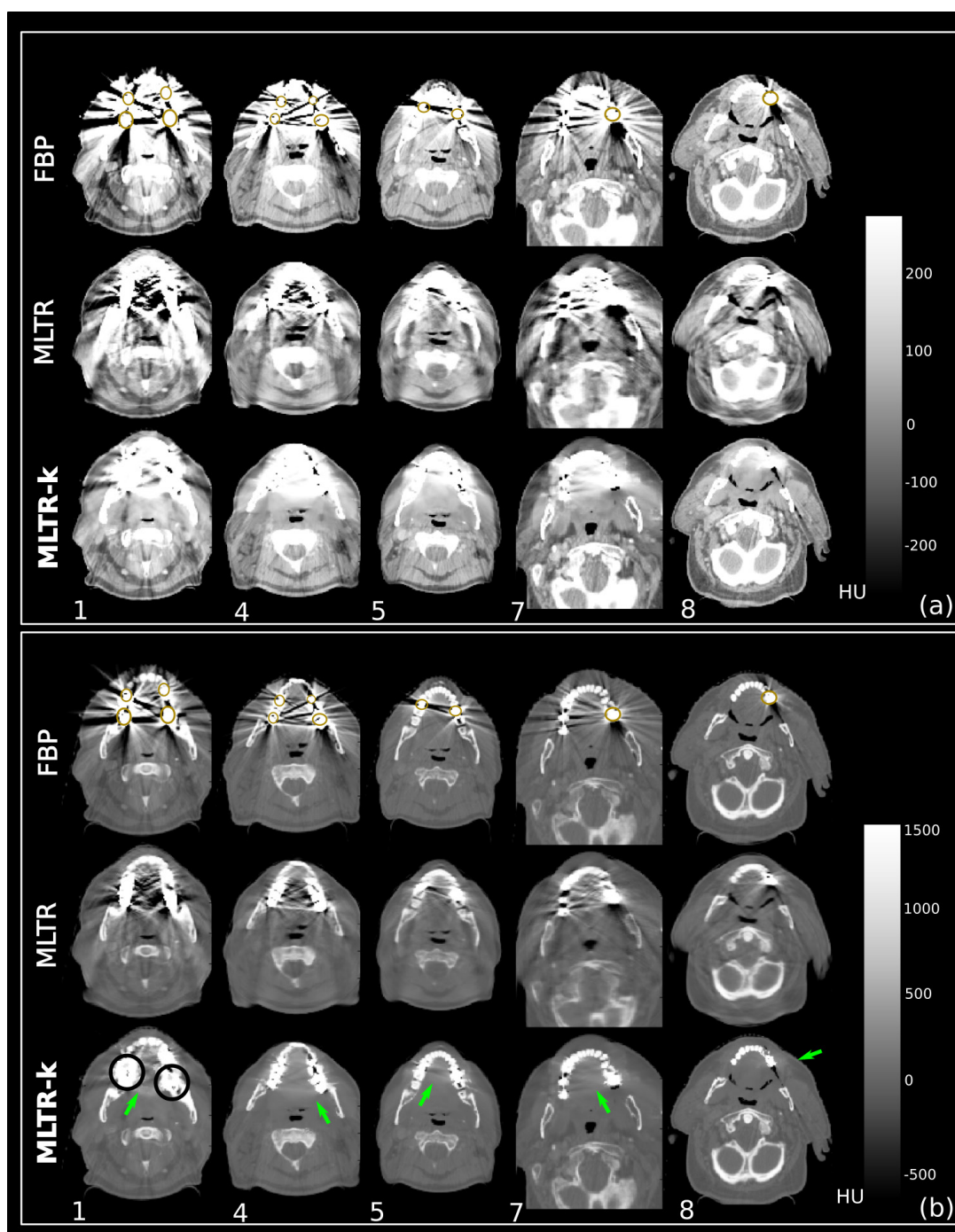
Figure 10. MLTR (MBIR without a prior, middle) and MLTR-k (MBIR with MR-based prior, bottom) head-and-neck image results shown alongside the FBP images (top) at different window levels. Golden circles indicate the metal implants, and patient numbers are shown beneath the images. (a) A narrow, soft tissue enhancing window level. (b) A wider window level. Arrows point to improved artifact reduction with MLTR-k, while the black rings in the teeth show a few artifacts introduced with MLTR-k.

In summary, the image and sinogram inpainting algorithms at least show benefits over the gold standard nMAR algorithm, while the MBIR results are more inconclusive.

5. Conclusion and discussion

We have presented a novel Bayesian approach to MR-based MAR, and in particular derived a predictive distribution of an ideal, uncorrupted CT from a corrupted FBP CT and a co-registered, conventional-sequence MR scan. We used the obtained predictive distribution to define three automatic MAR approaches:
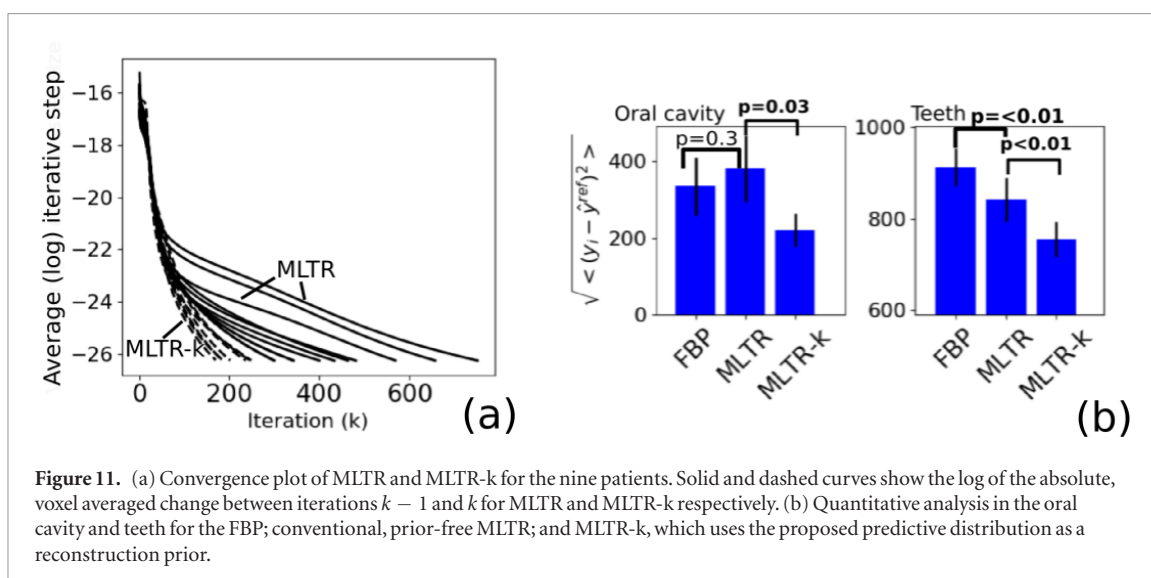
Figure 11. (a) Convergence plot of MLTR and MLTR-k for the nine patients. Solid and dashed curves show the log of the absolute, voxel averaged change between iterations $k - 1$ and k for MLTR and MLTR-k respectively. (b) Quantitative analysis in the oral cavity and teeth for the FBP; conventional, prior-free MLTR; and MLTR-k, which uses the proposed predictive distribution as a reconstruction prior.
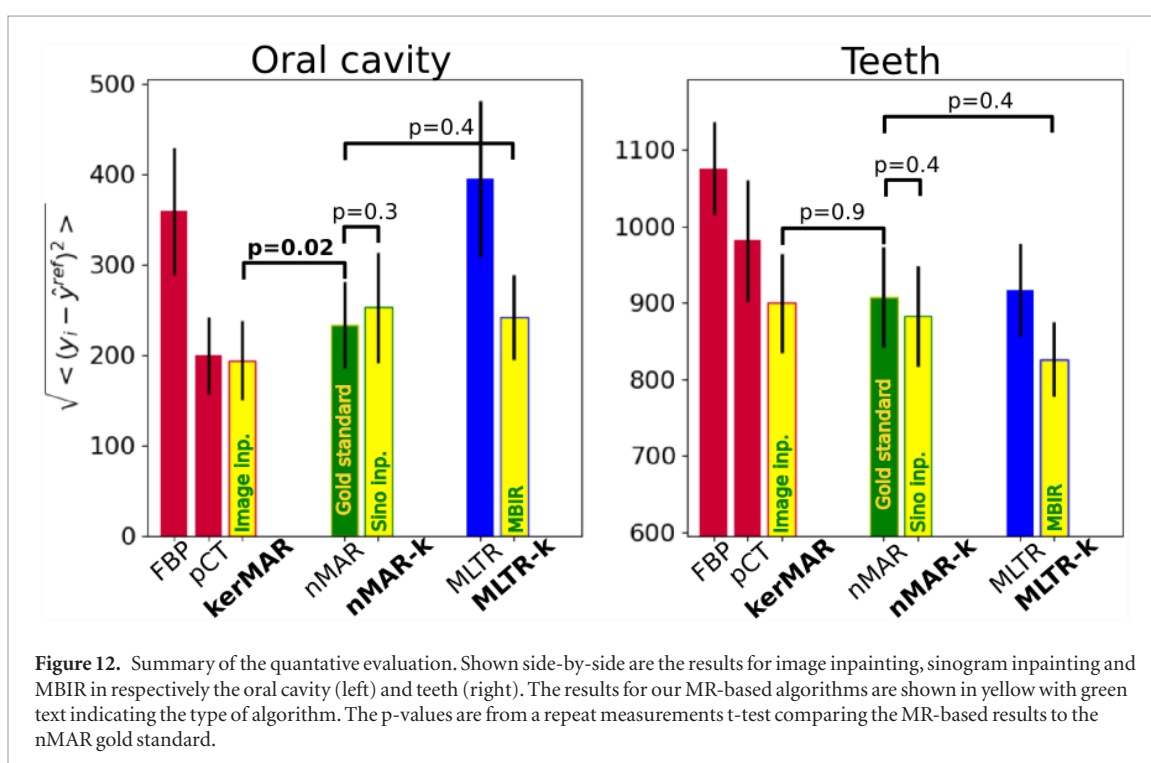

Figure 12. Summary of the quantative evaluation. Shown side-by-side are the results for image inpainting, sinogram inpainting and MBIR in respectively the oral cavity (left) and teeth (right). The results for our MR-based algorithms are shown in yellow with green text indicating the type of algorithm. The p-values are from a repeat measurements t-test comparing the MR-based results to the nMAR gold standard.

1. The image inpainting algorithm kerMAR seamlessly blends MR-derived predictions with information from the corrupted FBP CT, the former leading to substantial artifact reduction and the latter helping especially with bone/air disambiguation and correcting co-registration errors.

2. The sinogram inpainting algorithm nMAR-k uses the kerMAR result as the template in the well-known nMAR algorithm. In our experiments, this led to improvements over using a conventional, FBP-based template by introducing fewer artifacts and improving the artifact reduction.

3. MLTR-k uses the proposed predictive distribution as an image reconstruction prior in the MBIR algorithm MLTR. This led to improvements in terms of both speed (of $\sim 50\%$) and the quality of the reconstructed image, especially in terms of artifact reduction.

We conclude that the proposed approach provides a versatile way to use the anatomical information in the MR scan to boost the performance of MAR.

An important aspect of the proposed approach is that its parameters are automatically tuned in a patient-specific manner, which makes it well-suited for potential clinical implementation. The automatic tuning accounts for inter-patient variations, in particular the extent of the artifact corruption, which is an important robustness benefit that allows for application to diverse cases and especially reduces potential over-correction in less

corrupted cases. The tuning also accounts for variations in the imaging settings; we saw this in the consistent performance of our algorithms despite variations in the MR sequence parameters.

5.1. Study limitations

While our MR-based method is applicable whenever an MR scan is co-acquired with a CT, it should be noted that such cases may be of limited frequency outside of head-and-neck radiotherapy, as MR at the time of this study is only used infrequently in the case of metal artifacts at other sites, such as hip implants. Due to the continually increasing importance of MR in radiotherapy, this may however change in the future (Lagendijk *et al* 2014).

5.2. Future work

The inclusion of the FBP for the CT value prediction in our image inpainting algorithm kerMAR helped to mitigate co-registration and alignment issues between the MR and CT images, but for clinical adoption the associated errors in especially the windpipe should probably further reduced (see figures 4(a) and (b)). A straightforward way to achieve this is to improve the co-registration, e.g. by using a deformable registration methods, which are routinely used in modern radiotherapy clinics. Further accuracy gains may be achieved by improving the noise model, which in its current form simply assumes that the noise introduced by the artifacts decreases sigmoidally with the distance to the metal implants. This is a rough assumption as especially streak artifacts may persist throughout the entire CT image (see, for instance, figure 4(a)). Our current model may also not translate well to other types of implants, such as metallic dual hip prostheses, where the severity of the artifacts also depends heavily on whether or not they are located between the prostheses. Future experiments may therefore investigate different formulations of the function $f(\mathbf{x}_i)$ that better capture the spatial dependency of the artifact noise.

In order to successfully translate the proposed techniques into clinical routine, the required computation time will need to be further reduced. The most time-consuming part of our model is the calculation of patch distances over regression points, which is currently implemented on a subset of 200 points $\forall i \in \mathcal{T}$, found using a patch matching algorithm (Ta *et al* 2014). In our current Python implementation on a single CPU core (Intel Core i7-4712HQ @ 2.30 GHz), this process takes between 10–30 min. The algorithm is however parallelizable, and on a similar-sized dataset, Ta *et al* report results on the order of ∼1 min on a multi-CPU cluster (Ta *et al* 2014). In the future, we therefore intend to speed up our algorithm in a similar manner.

Our MBIR experiments used a relatively simple Poisson likelihood. While this model helps address the artifacts stemming from photon starvation of the metal projections (Nuyts *et al* 1998, Buzug 2008), it does not account for e.g. the important effect of beam hardening. This may have been the source of the artifacts that were sometimes introduced with MLTR-k (see e.g. the black rings in figure 10) (Nuyts *et al* 1998, De Man *et al* 2000, 2001, Buzug 2008, Van Slambrouck and Nuyts 2012). Future work may therefore consider using a more accurate likelihood model for the MBIR technique.

Acknowledgments

ORCID iDs

Jonathan Scharff Nielsen ⬦ https://orcid.org/0000-0001-9884-9886

References

Anderla A A, Culibrk D R and Delso G 2013 Metal artifact reduction from ct images using complementary MR images *11th Int. Conf. on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Telsiks* vol 1 p 6704943

Andersson K M, Ahnesjö A and Dahlgren C V 2014 Evaluation of a metal artifact reduction algorithm in CT studies used for proton radiotherapy treatment planning *J. Appl. Clin. Med. Phys.* **15** 112–9

Andreasen D, Van Leemput K and Edmund J M 2016 A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis *Med. Phys.* **43** 4742–52

Andreasen D *et al* 2015 Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain *Med. Phys.* **42** 1596–605

Bal M and Spies L 2006 Metal artifact reduction in CT using tissue-class modeling and adaptive prefiltering *Med. Phys.* **33** 2852–9

Beister M, Kolditz D and Kalender W A 2012 Iterative reconstruction methods in X-ray CT *Phys. Medica* **28** 94–108

Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)

Borman S 2009 The expectation maximization algorithm A short tutorial *Submitted for publication* **25** 1–9 (http://www.cs.utah.edu/~piyush/teaching/EM_algorithm.pdf)

Buzug T M 2008 *Computed Tomography* (Berlin: Springer) p 526

De Man B 2001 Iterative reconstruction for reduction of metal artifacts in computed tomography *PhD Thesis* Katholieke Universiteit Leuven

De Man B *et al* 2000 Reduction of metal streak artifacts in x-ray computed tomography using a transmission maximum *a posteriori* algorithm *IEEE Trans. Nucl. Sci.* **47** 977–81

De Man B *et al* 2001 An iterative maximum-likelihood polychromatic algorithm for CT *IEEE Trans. Med. Imaging* **20** 999–1008

Delso G *et al* 2013 MR-driven metal artifact reduction in PET/CT *Phys. Med. Biol.* **58** 2267–80

Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc.* B **39** 1–38

Dietlicher I *et al* 2014 The effect of surgical titanium rods on proton therapy delivered for cervical bone tumors: experimental validation using an anthropomorphic phantom *Phys. Med. Biol.* **59** 7181–94

Dowling J A *et al* 2012 An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **83** e5–11

Elbakri I A and Fessler J A 2002 Statistical image reconstruction for polyenergetic x-ray computed tomography *IEEE Trans. Med. Imaging* **21** 89–99

Fu L *et al* 2017 Comparison between pre-log and post-log statistical models in ultra-low-dose ct reconstruction *IEEE Trans. Med. Imaging* **36** 7740011

Giantsoudi D *et al* 2017 Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction *Phys. Med. Biol.* **62** R49–80

Gjesteby L *et al* 2016 Metal artifact reduction in CT: Where are we after four decades? *IEEE Access* **4** 7565564

Goldman L W 2007 Principles of CT and CT technology *J. Nucl. Med. Technol.* **35** 115–30

Hamelin B *et al* 2008 Iterative CT reconstruction of real data with metal artifact reduction *IEEE Int. Symp. on Biomedical Imaging* pp 1453–6

Han X 2017 MR-based synthetic CT generation using a deep convolutional neural network method *Med. Phys.* **44** 1408–19

Hofmann M *et al* 2008 MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration *J. Nucl. Med.* **49** 1875–83

Hsieh J *et al* 2000 An iterative approach to the beam hardening correction in cone beam CT *Med. Phys.* **27** 23–9

Jäkel O and Reiss P 2007 The influence of metal artefacts on the range of ion beams *Phys. Med. Biol.* **52** 635–44

Jin P, Bouman C A and Sauer K D 2015 A model-based image reconstruction algorithm with simultaneous beam hardening correction for x-ray CT *IEEE Trans. Comput. Imaging* **1** 200–16

Jonsson J H *et al* 2010 Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions *Radiat. Oncol.* **5** 62

Kalender W A, Hebel R and Ebersberger J 1987 Reduction of CT artifacts caused by metallic implants *Radiology* **164** 576–7

Kim Y and Tomé W A 2007 On the radiobiological impact of metal artifacts in head-and-neck IMRT in terms of tumor control probability (TCP) and normal tissue complication probability (NTCP) *Med. Biol. Eng. Comput.* **45** 1045–51

Koehler T, Brendel B and Brown K 2012 A new method for metal artifact reduction in CT *The Int. Conf. on Image Formation in X-ray Computed Tomography (June 24–27, Salt Lake City, Utah, USA)* (http://repository.tudelft.nl/assets/uuid:22d5815d-dcfe-48df-93a4-9d1c4e8e85fc/MS-33.229.pdf)

Kwon H *et al* 2015 Evaluation of a commercial orthopaedic metal artefact reduction tool in radiation therapy of patients with head and neck cancer *Br. J. Radiol.* **88** 13–22

Lagendijk J J *et al* 2014 MR guidance in radiotherapy *Phys. Med. Biol.* **59** R349

Lee M J *et al* 2017 Overcoming artifacts from metallic orthopedic implants at high-field-strength MR imaging and multi-detector *IEEE Trans. Nucl. Sci.* **27** 791–803

Lell M M *et al* 2012 Normalized metal artifact reduction in head and neck computed tomography *Investigative Radiol.* **47** 415–21

Lell M M *et al* 2013 Frequency split metal artefact reduction in pelvic computed tomography *Eur. Radiol.* **23** 2137–45

Maerz M, Koelbl O and Dobler B 2015 Influence of metallic dental implants and metal artefacts on dose calculation accuracy *Strahlentherapie Onkologie* **191** 234–41

Maes F *et al* 1997 Multimodality image registration by maximization of mutual information *IEEE Trans. Med. Imaging* **16** 187–98

Makeev A and Glick S J 2013 Investigation of statistical iterative reconstruction for dedicated breast CT *Med. Phys.* **40** 081904

MATLAB 2019 *Version 9.6.0 (R2019a)* (Natick, MA: MathWorks Inc)

Meyer E *et al* 2009 Normalized metal artifact reduction (nMAR) in computed tomography *IEEE Nuclear Science Symp. Conf. Record* p 5401721

Meyer E *et al* 2012 Frequency split metal artifact reduction (FSMAR) in computed tomography *Med. Phys.* **39** 1904–16

Nielsen J, Edmund J and Van Leemput K 2018 CT metal artifact reduction using MR image patches *Proc. SPIE* **10573** 105730P

Nuyts J *et al* 2013 Modelling the physics in the iterative reconstruction for transmission computed tomography *Phys. Med. Biol.* **58** R63–96

Nuyts J *et al* 1998 Iterative reconstruction for helical CT: a simulation study *Phys. Med. Biol.* **43** 729

Otsu N 1979 A threshold selection method from gray-level histograms *IEEE Trans. Syst. Man Cybern.* **9** 62–6

Pan Y *et al* 2010 TV-regularized iterative image reconstruction on a mobile C-ARM CT *Imaging* **7622** 76222L

Park P C *et al* 2015 MRI-based computed tomography metal artifact correction method for improving proton range calculation accuracy *Int. J. Radiat. Oncol. Biol. Phys.* **91** 849–56

Stayman J W *et al* 2012 Model-based tomographic reconstruction of objects containing known components *IEEE Trans. Med. Imaging* **31** 6200873

Stille M *et al* 2016 Augmented likelihood image reconstruction *IEEE Trans. Med. Imaging* **35** 7164320

Stonestrom J P, Alvarez R E and Macovski A 1981 A framework for spectral artifact corrections in x-ray CT *IEEE Trans. Biomed. Eng.* **BME-28** 128–41

Ta V T *et al* 2014 Optimized patchmatch for near real time and accurate label fusion *Medical Image Computing and Computer-assisted Intervention : Miccai ... Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* vol 17 pp 105–12

Thibault J B *et al* 2007 A three-dimensional statistical approach to improved image quality for multislice helical CT *Med. Phys.* **34** 4526–44

Tilley S, Siewerdsen J H and Stayman J W 2016 Model-based iterative reconstruction for flat-panel cone-beam CT with focal spot blur, detector blur, and correlated noise *Phys. Med. Biol.* **61** 296

Torrado-Carvajal A *et al* 2015 Fast patch-based pseudo-CT synthesis from T1-weighted MR images for PET/MR attenuation correction in Brain studies *J. Nucl. Med.* **57** 136–44

Torrado-Carvajal A *et al* 2019 Dixon-VIBE deep learning (DIVIDE) pseudo-CT synthesis for pelvis PET/MR attenuation correction *J. Nucl. Med.* **60** 429–35

van Aarle W *et al* 2016 Fast and flexible x-ray tomography using the astra toolbox *Opt. Express* **24** 25129–47

van Aarle W *et al* 2015 The ASTRA toolbox: a platform for advanced algorithm development in electron tomography *Ultramicroscopy* **157** 35–47

Van Slambrouck K and Nuyts J 2012 Local models for artefact reduction in iterative CT reconstruction *IEEE Nuclear Science Symp. Conf. Record* vol 2 pp 4346–50

Webster Stayman J and Fessler J 2001 Nonnegative definite quadratic penalty design for penalized-likelihood reconstruction *IEEE Nuclear Science Symp. Conf. Record (Cat. No.01CH37310)* vol 2 pp 1060–3

Wells W M I *et al* 1996 Multi-modal image registration by maximization of mutual information *Med. Image Anal.* **1** 35–51

Zhang Y *et al* 2011 A new CT metal artifacts reduction algorithm based on fractional-order sinogram inpainting *J. X-Ray Sci. Technol.* **19** 373–84